Genome-wide association study of resistance to *Mycobacterium tuberculosis* infection identifies a locus at 10q26.2 in three distinct populations

Jocelyn Quistrebert,^{1,2} Marianna Orlova,^{3,4,5} Gaspard Kerner,^{1,2,6} Le Thi Ton,⁷ Nguyễn Trong Luong,⁷ Nguyễn Thanh Danh,⁷ Fabienne Jabot-Hanin,^{8,9,10} Yoann Seeleuthner,^{1,2} Jacinta Bustamante,^{1,2} Stéphanie Boisson-Dupuis,¹¹ Nguyen Thu Huong,¹² Nguyen Ngoc Ba,¹² Jean-Laurent Casanova,^{1,2,11,13,14} Christophe Delacourt,^{2,15} Eileen G. Hoal,¹⁶ Alexandre Alcaïs,^{1,2} Vu Hong Thai,¹² Lai The Thành,⁷ Laurent Abel,^{1,2,11,17} Erwin Schurr,^{3,4,5,17**} Aurélie Cobat^{1,2,17*}

¹Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Paris, France

²University of Paris, Imagine Institute, Paris, France

³Infectious Diseases and Immunity in Global Health Program, Research Institute of the McGill University Health Centre, Montreal, Canada

⁴McGill International TB Centre, McGill University, Montreal, Canada,

⁵Departments of Medicine and Human Genetics, McGill University, Montreal, Canada

⁶Human Evolutionary Genetics Unit, Institut Pasteur, UMR2000, CNRS, Paris, France

⁷Center for Social Disease Control, Binh Duong, Vietnam

⁸Bioinformatics Core Facility, Paris Descartes University, Paris, France

⁹Bioinformatics Platform, INSERM UMR U1163, Imagine Institute, Paris, France

¹⁰Necker Federative Research Structure, INSERM US24, CNRS UMS3633, Paris, France

¹¹St Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY, USA

¹²Hospital for Dermato-Venereology, Ho Chi Minh City, Vietnam

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

¹³Howard Hughes Medical Institute, New York, NY, USA

¹⁴Paediatric Hematology-Immunology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France

¹⁵Paediatric Pulmonology and Allergology Department, Necker Hospital for Sick Children, Paris, France

¹⁶South African Medical Research Council Centre for Tuberculosis Research, DST-NRF Centre of Excellence for Biomedical Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

¹⁷These authors contributed equally to this work

*Corresponding author 1: aurelie.cobat@inserm.fr

**Corresponding author 2: erwin.schurr@mcgill.ca

Abstract

The natural history of tuberculosis (TB) is characterized by a large inter-individual outcome variability after exposure to *Mycobacterium tuberculosis*. Specifically, some highly exposed individuals remain resistant to *M. tuberculosis* infection, as inferred by tuberculin skin test (TST) or interferon-gamma release assays (IGRAs). We performed a genome-wide association study of resistance to *M. tuberculosis* infection in an endemic region of Southern Vietnam. We enrolled household contacts (HHC) of pulmonary TB cases and compared subjects who were negative for both TST and IGRA (n=185) with infected individuals (n=353) who were either positive for both TST and IGRA or had a diagnosis of TB. We found a genome-wide significant locus on chromosome 10q26.2 with a cluster of variants associated with strong protection against *M. tuberculosis* infection (OR=0.42, 95%CI 0.35-0.49, $P = 3.71 \times 10^{-8}$, for the genotyped variant rs17155120). The locus was replicated in a French multi-ethnic HHC cohort and a familial admixed cohort from a hyper-endemic area of South Africa, with an overall OR for rs17155120 estimated at 0.50 (95%CI 0.45-0.55, $P = 1.26 \times 10^{-9}$). The variants are located in intronic regions and upstream of *C10orf90*, a tumor suppressor gene which encodes an ubiquitin ligase activating the transcription factor p53. *In silico* analysis showed that the protective alleles were associated with a decreased expression in monocytes of the nearby gene *ADAM12* which could lead to an enhanced response of Th17 lymphocytes. Our results reveal a novel locus controlling resistance to *M. tuberculosis* infection across different populations.

Introduction

Tuberculosis (TB) remains a major public health threat worldwide. An estimated 10 million people developed TB disease in 2018, of whom 1.45 million died¹. The causative agent of TB is *Mycobacterium tuberculosis* which is transmitted by aerosol from contagious TB patients. However, not all persons encountering infectious aerosols will become infected with *M. tuberculosis*, defining the first line of human resistance against TB^{2,3}. Infection is inferred from the presence of anti-mycobacterial immunoreactivity, as shown by a positive result in tuberculin skin test (TST) and/or interferon-gamma (IFN- γ) release assay (IGRA). TST is done *in vivo* and consists of an intradermal injection of purified protein derivative (PPD) that provokes a delayed hypersensitivity reaction at the site of injection. IGRAs are performed *ex vivo* and measure the secretion of interferon-gamma (IFN- γ) by leukocytes in response to *M. tuberculosis*-specific antigens. Both tests have their own limitations and results are not fully concordant^{4,5,6}. Individuals who score positive by TST and/or IGRA are considered to suffer from asymptomatic latent TB infection (LTBI). Conversely, persons who score negative despite documented exposure to *M. tuberculosis* are considered resistant to infection. Based on TST and/or IGRA results, the intensity of exposure or the duration of follow-up, from 7% to 25% of subjects display the *M. tuberculosis* infection resistance phenotype^{2,7,8,9}.

The large inter-individual variability in exposure outcomes supports a major role for human genetic factors¹⁰. Various genome-wide approaches have confirmed this hypothesis by either considering TST and IGRA results as quantitative traits or relying on TST reactivity (positive/negative) as a surrogate marker for infection^{11,12,13,14}. Regarding this latter phenotype, persistent TST negativity was linked to loci on 2q21-2q24, further fine-mapped to *ZEB2*, and 5p13-5q22 in an Ugandan population^{14,15}. A major locus, named *TST1* on chromosome 11p14, employing stringently defined TST negativity (0 mm vs > 0 mm) as phenotype, was identified in a linkage analysis conducted in South Africa. *TST1* was later replicated in a household

contact study of French families^{11,16}. A genome-wide association study (GWAS) among highly *M. tuberculosis* exposed HIV-seropositive individuals from East Africa, identified a locus in the 5q31.1 region near *IL9* associated with negative TST¹⁷.

Here, we performed a GWAS of resistance to *M. tuberculosis* infection using a robust phenotype based on both TST and IGRA information. In addition, we used a household contact study design guaranteeing shared environmental effects and high intensity of exposure to *M. tuberculosis*. We found a locus on chromosome 10q26.2 associated with resistance to *M. tuberculosis* infection in an East Asian population from Southern Vietnam. Importantly, this locus was replicated in 2 other cohorts from France and South Africa, representing different ancestries and epidemiological settings.

Subjects and Methods

Ethics statement

Signed informed consent was obtained from all the participants, and from the parents of enrolled minors. The study was approved by the regulatory authorities in Binh Duong, Vietnam (1366/UBND-VX), the Research Ethics Board of McGill University Health Centre, Montreal, Canada (06-030 GEN, 2007-859), the French Consultative Committee for Protecting Persons in Biomedical Research of Henri Mondor Hospital (Créteil, France) and the Stellenbosch University Health Research Ethics Committee (Tygerberg, South Africa).

Study settings and populations

Vietnam

Vietnam is a middle-income country in South-East Asia with a high annual incidence of TB (130/100,000 at the time of the study¹⁸), high Bacille Calmette-Guérin (BCG) vaccination

coverage (with reported rates exceeding 95%¹⁹) and a low population prevalence of HIV (0.4% in 2015²⁰). From 2010 to 2015, we recruited, in a *M. tuberculosis* endemic region of Southern Vietnam, index pulmonary TB (PTB) adults with persistent cough before the start of treatment (mean duration of coughing = 3.2 months). PTB diagnosis was assessed by clinical presentation, chest X-ray and/or positive cultures for *M. tuberculosis*. A total of 1108 household contacts (HHCs) of 466 PTB index cases were invited to participate in the study and underwent both TST and QuantiFERON®-TB Gold In-Tube test (QFT-GIT) (Supplemental Methods).

France

As a first replication cohort, we used household TB contacts studied in Val-de-Marne, a suburban region of Paris, in the context of a general screening procedure. This multi-ethnic cohort has been previously described^{12,16}. Val-de-Marne is an area of low TB endemicity, displaying an annual incidence of 22/100,000 at the time of the study²¹ and BCG vaccination rates are high. For this study, 664 HHCs of 132 PTB index cases were investigated according to national guidelines that required 2 screening visits. HHCs were individuals sharing residence with an index during the 3 months before diagnosis. Briefly, the first visit (V1) included a physical examination, a chest radiograph, TST and in-house IGRA^{12,16} (Supplemental Methods). These investigations, except for IGRA, were repeated 8–12 weeks later (V2) if the contact subject did not meet the criteria for infection at V1.

South Africa

As a second replication cohort, we used a large sample (n = 415) of 153 nuclear families from a suburban area of Cape Town, South Africa, which has been previously described^{6,11,12,16}. All individuals belonged to the South African Coloured (SAC) group, a unique multi-way admixed population²². There was no specific requirement for subjects to be HHCs of PTB patients. Indeed, TB is hyper-endemic in this area with an incidence of ~800/100,000 at the time of the

study²³ and TB transmission occurs more often outside the household²⁴. BCG vaccination at birth is routine in this area²⁵. TST and in-house IGRA were performed as previously described⁶ (Supplemental Methods). Subjects who had clinical TB disease in the 2 years preceding the study were excluded.

Genotyping, quality control and imputation

A total of 724 individuals from Vietnam (Figure S1) and 573 from France were genotyped using the Illumina Infinium™ OmniExpressExome-8-v1 chip (960,212 single nucleotide polymorphisms, SNPs). For the South African cohort (n = 374), the Illumina HumanOmni2.5-8 BeadChip (~2 million of SNPs) was used. All quality control steps were done in each cohort with PLINK v1.9²⁶. Autosomal SNPs with a minor allele frequency (MAF) > 0.01, a genotype call rate > 0.99 and a Hardy-Weinberg (HWE) equilibrium $P > 1.00 \times 10^{-5}$ were retained. Individuals with a call rate < 95% were excluded (n = 2). Identity-by-descent (IBD) analysis was done to detect duplicated individuals and the members of the pairs with the lower call rate were excluded (n = 1). After the quality control, imputation was performed on 720 individuals and 598,090 variants from Vietnam and 573 individuals and 886,471 variants from France using the Michigan Imputation Server²⁷ with Eagle2²⁸ for the pre-phasing and the 1000 Genomes phase 3 (1000G) Project as reference panel²⁹. For the South African cohort (374 individuals and 1,347,846 variants), the imputation was done on the Sanger Imputation Server³⁰ with Eagle2 for the pre-phasing and the African Genome Resources as reference panel which includes ~2000 African samples in addition to the individuals from 1000G phase 3. Imputed SNPs with an imputation quality info score > 0.8 and MAF > 0.05 were retained for further association analyses (5,591,951 variants in Vietnam, 7,737,070 variants in France and 6,922,541 variants in South Africa).

For each cohort, principal component analysis (PCA) was conducted to evaluate population structure. Genotypes of the individuals from 1000G phase 3 were used to calculate principal components and data for subjects from the cohorts were projected onto the eigenvectors. The Vietnamese cohort, which was sampled from the Vietnamese Kinh group, was homogenous and clustered with the 1000G East Asian populations (Figure S2). By contrast, the families in the Val-de-Marne sample showed genetic diversity at the population level with a majority of individuals of European, North African and Sub-Saharan ancestries (Figure S3). The admixed SAC subjects, who exhibited genetic diversity at the individual level, were forming a distinct cluster close to the African populations of 1000G (Figure S4).

Definition of the *M. tuberculosis* infection phenotype

The definition of the *M. tuberculosis* infection phenotype relied on both TST and IGRA results. In particular, we used a 5 mm cut-off to determine TST status, based on previous studies in similar settings^{7,17} and published guidelines³¹. We explored covariates associated with our infection definition in the entire cohort of enrolled individuals and the subset of those with genotype information were retained for the GWAS (Figure S1).

For the Vietnamese study, resistance to *M. tuberculosis* infection was defined by the presence of a negative TST < 5mm and a negative QFT-GIT test result following a protocol suggested by the manufacturer (see Supplementary Methods). Infected individuals were defined as subjects presenting both a positive TST $\geq$ 5mm and a positive QFT-GIT test result. A total of 188 subjects were classified as double negative and 512 as double positive (Figure S5), among which 185 and 201 subjects were genotyped, respectively. We investigated covariates associated with our infection definition and no significant association was found (Table S1). Therefore, we conducted an unadjusted genetic association analysis of our binary infection phenotype. In order to increase the sample size, we also added 152 genotyped PTB patients,

consisting of 146 index cases and 6 subjects with a history of PTB, to the infected group. (Table S2).

For the French study, contacts could have had one or 2 screening visits (V1 and V2) with a TST measurement (Figure S6A). A TST was considered negative when the skin induration was i) < 5 mm at both V1 and V2, ii) < 5 mm at V1, when only one visit was done. A TST was considered positive when the skin induration was i) ≥ 5 mm at both V1 and V2, ii) < 5 mm at V1 and ≥ 10 mm at V2, which reflected true conversions. In-house IGRA was used in this study and provided quantitative levels of IFN- γ production upon early secretory antigenic target 6 (ESAT6) stimulation (Supplemental Methods). A negative IGRA result was defined by a null production of IFN- γ . To determine the optimal positivity cut-off, we built a receiver operating characteristic (ROC) curve with TST status as the observed outcome and the corrected IFN- γ levels (ESAT6 response minus non-stimulated control value) as the predicted outcome among all the contacts enrolled. We selected as positivity threshold the highest sum of sensitivity plus specificity, which was equal to 175 pg/mL (Figure S6B). Then, we defined uninfected subjects as HHCs with a negative TST and a null IFN- γ production ($n = 33$) and infected subjects as HHCs who presented both a positive TST and a positive IGRA result (IFN- γ production > 175 pg/mL) ($n = 147$) (Figure S6C). We also looked for covariates associated with our infection definition in this sample of 180 individuals (Table S3). Age was the only factor significantly associated with our infection definition and was included as covariate in the genetic association analysis that finally included 30 genotyped uninfected and 127 infected subjects (Table S4).

For the South African study, a 5 mm cut-off was used to identify negative and positive TSTs (Figure S7A). In-house IGRA was also based on the production of IFN- γ upon ESAT6 stimulation (Supplemental Methods). We determined the optimal IGRA positivity cut-off similarly to the French study by building a ROC curve, leading to a threshold of 20.9 pg/mL (Figure S7B). Then, we defined uninfected subjects as HHCs with a negative TST (< 5 mm)

and a null IFN- γ production ($n = 128$), and infected subjects as those with both positive TST and IGRA result (IFN- γ production > 20.9 pg/mL) ($n = 152$) (Figure S7C). Age was the only factor significantly associated with our infection definition in the whole cohort (Table S5) and was included as covariate in the genetic association analysis that finally included 118 genotyped uninfected and 136 infected subjects (Table S6).

Covariates associated with our infection definition were investigated using mixed-effects logistic regression with a random effect per family in each cohort. All the analyses were carried out using R software (version 3.5.2) and related packages “pROC” and “lme4”^{32,33,34}.

Genetic association analyses

We conducted genetic association analyses of uninfected vs. infected subjects in the 3 cohorts using a linear mixed-model (LMM) assuming an additive genetic model as implemented in GEMMA v0.98³⁵. To account for the familial relationships, a genetic relatedness matrix (GRM) was used as random effects. In each cohort, the GRM was estimated using centered genotypes after the quality control described above. P values from the likelihood ratio test were reported. For better interpretability, we reported odds ratios (OR) and their 95% confidence intervals (95%CI) after transforming the regression coefficients of the LMM³⁶. Manhattan plots of the $-\log_{10}(P \text{ values})$ and quantile-quantile (QQ) plots were generated using “CMplot” package in R³⁷. Regional plots were generated using LocusZoom Standalone v1.4³⁸. Haplotype plots were generated using Haploview³⁹. Replication of genome-wide associated variants ($P < 5 \times 10^{-8}$) in the primary cohort from Vietnam was assessed in the 2 cohorts from France and South Africa. We considered variants at a nominal one sided P value < 0.05 and with a consistent direction of the effect size as replicated.

We also conducted a trans-ethnic meta-analysis by using summary statistics (i.e. beta estimates and their standard errors) from the Vietnamese discovery cohort and the 2 replication datasets.

We used the random-effects model of Han and Eskin implemented in METASOFT⁴⁰. This model assumes effect sizes of exactly zero in all the studies (i.e. no heterogeneity) under the null hypothesis of no associations and allows the effect sizes to vary among studies (i.e. heterogeneity) under the alternative hypothesis. The effect size consistency across studies were determined using the Cochran's Q statistic. Allelic effect estimates were also derived on the log-odds scale.

Functional annotation

We used the UCSC Genome Browser⁴¹ to identify associated variants which may overlap with known regulatory regions: 1) histone marks from the ENCODE project⁴², 2) chromatin state annotated by ChromHMM on the basis of ROADMAP^{43,44} and 3) chromatin accessibility determined by assay for transposase-accessible chromatin using sequencing (ATAC-seq) from immune cell-types⁴⁵. We also looked at the associated variants in expression quantitative trait loci (eQTL) databases which focused on gene expression in monocytes^{46,47}, T cells⁴⁸, macrophages⁴⁹, and various types of immune cells⁵⁰.

Results

Genome-wide association study of resistance to *M. tuberculosis* infection in Vietnam

First, we conducted a GWAS in the Vietnamese sample using 185 uninfected and 353 infected subjects, consisting of 201 individuals with positive TST/QFT-GIT results and 152 PTB patients. A total of 5,591,951 high quality variants were tested with a genomic inflation factor (λ) at 0.997, suggesting that effects from the familial study design were well controlled (Figure S8). The corresponding Manhattan plot is shown in Figure 1. We observed a genome-wide

significant association on chromosome 10q26.2, corresponding to a cluster of 12 variants and 6 additional variants in high linkage disequilibrium (LD) with $P < 5.0 \times 10^{-7}$ in the intronic regions and upstream of *C10orf90* (or *FATS*, HGNC: 26563) (Figure 2). The top-associated variant was the imputed rs11245088 (OR = 0.42, 95%CI 0.39-0.45, $P = 1.58 \times 10^{-8}$) while the top-associated genotyped variant was rs17155120 ($P = 3.71 \times 10^{-8}$) (Table 1). Each copy of the minor allele T of rs17155120 conferred protection against *M. tuberculosis* infection with an OR of being infected for CT vs CC or TT vs CT at 0.42 (95%CI 0.35-0.49) (Figure 3). The intensity cluster plot for rs17155120 showed that the genotype calling was of high quality and separated clearly into 3 genotype groups (Figure S9). Since all 18 variants in the locus were in high LD (Figure S10), the imputed variants were likely to have a high imputation quality as suggested by their info score (Table S7).

We also performed a GWAS between the 185 uninfected and the 201 infected subjects, excluding the 152 PTB patients. Despite a smaller sample size, all the 18 variants of the locus were still associated with protection against infection with $P < 5.0 \times 10^{-6}$, with similar ORs (for rs17155120, OR = 0.40, 95%CI 0.32-0.49, $P = 2.55 \times 10^{-7}$) (Table S8). Similar findings were also observed when considering only the 152 PTB patients as infection reference, with an OR for rs17155120 estimated at 0.50 (95%CI 0.41-0.59, $P = 2.10 \times 10^{-4}$). These results indicate that PTB patients are an appropriate infection reference group in this analysis.

Replication of variants associated with resistance to *M. tuberculosis* infection in France and South Africa

We tested the effects of the variants of the 10q26.2 locus in a French multi-ethnic HHC cohort (30 uninfected vs. 157 infected subjects) and an admixed familial sample from South Africa (118 uninfected vs. 136 infected subjects). In the French cohort, 17 variants of the cluster could be genotyped or imputed, and 9 were replicated at the $P < 0.05$ level with effect sizes in the

same direction as in the Vietnamese cohort. The most significant associated variant was rs56106518 (OR = 0.40, 95%CI 0.30-0.51, $P = 2.98 \times 10^{-3}$) (Table 1). In the South African population, 12 variants of the locus were genotyped or successfully imputed and 9 of them showed evidence for replication. The most significant associated variant was rs118037357 (OR = 0.55, 95%CI 0.43-0.67, $P = 7.38 \times 10^{-3}$) (Table 1). The top genotyped variant rs17155120 was also replicated in the 2 cohorts ($P_{France} = 1.51 \times 10^{-2}$ and $P_{SouthAfrica} = 1.74 \times 10^{-2}$) (Table 1 and Figure 4).

Interestingly, the variants across the 2 replication cohorts were in high LD with each other presenting similar LD patterns as in Vietnam (Figure S11). The frequencies of the rs17155120 T allele for all 3 cohorts were also similar, ranging from 0.16 to 0.20. The rs17155120 T allele frequency of 0.20 in the Vietnamese cohort was consistent with the Kinh allele frequency of 0.18 in 1000G phase 3 (Figure S12 and Table S9). The rs17155120 T allele frequency of 0.18 in the multi-ethnic French cohort was also close to the global frequency of 0.16 in all 1000G populations. However, the frequency of 0.16 for rs17155120 T allele in the SAC cohort was higher than in any 1000G African population that ranged from 0.05 to 0.11. Discrepancy of allele frequencies between SAC subjects and African populations of 1000G was confirmed by different patterns of LD in a 50 kb region around rs17155120 (Figure S13), which could be explained by the specific ethnic origin of the SAC subjects.

Trans-ethnic meta-analysis

Next, we carried out a meta-analysis of resistance to *M. tuberculosis* infection using the GWAS data of the 3 cohorts. The combined analysis was carried out in 333 uninfected and 616 infected subjects. The only genome-wide significant result was observed with the variants of the 10q26.2 region (Figure S14). The most significant signal was observed at the genotyped variant

rs17155120 (OR = 0.50, 95%CI = 0.45-0.55, $P = 1.26 \times 10^{-9}$) and no heterogeneity was observed across the 3 studies ($P_{\text{het}} = 0.31$) (Figure 4, Table S10).

Functional annotation

The variants of the 10q26.2 region map to the intronic regions and upstream of the tumor suppressor gene *C10orf90* (or *FATS*, for Fragile-site Associated Tumor Suppressor, HGNC: 26563) (Figure 5). According to ENCODE, 2 associated intronic variants, rs28703703 and rs77513326, are located in a regulatory genomic region characterized by H3K4me1 and H3K27ac histone marks and an active enhancer signature in lymphocytes T helper 17 (Th17). The variant rs77513326 also overlaps ATAC peaks in Th17 cells, memory T cells, natural killer cells and CD8⁺ T cells. We further explored the variants in various eQTLs databases of relevant tissues for the phenotype and found an association between them and expression of the nearby gene *ADAM12*. In particular, rs28703703 and the genotyped variant rs17155120 displayed decreasing expression of *ADAM12* with each minor allele (having a protective effect against *M. tuberculosis* infection) in monocytes ($P = 2.10 \times 10^{-3}$ and $P = 4.70 \times 10^{-3}$ respectively) (Figure S15). No association was observed in other immune cell types.

Discussion

In this study, we explored the genetic determinants of natural resistance to *M. tuberculosis* infection after intense exposure. There are no direct tests for established infection because TST and IGRAs measure an immune response that does not allow to distinguish past or present infection with *M. tuberculosis* bacilli. Nevertheless, our results show a strong genetic effect on resistance to *M. tuberculosis* infection irrespective of infection being persistent or temporary. TST and IGRA tests have well-known limits⁴. To minimize misclassification of uninfected and infected subjects in a cross-sectional setting, we therefore relied on negative results in both tests

and defined stringent cut-offs. As exposure to *M. tuberculosis* is difficult to quantitate, yet a critical feature of TB studies⁵¹, we focused on individuals at high risk of infection. We recruited household contacts recently exposed for extended periods to a contagious PTB index, some of whom remained infection-free. We discovered a genome-wide significant association between a cluster of variants on chromosome 10q26.2 and resistance to *M. tuberculosis* infection in well-characterized Vietnamese subjects. Strikingly, this locus could be replicated in 2 independent cohorts with different epidemiological settings from France and South Africa, resulting in association of the C/T variant rs17155120 across the 3 populations at an estimated combined odds ratio of 0.50 (95%CI 0.45-0.55) for becoming infected for TC vs CC or TT vs TC individuals.

The cluster of associated variants overlaps intronic and 5' regions of *C10orf90* (or *FATS*, HGNC: 26563), a tumor suppressor gene. This gene lies within a common-fragile site, which is an evolutionarily conserved region among mammals and susceptible to DNA damage⁵². The protein encoded by *C10orf90* has been shown to promote p53 activation in response to DNA damage through an E3 ubiquitin ligase activity^{53,54}. Several E3 ubiquitin ligases have already been shown to participate in the defense against *M. tuberculosis*, in particular through autophagy^{55,56,57}. The p53 transcription factor is not only a master regulator of autophagy but also activates apoptosis, another key process for host infected cells to limit the spread of pathogens as *M. tuberculosis*^{58,59}. Recent studies demonstrated that p53-induced apoptosis plays a critical role in the inhibition of mycobacteria survival and the macrophage resistance, possibly mediated by IL-17^{60,61}. However, possible role of *C10orf90* during *M. tuberculosis* infection or any related immune process remains unknown.

In silico functional annotation of the 10q26.2 associated variants revealed additional regulatory features related to lymphocytes Th17. In particular, the A/G variant rs28703703 was identified as a likely cis-eQTL for *ADAM12* in monocytes, with decreased expression of *ADAM12*

associated with the G allele, protective against *M. tuberculosis* infection. *ADAM12* is located ~280 kb downstream *C10orf90*, and encodes a matrix metalloprotease linked to a broad range of biological processes⁶². *ADAM12* expression correlates with lung inflammation as it is overexpressed in cells issued from asthmatic sputum and in the airway epithelium during allergic inflammatory reaction^{63,64}. It was recently found to be expressed by alveolar macrophages which are among the first innate immunity defense cell encountering *M. tuberculosis* in the lung (unpublished data, E. S.). Previous studies also reported *ADAM12* expression in Th17 cells^{65,66} and *ADAM12* knockdown in human T cells was found to increase Th17 cytokine production (IL-17A, IL-17F, and IL-22)⁶⁵. In South India, TST-negative individuals produced significantly higher levels of Th17 cytokines than TST-positive individuals⁶⁷. Similarly, significant higher levels of Th17 cytokines were observed in persistent negative IGRA individuals as compared with IGRA converters in a recent study conducted in the Gambia⁶⁸. These findings are consistent with our present results showing that the rs28703703 G allele protective against *M. tuberculosis* infection is associated with lower *ADAM12* expression that could lead to higher Th17 cytokine production. Overall, these observations support the view that Th17 cytokines may have a protective role against early stages of *M. tuberculosis* infection.

The 3 enrolled samples were of modest size. However, the design of the 3 studies, and the definition of a stringent and homogenous phenotype across them enabled us to detect a significant association with large effects. Interestingly, adding the PTB Vietnamese patients, who are by definition infected, increased the power of the analysis. Strikingly, the top associated variants displayed large effects with similar allele frequencies across populations and were in high LD across multiple ancestries (Figures S9 and S10). This observation was not expected because of the various epidemiological settings and the genetic diversity of the cohorts that included Kinh Vietnamese, Europeans, North Africans, Sub-Saharan Africans and admixed

Western Cape Coloureds. Such populations are usually under-represented in studies of genetic association while they could provide valuable insights to understand complex diseases⁶⁹.

In conclusion, we demonstrated that rigorous epidemiological design and phenotype definition with seemingly limited sample sizes can reveal novel genetic factors that offer protection from major pathogens such as *M. tuberculosis*. We found that *C10orf90* and *ADAM12* are promising candidate genes involved in the natural resistance to *M. tuberculosis*. Further investigations are needed to elucidate their role in the process of the initial infection, which could be a major step to provide new opportunities in the fight against TB.

Supplemental Data

Supplemental Data include supplemental methods, 15 figures and 10 tables.

Declaration of Interests

The authors declare no competing interests.

Acknowledgements

This work was supported in part by the French Foundation for Medical Research (FRM) (EQU201903007798 ; FDM20170637664 to J.Q.), the Programme Hospitalier de Recherche Clinique (AOR-04-003); the Legs Poix (Chancellerie des Universités de Paris), the French National Research Agency (ANR) under the “Investments for the future” program (grant ANR-10-IAHU-01), the Integrative Biology of Emerging Infectious Diseases Laboratory of Excellence (ANR-10-LABX-62-IBEID), and the TBPATHEGEN project (ANR-14-CE14-0007-01), the SCOR Corporate Foundation for Science, the Rockefeller University, the Institut National de la Santé et de la Recherche Médicale (INSERM), Université de Paris, the St. Giles

Foundation, the Canadian Institutes of Health Research (CIHR, FDN-143332) and the Sequella/Aeras Global Tuberculosis Foundation.

We are grateful for the skilled technical assistance of Ho thi Ngoc, Binh Duong Center for Social Diseases Control, and Hanh La, Hospital for Dermato-Venereology, Ho Chi Minh City. We are indebted to the laboratory staff of the Hospital for Dermato-Venereology, Ho Chi Minh City, and the tuberculosis control teams of the Thu dau Mot, Thuận An, Ben Cat, and Di An districts. We sincerely thank all members of the community who participated in this work in South Africa and France and members of the lab of Human Genetics of Infectious Diseases for helpful discussions.

Web Resources

1000 Genomes Phase 3: <ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp>

Michigan Imputation Server : <https://imputationserver.sph.umich.edu/index.html>

Sanger Imputation Server : <https://imputation.sanger.ac.uk>

Softwares: PLINK, <https://www.cog-genomics.org/plink> ; R, <https://www.r-project.org> ; CMplot, <https://github.com/YinLiLin/R-CMplot> ; GEMMA, <https://github.com/genetics-statistics/GEMMA> ; Haploview, <https://www.broadinstitute.org/haploview/haploview> ; LocusZoom Standalone v1.4, https://genome.sph.umich.edu/wiki/LocusZoom_Standalone ; METASOFT, <http://genetics.cs.ucla.edu/meta/index.html>

Integrative Genomics Viewer: <https://igv.org/>

UCSC Genome Browser: <https://genome.ucsc.edu>

eQTLs: for monocytes, <http://immunpop.com/kim/eQTL>, <http://www.immunpop.com> ; for T cells, <https://genenetwork.nl/cd4cd8eqtlbrowser> ; for macrophages, <https://zenodo.org/communities/macrophage-gene-expression-genetics> ; Database of Immune

Cell Expression, Expression quantitative trait loci and Epigenomics, <https://dice-database.org/landing>

Data and Code Availability

GWAS summary statistics are available at: [10.5281/zenodo.3942126](https://zenodo.org/record/3942126)

References

1. WHO | Global tuberculosis report 2019.
2. Verrall, A.J., Alisjahbana, B., Apriani, L., Novianty, N., Nurani, A.C., van Laarhoven, A., Ussher, J.E., Indrati, A., Ruslami, R., Netea, M.G., et al. (2019). Early Clearance of Mycobacterium tuberculosis: The INFECT Case Contact Cohort Study in Indonesia. *J. Infect. Dis.* 221, 1351–1360.
3. Simmons, J.D., Stein, C.M., Seshadri, C., Campo, M., Alter, G., Fortune, S., Schurr, E., Wallis, R.S., Churchyard, G., Mayanja-Kizza, H., et al. (2018). Immunological mechanisms of human resistance to persistent Mycobacterium tuberculosis infection. *Nat. Rev. Immunol.* 18, 575–589.
4. Sharma, S.K., Vashishtha, R., Chauhan, L.S., Sreenivas, V., and Seth, D. (2017). Comparison of TST and IGRA in Diagnosis of Latent Tuberculosis Infection in a High TB-Burden Setting. *PLoS ONE* 12, e0169539.
5. Pai, M., and Menzies, D. (2007). The new IGRA and the old TST: making good use of disagreement. *Am. J. Respir. Crit. Care Med.* 175, 529–531.
6. Gallant, C.J., Cobat, A., Simkin, L., Black, G.F., Stanley, K., Hughes, J., Doherty, T.M., Hanekom, W.A., Eley, B., Beyers, N., et al. (2010). Tuberculin skin test and in vitro assays provide complementary measures of antimycobacterial immunity in children and adolescents. *Chest* 137, 1071–1077.
7. Mave, V., Chandrasekaran, P., Chavan, A., Shivakumar, S.V.B.Y., Danasekaran, K., Paradkar, M., Thiruvengadam, K., Kinikar, A., Murali, L., Gaikwad, S., et al. (2019). Infection free “resisters” among household contacts of adult pulmonary tuberculosis. *PLoS ONE* 14, e0218034.
8. Hanifa, Y., Grant, A.D., Lewis, J., Corbett, E.L., Fielding, K., and Churchyard, G. (2009). Prevalence of latent tuberculosis infection among gold miners in South Africa. *Int. J. Tuberc. Lung Dis.* 13, 39–46.
9. Stein, C.M., Zalwango, S., Malone, L.L., Thiel, B., Mupere, E., Nsereko, M., Okware, B., Kisingo, H., Lancioni, C.L., Bark, C.M., et al. (2018). Resistance and Susceptibility to Mycobacterium tuberculosis Infection and Disease in Tuberculosis Households in Kampala, Uganda. *Am. J. Epidemiol.* 187, 1477–1489.

10. Abel, L., Fellay, J., Haas, D.W., Schurr, E., Srikrishna, G., Urbanowski, M., Chaturvedi, N., Srinivasan, S., Johnson, D.H., and Bishai, W.R. (2018). Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. *Lancet Infect Dis* 18, e64–e75.
11. Cobat, A., Gallant, C.J., Simkin, L., Black, G.F., Stanley, K., Hughes, J., Doherty, T.M., Hanekom, W.A., Eley, B., Jaïs, J.-P., et al. (2009). Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. *J. Exp. Med.* 206, 2583–2591.
12. Jabot-Hanin, F., Cobat, A., Feinberg, J., Grange, G., Remus, N., Poirier, C., Boland-Auge, A., Besse, C., Bustamante, J., Boisson-Dupuis, S., et al. (2016). Major Loci on Chromosomes 8q and 3q Control Interferon γ Production Triggered by *Bacillus Calmette-Guerin* and 6-kDa Early Secretory Antigen Target, Respectively, in Various Populations. *J. Infect. Dis.* 213, 1173–1179.
13. Jabot-Hanin, F., Cobat, A., Feinberg, J., Orlova, M., Niay, J., Deswarte, C., Poirier, C., Theodorou, I., Bustamante, J., Boisson-Dupuis, S., et al. (2017). An eQTL variant of ZXDC is associated with IFN- γ production following *Mycobacterium tuberculosis* antigen-specific stimulation. *Sci Rep* 7, 1–8.
14. Stein, C.M., Zalwango, S., Malone, L.L., Won, S., Mayanja-Kizza, H., Mugerwa, R.D., Leontiev, D.V., Thompson, C.L., Cartier, K.C., Elston, R.C., et al. (2008). Genome scan of *M. tuberculosis* infection and disease in Ugandans. *PLoS ONE* 3, e4094.
15. Igo, R.P., Hall, N.B., Malone, L.L., Hall, J.B., Truitt, B., Qiu, F., Tao, L., Mupere, E., Schnell, A., Hawn, T.R., et al. (2019). Fine-mapping Analysis of a Chromosome 2 Region Linked to Resistance to *Mycobacterium tuberculosis* Infection in Uganda Reveals Potential Regulatory Variants. *Genes Immun* 20, 473–483.
16. Cobat, A., Poirier, C., Hoal, E., Boland-Auge, A., de La Rocque, F., Corrad, F., Grange, G., Migaud, M., Bustamante, J., Boisson-Dupuis, S., et al. (2015). Tuberculin skin test negativity is under tight genetic control of chromosomal region 11p14-15 in settings with different tuberculosis endemicities. *J. Infect. Dis.* 211, 317–321.
17. Sobota, R.S., Stein, C.M., Kodaman, N., Maro, I., Wieland-Alter, W., Igo, R.P., Magohe, A., Malone, L.L., Chervenak, K., Hall, N.B., et al. (2017). A chromosome 5q31.1 locus associates with tuberculin skin test reactivity in HIV-positive individuals from tuberculosis hyper-endemic regions in east Africa. *PLoS Genet.* 13, e1006710.
18. WHO | Global tuberculosis report 2015.
19. WHO | Immunization Country Profile.
20. World Bank. Prevalence of HIV, total (% of population ages 15-49) - Vietnam | Data. (access <https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?locations=VN>)
21. Antoine, D. (2008) Les cas de tuberculose maladie déclarés en France en 2006. *Bull Epidemiol Hebd.* 10-11
22. de Wit, E., Delport, W., Rugamika, C.E., Meintjes, A., Möller, M., van Helden, P.D., Seoighe, C., and Hoal, E.G. (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. *Hum. Genet.* 128, 145–153.

23. Naidoo, P., Theron, G., Rangaka, M.X., Chihota, V.N., Vaughan, L., Brey, Z.O., and Pillay, Y. (2017). The South African Tuberculosis Care Cascade: Estimated Losses and Methodological Challenges. *J Infect Dis* 216, S702–S713.
24. Mathema, B., Andrews, J.R., Cohen, T., Borgdorff, M.W., Behr, M., Glynn, J.R., Rustomjee, R., Silk, B.J., and Wood, R. (2017). Drivers of Tuberculosis Transmission. *J Infect Dis* 216, S644–S653.
25. Hesselning, A.C., Caldwell, J., Cotton, M.F., Eley, B.S., Jaspan, H.B., Jennings, K., Marais, B.J., Nuttall, J., Rabie, H., Roux, P., et al. (2009). BCG vaccination in South African HIV-exposed infants – risks and benefits. *S Afr Med J* 99, 88–91.
26. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., 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.
27. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. *Nat. Genet.* 48, 1284–1287.
28. Loh, P.-R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* 48, 1443–1448.
29. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74.
30. McCarthy, S., Das, S., Kretzschmar, W., Delaneau, O., Wood, A.R., Teumer, A., Kang, H.M., Fuchsberger, C., Danecek, P., Sharp, K., et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* 48, 1279–1283.
31. (2000). Diagnostic Standards and Classification of Tuberculosis in Adults and Children. This official statement of the American Thoracic Society and the Centers for Disease Control and Prevention was adopted by the ATS Board of Directors, July 1999. This statement was endorsed by the Council of the Infectious Disease Society of America, September 1999. *Am. J. Respir. Crit. Care Med.* 161, 1376–1395.
32. R: The R Project for Statistical Computing.
33. Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., and Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 12, 77.
34. Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R.H.B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., and Green, P. (2018). lme4: Linear Mixed-Effects Models using “Eigen” and S4.
35. Zhou, X., and Stephens, M. (2012). Genome-wide Efficient Mixed Model Analysis for Association Studies. *Nat Genet* 44, 821–824.

36. Lloyd-Jones, L.R., Robinson, M.R., Yang, J., and Visscher, P.M. (2018). Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics* 208, 1397–1408.
37. GitHub - YinLiLin/CMplot: Circular and Rectangular Manhattan Plot.
38. 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.
39. Barrett, J.C., Fry, B., Maller, J., and Daly, M.J. (2005). Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–265.
40. Han, B., and Eskin, E. (2011). Random-Effects Model Aimed at Discovering Associations in Meta-Analysis of Genome-wide Association Studies. *Am J Hum Genet* 88, 586–598.
41. Haeussler, M., Zweig, A.S., Tyner, C., Speir, M.L., Rosenbloom, K.R., Raney, B.J., Lee, C.M., Lee, B.T., Hinrichs, A.S., Gonzalez, J.N., et al. (2019). The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* 47, D853–D858.
42. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74.
43. Ernst, J., and Kellis, M. (2017). Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 12, 2478–2492.
44. Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al. (2015). Integrative analysis of 111 reference human epigenomes. *Nature* 518, 317–330.
45. Calderon, D., Nguyen, M.L.T., Mezger, A., Kathiria, A., Müller, F., Nguyen, V., Lescano, N., Wu, B., Trombetta, J., Ribado, J.V., et al. (2019). Landscape of stimulation-responsive chromatin across diverse human immune cells. *Nat Genet* 51, 1494–1505.
46. Kim-Hellmuth, S., Bechheim, M., Pütz, B., Mohammadi, P., Nédélec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E., et al. (2017). Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat Commun* 8, 1–10.
47. Nédélec, Y., Sanz, J., Baharian, G., Szpiech, Z.A., Pacis, A., Dumaine, A., Grenier, J.-C., Freiman, A., Sams, A.J., Hebert, S., et al. (2016). Genetic Ancestry and Natural Selection Drive Population Differences in Immune Responses to Pathogens. *Cell* 167, 657–669.e21.
48. Kasela, S., Kisand, K., Tserel, L., Kaleviste, E., Remm, A., Fischer, K., Esko, T., Westra, H.-J., Fairfax, B.P., Makino, S., et al. (2017). Pathogenic implications for autoimmune mechanisms derived by comparative eQTL analysis of CD4+ versus CD8+ T cells. *PLOS Genetics* 13, e1006643.
49. Alasoo, K., Rodrigues, J., Mukhopadhyay, S., Knights, A.J., Mann, A.L., Kundu, K., Hale, C., Dougan, G., and Gaffney, D.J. (2018). Shared genetic effects on chromatin and gene expression indicate a role for enhancer priming in immune response. *Nat Genet* 50, 424–431.

50. Schmiedel, B.J., Singh, D., Madrigal, A., Valdovino-Gonzalez, A.G., White, B.M., Zapardiel-Gonzalo, J., Ha, B., Altay, G., Greenbaum, J.A., McVicker, G., et al. (2018). Impact of Genetic Polymorphisms on Human Immune Cell Gene Expression. *Cell* 175, 1701-1715.e16.
51. Stein, C.M., Hall, N.B., Malone, L.L., and Mupere, E. (2013). The household contact study design for genetic epidemiological studies of infectious diseases. *Front Genet* 4, 61.
52. Ma, K., Qiu, L., Mrasek, K., Zhang, J., Liehr, T., Quintana, L.G., and Li, Z. (2012). Common Fragile Sites: Genomic Hotspots of DNA Damage and Carcinogenesis. *Int J Mol Sci* 13, 11974–11999.
53. Yan, S., Qiu, L., Ma, K., Zhang, X., Zhao, Y., Zhang, J., Li, X., Hao, X., and Li, Z. (2014). FATS is an E2-independent ubiquitin ligase that stabilizes p53 and promotes its activation in response to DNA damage. *Oncogene* 33, 5424–5433.
54. Song, F., Zhang, J., Qiu, L., Zhao, Y., Xing, P., Lu, J., Chen, K., and Li, Z. (2015). A functional genetic variant in fragile-site gene FATS modulates the risk of breast cancer in triparous women. *BMC Cancer* 15, 559.
55. Manzanillo, P.S., Ayres, J.S., Watson, R.O., Collins, A.C., Souza, G., Rae, C.S., Schneider, D.S., Nakamura, K., Shiloh, M.U., and Cox, J.S. (2013). The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature* 501, 512–516.
56. Franco, L.H., Nair, V.R., Scharn, C.R., Xavier, R.J., Torrealba, J.R., Shiloh, M.U., and Levine, B. (2017). The Ubiquitin Ligase Smurf1 Functions in Selective Autophagy of Mycobacterium tuberculosis and Anti-tuberculous Host Defense. *Cell Host Microbe* 21, 59–72.
57. Wang, L., Wu, J., Li, J., Yang, H., Tang, T., Liang, H., Zuo, M., Wang, J., Liu, H., Liu, F., et al. (2020). Host-mediated ubiquitination of a mycobacterial protein suppresses immunity. *Nature* 577, 682–688.
58. Kasthuber, E.R., and Lowe, S.W. (2017). Putting p53 in Context. *Cell* 170, 1062–1078.
59. Behar, S.M., Martin, C.J., Booty, M.G., Nishimura, T., Zhao, X., Gan, H.-X., Divangahi, M., and Remold, H.G. (2011). Apoptosis is an innate defense function of macrophages against Mycobacterium tuberculosis. *Mucosal Immunol* 4, 279–287.
60. Lim, Y.-J., Lee, J., Choi, J.-A., Cho, S.-N., Son, S.-H., Kwon, S.-J., Son, J.-W., and Song, C.-H. (2019). M1 macrophage dependent-p53 regulates the intracellular survival of mycobacteria. *Apoptosis* 5, 42-55.
61. Cruz, A., Ludovico, P., Torrado, E., Gama, J.B., Sousa, J., Gaifem, J., Appelberg, R., Rodrigues, F., Cooper, A.M., Pedrosa, J., et al. (2015). IL-17A Promotes Intracellular Growth of Mycobacterium by Inhibiting Apoptosis of Infected Macrophages. *Front. Immunol.* 6, 498.
62. Kveiborg, M., Albrechtsen, R., Couchman, J.R., and Wewer, U.M. (2008). Cellular roles of ADAM12 in health and disease. *Int. J. Biochem. Cell Biol.* 40, 1685–1702.

63. Paulissen, G., Rocks, N., Quesada-Calvo, F., Gosset, P., Foidart, J.-M., Noel, A., Louis, R., and Cataldo, D.D. (2006). Expression of ADAMs and Their Inhibitors in Sputum from Patients with Asthma. *Mol Med* 12, 171–179.
64. Estrella, C., Rocks, N., Paulissen, G., Quesada-Calvo, F., Noël, A., Vilain, E., Lassalle, P., Tillie-Leblond, I., Cataldo, D., and Gosset, P. (2009). Role of A Disintegrin And Metalloprotease-12 in Neutrophil Recruitment Induced by Airway Epithelium. *Am J Respir Cell Mol Biol* 41, 449–458.
65. Zhou, A.X., El Hed, A., Mercer, F., Kozhaya, L., and Unutmaz, D. (2013). The Metalloprotease ADAM12 Regulates the Effector Function of Human Th17 Cells. *PLoS ONE* 8,e81146.
66. Arlehamn, C.L., Seumois, G., Gerasimova, A., Huang, C., Fu, Z., Yue, X., Sette, A., Vijayanand, P., and Peters, B. (2014). Transcriptional profile of TB antigen-specific T cells reveals novel multifunctional features. *J Immunol* 193, 2931–2940.
67. Babu, S., Bhat, S.Q., Kumar, N.P., Kumaraswami, V., and Nutman, T.B. (2010). Regulatory T Cells Modulate Th17 Responses in Patients with Positive Tuberculin Skin Test Results. *J Infect Dis* 201, 20–31.
68. Coulter, F., Parrish, A., Manning, D., Kampmann, B., Mendy, J., Garand, M., Lewinsohn, D.M., Riley, E.M., and Sutherland, J.S. (2017). IL-17 Production from T Helper 17, Mucosal-Associated Invariant T, and $\gamma\delta$ Cells in Tuberculosis Infection and Disease. *Front. Immunol.* 8,1252.
69. Gurdasani, D., Barroso, I., Zeggini, E., and Sandhu, M.S. (2019). Genomics of disease risk in globally diverse populations. *Nat. Rev. Genet.* 20, 520–535.

Figure Titles and Legends

Figure 1. Manhattan plot of resistance to *M. tuberculosis* infection in Vietnam. Manhattan plot showing results from a genome-wide association study between 185 uninfected subjects (negative for both tuberculin skin test and QuantiFERON®-TB Gold In-Tube test) and 353 infected subjects (201 infected individuals positive for both tests and 152 patients with a history of pulmonary tuberculosis) for 5,591,951 variants (minor allele frequency > 5% and info > 0.8) with an unadjusted additive genetic model. The $-\log_{10}(P \text{ value})$ for each variant (y-axis) is presented according to its chromosomal position (x-axis, build hg19). The dashed line indicates the genome-wide significant threshold at $P = 5.0 \times 10^{-8}$.

Figure 2. Regional plot of resistance to *M. tuberculosis* infection in Vietnam. Locus zoom plot showing association for the 10q26.2 locus, in a 500 kb window surrounding the top imputed variant rs11245088 (purple diamond). Colors represent pairwise linkage disequilibrium (r^2) with rs11245088 as calculated for the Vietnamese Kinh population of 1000G phase 3.

Figure 3. Proportion of Vietnamese individuals resistant to *M. tuberculosis* infection by genotype for the variant rs17155120. Each bar represents the proportion of uninfected subjects among CC individuals (n= 93/344), CT individuals (n = 76/173) and TT individuals (n = 16/21) for the variant rs17155120 in Vietnam.

Figure 4. Forest plot of the association between an additive genetic effect of rs17155120 on chromosome region 10q26.2 and resistance to *M. tuberculosis* infection. Odds ratios and 95% confidence intervals derived from a linear mixed model, *P* values, sample sizes and frequency of the effect allele (EAF) are reported by individual cohort and for the random effects meta-analysis.

Figure 5. Genomic annotation of the locus on chromosome region 10q26.2. The upper panel is adapted from Integrative Genomics Viewer and the lower panel is adapted from UCSC Genome Browser. The 3 vertical grey lines represent the associated variants rs1715120, rs28703703 and rs77513326 (from left to right) that overlap regulatory regions. From top to bottom: H3K4me1 and H3K27ac histone marks from ENCODE, active enhancer in Th17 cells (chromHMM annotation from ROADMAP), chromatin accessibility as represented by ATAC peaks in Th17 cells, memory T cells, NK cells and CD8⁺ T cells⁵⁰.

Table 1. Association between an additive genetic effect of variants on chromosome region 10q26.2 and resistance to *M. tuberculosis* infection in Vietnam, France and South Africa under an additive genetic model

Variant	C10orf90	EA	Vietnam			France			South Africa		
			EAF	OR (95%CI)	P value	EAF	OR (95%CI)	P value	EAF	OR (95%CI)	P value
rs11245088	upstream	C	0.25	0.42 (0.39-0.45)	1.58×10^{-8}	-	-	-	0.44	0.99 (0.91-1.07)	4.75×10^{-1}
rs72163291	intron	ins	0.24	0.42 (0.35-0.49)	1.94×10^{-8}	0.37	0.88 (0.78-0.98)	3.54×10^{-1}	-	-	-
rs17155143	upstream	A	0.20	0.39 (0.31-0.47)	1.98×10^{-8}	0.18	0.63 (0.50-0.77)	1.29×10^{-1}	0.15	0.60 (0.48-0.72)	1.72×10^{-2}
rs7909756	upstream	G	0.24	0.42 (0.35-0.49)	2.03×10^{-8}	0.37	0.83 (0.73-0.93)	2.87×10^{-1}	0.36	0.90 (0.81-0.99)	2.77×10^{-1}
rs28703703	intron	G	0.20	0.41 (0.34-0.49)	2.52×10^{-8}	0.20	0.59 (0.48-0.69)	4.95×10^{-2}	0.17	0.66 (0.55-0.77)	2.95×10^{-2}
rs56106518	intron	C	0.23	0.41 (0.33-0.48)	3.07×10^{-8}	0.30	0.40 (0.30-0.51)	2.98×10^{-3}	-	-	-
rs75482972	intron	A	0.20	0.42 (0.34-0.49)	3.35×10^{-8}	0.20	0.58 (0.47-0.68)	4.38×10^{-2}	0.17	0.68 (0.57-0.78)	3.48×10^{-2}
rs17155120	intron	T	0.20	0.42 (0.35-0.49)	3.71×10^{-8}	0.18	0.48 (0.36-0.59)	1.51×10^{-2}	0.16	0.62 (0.52-0.73)	1.74×10^{-2}
rs73370887	intron	A	0.20	0.40 (0.33-0.48)	4.05×10^{-8}	0.31	0.78 (0.71-0.85)	1.43×10^{-1}	0.28	0.75 (0.67-0.83)	3.69×10^{-2}
rs79608098	intron	T	0.20	0.42 (0.35-0.49)	4.06×10^{-8}	0.20	0.58 (0.48-0.69)	4.50×10^{-2}	0.17	0.66 (0.55-0.76)	2.71×10^{-2}
rs61750007	upstream	C	0.24	0.43 (0.36-0.50)	4.30×10^{-8}	0.24	0.53 (0.42-0.64)	3.49×10^{-2}	0.20	0.73 (0.62-0.84)	7.62×10^{-2}
rs77513326	intron	A	0.20	0.42 (0.34-0.49)	4.93×10^{-8}	0.17	0.47 (0.36-0.59)	1.61×10^{-2}	0.16	0.63 (0.52-0.75)	2.35×10^{-2}
rs79918233	intron	A	0.20	0.41 (0.33-0.49)	5.61×10^{-8}	0.17	0.51 (0.39-0.63)	3.00×10^{-2}	0.16	0.63 (0.52-0.74)	1.95×10^{-2}
rs147584264	upstream	C	0.19	0.41 (0.33-0.49)	1.21×10^{-7}	0.22	0.77 (0.62-0.93)	3.00×10^{-1}	-	-	-
rs191820708	upstream	A	0.19	0.41 (0.33-0.49)	1.31×10^{-7}	0.19	0.63 (0.50-0.77)	1.40×10^{-1}	-	-	-
rs201178890	upstream	T	0.19	0.41 (0.33-0.49)	1.36×10^{-7}	0.22	0.79 (0.62-0.97)	3.36×10^{-1}	-	-	-
rs202189321	upstream	T	0.20	0.41 (0.33-0.49)	1.37×10^{-7}	0.18	0.65 (0.52-0.78)	1.44×10^{-1}	-	-	-
rs118037357	upstream	A	0.19	0.41 (0.33-0.49)	1.62×10^{-7}	0.15	0.47 (0.33-0.60)	2.83×10^{-2}	0.14	0.55 (0.43-0.67)	7.38×10^{-3}

CI, confidence intervals; EA, effect allele; EAF, effect allele frequency; OR, odds ratio