

Supplemental

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

Jocelyn Quistrebert, Marianna Orlova, Gaspard Kerner, Le Thi Ton, Nguyễn Trong Luong, Nguyễn Thanh Danh, Fabienne Jabot-Hanin, Yoann Seeleuthner, Jacinta Bustamante, Stéphanie Boisson-Dupuis, Nguyen Thu Huong, Nguyen Ngoc Ba, Jean-Laurent Casanova, Christophe Delacourt, Eileen G. Hoal, Alexandre Alcaïs, Vu Hong Thai, Lai The Thành, Laurent Abel, Erwin Schurr, Aurélie Cobat

Supplemental Methods

15 Supplemental Figures

10 Supplemental Tables

Supplemental Methods

Immune assays in Vietnam

For tuberculin skin test (TST), 5 units (TU) of purified protein derivative (PPD) solution per 0.1 mL test dose (Tubertest, Sanofi Pasteur, Lyon, France) were injected intradermally into the forearm of the household contacts (HHCs) using the Mantoux method. The diameter of the induration was measured by trained personal 48-72 hours after the injection. QuantiFERON®-TB Gold In-Tube test (QFT-GIT) assay was performed on blood samples collected before the TST was carried out, in accordance with the kit manufacturer's instructions. Briefly, peripheral blood was drawn directly into each of three blood collection tubes under vacuum: the nil tube or negative control, the mitogen tube containing phytohaemagglutinin (PHA) as a positive control and the TB antigen tube containing specific antigens for *M. tuberculosis* (early secretory antigenic target 6, ESAT6, culture filtrate protein 10 and TB 7.7 antigen). The contents of each tube were mixed and the tubes were incubated at 37°C for 16 to 24 hours within 8 hours of collection. Tubes were centrifuged at 2500 g for 15 minutes at room temperature, plasma was collected and stored at -20°C. The concentration of interferon-gamma (IFN- γ) in each sample was determined with the QFT-GIT ELISA (enzyme-linked immunosorbent assay) kit. Results were categorized by the manufacturer's software as positive, negative or indeterminate.

Immune assays in France

TSTs were performed on the volar surface of the forearm by intradermal injection of 0.1 mL (5 TU) of purified protein derivative tuberculin obtained from a human *M. tuberculosis* strain (Tubertest, Sanofi Pasteur, Lyon, France). The skin induration was read within 72 hours. IGRAs were performed once either at V1 or V2. Blood was drawn from each individual and peripheral blood mononuclear cells (PBMCs) were separated. They were activated with one of the following stimulations: ESAT-6 (rdESAT6, Statens-Serum-Institute, Denmark; 2.5 μ g/mL), PHA (6.25 μ g/mL) for positive control and medium alone for negative control. After 4 days of stimulation, supernatants were assayed for IFN- γ by ELISA according to the manufacturer's recommendations (Pelikin Compact; CLB). The final results were standardized per million PBMC's, in the unit pg/mL/10⁶ PBMC's.

Immune assays in South Africa

TSTs were performed by specially trained health care providers using the Mantoux method with PPD solution (2 T.U Tuberculin PPD RT 23, Statens Serum Institut, Copenhagen, Denmark) and skin induration was read 48-72 hours later. Whole-blood specimens were stimulated by ESAT-6 (rdESAT6, Statens-Serum-Institute, Denmark; 20µg/mL) and PHA (5 µg/mL) for positive control. Cultures incubated without stimulation in serum-free medium provided the negative control. IFN- γ levels were measured on day 3 after stimulation (limit of detection at 31 pg/mL).

Supplemental Figures

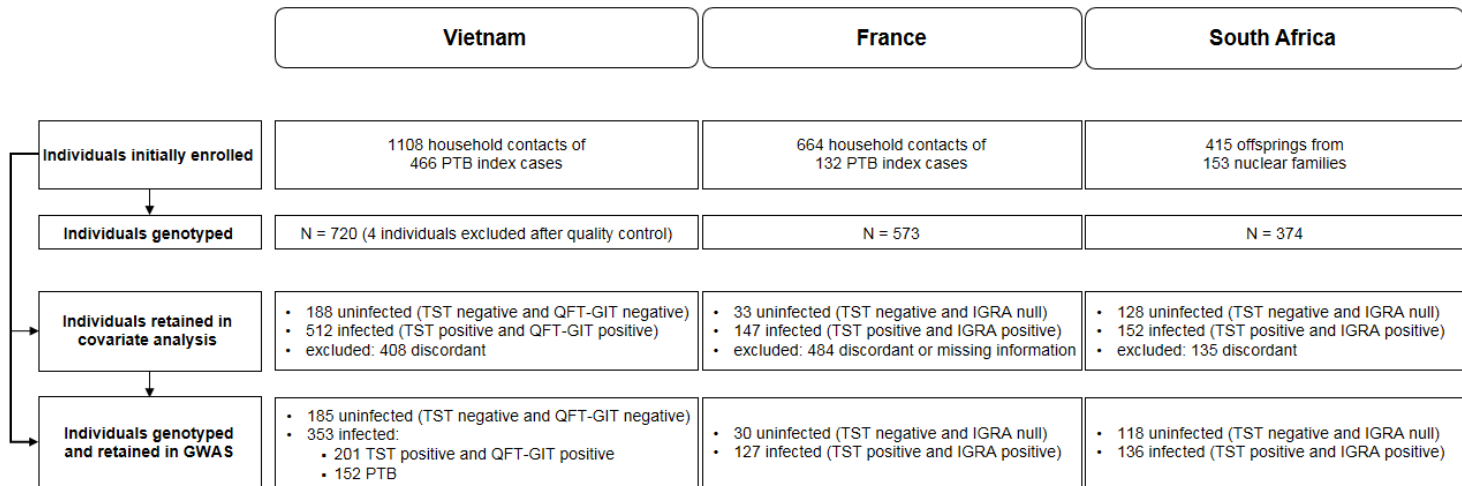
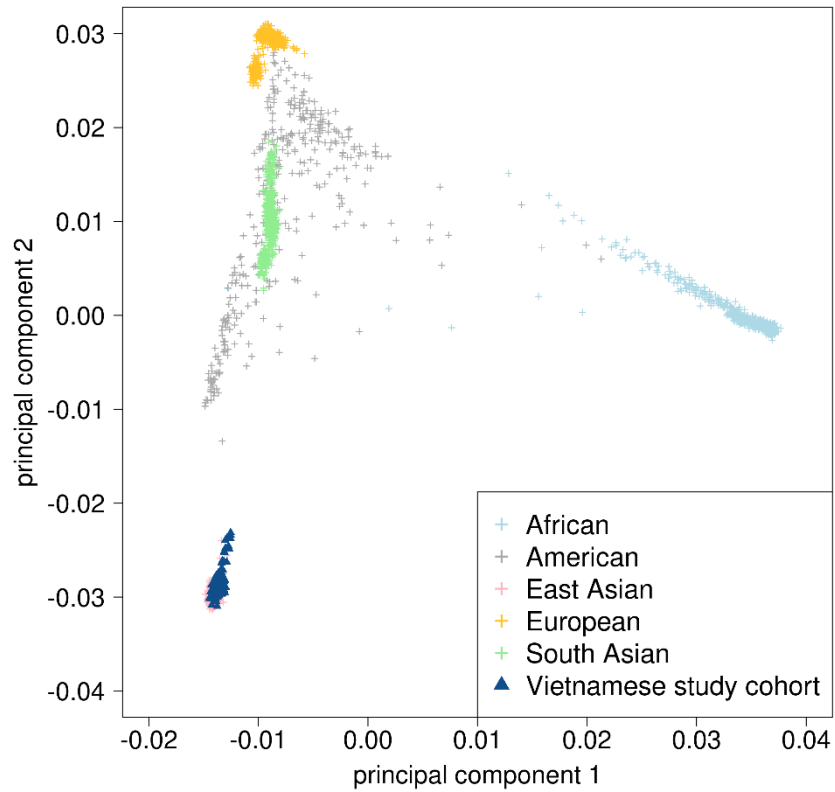


Figure S1. Flowchart of the study

A)



B)

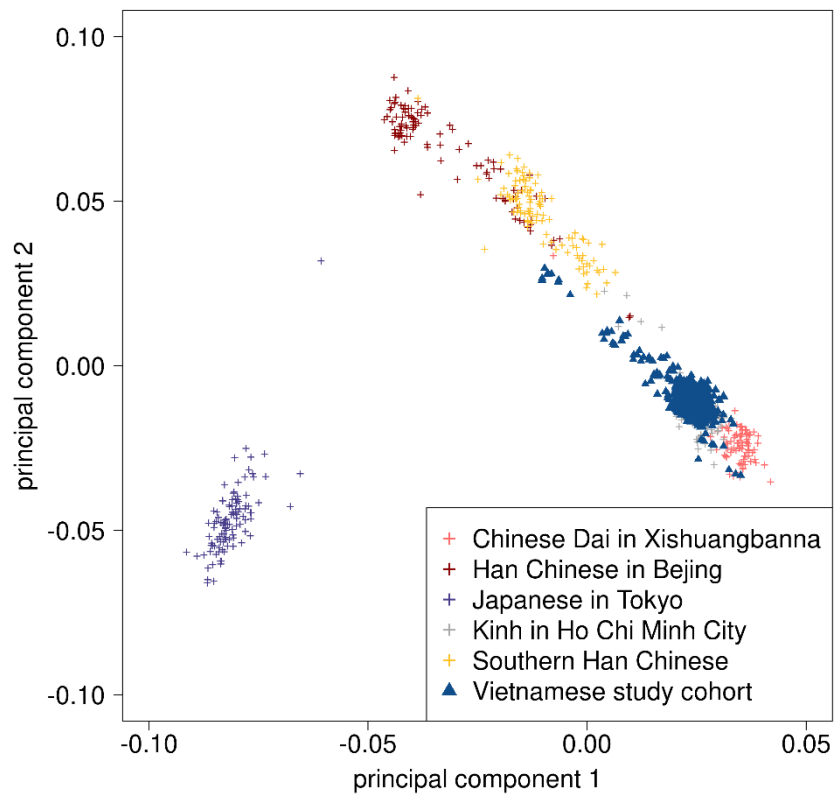
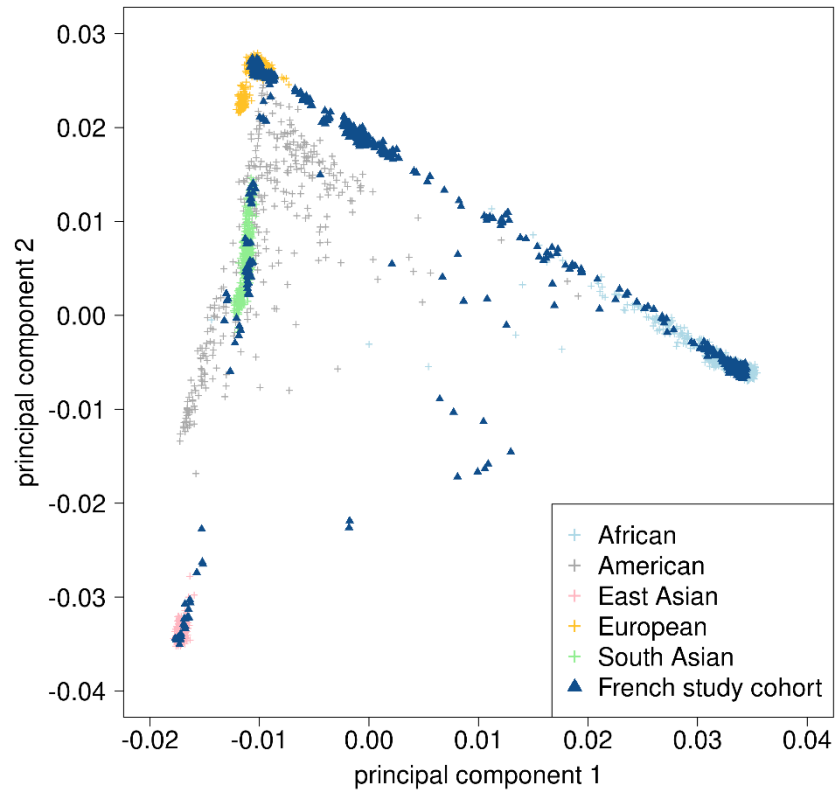


Figure S2. Principal component analysis of the Vietnamese study cohort. Plot of the first and second principal components of the 720 individuals from the Vietnamese cohort after projection A) on the 1000 Genomes phase 3 populations B) on the East Asian 1000 Genomes Phase 3 populations only.

A)



B)

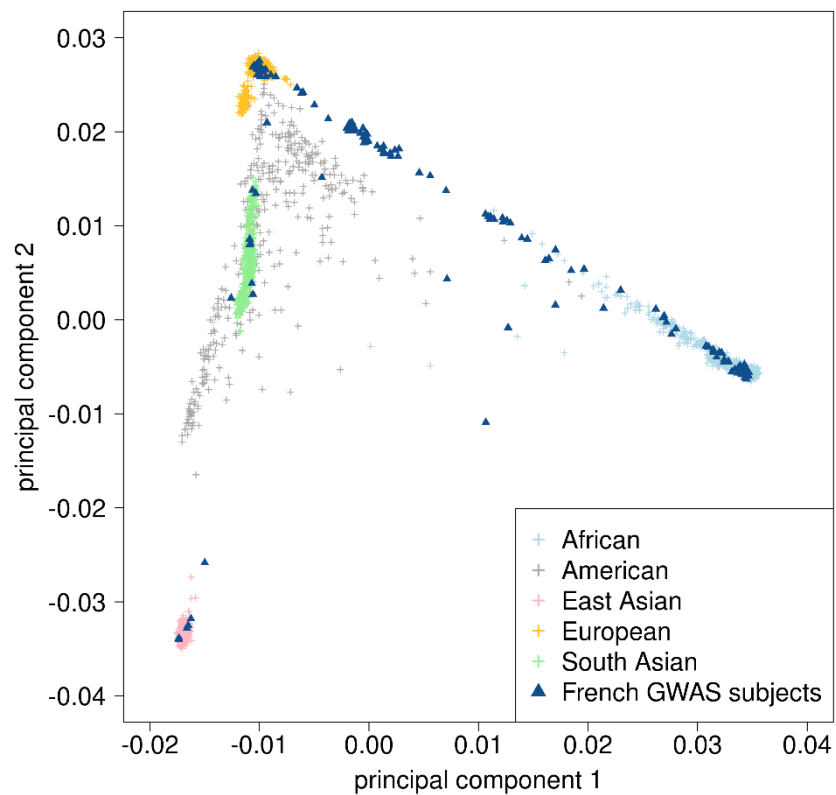


Figure S3. Principal component analysis of the French cohort. Plot of the first and second principal components of A) 573 individuals from the French cohort and B) the 157 subjects subsequently analyzed in the GWAS, after projection on the 1000 Genomes phase 3 populations. The majority of the GWAS subjects had European, North African and Sub-Saharan ancestries.

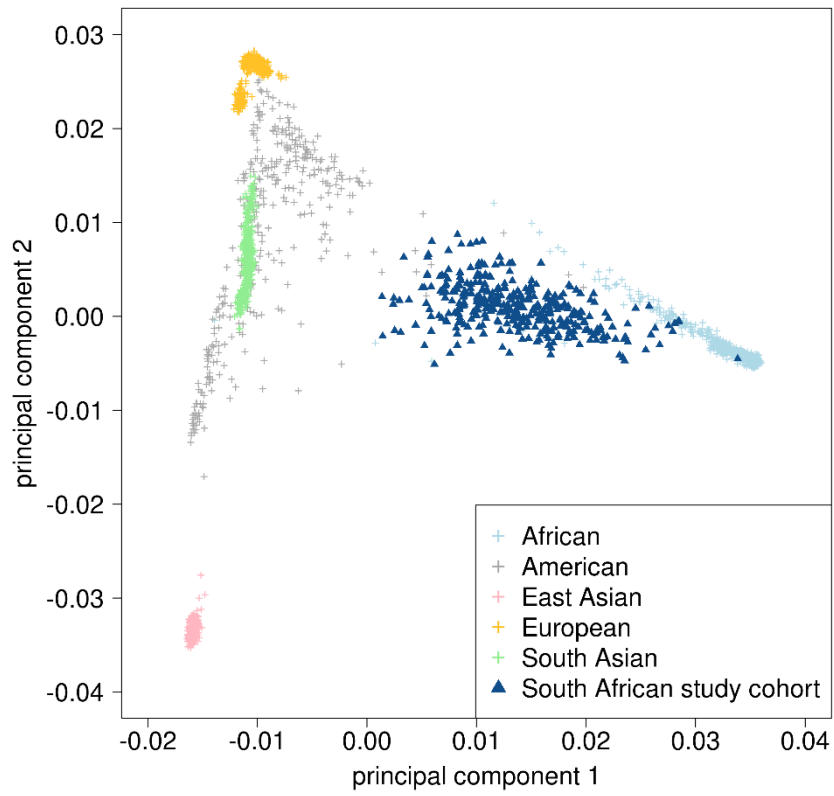


Figure S4. Principal component analysis of the South African cohort. Plot of the first and second principal components of 374 individuals from the South African cohort after projection on the 1000 Genomes phase 3 populations.

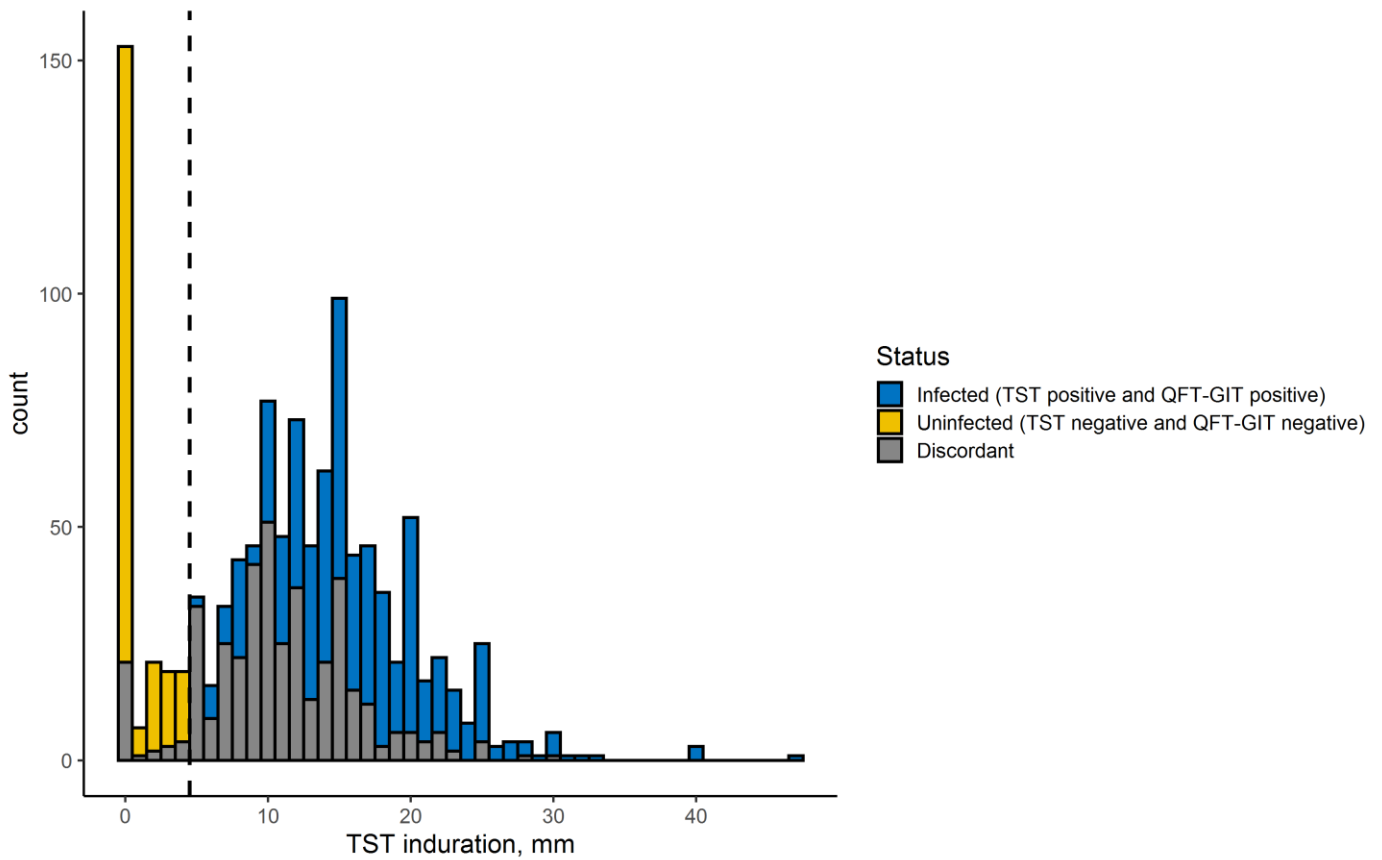


Figure S5. Stacked histogram of the tuberculin skin test (TST) distribution, stratified by our infection definition among 1108 household contacts in Vietnam. The dashed line represents a 5 mm cut-off for TST. Uninfected subjects (in yellow) were negative for TST and QuantiFERON-TB Gold In-Tube (QFT-GIT) and infected subjects (in blue) positive for both.

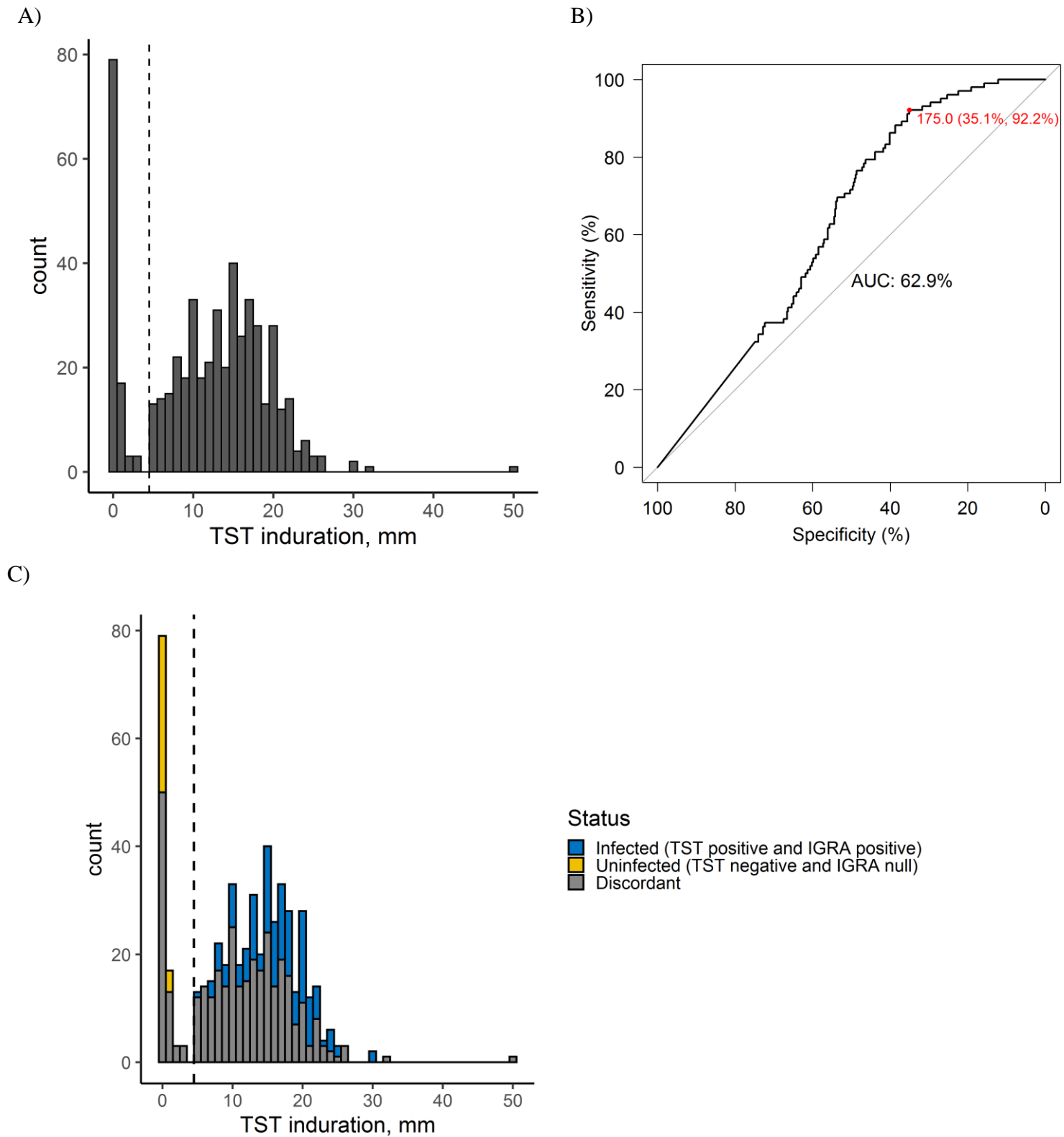


Figure S6. Definition of the GWAS phenotype in the household contacts (HHCs) study in France. A) Distribution of the tuberculin skin test (TST) induration among 516 HHCs (mean induration when 2 screening visits were done). The dashed line represents a 5 mm cut-off. TST was considered negative when the skin induration was i) < 5 mm at both V1 and V2, ii) < 5 mm at V1, when only one visit was done. 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. B) Construction of a ROC curve based on the defined TST status to determine the optimal interferon- γ release assay (IGRA) cut-off (175 pg/mL). C) Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- γ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- γ production > 175 pg/mL).

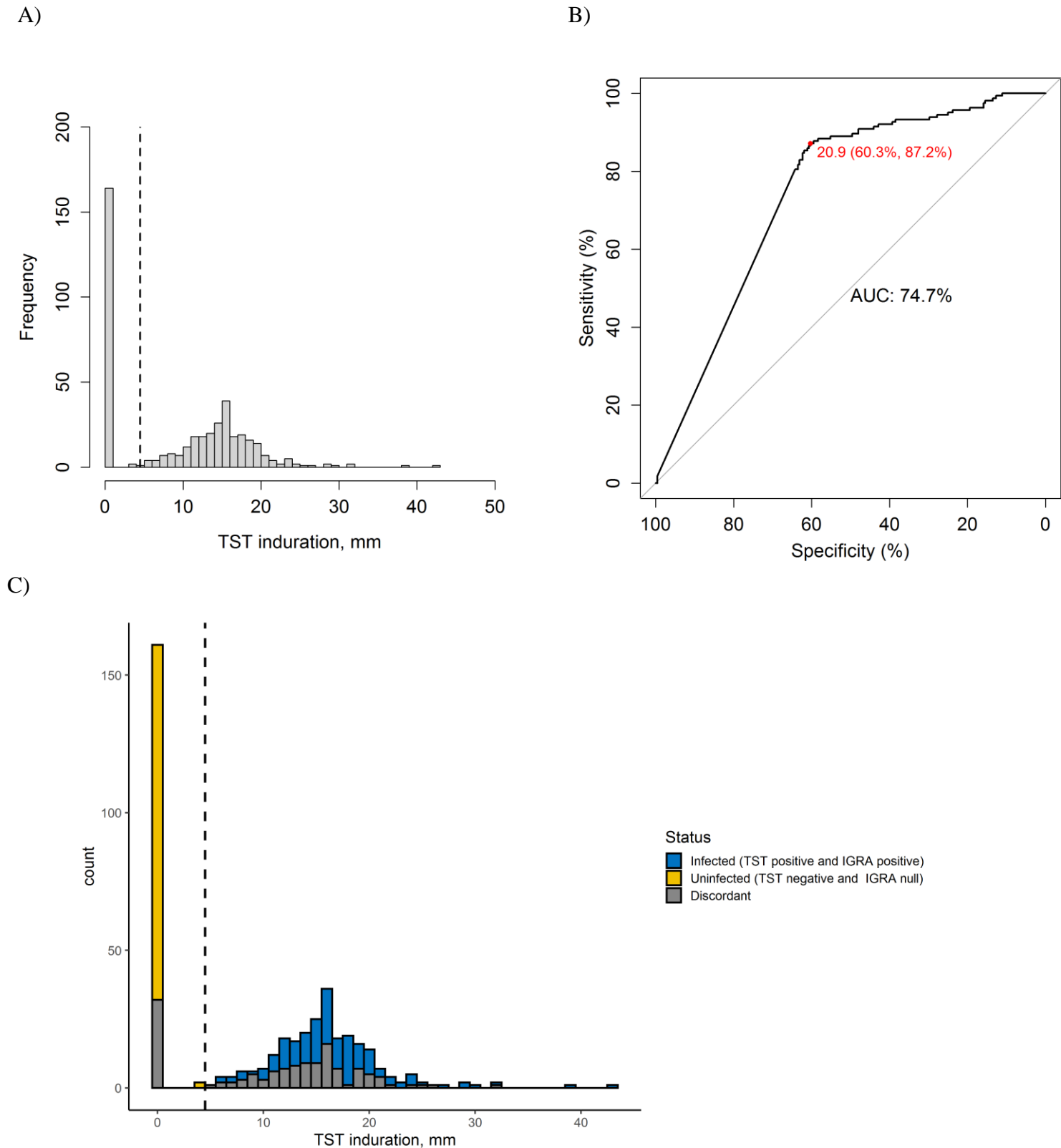


Figure S7. Definition of the GWAS phenotype in the family-based study in South Africa. A) Distribution of the tuberculin skin test (TST) induration among 415 participants. The dashed line represents a 5 mm cut-off. B) Construction of a ROC curve based on the TST status to determine the optimal interferon- γ release assay (IGRA) cut-off (20.9 pg/mL). C) Stacked histogram of the TST distribution stratified by our infection definition. Uninfected subjects (in yellow) presented a negative TST and a null production of IFN- γ . Infected subjects (in blue) presented a positive TST and a positive IGRA (IFN- γ production > 20.9 pg/mL).

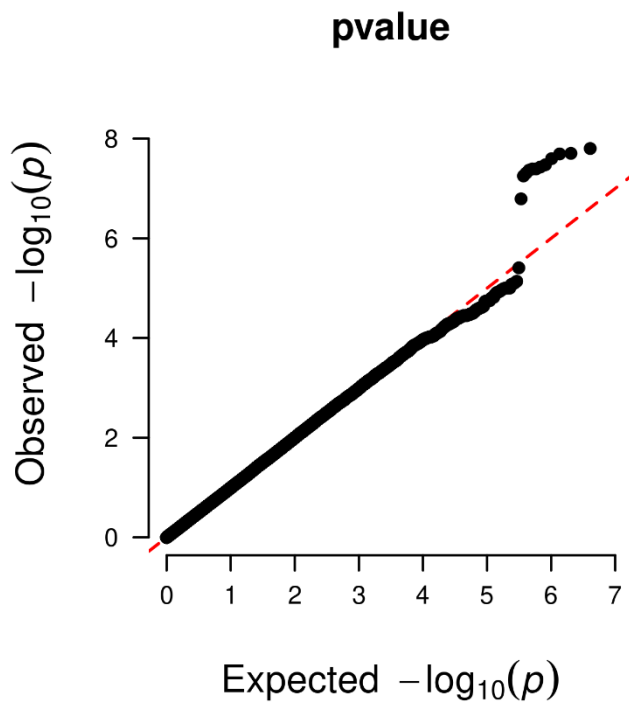


Figure S8. Quantile-quantile plot of GWAS resistance to tuberculosis infection in Vietnam. Plot of the expected distribution of association test statistics (x axis) for 5,591,951 variants compared to the observed values (y axis). Red line is the null hypothesis of no association ($y = x$).

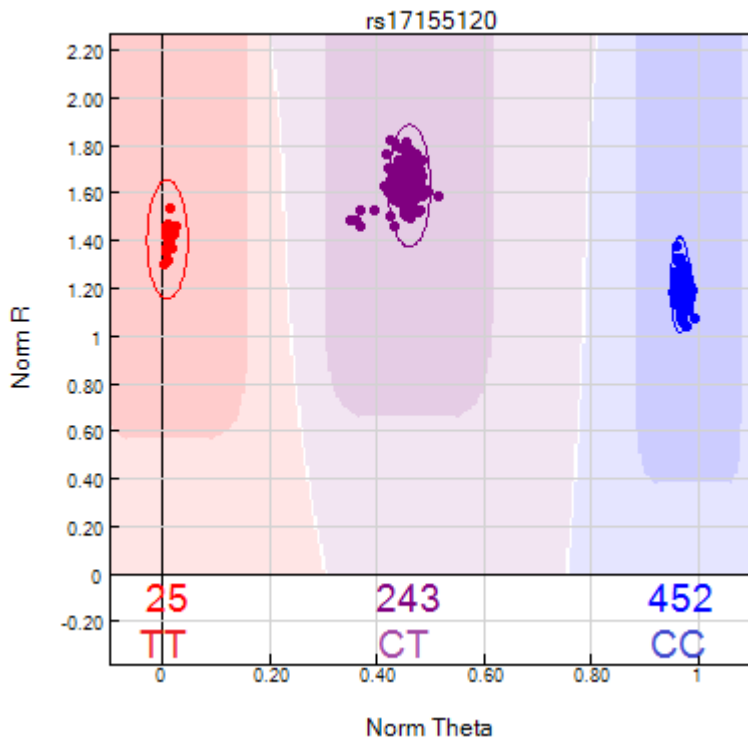


Figure S9. Intensity cluster plot for rs17155120 of the Illumina Infinium™ OmniExpressExome-8-v1 chip in the 720 Vietnamese individuals who were genotyped.

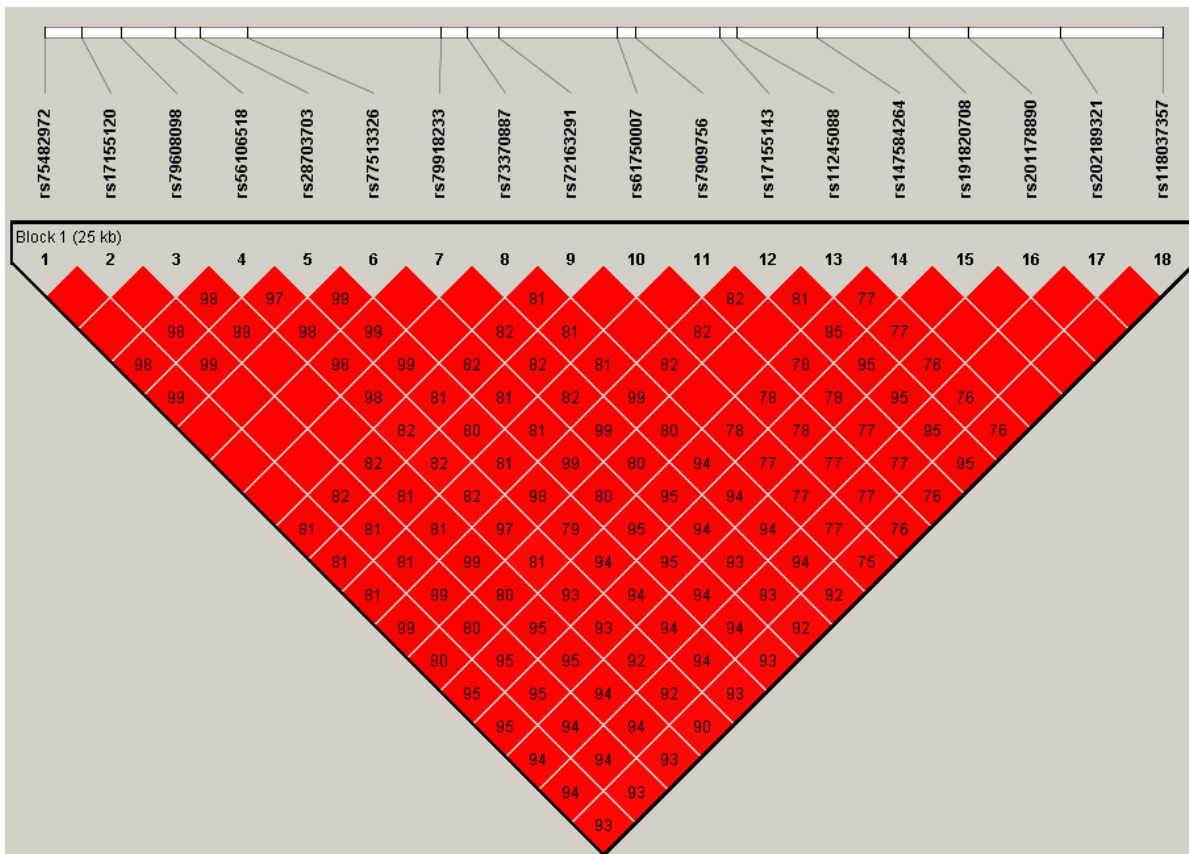


Figure S10. Haploview LD graph of the top associations on chromosome 10q26.2 in Vietnam. Pairwise LD coefficients r^2 are shown in each cell between 18 variants with $P < 5.0 \times 10^{-6}$. Top genotyped variant rs17155120 is second from left.

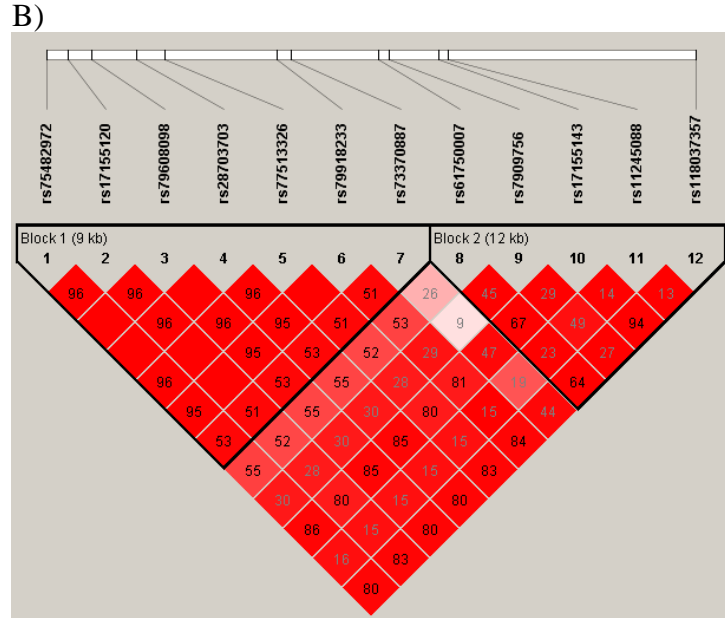
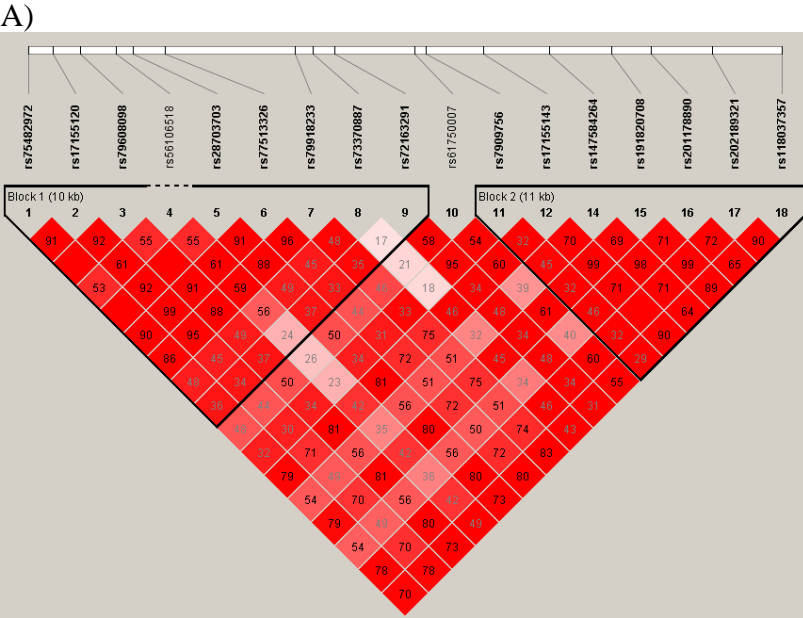


Figure S11. Haploview LD graphs of the locus on chromosome 10q26.2 in replication cohorts. Pairwise LD coefficients r^2 are shown in each cell A) between the 17 variants of the locus in the French cohort, and B) between 12 out of the 18 variants of the locus in the South African cohort. Top genotyped variant rs17155120 is second from left in each figure.

chr10:128348107 T/C

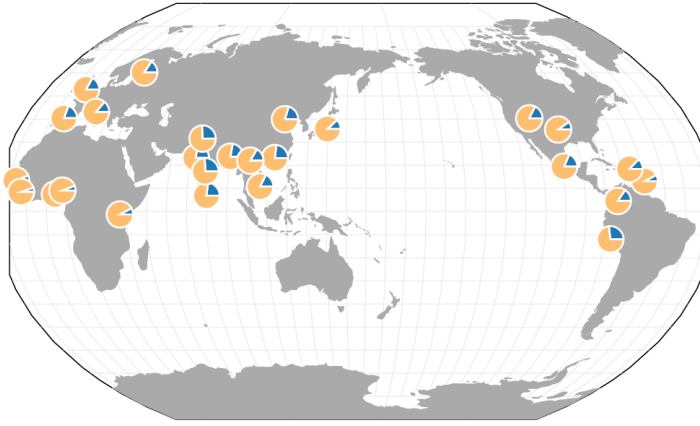


Figure S12. Geographic distribution of the variant rs17155120 in 1000G phase 3 populations. The T allele (in blue) is protective against *M. tuberculosis* infection. Map generated from the Geography of Genetic Variants Browser (<http://www.popgen.uchicago.edu/ggv>).

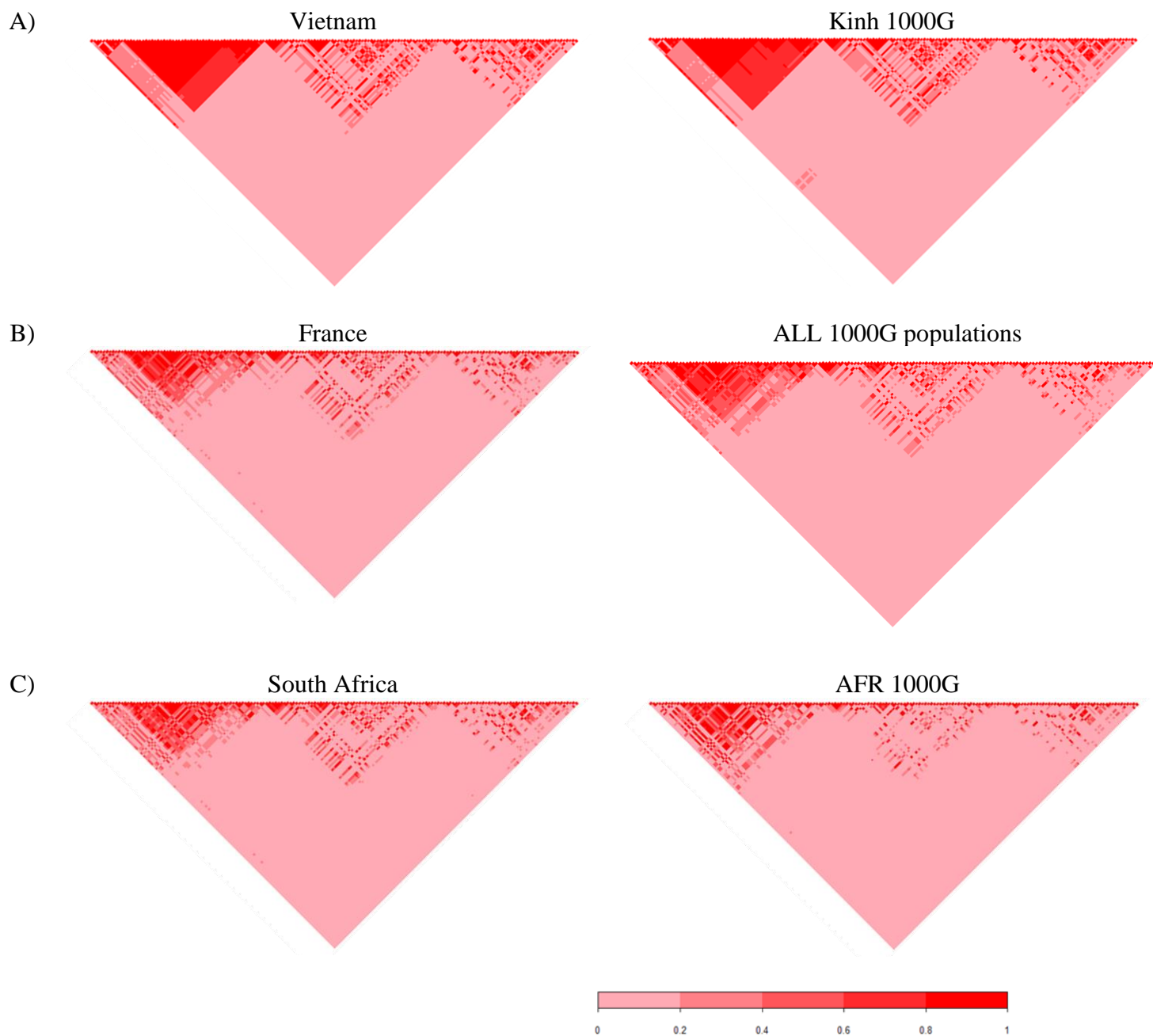


Figure S13. Pairwise linkage disequilibrium (r^2) in A) Vietnamese, B) French and C) South African study cohorts (left panels) between 162 variants in a 50 kb window around rs17155120 (middle of the triangle) as compared to 1000G population (right panels).

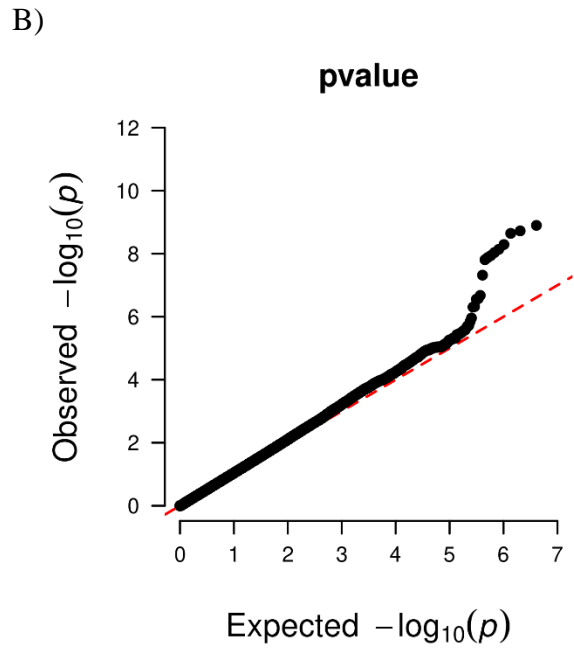
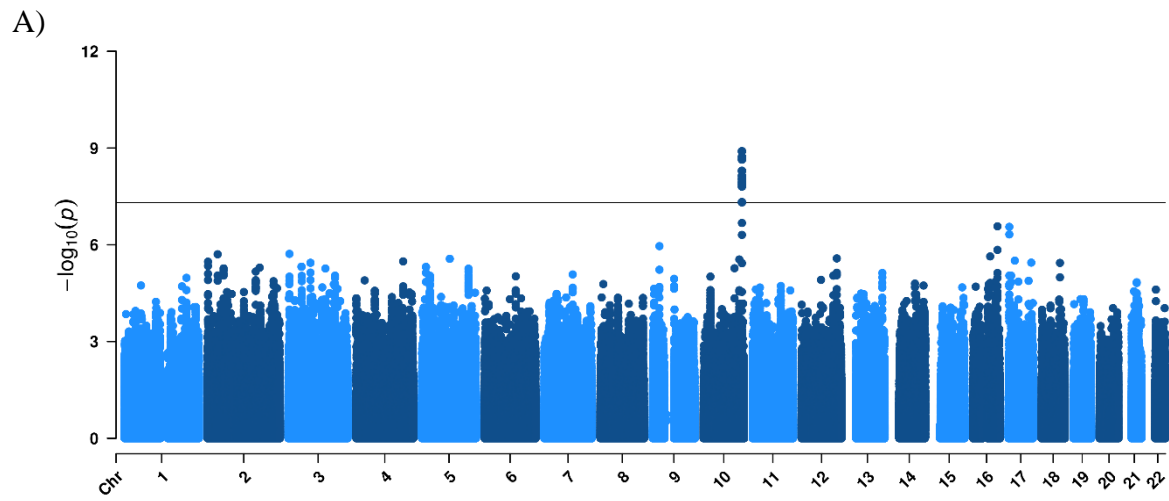


Figure S14. GWAS of resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects). A) Manhattan plot for 3,967,482 variants (minor allele frequency > 5% and info > 0.8). B) Quantile-quantile plot ($\lambda = 1.06$).

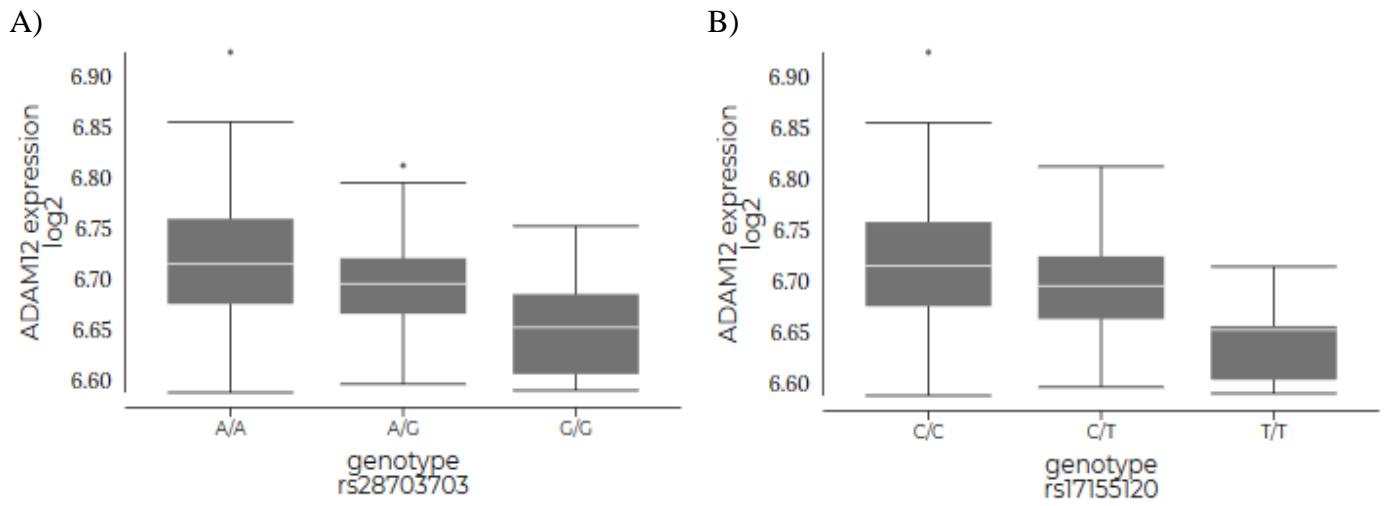


Figure S15. Expression levels of *ADAM12* in monocytes stratified by genotype for the associated variants A) best eQTL, rs28703703 ($P = 2.10 \times 10^{-3}$) B) top genotyped, rs17155120 ($P = 4.70 \times 10^{-3}$) (from <http://immunpop.com/kim/eQTL>).

Supplement Tables

Table S1. Risk factors associated with double positive tuberculin skin test (5 mm cut-off) and QuantiFERON-TB Gold In-Tube results compared to the reference group (double negative TST and QFT-GIT results) in household contacts in Vietnam

| Characteristics | Total n=1108 | TST+/QFT-GIT+ n=512 | OR ^a (95%CI) | OR ^b (95%CI) |
|--|-----------------|------------------------|-------------------------|-------------------------|
| Gender | | | | |
| Male | 427 | 189 | ref | ref |
| Female | 681 | 323 | 0.81 (0.53-1.24) | 0.84 (0.53-1.33) |
| Age group, years | | | | |
| <18 | 113 | 52 | ref | ref |
| 18-29 | 361 | 165 | 0.72 (0.33-1.59) | 0.75 (0.33-1.72) |
| 30-44 | 297 | 131 | 0.67 (0.30-1.49) | 0.69 (0.29-1.64) |
| 45-59 | 252 | 125 | 0.92 (0.40-2.10) | 0.95 (0.39-2.31) |
| ≥60 | 85 | 39 | 0.85 (0.31-2.38) | 0.76 (0.25-2.29) |
| BCG vaccination | | | | |
| No | 540 | 260 | ref | ref |
| Yes | 448 | 190 | 1.01 (0.65-1.55) | 1.01 (0.64-1.62) |
| Number of HHC per room, mean(sd) | 2.4(1.2) | 2.4(1.3) | 1.01 (0.83-1.21) | 0.99 (0.81-1.20) |
| Duration of index cough before treatment, months | | | | |
| <2 | 453 | 193 | ref | ref |
| 2-3 | 396 | 195 | 1.45 (0.85-2.46) | 1.64 (0.95-2.84) |
| >3 | 249 | 120 | 1.32 (0.72-2.41) | 1.07 (0.57-2.02) |

^aOdds ratio obtained in a univariable regression

^bOdds ratio obtained in a multivariable regression including all variables explored in the univariable analysis

Table S2. GWAS cohort in Vietnam

| | Uninfected | Infected | |
|----------------------|------------|---------------|--------|
| | | TST+/QFT-GIT+ | PTB |
| n | 185 | 201 | 152 |
| Age, years, mean(sd) | 36(16) | 32(14) | 48(12) |
| Sex, %F | 68 | 59 | 22 |
| TST, mm, mean(sd) | 1(1) | 16(6) | - |
| QFT-GIT | negative | positive | - |

PTB, pulmonary tuberculosis; QFT-GIT, QuantiFERON-TB Gold In-Tube; TST, tuberculin skin test

Table S3. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and positive interferon- γ release assay (IGRA) results compared to the reference group (negative TST and null IGRA) in household contacts in France

| Characteristics | Total n=180 | TST+/IGRA+ n=147 | OR ^a (95%CI) |
|--|----------------|---------------------|-------------------------|
| Gender | | | |
| Male | 86 | 72 | ref |
| Female | 94 | 75 | 0.69 (0.1-4.6) |
| Age, years, mean(sd) | 24(18) | 28(18) | 1.07 (1.01-1.15) |
| Contact duration with index, log hours, mean(sd) | 5(2) | 5(2) | 0.87 (0.46-1.51) |
| Index contagiousness | | | |
| High | 94 | 79 | ref |
| Low-intermediate | 86 | 68 | 0.57 (0.02-11.11) |
| Tuberculosis incidence of country of birth | | | |
| High | 56 | 51 | ref |
| ...Low-intermediate | 124 | 96 | 0.14 (0.05-1.37) |
| Complementary health insurance | | | |
| No | 55 | 45 | ref |
| Yes | 122 | 99 | 0.58 (0.01-8.33) |

^aOdds ratio obtained in a univariable regression

Table S4. GWAS cohort in France

| | Uninfected | Infected |
|-----------------------------|-------------------|-----------------|
| n | 30 | 127 |
| Age, years, mean(sd) | 17(19) | 28(18) |
| Sex, %F | 57 | 52 |
| TST, mm, mean(sd) | 0(1) | 16(4) |
| IGRA, IFN- γ , pg/mL | 0 | >175 |

IGRA, interferon- γ release assay; TST, tuberculin skin test

Table S5. Risk factors associated with both positive tuberculin skin test (5 mm cut-off) and interferon- γ release assay (IGRA) results compared to the reference group (both negative TST and IGRA results) in the family-based study in South Africa

| Characteristics | Total n=280 | TST+/IGRA+ n=152 | OR ^a (95%CI) |
|-------------------------|----------------|---------------------|-------------------------|
| Gender | | | |
| Male | 134 | 69 | ref |
| Female | 146 | 83 | 1.22 (0.67-2.17) |
| Age, years, mean(sd) | 12(6) | 15(6) | 1.30 (1.20-1.44) |

^aOdds ratio obtained in a univariable regression

Table S6. GWAS cohort in South Africa

| | Uninfected | Infected |
|-----------------------------|-------------------|-----------------|
| n | 118 | 136 |
| Age, years, mean(sd) | 10(4) | 15(6) |
| Sex, %F | 49 | 56 |
| TST, mm, mean(sd) | 0(0) | 17(5) |
| IGRA, IFN- γ , pg/mL | 0 | >20.9 |

IGRA, interferon- γ release assay; TST, tuberculin skin test

Table S7. Info score of the associated variants on 10q26.2

| Variant | Vietnam | France | South Africa |
|----------------|----------------|---------------|---------------------|
| rs11245088 | 0.91 | - | 0.92 |
| rs72163291 | 0.92 | 0.92 | - |
| rs17155143 | 0.93 | 0.94 | 0.93 |
| rs7909756 | 0.92 | 0.92 | 0.92 |
| rs28703703 | 0.99 | 0.98 | 0.99 |
| rs56106518 | 0.86 | 0.83 | - |
| rs75482972 | 0.99 | 0.98 | 0.96 |
| rs17155120 | genotyped | genotyped | genotyped |
| rs73370887 | 0.95 | 0.96 | 0.93 |
| rs79608098 | 0.99 | 0.99 | 0.99 |
| rs61750007 | 0.94 | 0.95 | 0.94 |
| rs77513326 | 0.99 | 0.99 | genotyped |
| rs79918233 | 0.97 | 0.98 | 0.97 |
| rs147584264 | 0.92 | 0.94 | - |
| rs191820708 | 0.91 | 0.93 | - |
| rs201178890 | 0.92 | 0.94 | - |
| rs202189321 | 0.90 | 0.93 | - |
| rs118037357 | 0.93 | 0.97 | 0.95 |

Table S8. Variants in the locus on chromosome 10q26.2 associated with resistance to *M. tuberculosis* infection in Vietnam (185 uninfected and 201 double positive infected subjects)

| Variant | EA | EAF | OR (95%CI) | P value |
|----------------|-----------|------------|-------------------|-----------------------|
| rs56106518 | C | 0.24 | 0.39 (0.30-0.47) | 1.37×10^{-7} |
| rs72163291 | ins | 0.26 | 0.41 (0.33-0.49) | 2.09×10^{-7} |
| rs28703703 | G | 0.21 | 0.40 (0.32-0.49) | 2.40×10^{-7} |
| rs17155120 | T | 0.21 | 0.40 (0.32-0.49) | 2.55×10^{-7} |
| rs11245088 | C | 0.26 | 0.41 (0.40-0.42) | 2.72×10^{-7} |
| rs7909756 | G | 0.26 | 0.41 (0.33-0.49) | 2.73×10^{-7} |
| rs75482972 | A | 0.21 | 0.41 (0.32-0.49) | 2.82×10^{-7} |
| rs17155143 | A | 0.21 | 0.39 (0.30-0.47) | 3.02×10^{-7} |
| rs79608098 | T | 0.21 | 0.41 (0.32-0.49) | 3.28×10^{-7} |
| rs77513326 | A | 0.21 | 0.40 (0.32-0.49) | 3.56×10^{-7} |
| rs73370887 | A | 0.21 | 0.39 (0.31-0.48) | 3.82×10^{-7} |
| rs61750007 | C | 0.25 | 0.42 (0.34-0.50) | 4.91×10^{-7} |
| rs79918233 | A | 0.21 | 0.40 (0.31-0.49) | 5.13×10^{-7} |
| rs147584264 | C | 0.21 | 0.41 (0.32-0.50) | 2.01×10^{-6} |
| rs191820708 | A | 0.21 | 0.41 (0.32-0.50) | 2.59×10^{-6} |
| rs201178890 | T | 0.21 | 0.41 (0.32-0.50) | 2.71×10^{-6} |
| rs202189321 | T | 0.21 | 0.41 (0.32-0.50) | 2.72×10^{-6} |
| rs118037357 | A | 0.20 | 0.42 (0.32-0.51) | 3.54×10^{-6} |

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

Table S9. Frequency of rs17155120 minor allele T in 1000 Genomes phase 3 populations¹

| 1000 Genomes population | rs17155120 T frequency |
|--|-------------------------------|
| ALL | 0.16 |
| African | 0.07 |
| Yoruba in Ibadan, Nigeria | 0.07 |
| Luhya in Webuye, Kenya | 0.08 |
| Gambian in Western Divisions in the Gambia | 0.06 |
| Mende in Sierra Leone | 0.05 |
| Esan in Nigeria | 0.06 |
| Americans of African Ancestry in SW USA | 0.11 |
| African Caribbeans in Barbados | 0.08 |
| American | 0.18 |
| Mexican Ancestry from Los Angeles USA | 0.19 |
| Puerto Ricans from Puerto Rico | 0.12 |
| Colombians from Medellin, Colombia | 0.15 |
| Peruvians from Lima, Peru | 0.27 |
| East Asian | 0.18 |
| Han Chinese in Beijing, China | 0.20 |
| Japanese in Tokyo, Japan | 0.12 |
| Southern Han Chinese | 0.22 |
| Chinese Dai in Xishuangbanna, China | 0.16 |
| Kinh in Ho Chi Minh City, Vietnam | 0.18 |
| European | 0.16 |
| Utah Residents with Northern and Western European Ancestry | 0.18 |
| Toscani in Italia | 0.14 |
| Finnish in Finland | 0.16 |
| British in England and Scotland | 0.16 |
| Iberian Population in Spain | 0.19 |
| South Asian | 0.24 |
| Gujarati Indian from Houston, Texas | 0.22 |
| Punjabi from Lahore, Pakistan | 0.26 |
| Bengali from Bangladesh | 0.22 |
| Sri Lankan Tamil from the UK | 0.23 |
| Indian Telugu from the UK | 0.25 |

1. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics* 31, 3555–3557

Table S10. Variants associated with resistance to *M. tuberculosis* infection in the 3 cohorts from Vietnam, France and South Africa (333 uninfected vs 616 infected subjects) with $P < 5.0 \times 10^{-8}$

| Variant | EA | OR (95%CI) | <i>P</i> value | <i>P</i>_{het} | Chromosome | Nearest Gene |
|----------------|-----------|-------------------|-----------------------|-------------------------------|-------------------|---------------------|
| rs17155120 | T | 0.50 (0.45-0.55) | 1.26×10^{-9} | 0.31 | 10 | <i>C10orf90</i> |
| rs118037357 | A | 0.47 (0.41-0.53) | 1.89×10^{-9} | 0.50 | 10 | <i>C10orf90</i> |
| rs77513326 | A | 0.50 (0.45-0.56) | 2.26×10^{-9} | 0.29 | 10 | <i>C10orf90</i> |
| rs79918233 | A | 0.51 (0.46-0.56) | 5.12×10^{-9} | 0.24 | 10 | <i>C10orf90</i> |
| rs17155143 | A | 0.53 (0.48-0.58) | 7.30×10^{-9} | 0.10 | 10 | <i>C10orf90</i> |
| rs28703703 | G | 0.55 (0.50-0.60) | 9.18×10^{-9} | 0.11 | 10 | <i>C10orf90</i> |
| rs79608098 | T | 0.55 (0.50-0.60) | 1.15×10^{-8} | 0.14 | 10 | <i>C10orf90</i> |
| rs75482972 | A | 0.55 (0.50-0.60) | 1.32×10^{-8} | 0.11 | 10 | <i>C10orf90</i> |
| rs61750007 | C | 0.56 (0.51-0.61) | 1.55×10^{-8} | 0.12 | 10 | <i>C10orf90</i> |

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