Supplementary Information

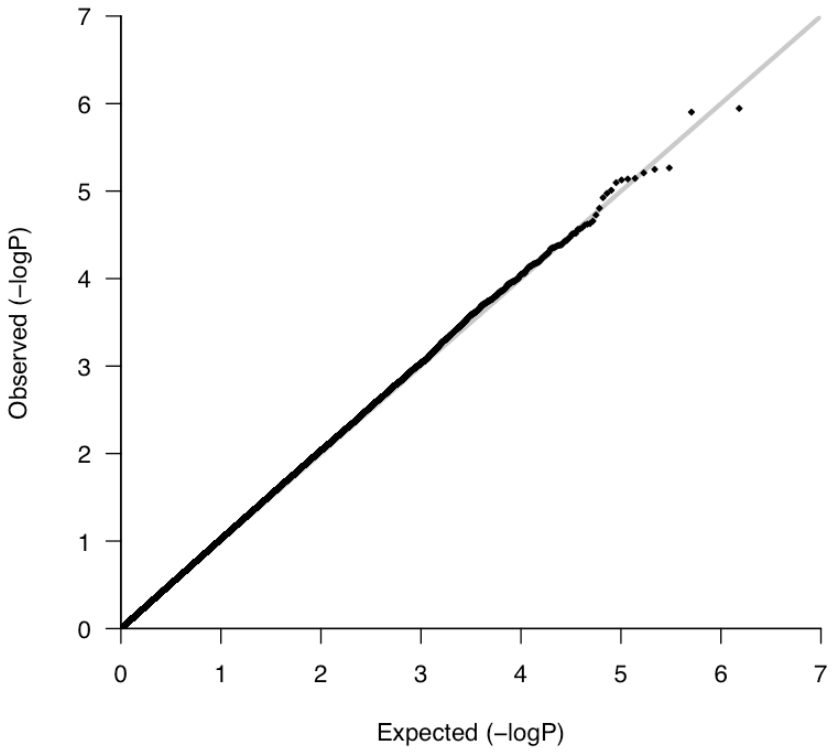
A novel susceptibility locus for tuberculosis on chromosome 11p13 downstream the *WT1* gene

Thorsten Thye[1,2], Ellis Owusu-Dabo[3,4], Fredrik O. Vannberg[5], Reinout van Crevel[6], James Curtis[7], Edhyana Sahiratmadja[8], Yanina Balabanova[9,10], Christa Ehmen[1], Birgit Muntau[1], Gerd Ruge[1], Jürgen Sievertsen[1], John Gyapong[11], Vladyslav Nikolayevskyy[9], Philip C. Hill[12], Giorgio Sirugo[12], Francis Drobniewski[9], Esther van de Vosse[13], Melanie Newport[12,14], Bachti Alisjahbana[15], Sergey Nejentsev[7], Tom H.M. Ottenhoff[13], Adrian V.S. Hill[5], Rolf D. Horstmann[1], Christian G. Meyer[1]

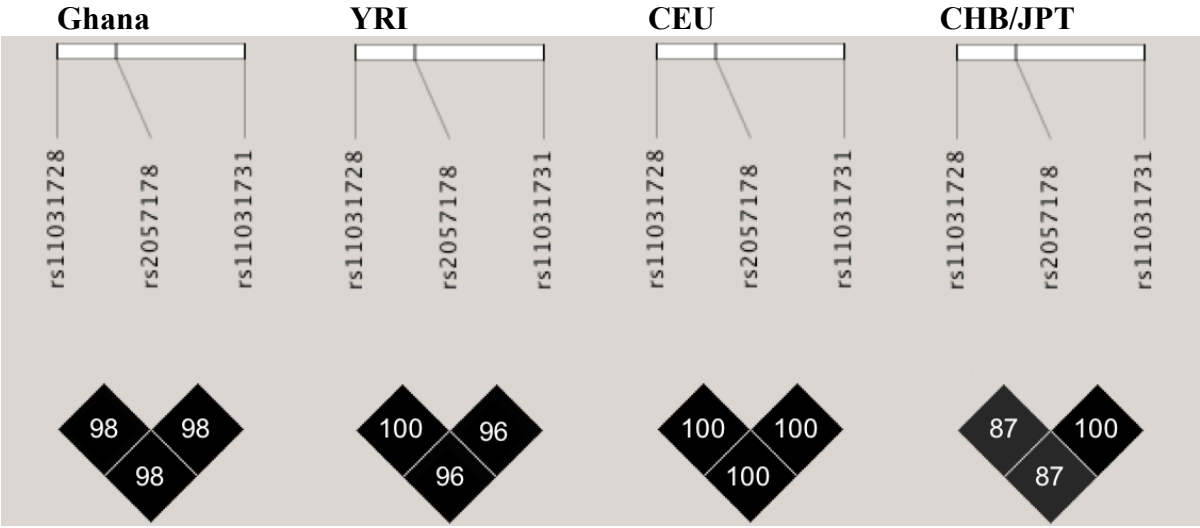
[1] Department of Molecular Medicine, Bernhard Nocht Institute for Tropical Medicine, Hamburg, Germany, [2] Institute of Medical Biometry and Statistics, University Hospital Schleswig-Holstein, Campus Lübeck, Lübeck, Germany, [3] Kumasi Centre for Collaborative Research in Tropical Medicine, Kumasi, Ghana, [4] Department of Community Health, College of Health Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, [5] The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, United Kingdom, [6] Department of Internal Medicine, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands, [7] Department of Medicine, University of Cambridge, Cambridge, United Kingdom, [8] Health Research Unit, Faculty of Medicine, University of Padjadjaran, Bandung, Indonesia, [9] Clinical TB and HIV Group and HPA National Mycobacterium Reference Laboratory, Blizard Institute, Barts and the London School of Medicine, Queen Mary College, University of London, London, UK, [10] Samara Oblast Tuberculosis Dispensary, Samara City, Samara, Russian Federation, [11] School of Public Health, College of Health Sciences, University of Ghana, Legon, Ghana, [12] Medical Research Council Laboratories, Fajara, Banjul, The Gambia, [13] Department of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands, [14] Brighton and Sussex Medical School, UK, [15] Division of Tropical and Infectious Diseases, Department of Internal Medicine, Faculty of Medicine, University of Padjadjaran, Bandung, Indonesia

Correspondence should be addressed to C.G.M (c.g.meyer@bnitm.de)

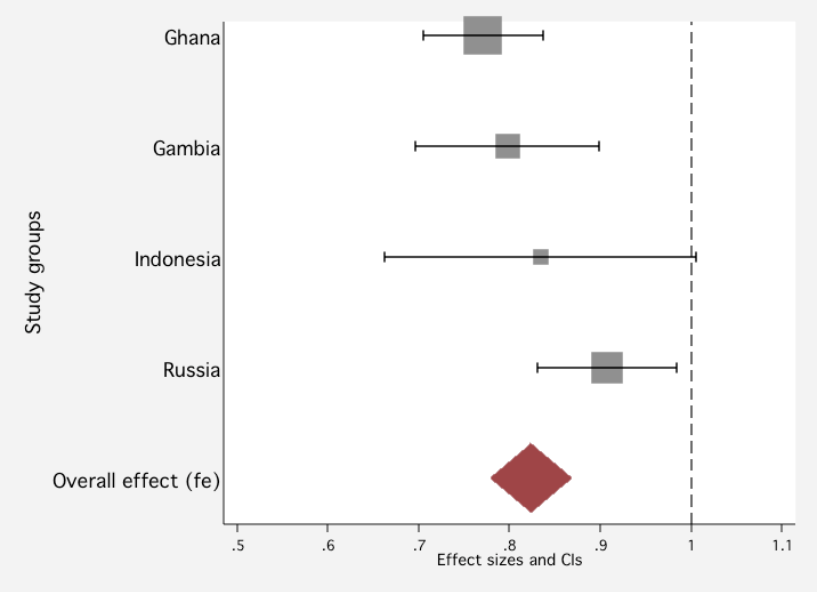
Contents:



Supplementary Figure 1. Quantile-quantile plot of the Ghanaian association statistics corrected for the first three principal components ($\lambda = 1.03$).



Supplementary Figure 2. Pairwise linkage disequilibrium (r^2) of the chromosome 11 hit SNPs in Ghanaian, Nigerian (YRI), Caucasian (CEU) and Chinese/Japanese (CHB/JPT) populations.



Supplementary Figure 3. Fixed effect (fe) meta-analysis of SNP rs2057178 including the Ghanaian, Gambian, Indonesian and Russian study groups.

Supplementary Table 1. Association statistic of 46 SNPs after imputation and genotyping, all *P* values < 10^{-5}

CHR	rs#	Position (hg19)	Major / minor allele	Imputation				Genotyping					
								GWAS group			Entire Ghanaian study group		
				MAF	RSQ	OR	LR-P	P	N	HWE-P	OR (95% CI)	P	N
1	rs1995311	19820119	G/T	0.46	0.80	1.29	9.72E-06	1.87E-04	3096	0.32			
1	rs9440631	109394524	G/A	0.08	0.96	1.52	4.63E-06	1.45E-05	3073	0.02			
1	rs7542223	180420271	T/G	0.20	0.99	0.74	4.29E-06	1.46E-05	3083	0.98			
1	rs17739539	218054687	C/T	0.04	0.99	1.77	6.67E-06	4.66E-05	3140	0.27			
2	rs77088208	112811013	C/T	0.04	0.66	2.03	3.30E-06	3.27E-06	3060	0.98	1.44 (1.21-1.73)	5.81E-05	7568
2	rs2356081	159118048	T/A	0.45	0.79	1.31	3.26E-06	1.07E-04	3072	0.93			
2	rs2646157	175724102	G/A	0.50	1	0.79	4.18E-06	6.68E-06	3085	0.50	1.05 (1.00-1.11)	3.10E-02	7739
2	rs1400859	215451185	G/C	0.15	0.93	0.71	5.88E-06	1.42E-03	2997	0.17			
2	chr2:230377024	230377024	A/G	0.04	0.67	0.43	4.64E-06	4.12E-04	3049	0.43			
3	rs73048609	23746535	C/T	0.08	0.78	1.60	6.43E-06	4.25E-04	3086	0.64			
3	chr3:54882257	54882257	C/A	0.03	0.79	2.21	4.10E-06	1.52E-01	3074	0.38			
4	rs79676281	54441597	G/A	0.07	0.78	0.46	3.30E-10	1.37E-01	3023	0.29			
4	rs76552024	118025780	G/A	0.01	0.63	0.22	1.17E-06	1.76E-03	2896	0.19			
4	rs13129961	188022194	G/A	0.21	1	1.32	6.61E-06	9.66E-05	3055	0.94			
5	rs7714932	4909055	G/A	0.23	0.77	0.73	5.14E-06	1.29E-04	3084	0.04			
5	rs294582	163018061	C/T	0.11	0.76	0.65	7.18E-06	1.76E-04	2859	0.29			
6	rs116682106	8682530	T/G	0.02	0.36	0.15	4.95E-07	9.63E-01	3050	0.66			
6	rs111415789	52463455	A/G	0.01	0.73	3.37	8.46E-06	mono-allelic					
6	chr6:109669279	109669279	G/T	0.04	0.83	2.08	5.69E-07	4.32E-03	3077	0.73			
7	rs115912309	2613564	A/G	0.02	0.64	0.30	5.16E-06	3.74E-01	3072	0.68			
7	rs111399899	25277300	A/G	0.05	0.73	1.82	4.45E-06	8.22E-06	3073	0.85	1.29 (1.10-1.50)	1.34E-03	7594
7	rs73694499	46987740	G/T	0.11	0.85	0.68	9.61E-06	1.26E-05	3081	0.54			
8	rs12156380	5099914	C/A	0.46	0.99	1.27	5.58E-06	4.86E-06	3118	0.72	1.15 (1.06-1.24)	7.79E-04	7584
8	rs116496006	16577542	G/A	0.02	0.41	0.19	8.84E-06	5.20E-01	3083	0.38			
8	rs28439133	73016162	C/T	0.08	0.96	0.62	2.25E-06	9.89E-05	3088	0.69			
10	rs77751818	59785350	C/T	0.08	0.73	0.59	8.36E-06	2.42E-05	3088	0.47			
10	chr10:119991669	119991669	C/T	0.03	0.54	2.68	5.48E-06	1.49E-04	3084	0.58			
11	rs2057178	32364187	G/A	0.33	0.89	0.72	2.97E-08	1.74E-08	3173	0.74	0.77 (0.71-0.84)	2.63E-09	7763
11	rs77576197	116034074	C/G	0.08	0.89	1.57	9.22E-06	8.81E-05	3077	0.58			
12	rs2058784	15235173	T/C	0.47	0.78	1.30	9.33E-06	9.92E-06	3094	0.22	1.12 (1.04-1.21)	4.15E-03	7610

Supplementary Table 1 (continued)

CHR	rs#	Position (hg19)	Major / minor allele	Imputation				Genotyping					
								GWAS group			Entire Ghanaian study group		
				MAF	RSQ	OR	LR-P	P	N	HWE-P	OR (95% CI)	P	N
12	rs10848180	131157748	C/T	0.46	0.93	0.79	9.72E-06	4.58E-06	3096	0.83	0.88 (0.81-0.95)	7.27E-04	7755
13	rs1887474	39943576	C/T	0.31	0.93	1.30	8.88E-06	1.12E-04	2995	0.15			
13	rs116261290	84260291	G/A	0.03	0.83	2.01	8.62E-06	3.85E-04	2913	0.95			
14	rs17429659	30872069	T/A	0.02	0.89	2.87	3.40E-06	mono-allelic					
14	rs17110454	40771636	C/A	0.08	0.91	1.52	9.73E-06	6.66E-06	3095	0.60	1.34 (1.15-1.54)	9.48E-05	7627
14	rs4906296	103658864	A/G	0.33	0.88	1.30	9.25E-06	2.83E-05	3046	0.14			
15	rs28673058	78966422	C/T	0.04	0.84	1.86	8.94E-06	1.23E-04	3089	0.35			
15	rs114416857	101749385	G/A	0.02	0.69	0.33	9.55E-06	1.18E-03	3085	0.73			
16	rs116282608	220340	T/G	0.08	0.59	0.57	3.25E-06	8.81E-01	2828	0.16			
16	rs55681758	87729743	T/C	0.42	0.49	1.42	2.73E-06	1.61E-02	3087	0.10			
17	rs11657700	30893571	G/A	0.39	0.99	1.27	4.57E-06	5.48E-06	3109	0.76	1.14 (1.05-1.24)	1.21E-03	7552
18	chr18:5451523	5451523	C/T	0.01	0.42	0.16	9.07E-06	5.38E-03	3085	0.52			
18	rs35687920	20154558	G/A	0.24	0.95	1.33	3.79E-06	1.97E-06	3083	0.12	1.25 (1.14-1.36)	5.03E-07	7714
18	rs7235966	30322511	G/A	0.29	0.73	1.38	7.13E-07	1.09E-02	3097	0.32			
18	rs2543007	43051699	T/C	0.08	0.81	1.63	4.04E-06	6.53E-08	3167	0.96	1.37 (1.19-1.58)	1.10E-05	7625
19	rs642776	56605454	G/A	0.02	0.39	4.31	5.47E-06	3.30E-02	3033	0.34			

Association results filtered for those SNPs with imputed *P* values < 10^{-5} and imputation quality scores of R-squared (RSQ) > 0.3 as calculated by the minimac software. Genotyping analyses in the entire study group were performed with SNPs with *P* values < 10^{-5} after genotyping in the GWAS group. rs#. reference SNP number; Position. chromosomal position (bp); hg19. Human Genome built 19; MAF. minor allele frequency in the Ghanaian data set; OR. odds ratio; HWE-P. Hardy Weinberg equilibrium *P* value in controls; Imputation. LR-P. likelihood ratio *P* values as calculated by logistic regression analysis with allelic dosages derived by the minimac software and adjusted by the first three principal components (Eigenstrat); Genotyping. GWAS group P. *P* values calculated by logistic regression analysis adjusted by the first three principal components (Eigenstrat); Genotyping. entire Ghanaian study group P. P values calculated by logistic regression analysis adjusted by ethnic groups of Ghana. sex and age. SNPs rs111415789 and rs17429659 were found to be mono-allelic in the Ghanaian study group.

Supplementary Table 2. Association statistic of imputed SNPs in the GWAS and entire study group at the chromosome 11 hit locus

CHR	rs#	Position (hg19)	Imputation results				Genotyping results					
							GWAS group			Entire Ghanaian study group		
			MAF	RSQ	OR	LR-P	P	N	HWE-P	OR (95% CI)	P	N
11	rs7116974	32348789	0.31	1	0.78	9.61E-06	6.88E-05	3125	0.90	0.9 (0.82-0.98)	0.017	7611
11	chr11:32358589	32358589	0.21	0.95	0.70	5.58E-06	3.23E-07	3109	0.26	0.79 (0.72-0.88)	5.74E-06	7564
11	rs11031728	32363616	0.32	0.89	0.72	2.97E-08	3.44E-08	3139	0.70	0.77 (0.71-0.84)	5.25E-09	7762
11	rs11031731	32365430	0.31	0.87	0.70	8.88E-06	1.85E-08	3136	0.64	0.78 (0.71-0.85)	7.01E-09	7764

Association results filtered for those SNPs with *P* values < 10^{-5} and imputation quality scores of R-squared (RSQ) > 0.8 as calculated by the minimac software. rs#, reference SNP number; Position, chromosomal position (bp); hg19, Human Genome built 19; MAF, minor allele frequency in the Ghanaian data set; OR, odds ratio; HWE-P, Hardy Weinberg equilibrium *P* value in controls; Imputation LR-P, likelihood ratio *P* values as calculated by logistic regression analysis with allelic dosages derived by the minimac software and adjusted by the first three principal components (Eigenstrat); Genotyping, GWAS group P, *P* values calculated by logistic regression analysis adjusted by the first three principal components (Eigenstrat); Genotyping, entire Ghanaian study group; P, *P* values calculated by logistic regression analysis adjusted by ethnic groups of Ghana, sex and age.

Supplementary Table 3. Mantel-Haenszel statistic of SNP rs2057178 adjusted for Ghanaian ethnic groups

Ghanaian ethnic groups	OR	95% CI	*P*-value
Akan	0.80	0.72-0.88	2.99×10^{-6}
Ewe	0.77	0.56-1.05	0.10
Gaa	0.78	0.63-0.96	0.02
Northerners	0.81	0.66-1.00	0.056
Others	0.82	0.55-1.23	0.34
Test of homogeneity of ORs: $P = 0.9997$			

Supplementary Note

Ghanaian study group and population structure

TB patients were enrolled in Ghana, West Africa, between September 2001 and July 2004 at Komfo Anokye Teaching Hospital in Kumasi, and Korle Bu Teaching Hospital in Accra as well as at additional hospitals/outpatient departments in Accra and Kumasi and at regional district hospitals. The case group consisted of 2146 individuals with a median age of 32 years. The proportion of males was 66.7%. Phenotyping of patients included documentation of major TB-related symptoms on structured questionnaires, medical histories and physical examination, HIV-1/2 testing (Capillus, Trinity Biotech, Bray, Co Wicklow, Ireland), posterior-anterior chest X-rays, Ziehl-Neelsen staining of two independent sputum smears, and culturing of *M. tuberculosis* complex isolates on solid Loewenstein-Jensen medium. All cases were HIV-negative and presented with characteristic radiological lesions of pulmonary TB assessed by two independent radiologists. The patients were treated in the framework of the DOTS programme (Directly Observed Treatment Short-Course strategy) organized by the Ghanaian National Tuberculosis Programme.

Out of 5652 control individuals who were included in the study, 3557 were part of the original Ghanaian TB study. Of those, 2589 persons were characterized by their medical histories and clinical examination as well as posterior-anterior chest X-rays. An additional 968 clinically healthy control individuals were phenotyped by medical history and clinical examination only. The age range of the 3557 control individuals was 5-76 years, with a median age of 28 years and a male proportion of 51.4%. An additional 2095 Ghanaian population controls belonged to a control group recruited for a malaria association study in the Kumasi

region of Ghana. All individuals of this group were children with a median age of 2.5 years and a proportion of males of 52.6%. The control children of this group were healthy as assessed by physical examination.

Study participants belonged to the ethnic groups of Akan, Ga-Adangbe, Ewe and several other ethnic groups from northern Ghana (Dagomba, Sissala, Gonja, and Kusasi; herein denoted as Northerners).

The study protocol was approved by the Committee on Human Research, Publications and Ethics, School of Medical Sciences, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, and the Ethics Committee of the Ghana Health Service, Accra, Ghana. Full venous blood samples were taken only after a detailed explanation of the aims of the study, and consent was obtained of individuals enrolled or their parents/guardians by signature or by thumbprint in case of illiteracy.

Gambian study group and population structure

The Gambia tuberculosis cohort consists of 1,498 pulmonary tuberculosis cases and 1,496 controls recruited from The Gambia. The diagnosis of tuberculosis was based on compatible clinical features, a positive chest x-ray, and by culture or smear positivity. All cases were collected at major urban health centres. The controls were recruited from routine births at local clinics[1]. The majority (>95%) of the cases was screened for HIV-1, with positive cases excluded from the study because HIV infection increases the risk of tuberculosis. All samples were obtained with informed consent. Ethical approval for the study was granted by the Medical Research Council (MRC) and the Gambian government joint ethical committee.

Indonesian study group and population structure

1050 new pulmonary tuberculosis patients above 15 years of age were recruited from an outpatient tuberculosis clinic, Perkumpulan Pemberantasan Tuberkulosis Indonesia (PPTI) Jakarta and from the Hasan Sadikin Hospital in Bandung, two cities at Java, Indonesia. Diagnosis was based on clinical presentation, chest X-ray examination, and confirmed by sputum microscopy positive for mycobacteria. All patients were provided with free anti-TB therapy according to the national TB program (2HRZE/4H3R3). 53.9 % were males; the median age of the patients was 30 years (range 14-75). Patients diagnosed with HIV co-infection (1.6%) were excluded from the study.

Nine-hundred-eighty-seven randomly selected control subjects were recruited from neighboring households. First-degree relatives of patients were excluded. Control subjects with signs and symptoms suggesting active tuberculosis or a history of prior anti-TB treatment were also excluded. 53.6 % were males, and the median age of the controls was 30 years (range 14-76).

Self and parental ethnicities were recorded upon recruitment. A Javanese origin characterized three groups - the Jawa, Betawi, and Sunda - and altogether comprised more than 80% of the total sample. The non-Javanese category included individuals from ethnicities other than Javanese or born on other Indonesian islands.

Written informed consent was obtained from all subjects, and the study was approved by the Ethical Committee of the Eijkman Institute of Molecular Biology, Jakarta, and of the Faculty of Medicine, Padjadjaran University, Hasan Sadikin Hospital, Bandung, Indonesia.

Russian study group and population structure

Pulmonary TB patients attending civilian TB dispensaries and TB clinics along with healthy subjects attending the blood transfusion service were recruited in two Russian cities, St. Petersburg and Samara, as described previously[2]. Patients were diagnosed by the local health care service using information about TB contact, clinical symptoms, evidence from X-rays and sputum smear. HIV-positive subjects, patients with extra-pulmonary TB and TB patients who had negative sputum cultures of *M. tuberculosis* were excluded. Control individuals were healthy adult blood bank donors with no history of TB. In total 4,441 culture-confirmed pulmonary TB patients and 5,874 controls were studied, including 2,432 patients and 3,126 controls from St. Petersburg and 2,009 patients and 2,748 controls from Samara. The study was approved by the Human Biology Research Ethics Committees of the University of Cambridge and of Queen Mary College, London and the local Ethics Committees in St.-Petersburg and Samara, Russia. All participants provided written informed consent before being enrolled into the study.

Ghana GWA study; genotyping and quality control of samples and SNPs

Genotyping. DNA was isolated from venous blood samples of study participants (AGOWA® mag Maxi DNA Isolation Kit, Macherey & Nagel, Germany) following the manufacturer´s instructions. Genotyping was performed with the Affymetrix SNP 6.0 array at Affymetrix Services Laboratory (South San Francisco, CA 9408, USA) and ATLAS Biolabs GmbH (10117, Berlin, Germany) as previously described[3]. The genotype calling algorithm Birdseed (software v2) was applied on all samples.

Quality control of samples. Genotype data of 25 individuals exhibiting subject-wise heterogeneity rates > 33% and with SNP missing rates > 4% were removed from further

analyses. Genetic outliers were eliminated by the nearest neighbour allele sharing algorithm of PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). Two individuals with Z scores < -3 and 5 subjects with N > 1 were excluded from further analyses. Only one subject of related individuals was kept in the analyses (proportion of identity by descent ≤ 0.25). Among individuals of 77 related pairs, each one individual with the higher SNP missing frequencies was excluded. One-hundred-four individuals were excluded from the data set, resulting in the total of 1329 cases and 1847 controls.

Quality control of SNPs. SNPs with minor allele frequencies (MAF) < 1% and deviation from Hardy Weinberg Equilibrium (HWE) as defined by a P value $<10^{-4}$ were removed from the data set. In addition, a SNP-wise genotyping rate <98% was applied. From all SNPs that met these criteria and yielded association signals of $< 10^{-5}$ in the analysis, signal intensity plots were visually inspected to exclude SNPs with ambiguous or erroneous calling results.

Genotyping of imputation signals

Genotyping of the imputed variants in the Ghanaian GWAS group and replication samples as well as the Gambian and Indonesian samples were performed by dynamic allele specific hybridization with fluorescence resonance energy transfer in a LightTyper device (Roche Diagnostics, Mannheim, Germany) in Hamburg, Germany. Genotyping of the Russian cohort was done in Cambridge, UK, using a Taqman assay (Applied Biosystems, Warrington, UK).

Imputation and Statistical Analyses

In order to identify further associated SNPs on a genome-wide scale in the Ghanaian study group we performed genome wide imputation analyses. The most recent genotype release 2010-08 of

the African data set of the "1000 Genomes Project" including 78 Yoruban (Nigeria) individuals, 67 Luhya (Kenya), 24 American individuals with African ancestry (Southwest US) and 5 Puerto Rican individuals were used as reference to impute untyped SNPs into the Ghanaian GWAS group. In a first step the Ghanaian GWAS SNP data were haplotyped (MACH software; http://www.sph.umich.edu/csg/abecasis/MACH/). The preformed haplotypes inferred from the 1000 Genomes reference data were then imputed in the phased genotypes of the Ghanaian sample using the minimac software, which is an implementation of the MACH algorithm for genotype imputation to handle phased genotypes and large reference panels with large amounts of haplotypes. The imputed SNPs were tested for association with the disease phenotype in the Ghanaian GWAS samples by logistic regression analyses (mach2dat software; (www.sph.umich.edu/csg/yli/software.html)). To control for population stratification, we performed a principal component analysis (PCA) with the Eigenstrat software (http://genepath.med.harvard.edu/~reich/Software.htm) and used the first three components as covariates in logistic regression analyses (mach2dat). As strong linkage disequilibria (LD) may bias the correct identification of principal components, a pruned set of SNPs was generated by removing one SNP of a pair with LD values (R^2) > 0.2 within a window of 1500 SNPs and by discarding SNPs of 4 chromosomal regions with known strong LD (chr8: 8..12Mb, chr6: 25..33.5Mb, chr11: 45..57Mb, chr5 :44..51.5Mb).

Imputed SNPs in the GWAS sample with minor allele frequencies (MAF) > 10%, RSQ values > 0.8 and *P* values < $1x10^{-5}$ were selected to be genotyped. These thresholds were chosen in order to reduce the risk of false positive test results due to known low imputation accuracy in African GWAS data sets. SNPs yielding equivalent association results after genotyping were further tested in a replication sample of 817 TB cases and 3805 control individuals. Statistical

analyses of SNPs in the entire data set including 2146 cases and 5652 controls were done by logistic regression tests adjusted for age, gender and ethnicity under an additive mode of inheritance. For each SNP yielding significant replication results, Mantel-Haenszel and Breslow-Day statistics were performed to test for heterogeneity of genetic effects across ethnic groups. To assess the amount of between-study heterogeneity, the Cochran Q statistic was calculated with odds ratios (OR) and standard errors derived from the logistic regression. All statistical calculations of replication data were performed using the STATA 10.0 software (Stata Corporation, College Station, TX, USA).

Power Calculations

Power calculations were performed for the Ghanaian GWAS group with the CaTS software (http://www.sph.umich.edu/csg/abecasis/CaTS/). A multiplicative model was assumed with a TB prevalence of 0.003, 20 % disease allele frequency, a genotype relative risk of 1.5 and *P* value level of $5x10^{-8}$. Applying the CaTS software on 1329 TB cases and 1847 controls resulted in 90% power to detect existing effects.

Supplementary References

1. Bellamy, R. *et al*. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. *N Engl J Med*. **338** 640-4 (1998).
2. Szeszko, J.S. *et al*. Resequencing and association analysis of the SP110 gene in adult pulmonary tuberculosis. *Hum Genet*. **121**, 155-60 (2007).
3. Thye, T. *et al*. MCP-1 promoter variant -362C associated with protection from pulmonary tuberculosis in Ghana, West Africa. *Hum. Mol. Genet*. **18**, 381-8 (2009).