ARTICLE

A Locus at 5q33.3 Confers Resistance to Tuberculosis in Highly Susceptible Individuals

Rafal S. Sobota,[1,2] Catherine M. Stein,[3,4] Nuri Kodaman,[1,2,3] Laura B. Scheinfeldt,[5,6] Isaac Maro,[7] Wendy Wieland-Alter,[2] Robert P. Igo, Jr.,[3] Albert Magohe,[8] LaShaunda L. Malone,[4] Keith Chervenak,[4] Noemi B. Hall,[3] Chawangwa Modongo,[9,10,11] Nicola Zetola,[9,10,11] Mecky Matee,[8] Moses Joloba,[12,13] Alain Froment,[14] Thomas B. Nyambo,[15] Jason H. Moore,[2,16] William K. Scott,[17] Timothy Lahey,[2] W. Henry Boom,[4,12] C. Fordham von Reyn,[2] Sarah A. Tishkoff,[5,18] Giorgio Sirugo,[19] and Scott M. Williams[2,3,*]

Immunosuppression resulting from HIV infection increases the risk of progression to active tuberculosis (TB) both in individuals newly exposed to *Mycobacterium tuberculosis* (MTB) and in those with latent infections. We hypothesized that HIV-positive individuals who do not develop TB, despite living in areas where it is hyperendemic, provide a model of natural resistance. We performed a genome-wide association study of TB resistance by using 581 HIV-positive Ugandans and Tanzanians enrolled in prospective cohort studies of TB; 267 of these individuals developed active TB, and 314 did not. A common variant, rs4921437 at 5q33.3, was significantly associated with TB (odds ratio = 0.37, $p = 2.11 \times 10^{-8}$). This variant lies within a genomic region that includes *IL12B* and is embedded in an H3K27Ac histone mark. The locus also displays consistent patterns of linkage disequilibrium across African populations and has signals of strong selection in populations from equatorial Africa. Along with prior studies demonstrating that therapy with IL-12 (the cytokine encoded in part by *IL12B*, associated with longer survival following MTB infection in mice deficient in CD4 T cells), our results suggest that this pathway might be an excellent target for the development of new modalities for treating TB, especially for HIV-positive individuals. Our results also indicate that studying extreme disease resistance in the face of extensive exposure can increase the power to detect associations in complex infectious disease.

Introduction

Mycobacterium tuberculosis (MTB) and HIV are major causes of infectious-disease mortality worldwide.[1,2] In 2014, more than 25% of the 1.5 million people who died from active tuberculosis (TB) were HIV positive, and the great majority were in sub-Saharan Africa.[1,2] The immunosuppression resulting from HIV infection increases the risk of progression of an MTB infection to active disease both in individuals newly exposed to MTB and in those with latent infections.[3,4]

Various approaches to studying the relationship between host genetics and TB, including genome-wide association studies (GWASs) in sub-Saharan Africa, have been published.[5–7] The GWAS results have not explained a substantial portion of the genetic risk, possibly owing to the cross-sectional ascertainment of control individuals, which does not guarantee MTB exposure; this will most likely reduce estimated effect sizes as a result of increased noise. These studies compared immunocompetent case and control individuals in search of common variation that affects disease risk in the general population. An alternative approach, which we have taken, is to identify cohorts at extreme risk and focus on genetic factors that prevent the development of active disease despite this increased vulnerability.

All previous studies of human genetic susceptibility to TB have used HIV-seropositive status as an exclusion criterion or adjusted for it as a potential confounder in the analyses.[6–9] Here, we hypothesized that HIV-positive individuals who do not develop TB, despite living in areas hyperendemic for MTB, are highly resistant. Using prospective cohorts of TB in East Africa, we compared this extreme control group to HIV-positive individuals with TB. We identified a genome-wide-significant variant in a regulatory region near *IL12B* (MIM: 161561), a gene involved in cell-mediated immunity against intracellular bacteria.

Subjects and Methods

Study Populations

Samples were analyzed from three prospective studies, two from Tanzania and one from Uganda. All three were designed to assess different aspects of TB risk or progression but had comparable data

[1]Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN 37232, USA; [2]Geisel School of Medicine, Dartmouth College, Hanover, NH 03755, USA; [3]Department of Epidemiology and Biostatistics, Case Western Reserve University, Cleveland, OH 44106, USA; [4]Tuberculosis Research Unit, Case Western Reserve University, Cleveland, OH 44106, USA; [5]Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA; [6]Coriell Institute for Medical Research, Camden, NJ 08103, USA; [7]Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan; [8]Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; [9]Division of Infectious Diseases, University of Pennsylvania, Philadelphia, PA 19104, USA; [10]Botswana-UPenn Partnership, 214 Independence Avenue, Gaborone, Botswana; [11]Department of Medicine, University of Botswana, Gaborone, Botswana; [12]Uganda-CWRU Research Collaboration, Kampala, Uganda; [13]College of Health Sciences Makerere University and Mulago Hospital, Kampala, Uganda; [14]UMR 208, Institut de Recherche pour le Développement and Musée de l'Homme, Muséum National d'Histoire Naturelle, 75116 Paris, France; [15]Department of Biochemistry, Muhimbili University of Health and Allied Sciences, Dar es Salaam, Tanzania; [16]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; [17]John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL 33136, USA; [18]Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA; [19]Centro di Ricerca, Ospedale San Pietro Fatebenefratelli, 00189 Rome, Italy
*Correspondence: smw154@case.edu
http://dx.doi.org/10.1016/j.ajhg.2016.01.015.

available for most variables. All individuals who participated in our study provided written informed consent.

Tanzania

Participants were recruited from the extended follow-up cohort of the DarDar vaccine trial. This cohort has been previously described.[10] In brief, the DarDar trial was a randomized double-blind phase 3 trial of an inactivated whole-cell mycobacterial vaccine (SRL 172). Subject enrollment occurred between 2001 and 2005, and the study was concluded in 2008. All enrolled individuals were HIV-positive adults (>18 years old) with a CD4 count > 200 cells/μl, had a Bacille Calmette-Guérin childhood vaccination scar, and were TB negative at the time of enrollment. The final study population consisted of 1,007 individuals in the placebo arm and 1,006 in the vaccine group. Participants were routinely followed up every 3 months and evaluated for active TB through a physical examination, a chest radiograph, sputum samples for culture and acid-fast bacilli (AFB) stain, and phlebotomy for an automated mycobacterial blood culture. At the conclusion of the study, a cohort of 800 total participants in both the placebo and vaccine arms were selected for extended follow-up. These individuals have been evaluated for active TB once a year since the conclusion of the study. We recruited 304 of the extended-follow-up participants between September and December 2013 during their routine visits. 36 of the individuals had been diagnosed with definite or probable TB since the onset of the trial. Diagnostic criteria are described in von Reyn et al.[10] Individuals who did not develop TB during follow-up but stated that they had previous active TB were excluded from the study because those diagnoses could not be confirmed (exclusion criteria and the remaining sample size are provided in Table S1A). As with most genetic association studies of TB, it was not possible to determine with certainty whether our control subjects were exposed to MTB during the course of the study. However, such exposure in Dar es Salaam is highly likely. This conclusion is based on the length of follow-up (at least 8 years) and the high prevalence of TB in Tanzania, which is among the highest in the world at 528 per 100,000 individuals.[11] Additionally, this prevalence estimate is based on a country average, and it has been shown that such national-surveillance values underestimate urban incidence.[12]

We also used previously collected samples from the DarDar nutrition study in Dar es Salaam, Tanzania.[13] The DarDar nutrition study was a randomized, controlled trial assessing the effectiveness of a protein-calorie supplement (PCS) on standard TB and HIV treatment in women. All enrolled women were HIV-positive adults (>18 years old) and had newly diagnosed active TB. 150 participants were randomized to either a PCS arm or a multivitamin control group. Treatment protocols recommended by the World Health Organization were followed through the Tanzanian Ministry of Health National Tuberculosis and Leprosy Program. Subjects were followed monthly until the completion of the study in 2014. Available samples from 85 participants of the DarDar nutrition study were used.

Uganda

We used samples from participants of the household-contact (HHC) study conducted in Kampala, Uganda, from 2002 to 2014. This cohort has been previously described.[14,15] In brief, the HHC combined case-control and family-based designs to analyze the genetic epidemiology of TB. Individuals diagnosed with new active TB were referred to the study through the Uganda National Tuberculosis and Leprosy Programme, and those who consented were enrolled as index subjects. Relatives and unrelated individuals living within the same household were subsequently enrolled and evaluated for active TB, latent TB, and HIV. Both index and incident subjects of active TB were given the recommended therapy.[16] Importantly, selecting our TB case and control individuals from the incident contacts guaranteed exposure of the control individuals because of their proximity to the index individuals.[14] We analyzed samples from a subset of 263 HIV-positive individuals. To assure a similar age distribution between the Ugandan and Tanzanian cohorts, we removed individuals under the age of 18 years from the analysis. Also, individuals who did not develop TB during the study follow-up but stated that they had previous active TB were removed (Table S1A). Data for 69 previously genotyped samples from HIV-positive individuals from this cohort were also available. To determine whether HIV status of the original index individuals affected TB transmissibility, we evaluated whether the HIV status of index individuals in the entire HHC study associated with TB status in the contacts, and it did not ($p = 0.91$).

Equatorial African Populations Used in Selection Analyses

Individuals from a previous study of six populations in Cameroon were included in this study (n = 125).[17] The Baka (n = 25), Bakola (n = 29), and Bedzan (n = 13) are Niger-Kordofanian Bantu-speaking Western Pygmy hunter-gatherer populations. The Ngumba (n = 20), Southern Tikar (n = 19), and Lemande (n = 19) are neighboring Niger-Kordofanian Bantu-speaking agricultural populations. Data were also available for 12 Yoruba individuals living in Nigeria. The Yoruba participants were analyzed with the Ngumba, Southern Tikar, and Lemande groups as a Niger-Kordofanian-West population sample. An additional 18 Datog from Tanzania comprised the third equatorial African grouping.

These studies were approved by the institutional review boards at Muhimbili University for Health and Allied Sciences, Dartmouth College, Uganda National Council for Science and Technology, University Hospitals of Cleveland, and the University of Pennsylvania for the appropriate samples.

DNA Isolation and Genotyping

For participants from the DarDar vaccine trial, 5 ml of whole blood was drawn into EDTA-coated tubes (BD Biosciences) and immediately stored at 4°C. DNA was extracted the day of the phlebotomy with the Gentra Puregene Blood Kit (QIAGEN) in accordance with the manufacturer's recommendations in Dar es Salaam. For participants of the DarDar nutrition and HHC studies, buffy coats were isolated on site and shipped to Dartmouth College for DNA extraction. DNA was isolated from buffy coats with the QIAamp DNA Blood Mini Kit (QIAGEN). All DNA samples were stored at −80°C prior to genotyping.

All available samples from the DarDar vaccine trial (n = 304), the DarDar nutrition study (n = 85), and the HHC study (n = 263) were selected for genotyping. DNA quality was evaluated according to the 260/280 ratio with a NanoDrop 2000 spectrophotometer (Thermo Scientific) and an electrophoresis quality score. After quality control, a total of 639 samples were genotyped with the Human Core Exome BeadChip (542,585 SNPs) at the Hussman Institute for Human Genetics in Miami. After applying the exclusion criteria of adult age (≥18 years old) and no history of prior active TB, we analyzed the remaining participants, who consisted of 283 participants from the DarDar vaccine trial, 75 participants from the DarDar nutrition study, and 234 participants from the HHC study (Table S1A). SNPs with a genotyping call rate > 0.95 and a Hardy-Weinberg equilibrium p value > 1.00×10^{-4} were retained. All remaining participants had a per-individual genotyping call rate > 0.95. A sex check was performed, and cryptic relatedness was

evaluated in PLINK (v.1.07).[18] From each pair of related study participants, one individual was randomly removed (pi-hat > 0.20). The final study population genotyped on the Exome BeadChip included 278 participants from the extended follow-up of the DarDar vaccine trial, 65 participants from the DarDar nutrition study, and 213 participants from the HHC study.

Samples from 64 additional HIV-positive Ugandan HHC participants were genotyped with an Illumina HumanOmni5-Quad BeadChip (4.8 million SNPs). DNA was extracted from buffy coats with the QIAamp DNA Mini Kit (QIAGEN) and quantified with NanoDrop and Qubit. Genotyping and DNA quality checking were done at the Genomics Core at Case Western Reserve University. Further quality-control steps for samples included checking for sex-mismatch errors, relationship errors (within the larger dataset of 483 samples), consistency of blind duplicates, call rate (>95%), tenth percentile GenCall score (>0.42), visual inspection of the B-allele-frequency plot of samples with a <98% call rate, and unusually high autosomal heterozygosity. Individuals < 18 years old and those related to the participants genotyped with the Exome BeadChip (pi-hat > 0.20) were excluded. 25 of the additional 64 individuals remained after we applied these exclusion criteria.

For the samples used in the selection analyses, after phlebotomy, white blood cells were isolated by a modified salting-out procedure,[19] and DNA was extracted with a Puregene DNA Extraction Kit (Gentra Systems).[17] The samples were genotyped with the Illumina Human 1M-Duo BeadChip.[17] We evaluated cryptic relatedness by randomly removing individuals with pi-hat > 0.25, and we removed SNPs with call rates less than 95%. All quality-control measures were done in PLINK (v.1.07).[18] Admixture was evaluated by principal-component analysis in R[20] and STRUCTURE as previously described.[17]

Immunological Data

Enrollment CD4 counts were available for 437 subjects, including all participants of the extended follow-up of the DarDar vaccine trial and nutrition study and HHC study participants who enrolled after 2004. CD4 counts were not ascertained for individuals who entered the HHC study before 2004 (n = 144) because antiretroviral therapy was unavailable in Uganda before then.

All participants of the DarDar nutrition study were affected by TB, and the design of the HHC study guaranteed exposure to MTB. To address possible confounding of our results by exposure in the extended follow-up of the DarDar vaccine trial, we leveraged available interferon-γ (IFN-γ) release assay (IGRA) data. Immune response to *Mycobacteria* was assessed with an IFN-γ ELISA, a tritiated thymidine lymphocyte proliferation assay (LPA), and ELISA for antibodies to the MTB glycolipid lipoarabinomannan (LAM). The assays used in this study have been previously described.[21] In brief, IFN-γ and LPA assays used four different antigens in the samples from Tanzania: *Mycobacterium vaccae* sonicate (2 mcg/ml), *MTB* antigen 85 (Ag85; 1 mcg/ml), *MTB* early secretory antigenic target 6 (ESAT-6; 2 mcg/ml), and *MTB* whole-cell lysate (1 mcg/ml).[21] Media alone served as a negative control, and phytohemagglutinin (2.5 mcg/ml; Sigma) was used as a positive control.[21]

Statistical Analyses

Logistic regression in PLINK (v.1.07)[18] was used for testing the association between single SNPs in an additive model and active TB case-control status. A priori power analyses in QUANTO revealed that a minor allele frequency (MAF) > 0.20 should be used in both cohorts in the single-SNP association analyses.[22] Summary statistics for available covariates were calculated in STATA (v.11.2).[23] Covariate data were analyzed for differences between TB case and control individuals. For categorical variables, analyses were performed with χ^2 tests. For continuous variables, either t tests or non-parametric Wilcoxon rank-sum tests were used, where appropriate. Covariates significantly associated with case-control status were included in the final models (Tables S1B–S1E). Vaccine status did not associate with active TB in the voluntary DarDar extended follow-up (p value = 0.122; Table S1B); therefore, it was not adjusted for.

Because the Tanzanian cohorts were recruited in Dar es Salaam and the Ugandan cohort was recruited in Kampala, both large urban centers, admixture was a likely confounder in this study. To adjust for population structure within each country, we used the SNPRelate package in R[20,24] to calculate principal components separately for participants from Uganda and Tanzania by using SNPs with $r^2 < 0.1$, a MAF > 0.2, and a genotype call rate > 0.95. Analyses of the combined cohorts were also adjusted for country of origin. Data on self-reported tribal identity were available for the individuals from Uganda. The first, sixth, and seventh principal components were significant in predicting membership from the predominant tribe, Muganda (63% of the participants), versus all others. Because self-described ethnicity was not available for Tanzania, we conservatively chose to use ten principal components (Table S2). We used locus zoom to plot the region of the SNP with the strongest association[25] (Figure 1). SNPs in the region of interest (rs4921437 position ± 1 Mb) were imputed with IMPUTE2 (v.2.3.1) and one multi-population phased reference panel from the 1000 Genomes Project.[26–28]

We performed logistic-regression analysis of TB in the combined cohort while adjusting for CD4 count at enrollment, sex, country of origin, and ten principal components in 437 individuals for whom CD4 data were available (Table S3A). To assess the effect of variable immunosuppression of case and control individuals on the association results, we repeated this analysis without adjustment for CD4 counts in the same subset of individuals (Table S3B). We performed logistic-regression analyses of TB with IGRA results in the extended follow-up of the DarDar vaccine trial (Table S4).

We examined all pairwise haplotypes within the locus of interest, i.e., the region that included our most significant SNP, rs4921437. The additional SNPs were selected on the basis of MAF > 0.05. 11 unimputed SNPs passed quality control and spanned *IL12B* and *UBLCP1* (MIM: 609867), and we used these to test for haplotype association (Figures S1A and 1B). This region was defined with Haploview.[29] We performed haplotype association analyses with UNPHASED (v.3.1.7)[30] while adjusting for the same covariates as in the single-SNP association analyses. We then performed a single three-SNP haplotype analysis with SNPs from the two most significant pairwise haplotypes (Table S5).

Additionally, for each cohort, we used a χ^2 test in STATA (v.11.2)[23] to compare haplotypes from the case group only, the control group only, and the entire dataset to corresponding haplotypes in phase 3 samples of the HapMap Project.[31,32] MAFs were obtained from the 1000 Genomes Project and the Human Genome Diversity Project.[26,33,34]

Selection Analysis

Phase was inferred with the Beagle software package.[35] Using LDhat (v.2.1), we generated fine-scale recombination maps with 25 males and 25 females from each of the following groups: unrelated individuals from two HapMap3 release 2 populations,[32] Yoruba from Ibadan, Nigeria (YRI), and Luhya from Webuye,

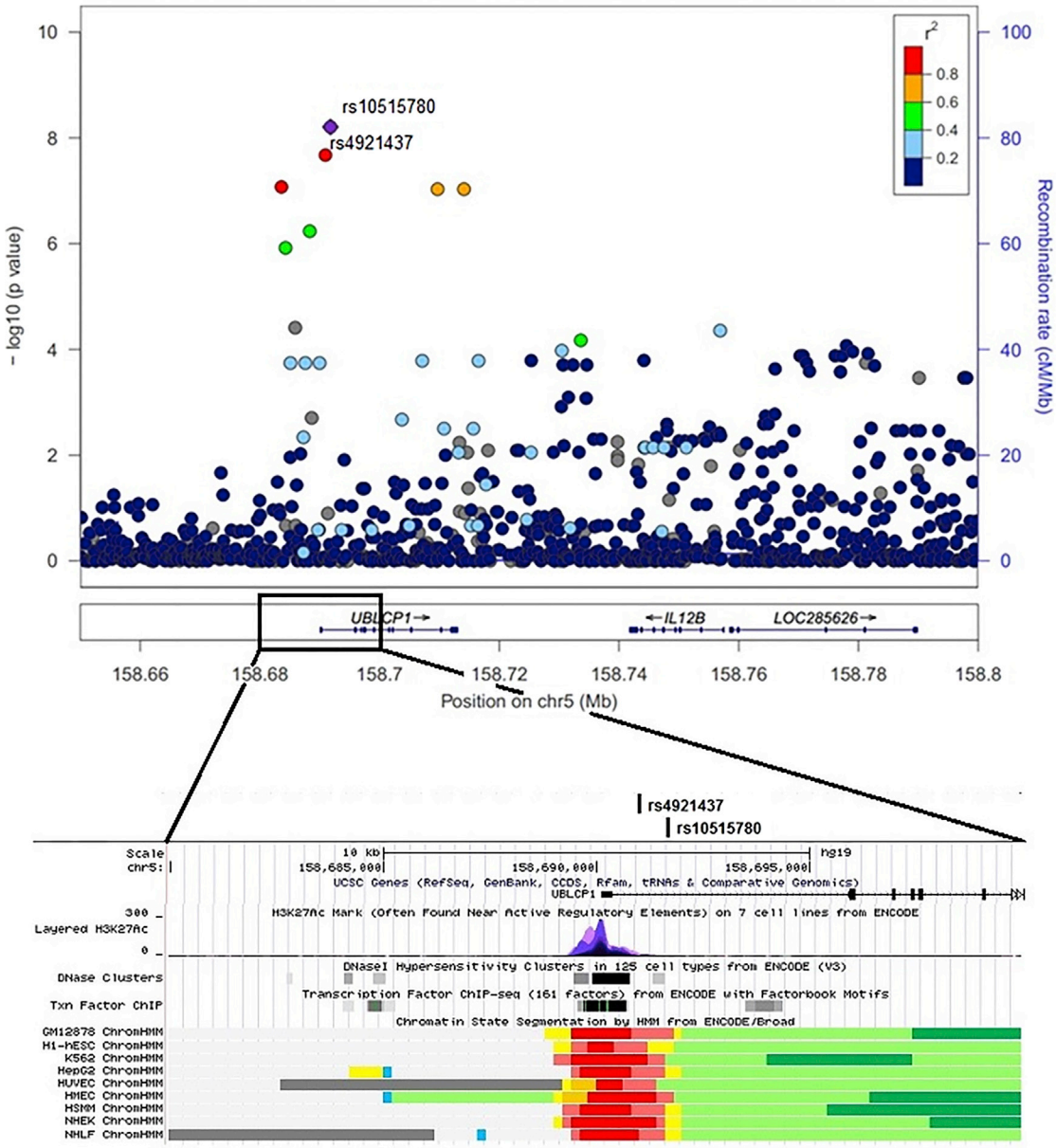

Figure 1. Results for Imputed SNPs in a 50 kb Region around rs4921437 from an Additive Genetic Model for Association with TB in the Combined Cohort

The imputed SNP with the most significant association, rs10515780, is in a purple. Colors represent LD (measured with r^2) between available variants and rs10515780. Annotation of chromatin-state segmentation in the zoomed region around rs4921437 is from the UCSC Genome Browser (hg19) for nine human cell types (chr5: 158,680,000–158,700,000). Bright red indicates an active promoter, light red indicates a weak promoter, yellow represents a weak or poised enhancer, light green represents a weak transcribed state, and dark green indicates transcriptional transition or elongation. An H3K27Ac histone mark aligns with the most significant SNPs.

Kenya (LWK).[36] All ancestral alleles for the SNPs included in this analysis were established with genome-wide sequences of non-human primates, chimpanzee, orangutan, and rhesus macaque, downloaded from the UCSC Genome Browser.[37] Approximately 5% of available SNPs could not be assigned an unambiguous ancestral state and were thus removed prior to selection analysis, as were SNPs with a MAF < 0.05.[38] Selection was assessed with the integrated haplotype score (iHS) test statistic, as described by Voight et al.[38] iHS values were standardized to a mean of 0 and a unit variance with respect to SNPs with similar derived allele frequencies.[38] iHS values in the top 0.1% of the distribution of absolute values were considered top candidates for selection.

Functional Annotation

The ENCODE Project[39] was accessed via the UCSC Genome Browser,[37] and data provided by the Bernstein lab were used for functional annotation.

Results

The combined study population consisted of 267 individuals with TB (100 from Tanzania and 167 from Uganda) and 314 control individuals (243 from Tanzania and 71 from Uganda; Table S1). Male gender associated with active

Table 1. SNPs Associated with Active TB at $p < 5.00 \times 10^{-5}$

SNP	Chr	Minor Allele	MAF	OR	95% CI	p Value	Gene
Uganda							
rs4921437	5	T	0.210	0.279	(0.167, 0.467)	1.18×10^{-6}	*UBLCP1, IL12B*
rs7297313	12	C	0.337	0.349	(0.22, 0.555)	8.6×10^{-6}	*GRIN2B*
rs859063	1	G	0.345	3.057	(1.82, 5.135)	2.40×10^{-5}	*SLC44A3*
Tanzania							
rs2681052	7	T	0.387	2.649	(1.717, 4.087)	1.07×10^{-5}	*THSD7A*
rs17048476	3	A	0.419	2.433	(1.618, 3.657)	1.91×10^{-5}	*AK124857*
rs2587469	10	C	0.457	0.423	(0.284, 0.63)	2.26×10^{-5}	*ADAMTS14*
rs9893385	17	G	0.467	0.414	(0.275, 0.622)	2.26×10^{-5}	*ABCA8*
rs3860173	9	T	0.361	2.276	(1.556, 3.329)	2.27×10^{-5}	*RGS3*
rs1524713	1	T	0.277	2.427	(1.595, 3.693)	3.47×10^{-5}	*DAB1*
rs2807348	1	C	0.255	0.354	(0.217, 0.579)	3.48×10^{-5}	*WNT4*
rs4700255	5	G	0.323	2.337	(1.56, 3.502)	3.87×10^{-5}	*ACTBL2*
rs16870583	5	C	0.251	2.317	(1.549, 3.466)	4.35×10^{-5}	*IRX2*
rs930205	5	G	0.294	2.393	(1.572, 3.642)	4.64×10^{-5}	*SH3TC2*
rs557438[a]	5	T	0.368	0.412	(0.269, 0.632)	4.70×10^{-5}	*MCC*
rs4921437	5	T	0.209	0.484	(0.296, 0.792)	3.84×10^{-3}	*UBLCP1, IL12B*
Combined Cohorts							
rs4921437	5	T	0.209	0.374	(0.266, 0.528)	2.11×10^{-8}	*UBLCP1, IL12B*
rs8028149	15	C	0.416	0.509	(0.385, 0.672)	1.96×10^{-6}	*VPS13C*
rs1616723	6	G	0.201	0.439	(0.309, 0.624)	4.13×10^{-6}	*GLO1*
rs955263	4	T	0.343	1.848	(1.391, 2.454)	2.26×10^{-5}	*SORBS2*
rs12636260	3	T	0.257	1.977	(1.442, 2.709)	2.26×10^{-5}	*ZPLD1*[b]
rs4768760	12	C	0.407	0.562	(0.429, 0.735)	2.57×10^{-5}	*SLC38A4*
rs844669	7	G	0.301	0.521	(0.385, 0.707)	2.68×10^{-5}	*CALN1*
rs4236914	8	T	0.237	1.912	(1.412, 2.591)	2.84×10^{-5}	*SFRP1*
rs4860106	4	G	0.460	0.580	(0.449, 0.749)	3.03×10^{-5}	*LPHN3*
rs1482868	6	T	0.279	0.544	(0.407, 0.728)	4.29×10^{-5}	*F13A1*
rs2346943	16	G	0.326	1.793	(1.354, 2.373)	4.48×10^{-5}	*RBFOX1*

Results are presented for the Ugandan cohort, the Tanzanian cohorts, and the combined cohorts. All analyses were adjusted for principal components and sex; the combined analyses were also adjusted for country of origin. Abbreviations are as follows: Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; and OR, odds ratio.
[a]Other SNPs between rs557438 and rs4921437 are statistically significant but are not relevant here.
[b]Closest protein-coding gene.

TB in univariate logistic regression (odds ratio [OR] = 1.78, 95% confidence interval [CI] = 1.23–2.57, p = 0.002). Age did not significantly associate with TB (p = 0.186).

Using logistic regression, we observed a genome-wide-significant association between common variant rs4921437 and TB (OR = 0.37, 95% CI = 0.27–0.53, $p = 2.11 \times 10^{-8}$; Table 1 and Figure S2A). Results were consistent when the samples were stratified by country (Uganda: OR = 0.28, 95% CI = 0.17–0.47, $p = 1.18 \times 10^{-6}$; Table 1; Tanzania: OR = 0.48, 95% CI = 0.30–0.79, $p = 3.84 \times 10^{-3}$; Table 1 and Figures S2B and S2C). The genomic inflation factor (λ) for the final adjusted logistic-regression model was 1.029 for analysis of the cohorts combined, 1.023 for analysis of samples from Tanzania, and 1.056 for analysis of samples from Uganda. The cluster plot for rs4921437 shows that genotype calling for this SNP was unambiguous for virtually all samples (Figure S3). These results indicate not only that the combined association is highly significant but also that each cohort alone provides strong evidence of association; rs4921437 was the most significant SNP in

Uganda and was among the most significant in Tanzania (Figures S2A–S2C).

We imputed all SNPs within 1 Mb of rs4921437. Imputed variant rs10515780, only 842 bp from rs4921437, displayed the most significant association with TB (combined cohort OR = 0.34, 95% CI = 0.24–0.49, $p = 6.21 \times 10^{-9}$; Figure 1 and Table S6). This SNP was in very high linkage disequilibrium (LD) with the index SNP ($D' = 1$ in both cohorts, $r^2 = 0.84$ in Tanzania and 0.89 in Uganda) and was imputed with a high degree of certainty (imputation certainty was 0.993 in both Tanzania and Uganda).

Variants rs4921437 and rs10515780 map to an intron in *UBLCP1*, ubiquitin-like-domain-containing C-terminal domain phosphatase 1. Variant rs4921437 is 51 kb away from the 3′ UTR of *IL12B*. According to ENCODE,[39] both rs4921437 and rs10515780 are located in an H3K27Ac histone mark that most likely represents a regulatory region (Figure 1).

Adjustment for CD4 counts in the subset of participants for whom we had data (n = 437) demonstrated that variable immunosuppression among case and control individuals was not a confounding factor (Table S3). Prior MTB exposure of the control group is essential, and our cohorts virtually assured it. All individuals from the DarDar nutrition study were affected by TB. The HHC study design guarantees exposure of the TB-negative individuals to MTB, given that co-habitation with an individual with active TB was a prerequisite for enrollment. Although direct contact with index individuals with active TB was not assessed in the DarDar vaccine trial, 8 years had elapsed between the completion of enrollment for the vaccine trial and recruitment into our study, and TB prevalence in Tanzania is among the highest in the world (see Subjects and Methods). Therefore, participants were both HIV positive and living in an area hyper-endemic for TB at least this long, virtually guaranteeing MTB exposure. Nevertheless, to address possible confounding by lack of exposure, we assessed response of IGRAs and LPAs by case-control status and observed no statistically significant differences (Table S4). Although the Ugandan and Tanzanian studies were designed differently, the consistency in genetic association results underscores the robustness of this finding.

Analysis of the *UBLCP1-IL12B* region in the unimputed dataset revealed blocks of LD between rs4921437 and SNPs in *IL12B* in the Ugandan and Tanzanian cohorts, as well as in African samples from phase 3 of the HapMap Project (Figure S1). A three-variant haplotype of rs4921437, rs4921468 (intergenic), and rs3213094 (intronic in *IL12B*) had a more significant association with TB than rs4921437 alone in the combined cohort ($p = 4.56 \times 10^{-15}$) (Table S5). Results were consistent when the Ugandan and Tanzanian cohorts were analyzed separately ($p = 1.98 \times 10^{-11}$ and 4.56×10^{-5}, respectively). When we conditioned an rs4921468-rs3213094 haplotype analysis on the main-effect variant, rs4921437, the combined cohort haplotype association was significant at a multiple-testing-corrected threshold based on the number of SNPs remaining after quality control (p value 1.19×10^{-7}; threshold for 175,906 variants $= 2.84 \times 10^{-7}$).

Frequencies of the three-variant haplotype were not significantly different between the Ugandan and Tanzanian cohorts, nor among founders of the LWK or ASW (African ancestry in Southwest USA) from phase 3 of HapMap (Table S7).[32] This conservation of haplotype frequencies among these Africa-derived samples was unexpected given the known diverse history of these populations and genomic differentiation between them.[40] Frequencies of the three-variant haplotype in Uganda and Tanzania were significantly different from those in all other HapMap phase 3 populations (Table S7).[32]

We assessed whether allele-frequency distributions were consistent with selection. The ancestral, minor allele of rs4912437 (T), which associated with TB resistance, had a MAF of 0.21 in our combined cohort (0.21 in both Tanzania and Uganda). In the African, European, and South Asian origin populations from phase 3 of the HapMap Project and the 1000 Genomes Project, the T allele frequency varied between 0.11 and 0.36 (Table S8).[32] The MAFs in the Biaka Pygmy and Mbuti Pygmy populations of the Human Genome Diversity Project were 0.45 and 0.46, respectively. Of note, the ancestral variant had a very low frequency in East Asian populations for which data are available (MAF < 0.025), suggesting either a loss due to drift or its maintenance due to selection where the allele remained.

We found highly similar LD patterns over the entire *IL12B* region in our Tanzanian and Ugandan cohorts (Figure S1). Considering previously documented patterns of less LD and greater variation of LD among African populations,[41] we did not expect to find such striking similarities. We then evaluated signatures of selection by using an iHS[38] in equatorial African populations (West Pygmy, Datog, and Niger-Kordofanian-West) for which we had available data with coverage adequate for these analyses.[42] The iHS test is most sensitive to recent adaptive events (<25,000 years ago), appropriate here because TB-related selective pressure is thought to have increased after agriculture-associated growth in population density.[38,42] Two intronic SNPs in *IL12B* (4,672 bp apart) were in the top 0.1% of the distribution of absolute values of iHSs for one million genotyped SNPs in at least one equatorial African population (Table 2). The region surrounding rs3213093 had a signature of selection in the top 0.1% of the distribution in the Datog population sample, whereas the region centered on rs2421047 had a signature of selection in the top 0.1% of the distribution in the Niger-Kordofanian-West population samples (Table 2). In our study cohorts, rs3213093 and rs2421047 were in high LD with each other ($D' = 0.98$–1.0) and with rs4921437 ($D' = 0.79$–0.92) (Figure S4). The iHS values of rs4921437 in the Datog, Western Pygmy, and Niger-Kordofanian-West populations were not in the top 0.1% of the respective distributions.

We replicated several variants that have been previously reported to associate with TB. A SNP in *PTX3*

Table 2. Analyses of iHS Values

SNP	|iHS| in the Datog	|iHS| Percentage in the Datog	|iHS| in the Niger Kordofanian West	|iHS| Percentage in the Niger Kordofanian West	|iHS| in the Western Pygmy	|iHS| Percentage in the Western Pygmy
rs3213093	3.915	0.054%	3.591	0.104%	2.599	1.23%
rs2421047	3.088	0.371%	3.616	0.097%	2.636	1.13%
rs4921437	0.391	68.84%	0.742	45.40%	0.997	31.44%

Absolute values of normalized iHS values are provided for two *IL12B* SNPs that were in the top 0.1% of the |iHS| distribution in any of the available equatorial African populations. Also shown are the percentages and corresponding values for rs4921437. The following abbreviation is used: iHS, integrated haplotype score.

(MIM: 602492), rs3816527, associated with TB at a p value of 4.69×10^{-4}. This SNP is 715 bp away from rs1840680, a previously associated variant in a West African population,[43] and it is in strong LD with it in the LWK 1000 Genomes population ($r^2 = 0.86$). We imputed rs1840680, and it associated with a p value of 0.0013 and an OR of 1.63 (Table 3). Variants rs2057178, rs4331426, and rs4733781, which previously associated with TB in genome-wide analyses, had ORs similar to those previously reported (rs2057178: OR = 0.84 in our combined cohort versus 0.77 in Thye et al.[6]; rs4331426: OR = 1.16 in our combined cohort versus 1.18 in Thye et al.[7]; and rs4733781: OR = 0.92 in our combined cohort versus 0.84 in Curtis et al.[5]). However, our sample size was smaller than those in the reported studies, and the SNPs were not significant in our analyses (Table 3). These results, even though not significant, support the validity of our cohort design and phenotype definitions.

Discussion

Our approach to the study of TB genetics focuses on resistance as opposed to susceptibility. Specifically, we hypothesized that HIV-immunocompromised individuals who resist TB despite prolonged exposure represent an extreme phenotype that can provide new insights into TB pathogenesis. We identified a region encompassing *IL12B*, a gene originally shown to underpin a very rare monogenic immunodeficiency that promotes high susceptibility to several mycobacterial infections.[44] Unlike Mendelian immunodeficiency-causing mutations identified in subjects presenting with hyper-susceptibility to infection, resistance variants only emerge as conferring noticeable protection in high-risk yet disease-free subjects. Our HIV-positive cohorts enabled us to find a variant that has a much larger effect size than previously associated common SNPs, one detected with a sample size well below that of most published GWASs. Also, in contrast to most previous TB association studies that have used cross-sectional selection of the control group, in our study the control individuals were ascertained as part of long-term, well-characterized, prospective cohorts. This is a distinct strength of our design; individuals were routinely tested for TB. Furthermore, in the subset of individuals for whom CD4-count data and IGRA data were available, we did not observe any confounding by extent of immunosuppression or exposure to MTB.

Our SNP of primary interest is located in an area enriched with a histone acetylation mark often found in active regulatory elements, suggesting functionality. The H3K27Ac annotation indicates a modification that acts in concert with an H3K4Me1 mark to differentiate active enhancer elements from poised or inactive ones.[45,46] Such chromatin modifications are associated with higher regulatory responses at enhancer transcription factor binding sites,[45] possibly potentiating the effect of DNA polymorphisms in the region.

The variants with the strongest association with TB are located in introns of *UBLCP1*, previously shown to regulate nuclear proteasome activity.[47] Variants in this gene have been previously associated with carotid stenosis,[48] psoriasis,[49] and platelet aggregation.[50] Importantly, these SNPs are also in proximity to *IL12B*, a gene previously associated with both TB[9,51] and leprosy.[52] *IL12B* variants have also been associated with carotid stenosis[53] and psoriasis,[54] possibly owing to LD with SNPs in *UBLCP1*. *IL12B* encodes p40, one of two subunits of interleukin-12 (IL-12)[55] and IL-23.[56] IL-12 is a cytokine secreted by phagocytes and dendritic cells, and it causes differentiation of naive T cells into T helper 1 (Th1) cells and stimulates IFN-γ production from T cells and natural killer cells.[57,58] In experimental models, IL-12 restricted MTB proliferation throughout the course of infection and was a necessary factor in granuloma formation and antigen-specific delayed-type hypersensitivity.[59–61] Previous studies in humans have reported associations between IL-12 deficiency and predisposition to infections with *Mycobacteria* but have not demonstrated strong association with common functional variation.[8,62] IL-23 also plays a role in controlling MTB infection; it is essential for the differentiation of naive T cells into Th17 cells, which are responsible for the production of IL-17.[63] IL-17 is necessary for the targeting of neutrophils to the site of infection, and IL-17-deficient mice are unable to control highly virulent MTB strains.[63,64] Therefore, although our analyses do not provide direct evidence of association with *IL12B*, it remains a gene of substantial interest on the basis of both our results and prior data relating this locus to TB and related phenotypes.

Additional evidence supporting the detected association derives from the high concordance of LD patterns across several African-descent populations in the *IL12B* region. Analyses using other sub-Saharan populations demonstrated that the *IL12B* region has undergone positive selection. These results demonstrate strong selection in

Table 3. Logistic-Regression Results for SNPs Reported in Other TB Studies

SNP	Study	Chr	Minor Allele	n	OR	95% CI	p Value	Gene
rs2057178	Thye et al.[6]	11	A	581	0.836	(0.64, 1.10)	0.2004	*WT1*
rs4331426	Thye et al.[7]	18	G	556	1.162	(0.68, 1.51)	0.259	*GATA6*
rs4733781[a]	Curtis et al.[5]	8	C	581	0.916	(0.67, 1.25)	0.578	*ASAP1*
rs1840680[a]	Olesen et al.[43]	3	A	556	1.635	(1.21, 2.21)	0.0013	*PTX3*

Results are from the combined cohort, adjusted for principal components, sex, and country of origin. Abbreviations are as follows: Chr, chromosome; CI, confidence interval; and OR, odds ratio.
[a]Imputed genotypes.

the region of interest in populations related to our study cohorts, although the selection signature appears to be strongest within *IL12B* and not in the regulatory region in which rs4921437 is located. This lower value might be due to the SNP location, near the end of the *UBLCP1-IL12B* haplotype block.[38] Because the *IL12B* locus has pleiotropic effects, we cannot rule out that other historical intra-cellular pathogens, such as *Plasmodium falciparum* or *Yersinia pestis*,[65] have exerted selective pressure on this gene.

The high prevalence of HIV-MTB co-infection in sub-Saharan Africa, along with persisting rates of multi-drug-resistant TB, necessitates development of new treatment modalities tailored to this population.[2] Previous studies of exogenous IL-12 administration in murine models have shown efficacy in limiting disease progression. Cooper et al. demonstrated that survival of MTB-infected mice with severe combined immunodeficiency was extended with IL-12 treatment.[66] Nolt et al. reported improved survival, diminished bacterial loads, and decreased pulmonary inflammation in CD4-deficient mice 8 weeks after IL-12 therapy.[67] Reduced granuloma formation and decreased inflammation were also observed in the lungs of wild-type mice.[67] The strong association between TB and a common variant near the gene encoding an IL-12 subunit in an HIV-positive host population links the murine observations to the human phenotype. IL-12 and members of its pathway are therefore promising candidates for development of new TB treatments, which could be efficacious in both HIV-positive individuals and the general population.

In conclusion, we have identified a locus conferring strong resistance to TB, thereby demonstrating that use of extreme phenotypes can be a powerful strategy for unraveling the host genetics of complex infectious diseases. This approach reduces the noise caused by both phenotypic and genetic heterogeneity, allowing detection of signals from major genes even in seemingly underpowered studies.

Accession Numbers

Institutional-review-board restrictions prohibit the deposition of our full genotype data, so accession numbers are unavailable. However, the authors will try to accommodate all legitimate requests for data sharing. To request access to the data, please contact either C.M.S. (cmj7@case.edu) or S.M.W. (corresponding author).

Supplemental Data

Supplemental Data include four figures and eight tables and can be found with this article online at http://dx.doi.org/10.1016/j.ajhg.2016.01.015.

Acknowledgments

R.S.S. was supported by Public Health Service award T32 GM07347 from the National Institute of General Medical Studies for the Vanderbilt Medical-Scientist Training Program. R.S.S., N.K., and S.M.W. were partially supported by NIH grant P20 GM103534. C.M.S., L.L.M., K.C., and W.H.B. were partially supported by the Tuberculosis Research Unit (grants N01-AI95383 and HHSN266200700022C/ N01-AI70022 from the National Institute of Allergy and Infectious Diseases). C.M.S., W.H.B., and R.P.I. were supported by National Heart, Lung, and Blood Institute (NHLBI) grant R01HL096811. C.M.S. and W.H.B. were partially supported by NHLBI grant R01HL10566113, and N.B.H. was supported by NHLBI grant T32HL007567. J.H.M. and S.M.W. were supported by NIH grant LM010098. J.H.M. was supported by NIH grants AI59694, AI116794, and LM009012. S.A.T. was partially supported by NIH grants 1R01GM113657-01 and 8DP1ES022577-04.

Received: October 1, 2015
Accepted: January 26, 2016
Published: March 3, 2016

Web Resources

The URLs for data presented herein are as follows:

OMIM, http://www.omim.org/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/
Quanto, http://biostats.usc.edu/Quanto.html
UCSC Genome Browser, https://genome.ucsc.edu/

References

1. UNAIDS. (2013). Global Report: UNAIDS report on the global AIDS epidemic 2013. http://www.unaids.org/sites/default/files/media_asset/UNAIDS_Global_Report_2013_en_1.pdf.
2. World Health Organization. (2015). Global Tuberculosis Report 2015. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1.
3. Selwyn, P.A., Schoenbaum, E.E., Davenny, K., Robertson, V.J., Feingold, A.R., Shulman, J.F., Mayers, M.M., Klein, R.S., Friedland, G.H., and Rogers, M.F. (1989). Prospective study of human immunodeficiency virus infection and

pregnancy outcomes in intravenous drug users. JAMA *261*, 1289–1294.
4. Di Perri, G., Cruciani, M., Danzi, M.C., Luzzati, R., De Checchi, G., Malena, M., Pizzighella, S., Mazzi, R., Solbiati, M., Concia, E., et al. (1989). Nosocomial epidemic of active tuberculosis among HIV-infected patients. Lancet *2*, 1502–1504.
5. Curtis, J., Luo, Y., Zenner, H.L., Cuchet-Lourenço, D., Wu, C., Lo, K., Maes, M., Alisaac, A., Stebbings, E., Liu, J.Z., et al. (2015). Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. Nat. Genet. *47*, 523–527.
6. Thye, T., Owusu-Dabo, E., Vannberg, F.O., van Crevel, R., Curtis, J., Sahiratmadja, E., Balabanova, Y., Ehmen, C., Muntau, B., Ruge, G., et al. (2012). Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat. Genet. *44*, 257–259.
7. Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I., Gyapong, J., Sirugo, G., Sisay-Joof, F., Enimil, A., Chinbuah, M.A., et al.; African TB Genetics Consortium; Wellcome Trust Case Control Consortium (2010). Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat. Genet. *42*, 739–741.
8. Picard, C., Fieschi, C., Altare, F., Al-Jumaah, S., Al-Hajjar, S., Feinberg, J., Dupuis, S., Soudais, C., Al-Mohsen, I.Z., Génin, E., et al. (2002). Inherited interleukin-12 deficiency: IL12B genotype and clinical phenotype of 13 patients from six kindreds. Am. J. Hum. Genet. *70*, 336–348.
9. Morris, G.A., Edwards, D.R., Hill, P.C., Wejse, C., Bisseye, C., Olesen, R., Edwards, T.L., Gilbert, J.R., Myers, J.L., Stryjewski, M.E., et al. (2011). Interleukin 12B (IL12B) genetic variation and pulmonary tuberculosis: a study of cohorts from The Gambia, Guinea-Bissau, United States and Argentina. PLoS ONE *6*, e16656.
10. von Reyn, C.F., Mtei, L., Arbeit, R.D., Waddell, R., Cole, B., Mackenzie, T., Matee, M., Bakari, M., Tvaroha, S., Adams, L.V., et al.; DarDar Study Group (2010). Prevention of tuberculosis in Bacille Calmette-Guérin-primed, HIV-infected adults boosted with an inactivated whole-cell mycobacterial vaccine. AIDS *24*, 675–685.
11. World Health Organization. (2015). Global Tuberculosis Report 2015, p. 143. http://apps.who.int/iris/bitstream/10665/191102/1/9789241565059_eng.pdf?ua=1.
12. Guwatudde, D., Zalwango, S., Kamya, M.R., Debanne, S.M., Diaz, M.I., Okwera, A., Mugerwa, R.D., King, C., and Whalen, C.C. (2003). Burden of tuberculosis in Kampala, Uganda. Bull. World Health Organ. *81*, 799–805.
13. Kim, F., Neke, N.M., Hendricks, K., Wamsele, J., Lukmanji, Z., Waddell, R., Maro, I., Connor, R., Mackenzie, T., Matee, M., et al. (2014). Deficiencies of macronutrient intake among HIV-positive breastfeeding women in Dar es Salaam, Tanzania. J. Acquir. Immune Defic. Syndr. *67*, 569–572.
14. 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.
15. Hall, N.B., Igo, R.P., Jr., Malone, L.L., Truitt, B., Schnell, A., Tao, L., Okware, B., Nsereko, M., Chervenak, K., Lancioni, C., et al.; Tuberculosis Research Unit (2015). Polymorphisms in TICAM2 and IL1B are associated with TB. Genes Immun. *16*, 127–133.
16. Blumberg, H.M., Burman, W.J., Chaisson, R.E., Daley, C.L., Etkind, S.C., Friedman, L.N., Fujiwara, P., Grzemska, M., Hopewell, P.C., Iseman, M.D., et al.; American Thoracic Society, Centers for Disease Control and Prevention and the Infectious Diseases Society (2003). American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. Am. J. Respir. Crit. Care Med. *167*, 603–662.
17. Jarvis, J.P., Scheinfeldt, L.B., Soi, S., Lambert, C., Omberg, L., Ferwerda, B., Froment, A., Bodo, J.M., Beggs, W., Hoffman, G., et al. (2012). Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. PLoS Genet. *8*, e1002641.
18. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
19. Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215.
20. R Development Core Team (2007). R: A language and environment for statistical computing (R Foundation for Statistical Computing).
21. Lahey, T., Mitchell, B.K., Arbeit, R.D., Sheth, S., Matee, M., Horsburgh, C.R., MacKenzie, T., Mtei, L., Bakari, M., Vuola, J.M., et al. (2011). Polyantigenic interferon-γ responses are associated with protection from TB among HIV-infected adults with childhood BCG immunization. PLoS ONE *6*, e22074.
22. Gauderman, W.J. (2002). Sample size requirements for matched case-control studies of gene-environment interaction. Stat. Med. *21*, 35–50.
23. StataCorp (2009). Stata Statistical Software: Release 11 (College Station, TX: StataCorp LP).
24. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics *28*, 3326–3328.
25. 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.
26. Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A.; 1000 Genomes Project Consortium (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65.
27. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
28. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda) *1*, 457–470.
29. 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.
30. Dudbridge, F. (2008). Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum. Hered. *66*, 87–98.
31. International HapMap Consortium (2003). The International HapMap Project. Nature *426*, 789–796.
32. International HapMap Consortium (2005). A haplotype map of the human genome. Nature *437*, 1299–1320.

33. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.
34. Rosenberg, N.A., Pritchard, J.K., Weber, J.L., Cann, H.M., Kidd, K.K., Zhivotovsky, L.A., and Feldman, M.W. (2002). Genetic structure of human populations. Science *298*, 2381–2385.
35. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
36. McVean, G.A., Myers, S.R., Hunt, S., Deloukas, P., Bentley, D.R., and Donnelly, P. (2004). The fine-scale structure of recombination rate variation in the human genome. Science *304*, 581–584.
37. Fujita, P.A., Rhead, B., Zweig, A.S., Hinrichs, A.S., Karolchik, D., Cline, M.S., Goldman, M., Barber, G.P., Clawson, H., Coelho, A., et al. (2011). The UCSC Genome Browser database: update 2011. Nucleic Acids Res. *39*, D876–D882.
38. Voight, B.F., Kudaravalli, S., Wen, X., and Pritchard, J.K. (2006). A map of recent positive selection in the human genome. PLoS Biol. *4*, e72.
39. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.
40. Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., Ehret, C., Ranciaro, A., Froment, A., Hirbo, J.B., Awomoyi, A.A., Bodo, J.M., Doumbo, O., et al. (2009). The genetic structure and history of Africans and African Americans. Science *324*, 1035–1044.
41. Sobota, R.S., Shriner, D., Kodaman, N., Goodloe, R., Zheng, W., Gao, Y.T., Edwards, T.L., Amos, C.I., and Williams, S.M. (2015). Addressing population-specific multiple testing burdens in genetic association studies. Ann. Hum. Genet. *79*, 136–147.
42. Scheinfeldt, L.B., and Tishkoff, S.A. (2013). Recent human adaptation: genomic approaches, interpretation and insights. Nat. Rev. Genet. *14*, 692–702.
43. Olesen, R., Wejse, C., Velez, D.R., Bisseye, C., Sodemann, M., Aaby, P., Rabna, P., Worwui, A., Chapman, H., Diatta, M., et al. (2007). DC-SIGN (CD209), pentraxin 3 and vitamin D receptor gene variants associate with pulmonary tuberculosis risk in West Africans. Genes Immun. *8*, 456–467.
44. Bustamante, J., Boisson-Dupuis, S., Abel, L., and Casanova, J.L. (2014). Mendelian susceptibility to mycobacterial disease: genetic, immunological, and clinical features of inborn errors of IFN-γ immunity. Semin. Immunol. *26*, 454–470.
45. Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.
46. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. USA *107*, 21931–21936.
47. Guo, X., Engel, J.L., Xiao, J., Tagliabracci, V.S., Wang, X., Huang, L., and Dixon, J.E. (2011). UBLCP1 is a 26S proteasome phosphatase that regulates nuclear proteasome activity. Proc. Natl. Acad. Sci. USA *108*, 18649–18654.
48. O'Donnell, C.J., Cupples, L.A., D'Agostino, R.B., Fox, C.S., Hoffmann, U., Hwang, S.J., Ingellson, E., Liu, C., Murabito, J.M., Polak, J.F., et al. (2007). Genome-wide association study for subclinical atherosclerosis in major arterial territories in the NHLBI's Framingham Heart Study. BMC Med. Genet. *8* (*Suppl 1*), S4.
49. Nair, R.P., Duffin, K.C., Helms, C., Ding, J., Stuart, P.E., Goldgar, D., Gudjonsson, J.E., Li, Y., Tejasvi, T., Feng, B.J., et al.; Collaborative Association Study of Psoriasis (2009). Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. Nat. Genet. *41*, 199–204.
50. Yang, Q., Kathiresan, S., Lin, J.P., Tofler, G.H., and O'Donnell, C.J. (2007). Genome-wide association and linkage analyses of hemostatic factors and hematological phenotypes in the Framingham Heart Study. BMC Med. Genet. *8* (*Suppl 1*), S12.
51. Velez Edwards, D.R., Tacconelli, A., Wejse, C., Hill, P.C., Morris, G.A., Edwards, T.L., Gilbert, J.R., Myers, J.L., Park, Y.S., Stryjewski, M.E., et al. (2012). MCP1 SNPs and pulmonary tuberculosis in cohorts from West Africa, the USA and Argentina: lack of association or epistasis with IL12B polymorphisms. PLoS ONE *7*, e32275.
52. Morahan, G., Kaur, G., Singh, M., Rapthap, C.C., Kumar, N., Katoch, K., Mehra, N.K., and Huang, D. (2007). Association of variants in the IL12B gene with leprosy and tuberculosis. Tissue Antigens *69* (*Suppl 1*), 234–236.
53. Swaminathan, B., Goikuria, H., Vega, R., Rodríguez-Antigüedad, A., López Medina, A., Freijo, Mdel.M., Vandenbroeck, K., and Alloza, I. (2014). Autophagic marker MAP1LC3B expression levels are associated with carotid atherosclerosis symptomatology. PLoS ONE *9*, e115176.
54. Hüffmeier, U., Uebe, S., Ekici, A.B., Bowes, J., Giardina, E., Korendowych, E., Juneblad, K., Apel, M., McManus, R., Ho, P., et al. (2010). Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat. Genet. *42*, 996–999.
55. Sieburth, D., Jabs, E.W., Warrington, J.A., Li, X., Lasota, J., LaForgia, S., Kelleher, K., Huebner, K., Wasmuth, J.J., and Wolf, S.F. (1992). Assignment of genes encoding a unique cytokine (IL12) composed of two unrelated subunits to chromosomes 3 and 5. Genomics *14*, 59–62.
56. Capon, F., Di Meglio, P., Szaub, J., Prescott, N.J., Dunster, C., Baumber, L., Timms, K., Gutin, A., Abkevic, V., Burden, A.D., et al. (2007). Sequence variants in the genes for the interleukin-23 receptor (IL23R) and its ligand (IL12B) confer protection against psoriasis. Hum. Genet. *122*, 201–206.
57. Chan, S.H., Perussia, B., Gupta, J.W., Kobayashi, M., Pospísil, M., Young, H.A., Wolf, S.F., Young, D., Clark, S.C., and Trinchieri, G. (1991). Induction of interferon gamma production by natural killer cell stimulatory factor: characterization of the responder cells and synergy with other inducers. J. Exp. Med. *173*, 869–879.
58. Gately, M.K., Desai, B.B., Wolitzky, A.G., Quinn, P.M., Dwyer, C.M., Podlaski, F.J., Familletti, P.C., Sinigaglia, F., Chizonnite, R., Gubler, U., et al. (1991). Regulation of human lymphocyte proliferation by a heterodimeric cytokine, IL-12 (cytotoxic lymphocyte maturation factor). J. Immunol. *147*, 874–882.
59. Cooper, A.M., Magram, J., Ferrante, J., and Orme, I.M. (1997). Interleukin 12 (IL-12) is crucial to the development of protective immunity in mice intravenously infected with mycobacterium tuberculosis. J. Exp. Med. *186*, 39–45.
60. Flynn, J.L., Goldstein, M.M., Triebold, K.J., Sypek, J., Wolf, S., and Bloom, B.R. (1995). IL-12 increases resistance of BALB/c

mice to Mycobacterium tuberculosis infection. J. Immunol. *155*, 2515–2524.
61. Hölscher, C., Atkinson, R.A., Arendse, B., Brown, N., Myburgh, E., Alber, G., and Brombacher, F. (2001). A protective and agonistic function of IL-12p40 in mycobacterial infection. J. Immunol. *167*, 6957–6966.
62. Altare, F., Durandy, A., Lammas, D., Emile, J.F., Lamhamedi, S., Le Deist, F., Drysdale, P., Jouanguy, E., Döffinger, R., Bernaudin, F., et al. (1998). Impairment of mycobacterial immunity in human interleukin-12 receptor deficiency. Science *280*, 1432–1435.
63. Gaffen, S.L., Jain, R., Garg, A.V., and Cua, D.J. (2014). The IL-23-IL-17 immune axis: from mechanisms to therapeutic testing. Nat. Rev. Immunol. *14*, 585–600.
64. Gopal, R., Monin, L., Slight, S., Uche, U., Blanchard, E., Fallert Junecko, B.A., Ramos-Payan, R., Stallings, C.L., Reinhart, T.A., Kolls, J.K., et al. (2014). Unexpected role for IL-17 in protective immunity against hypervirulent Mycobacterium tuberculosis HN878 infection. PLoS Pathog. *10*, e1004099.
65. Rasmussen, S., Allentoft, M.E., Nielsen, K., Orlando, L., Sikora, M., Sjögren, K.G., Pedersen, A.G., Schubert, M., Van Dam, A., Kapel, C.M., et al. (2015). Early divergent strains of Yersinia pestis in Eurasia 5,000 years ago. Cell *163*, 571–582.
66. Cooper, A.M., Roberts, A.D., Rhoades, E.R., Callahan, J.E., Getzy, D.M., and Orme, I.M. (1995). The role of interleukin-12 in acquired immunity to Mycobacterium tuberculosis infection. Immunology *84*, 423–432.
67. Nolt, D., and Flynn, J.L. (2004). Interleukin-12 therapy reduces the number of immune cells and pathology in lungs of mice infected with Mycobacterium tuberculosis. Infect. Immun. *72*, 2976–2988.