MAJOR ARTICLE

Association Study of Genes Controlling IL-12-dependent IFN-γ Immunity: *STAT4* Alleles Increase Risk of Pulmonary Tuberculosis in Morocco

Ayoub Sabri,[1,2,a] Audrey V. Grant,[3,4,a] Kristel Cosker,[3,4] Safa El Azbaoui,[1,2] Ahmed Abid,[5] Ismail Abderrahmani Rhorfi,[5] Hicham Souhi,[5] Hicham Janah,[5] Kebir Alaoui-Tahiri,[5] Yasser Gharbaoui,[5] Majid Benkirane,[6] Marianna Orlova,[7] Anne Boland,[8] Caroline Deswarte,[3,4] Melanie Migaud,[3,4] Jacinta Bustamante,[3,4,9] Erwin Schurr,[7] Stephanie Boisson-Dupuis,[3,4,10] Jean-Laurent Casanova,[3,4,10,11] Laurent Abel,[3,4,10,a] and Jamila El Baghdadi[1,a]

[1]Genetics Unit, Military Hospital Mohamed V, Hay Riad, Rabat, Morocco; [2]Faculty of Science-Kenitra, Ibn Tofail University, Kenitra, Morocco; [3]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale U1163, Paris, France; [4]Paris Descartes University, Sorbonne Paris Cité, Imagine Institute, Paris, France; [5]Department of Pneumology, Military Hospital Mohamed V, Hay Riad, Rabat, Morocco; [6]Blood Transfusion Center, Military Hospital Mohamed V, Hay Riad, Rabat, Morocco; [7]McGill International TB Centre, The Research Institute of the McGill University Health Centre, Montreal, Canada; [8]CEA, Institut de Génomique, Centre National de Génotypage, Evry, France; [9]Center for the Study of Primary Immunodeficiencies, AP-HP, Necker hospital, Paris France; [10]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, New York and [11]Pediatric Hematology-Immunology Unit, AP-HP, Necker Hospital, Paris, France

Background. Only a minority of individuals infected with *Mycobacterium tuberculosis* develop clinical tuberculosis. Genetic epidemiological evidence suggests that pulmonary tuberculosis has a strong human genetic component. Previous genetic findings in Mendelian predisposition to more severe mycobacterial infections, including by *M. tuberculosis*, underlined the importance of the interleukin 12 (IL-12)/interferon γ (IFN-γ) circuit in antimycobacterial immunity.

Methods. We conducted an association study in Morocco between pulmonary tuberculosis and a panel of single-nucleotide polymorphisms (SNPs) covering 14 core IL-12/IFN-γ circuit genes. The analyses were performed in a discovery family-based sample followed by replication in a case-control population.

Results. Out of 228 SNPs tested in the family-based sample, 6 *STAT4* SNPs were associated with pulmonary tuberculosis ($P = .0013–.01$). We replicated the same direction of association for 1 cluster of 3 SNPs encompassing the promoter region of *STAT4*. In the combined sample, the association was stronger among younger subjects (pulmonary tuberculosis onset <25 years) with an odds ratio of developing pulmonary tuberculosis at rs897200 for GG vs AG/AA subjects of 1.47 (1.06–2.04). Previous functional experiments showed that the G allele of rs897200 was associated with lower *STAT4* expression.

Conclusions. Our present findings in a Moroccan population support an association of pulmonary tuberculosis with *STAT4* promoter-region polymorphisms that may impact *STAT4* expression.

Keywords. genetic association; family-based study; candidate pathway; IL-12; IFN-γ; *STAT4*; pulmonary tuberculosis; eQTL; Behcet disease; common variant.

Tuberculosis remains a major global public health problem. Incidence and mortality estimates by the World Health Organization were 8.7 million new cases and 1.4 million deaths from tuberculosis for 2011 [1], and one-third of the world's population is estimated to be infected by the causative agent *Mycobacterium tuberculosis*. Most infected subjects develop latent tuberculosis infection, with only approximately

Received 10 December 2013; accepted 2 March 2014; electronically published 8 March 2014.

[a]These authors contributed equally to this study.

Meetings: This work has not been presented previously at a scientific meeting.

Correspondence: Laurent Abel, MD, PhD, Human Genetics of Infectious Diseases, Inserm U1163, Imagine Institute, 24 Bd du Montparnasse, 75015 Paris, France (laurent.abel@inserm.fr).

The Journal of Infectious Diseases 2014;210:611–8

© The Author 2014. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (http://creativecommons.org/licenses/by-nc-nd/4.0/), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com.

DOI: 10.1093/infdis/jiu140

5% going on to develop clinical tuberculosis within 2 years of infection [2, 3]. This primary tuberculosis mostly affects children, in whom it is often associated with extrapulmonary dissemination of the bacilli. In about 5% of patients with latent infection, tuberculosis develops later in life, principally as a pulmonary disease in adults, typically due to reactivation of the original infection. Besides environmental (eg, microbial) and nongenetic host factors (eg, acquired immunodeficiency), a variety of studies show that human genetic factors contribute to the striking heterogeneity in clinical response to *M. tuberculosis* [4–6]. The human genetic background affects susceptibility or resistance to infection by *M. tuberculosis* [7–9] and the development of disseminated tuberculosis in children [3, 4, 10] and pulmonary tuberculosis in adults [3, 4, 11, 12].

Although the genetic basis for pulmonary tuberculosis susceptibility has been demonstrated by the greater concordance of pulmonary tuberculosis among monozygotic than dizygotic twins [12], only a few replicated pulmonary tuberculosis susceptibility alleles have been identified to date. Most previous genetic association studies employing a candidate gene approach showed a lack of consistency across independent studies, as reviewed recently [11, 13]. One of the most convincing findings was the identification of associated polymorphisms in *NRAMP1* [14, 15]. Using a genome-wide association study (GWAS) approach, 3500 cases and 7500 controls from Ghana and the Gambia led to the identification of an intergenic variant associated with pulmonary tuberculosis on chromosome region 18q11.2 [16]. In an enlarged sample including cases from Indonesia and Russia, a protective variant on region 11p13 was detected [17]. In both studies, the odds ratios (ORs) were modest (OR = 1.19 and 0.80, respectively) [16, 17]. The 11p13 locus was recently replicated in a GWAS from South Africa (OR = 0.62) [18]. Another GWAS in Asian populations identified an independent tuberculosis risk locus on chromosome region 20q12 only in the younger cases (OR = 1.73) [19]. Employing a third strategy, a positional cloning approach in a family-based population from Morocco led to the identification of a major locus on chromosome 8q12-q13 [20]. Fine mapping of the linkage region recently led to the identification of susceptibility single-nucleotide polymorphism (SNP) alleles in the *TOX* gene [21]. Remarkably, the associated SNP alleles conferred the highest risk (OR approximately 3) for pulmonary tuberculosis among the subgroup with an age of onset under 25 years [21]. Overall, these studies suggest that the human genetic component of pulmonary tuberculosis is characterized by high genetic heterogeneity. This may in part be attributable to the complexity of the natural history of pulmonary tuberculosis patients including a highly variable latency period during which immunological mechanisms maintain latency until the point at which the equilibrium is disturbed resulting in clinical symptoms [5, 6].

Rare children with Mendelian predisposition to severe tuberculosis have also been described. This followed the study of the rare syndrome of Mendelian Susceptibility to Mycobacterial Disease (MSMD), which is characterized by severe infections caused by weakly virulent mycobacteria such as BCG vaccines and environmental bacteria [4, 22]. MSMD is a collection of monogenic disorders, not all of which display full clinical penetrance, with mutations in 9 genes resulting in impaired interleukin 12 (IL-12)-dependent interferon γ (IFN-γ) immunity [4, 22–25]. Mutations in one of these genes, *IL12RB1*, have been identified in several children with severe tuberculosis [4, 10, 25]. More recently, we found heterozygous mutations in the gene encoding the β2 chain of the IL-12 receptor (*IL12RB2*) in several subjects with severe tuberculosis (unpublished results). Thus, genes controlling the IL-12-IFN-γ circuit are plausible pulmonary tuberculosis susceptibility candidate genes. *IL12B* and *IFNG* are the most widely studied genes by previous candidate gene association studies focusing on a single functional polymorphism, or on one or few genes from this circuit [11, 13] (see Supplementary Materials 1). Overall, as for the candidate gene approach applied to pulmonary tuberculosis in general, the variability in the results of IL-12/IFN-γ circuit association studies probably attests of the genetic heterogeneity underlying susceptibility to pulmonary tuberculosis [11, 13]. The objective of the present study was to carry out an association study between pulmonary tuberculosis and a set of 14 genes controlling the core IL-12-IFN-γ circuit, using a panel of SNPs providing comprehensive coverage of these genes. The association study was conducted in Moroccan subjects using a primary family-based population, followed by replication in a case-control population.

MATERIALS AND METHODS

Samples From Morocco

Study subjects were recruited from hospital Mohamed V of Rabat and tuberculosis diagnostic centers located in highly endemic areas of Casablanca and Salé, where the annual incidence of tuberculosis is estimated at approximately 150 cases/100 000 inhabitants [1]. Participants presenting with pulmonary tuberculosis were enrolled in the primary family-based sample if their 2 parents or any number of unaffected siblings were also willing to participate and otherwise were enrolled in the replication case-control study. Among subjects given a diagnosis of pulmonary tuberculosis on the basis of clinical symptoms and pathologic findings on chest radiographs, only those with positive sputum smear microscopy results (Ziehl–Neelsen staining) and/or positive sputum culture examination (Lowenstein–Jensen medium) were recruited. Siblings were considered to be unaffected on the basis of normal findings of clinical examination, and normal findings on chest radiographs, and were otherwise considered to have unknown affection status. A total of 185 nuclear families with 260 affected pulmonary tuberculosis offspring were recruited including 170 families (92%)

with at least one available parent (Supplementary Tables 1 and 2). Controls for the case-control replication population (300 cases and 624 controls) were recruited as described elsewhere [21] from healthy blood donors, and only those with a normal clinical exam and without any history of tuberculosis or pulmonary disease were retained. The combined sample of affected offspring from the primary and replication family-based studies and the cases from the case-control replication study consisted of 560 pulmonary tuberculosis subjects with 64.8% of males and a mean (SD) age of tuberculosis onset of 26 (10.4) years (Table 1).

Gene and SNP Selection -Genotyping Methods

Given the implication of the IL-12/IFN-γ circuit in antimycobacterial immunity, the 14 core genes *IFNG, IFNGR1, IFNGR2, IL12A, IL12B, IL12RB1, IL12RB2, IL23A, IL23R, JAK1, JAK2, STAT1, STAT4* and *TYK2* [22] were selected as the focus of our study as described in more detail in Supplementary Materials 1. Tagging SNPs were selected within the 14 genes with borders +/− 3 kb from start and stop codons (human genome assembly GRCh37.p5) using data from the International HapMap Project for the CEPH (Utah residents with ancestry from northern and western Europe) (abbreviation: CEU) (Supplementary Data 1). This procedure resulted in 250 SNPs that provided >80% coverage of CEU tagging SNPs at an R^2 cutoff of 0.80 (Table 2). Genotyping in the familial sample was performed in two steps. The whole panel of SNPs was genotyped in a first subsample of 95 families, and the SNPs selected from the analyses in this first subsample were then genotyped in the remaining 90 families (Supplementary Table 1). The 250 SNPs were genotyped using the ultrahigh throughput Illumina platform, which uses the GoldenGate assay followed by a bead-based technology to resolve individual SNP genotypes (Illumina Inc, San Diego, CA, USA). SNPs selected for testing in the family-based and case-control replication samples were genotyped using a custom oligo pool assay (OPA) also based on the GoldenGate Illumina platform which included other SNPs genotyped in the context of other projects. Data quality control was performed with PLINK software (http://pngu.mgh.harvard.edu/~purcell/plink/) as described in Supplementary Materials 1. These measures resulted in 228 high-quality SNPs that were included in subsequent analyses in the first familial subsample. Seven SNPs selected from this first analysis were genotyped successfully in the remaining families. In the case-control replication study, genotyping conducted on the 6 SNPs associated with pulmonary tuberculosis in the familial sample failed for 1 SNP. All allele frequencies were calculated among founders using PLINK. Pairwise linkage disequilibrum measures (R^2) were calculated across the region using Haploview (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/).

Table 1. Clinical Characteristics of Pulmonary tuberculosis Study Populations From the Moroccan Family-based and Case-control Studies

	Family-Based (Part 1)		Family-Based (Part 2)		Case-Control		Combined
Characteristic	Founders	Offspring	Founders	Offspring	Cases	Controls[a]	Affected Offspring and Cases
N	151	194	138	156	300	624	560
N by status (affected; unaffected/ unknown)	39; 112	141; 53	31; 107	119; 37	. . .	. . .	. . .
Age in years: mean (SD, range) by status							
Affected	44 (13.3, 16–70)	22.2 (8.5, 2–51)	48.5 (12.1, 21–66)	20.7 (7.4, 1–42)	29.9 (10.6, 8–69)	. . .	26 (10.4, 1–69)
Unaffected/Unknown	52 (9, 30–73)	27.2 (10.1, 10–50)	49.9 (10.4, 24–73)	24.4 (5.3, 18–39)	. . .	32.5 (8.9, 20–68)	. . .
% Males by status							
Affected	61.5	56.7	51.6	48.7	75	. . .	64.8
Unaffected/Unknown	39.2	41.5	42.1	43.2	. . .	62.5	. . .

[a] Controls were recruited from among healthy blood donors. Although we cannot exclude the possibility that some of these controls may develop pulmonary tuberculosis later in life (in any case no more than 5%, the expected proportion of infected individuals who develop pulmonary tuberculosis after infection by *Mycobacterium tuberculosis*), the level of misclassification should be quasi-negligible (<5%) as the control population is older than the cases and do not have a history of tuberculosis. In any case, this slight possible misclassification of controls can only affect the power of our analysis and could not lead to false positive results.

Table 2. Coverage of the 14 IL-12/IFN-γ Circuit Candidate Genes Based on Common SNPs Included in The International Hapmap Project CEU Population

Gene	Size/kb	Coverage[a]	N (selected)
IFNG	4.97	1	5
IFNGR1	21.95	0.92	7
IFNGR2	34.63	0.95	12
IL12A	7.18	1	6
IL12B	15.69	0.89	11
IL12RB1	27.33	0.85	9
IL12RB2	8.95	0.92	23
IL23A	1.53	1	1
IL23R	93.48	0.94	35
JAK1	133.28	0.89	37
JAK2	142.94	0.98	34
STAT1	45.22	0.84	21
STAT4	18.65	0.9	38
TYK2	30.04	0.89	11
			250

Abbreviations: IFN-γ, interferon γ; IL-12, interleukin 12; SNP, single-nucleotide polymorphism.

[a] Proportion of SNPs from the International Hapmap Project based on the CEU (Utah residents with ancestry from northern and western Europe) population that are tagged by a selected SNP at an R^2 cutoff of 0.8.

Statistical Methods

Family-based association tests (FBATs) were performed using FBAT v2.0.3 software in the discovery familial sample [26]. These family data can also be analyzed by conditional logistic regression after recoding genotype data for each affected child and up to 3 unaffected pseudosiblings as described elsewhere [21, 27]. An α-level of 0.01 was set for 2-sided FBATs performed in the discovery sample of 95 families (part 1). SNPs selected on the basis of analyses in the discovery sample (part 1) were genotyped in the remaining 90 families of the discovery sample (part 2). FBATs were then performed in the combined familial sample. SNPs showing a $P \leq .01$ in the combined family-based population were retained for further analysis and for genotyping the 300 cases and 624 healthy controls. In the case-control replication sample, the risk allele frequency was calculated among cases and controls, and a 1-sided test of difference of proportions was performed ($\alpha = 0.05$) based on the risk allele identified in the discovery sample. Finally, the conditional logistic regression framework was used to perform a combined analysis including data from the full discovery family-based study and the case-control replication study as described elsewhere [21]. The combined (family-based and case-control) sample was also stratified according to sex and age, with the same age cutoff of 25 years as used previously in a similar study design conducted in Morocco [21] and also appropriate in the present study given the mean age among affected subjects of 26 years (Supplementary Table 3). For the case-control replication study, only relevant cases (eg, <25 years or ≥25 years) were included whereas the full control group was always used. We tested for heterogeneity between the strata using the χ^2 test for heterogeneity (Cochran Q test) [28], which has been used in meta-analyses of GWAS as implemented in GWAMA v.2.1 (http://www.well.ox.ac.uk/gwama/download.shtml) [29]. All classical and conditional logistic regression analyses were performed using the LOGISTIC and PHREG procedures of the SAS software (SAS, Cary, NC). The forward and backward options were used for the multivariate analyses.

RESULTS

We performed FBATs for each of the 228 genotyped high-quality SNPs among the 95 Moroccan families from the discovery sample (see Supplementary Table 4). Four SNPs displayed *P*-values ≤ .01, which all belonged to the *STAT4* gene (Table 3). The SNPs rs6752770, rs3024861, rs7572482 are located within *STAT4* introns, and rs897200 is located in the promoter region of *STAT4* (Figure 1). We thus examined association test results

Table 3. Genetic Association Results for *STAT4* SNPs in the Discovery Moroccan Family-Based Study

					Discovery–Part 1		Discovery–Full		
SNP	m	M	Freq (Risk allele)	Model	OR (95% CI)[a]	*P* Value[a]	OR (95% CI)[a]	*P* Value[a]	-*P P* Value[b]
rs1400654	T	A	0.76	ADD	1.64 (1.02–2.63)	.047	1.67 (1.18–2.33)	.0047	.0045
rs3024861	A	T	0.57	ADD	1.69 (1.07–2.65)	.0096	1.59 (1.2–2.2)	.0043	.0042
rs6752770	G	A	0.72	ADD	1.75 (1.15–2.7)	.0085	1.69 (1.22–2.33)	.0013	.0013
rs7596818	A	G	0.86	ADD	1.89 (1.08–3.33)	.043	1.59 (.96–2.63)	.17	.17
rs1031509	A	C	0.59	REC	1.69 (1.02–2.86)	.02	1.72 (1.15–2.56)	.0022	.0029
rs7572482	G	A	0.48	REC	1.85 (1.07–3.18)	.0046	1.52 (1.01–2.31)	.0058	.0058
rs897200	G	A	0.48	REC	1.75 (1.01–3.00)	.0083	1.49 (1.01–2.26)	.01	.01

Abbreviations: ADD, additive; CI, confidence interval; M, major allele; m, minor allele; OR, odds ratio; REC, recessive; SNP, single-nucleotide polymorphism.

[a] 2-sided test in reference to the risk allele (FBAT); risk alleles are underlined.

[b] FBAT *P*-values obtained using the FBAT permutation test (100 000 permutations).

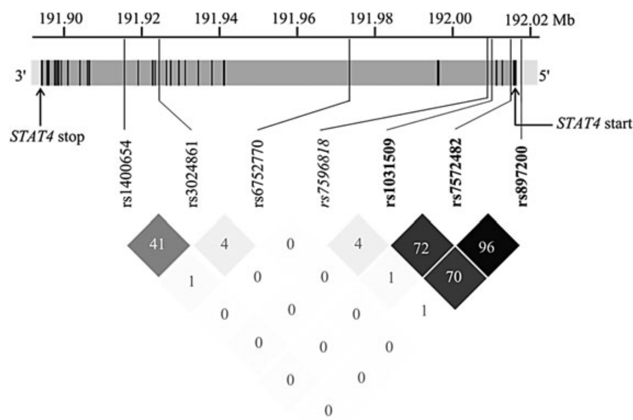

Figure 1. Chromosome 2 map of the 7 SNPs genotyped in the full family-based study. Chromosome 2 location at 191,894,306–192,015,925 bp (2q32.2-q32.3) of *STAT4* (121.62 kb, and the translated product comprises 748 amino acids) is presented. The 24 exons are shown in black, introns in gray, and the promoter region and 3′ UTR in light gray. Locations of the *STAT4* start and stop codons are indicated by arrows. Out of the 7 SNPs, the SNP that is not significantly associated with pulmonary tuberculosis in the full familial sample is in italics, and the 3 SNPs significantly associated with pulmonary tuberculosis in the combined familial and case-control samples are shown in bold. Below, pairwise R^2 values for all pairs of SNPs are given as percentages, and shading from white to black indicates intensity, from an R^2 of 0 to 1. Abbreviation: SNP, single-nucleotide polymorphism.

among other SNPs in *STAT4* at $P < .05$ and identified 3 additional SNPs: rs1031509, rs1400654, and rs7596818 (Table 3, Figure 1). These 7 SNPs were successfully genotyped among 90 additional families, and FBATs were performed in the combined familial sample.

Six SNPs showed a combined $P \leq .01$, whereas 1 SNP, rs7596818, displayed $P = .17$ given an opposite effect in part 2 of the discovery sample (Table 3). Considering the span of *STAT4*, these 6 SNPs can be grouped in 3 clusters (Figure 1): (a) 3 SNPs in high LD (pairwise $R^2 = 0.70–0.96$) close to the 5′ end of *STAT4,* rs7572482, rs897200 and rs1031509, which were associated with pulmonary tuberculosis under a recessive model—the most significant was rs1031509 ($P = .0022$) with an OR of developing pulmonary tuberculosis for CC homozygous subjects vs those with an AC/AA genotype estimated at 1.72 (1.15–2.56); (b) 2 SNPs, rs1400654 and rs3024861, in moderate pairwise LD ($R^2 = 0.41$), and situated near the 3′ end of the gene were associated with pulmonary tuberculosis under an additive or a dominant model, with an OR of developing pulmonary tuberculosis for the replicated SNP rs1400654 of 1.67 (1.18–2.33) for AA vs AT subjects or for AT vs TT subjects ($P = .0047$); (c) a single SNP, rs6752770, situated in the largest *STAT4* intron between the 2 previously described clusters, in low LD with the other 5 SNPs (pairwise $R^2 < 0.05$), and showing association under the additive model (OR = 1.69 (1.22–2.33) for AA vs AG subjects, or for AG vs GG subjects; $P = .0013$).

Table 4. Case-control Association Results for *STAT4* SNPs Among 300 Cases With Pulmonary Tuberculosis and 624 Healthy Controls

SNP	m[a]	M[a]	Freq (Risk)[b] (Cases)	Freq (Risk)[b] (Controls)	Allelic *P* Value[c]
rs1400654	T	A (underlined)	0.77	0.78	>.5
rs3024861	G	A	. . .	. . .	. . .
rs6752770	G	A (underlined)	0.64	0.66	>.5
rs1031509	A	C (underlined)	0.57	0.55	.25
rs7572482	G (underlined)	A	0.52	0.47	.036
rs897200	G (underlined)	A	0.51	0.46	.03

Abbreviation: SNP, single-nucleotide polymorphism.

[a] Minor (m) and Major (M) allele; the risk allele is underlined.

[b] Risk allele frequency.

[c] 1-sided test for the allelic χ^2 test.

Multivariable analysis performed on the combined family sample using the 6 SNPs significant in univariate analyses confirmed the presence of 3 independent *STAT4* association signals. The best multivariable model included rs6752770, rs1400654, and rs1031509 (data not shown). However, the models replacing rs1031509 for rs7572482 or rs897200 provided similar fit indicating that the 3 SNPs of cluster (a) were interchangeable in a multivariable model. Finally, a total of 113 SNPs were successfully imputed in the target region of *STAT4* in the full familial sample (see Supplementary Materials 1). A total of 4 imputed SNPs, rs10931481 ($R^2 = 0.88$ with rs3024861 in the CEU population, $P = .0063$), rs16833260 ($R^2 = 0.74$ with rs3024861, $P = .01$), rs12327969 ($R^2 = 0.78$ with rs1031509, $P = .0052$), and rs10208033 ($R^2 = 0.34$ with rs3024861, $P = .0064$) displayed association *P*-values $\leq .01$, and all of these were less significant than the original genotyped SNP in highest LD with the corresponding proxy SNP.

Thus, only the 6 genotyped SNPs significantly associated with pulmonary tuberculosis in the combined familial sample were selected for genotyping in the case-control replication sample. One of these failed (rs3024861), but the other SNP of cluster (b), rs1400654, provided a 1-sided *P*-value > .5 showing a lower frequency of the pulmonary tuberculosis risk allele among cases than controls (Table 4). Similarly, single SNP rs6752770 showed a 1-sided *P*-value > .5. Only the 3 SNPs of cluster (a) displayed a higher frequency of the pulmonary tuberculosis risk allele among cases than controls, and this difference was significant for 2 of them, rs7572482 ($P = .036$) and rs897200 ($P = .03$; Table 4). When testing for a recessive genetic model, the highest OR (1.28) was observed for rs897200 although borderline nonsignificant ($P = .065$). Next, we performed association tests for the 3 SNPs of cluster (a) with pulmonary tuberculosis in the whole sample by combining the familial and case-control data. The 3 SNPs were significantly associated ($P < .05$) under a recessive model (Table 5).

Table 5. Genetic Association Results for the Cluster of Three *STAT4* SNPs in the Discovery Moroccan Family-based Study, and Moroccan Case-control Replication Study

					Family and Case Control		<25 y		≥25 y	
SNP	m[a]	M[a]	Freq (Risk)[b]	Model[c]	OR (95% CI)[d]	*P* Value[d]	OR (95% CI)[d]	*P* Value[d]	OR (95% CI)[d]	*P* Value[d]
rs1031509	A	C	0.59	REC	1.27 (1.01–1.61)	.045	1.47 (1.1–2)	.011	1.06 (.78–1.45)	.70
rs7572482	G	A	0.48	REC	1.32 (1.03–1.70)	.030	1.49 (1.08–2.07)	.016	1.19 (.86–1.65)	.30
rs897200	G	A	0.48	REC	1.35 (1.05–1.74)	.019	1.47 (1.06–2.04)	.019	1.28 (.92–1.78)	.15

Abbreviations: CI, confidence interval; OR, odds ratio; SNP, single-nucleotide polymorphism.

[a] Minor (m) and Major (M) allele; the risk allele is underlined.

[b] Risk allele frequency.

[c] Genetic model (REC, recessive).

[d] Odds ratio (95% confidence interval) and *P*-value for the 2-sided Wald test in reference to the risk allele from combined analysis using the conditional logistic regression framework.

The most significant finding was observed at rs897200 ($P = .019$) with an OR of developing pulmonary tuberculosis for GG subjects vs those with an AG/AA genotype estimated at 1.35 (1.05–1.74). Stratified analyses were conducted on these 3 SNPs based on age at pulmonary tuberculosis onset (dividing the population into younger (<25 years) and older (≥25 years) groups; Table 4) and sex (data not shown) after verifying the structure of informative families within the 2 age strata (Supplementary Table 2). Although no sex effect was found, we observed that the association with pulmonary tuberculosis was stronger for the <25 years stratum. Specifically, the OR estimates increased from 1.27–1.35 in the full population to 1.47–1.49 in the <25 years stratum, also giving lower *P*-values (.019–.011), although the Cochran Q test for heterogeneity across the 2 strata remained nonsignificant (the minimum *P*-value was .13 at rs1031509). Overall, our analyses identified 3 correlated *STAT4* SNPs, with approximately 25%–30% of Moroccan subjects bearing the risk genotype to develop pulmonary tuberculosis, in particular at a young age.

DISCUSSION

The present candidate pathway association study of 14 IL-12/IFN-γ genes identified a cluster of three SNPs in the promoter region of *STAT4* as associated with pulmonary tuberculosis by employing an initial family-based study followed by a case-control replication study, all in Moroccan patients. SNPs tested in the 13 other genes did not show any detectable association at the 0.01 level in the first discovery sample. Within the IL-12/IFN-γ pathway, the *STAT4* protein encoded by *STAT4* is a signal transducer phosphorylated by the kinases JAK2 and TYK2 upon binding of IL-12 to its receptor. Nuclear translocation of phosphorylated *STAT4* dimers drives the transcription of multiple target genes, in particular *IFNG* [30, 31]. The family-based discovery study identified 6 *STAT4* SNPs grouped in 3 clusters, 1 including 3 SNPs in the promoter region and 5′ end of the gene (denoted as cluster (a)), another including 2 SNPs in introns toward the 3′ end of the gene, and finally an independent intronic SNP. However, the case-control study provided replication evidence for only 2 SNPs of the first cluster (a) of 3 SNPs, with a weaker magnitude of effect than in the family-based study. In spite of the older age of controls and the absence of pulmonary tuberculosis history for controls, it is possible that some control subjects will go on to develop pulmonary tuberculosis at a later time. Such misclassification of controls would bias OR estimates toward the null if the association is true and could lead to reduced significance.

To our knowledge, only one pulmonary tuberculosis association study investigated *STAT4* as a candidate gene, by focusing on a single microsatellite marker, and without any significant results [32]; this microsatellite marker with 4 alleles is in strong LD with SNP rs1551443, based on 1000 Genomes Project data, which was also not associated with pulmonary tuberculosis in our sample. [32]. However, variants in *STAT4* have been found to be associated through GWAS with a number of autoimmune or inflammatory disorders such as systemic lupus erythematosus, rheumatoid arthritis, primary biliary cirrhosis, and systemic sclerosis [33]. Of particular interest, 2 SNPs of our cluster (a), rs7572482 and rs897200, were found to be associated with Behcet disease, which is a rare immune-mediated small-vessel systemic vasculitis. In a first study performed in a Turkish population, rs7572482 was associated with Behcet disease [34]. This SNP was also identified as a part of a cluster of 3 SNPs, which included rs897200, in an independent Chinese Behcet disease GWAS [35]. Markers rs7572482 and rs897200 are in strong LD ($R^2 = 0.96$) at the edge of the 5′ promoter region (Figure 1), and using the CEU Hapmap and 1000 Genomes Project population, we identified 7 additional SNPs (rs55925192, rs16833437, rs7561569, rs1031507, rs6736458, rs16833453, and rs57081321) highly correlated ($R^2 = 0.97–1$) with these 2 SNPs, and located within 7.3 kb 5′ of rs897200. We investigated the Regulomedb [36] and found that both rs7572482 and rs897200, as well as rs1031507, are likely to belong to transcription factor binding sites (TFBS). In particular for rs897200,

using the software ALGGEN PROMO [37] which provides an in silico prediction model for transcription factor binding, we found that several transcription factors bound to the DNA sequence overlapping rs897200 when the A allele was present, whereas none of these factors were predicted to bind to the region when the G allele was present. These data indicate that these SNPs, in particular rs897200, may impact on regulatory functions.

Furthermore, SNP rs897200 was reported as an expression quantitative trait locus (eQTL) of *STAT4* in lymphoblastoid cells, as evidenced by a posterior probability of 0.60 using Bayesian hierarchical modeling (and therefore higher than the 0.5 cutoff for establishing a SNP-gene pair as an eQTL) [38]. In addition, the Chinese Behcet disease study performed experiments evaluating transcription level differences by genotype at rs897200, providing further evidence for the potential functional role of this SNP. Among 19 normal controls, subjects with the AA genotype had significantly higher *STAT4* mRNA levels in PBMCs and skin cells than GG subjects. Luciferase reporter assays showed that luciferase activity was significantly increased in cells carrying the A allele as compared with those carrying the G allele [35]. In the Chinese study, the A allele was associated with increased risk of Behcet disease, whereas in our study, the GG subjects are at increased risk of pulmonary tuberculosis, thus suggesting pleiotropic effects with an inverse relationship between Behcet disease risk and pulmonary tuberculosis. Interestingly, down-regulation of *STAT4* expression was reported in PBMCs of subjects with active tuberculosis stimulated with purified protein derivative of tuberculin (PPD) [39]. These expression data combined with our association results suggest that *STAT4* may be implicated in host defense against *M. tuberculosis*, with lower *STAT4* expression associated with active disease. Of note, *STAT4* was not part of the 393 transcript signature for active tuberculosis observed in whole-blood, and dominated by a neutrophil-driven IFN-α/β-inducible gene profile [40]. Further studies investigating different cell types under different stimulation conditions are needed to confirm and elaborate patterns of expression according to genotype at the pulmonary tuberculosis-associated SNPs.

Interestingly, the role of *STAT4* variants appears to be more pronounced in patients <25 years with pulmonary tuberculosis. This is consistent with previous studies of tuberculosis and other infections [41, 42], in particular our recent finding of variants in *TOX* influencing pulmonary tuberculosis risk in subjects <25 years [21]. Younger age is likely to be a phenotypic indicator for those who more rapidly exit from latency to enter into the state of active pulmonary disease given endemic exposure to *M. tuberculosis*. The present study further underlines the importance of age at tuberculosis onset as a critical factor to consider in any future pulmonary tuberculosis association studies to reduce pulmonary tuberculosis phenotypic heterogeneity. The functional data (including from a previous study of Behcet disease performed on healthy Chinese controls, expression studies and databases) provide strong support for our association findings with our 3 SNP cluster, especially rs897200. Further genetic association studies of these variants are needed in pulmonary tuberculosis study populations of other ethnicities, especially in settings with a substantial proportion of early-onset pulmonary tuberculosis patients.

Supplementary Data

Supplementary materials are available at *The Journal of Infectious Diseases* online (http://jid.oxfordjournals.org/). Supplementary materials consist of data provided by the author that are published to benefit the reader. The posted materials are not copyedited. The contents of all supplementary data are the sole responsibility of the authors. Questions or messages regarding errors should be addressed to the author.

Notes

Acknowledgments. We thank all patients and family members from Morocco and all the healthy Moroccan subjects for their participation in this study. We thank the late Dr M. Chentoufi from the Tuberculosis Diagnostic Center (CDST) of Hay Mohammadi, and Dr I. Sentissi from the CDST of Salé for helping to recruit tuberculosis patients. We thank F. El Bordi, the nurse from the CDST of Salé, and Dr N. El Amraoui and Dr M. Hakam from the Blood Transfusion Center of Rabat for recruiting healthy donors. We thank the *Centre National de Génotypage* for carrying out the genotyping.

Financial support. This work was supported by grants from the French National Agency for Research (ANR, grant ANR-08-MIEN-014-02), EU-grant HOMITB (HEALTH-F3-2008-200732), the European Research Council (ERC-2010-AdG-268777), the St. Giles Foundation, the Rockefeller University grant 8ULTR000043 from the National Center for Research Resources and the National Center for Advancing Sciences (NCATS) of the National Institutes of Health, the Rockefeller University, the National Institute of Allergy and Infectious Diseases grants R01AI089970, and U01AI088685. A. V. G. was partly supported by the *Fondation pour la Recherche Médicale* and *Fondation BNP Paribas*.

Potential conflicts of interest. All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. World Health Organization (WHO). Global tuberculosis control 2011. http://www.who.int/tb/publications/global_report/en/. Accessed 23 April 2012.
2. Zumla A, Raviglione M, Hafner R, von Reyn CF. Tuberculosis. N Engl J Med **2013**; 368:745–55.
3. Alcaïs A, Fieschi C, Abel L, Casanova J-L. Tuberculosis in children and adults: two distinct genetic diseases. J Exp Med **2005**; 202:1617–21.
4. Casanova J-L, Abel L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol **2002**; 20:581–620.
5. Ernst JD. The immunological life cycle of tuberculosis. Nat Rev Immunol **2012**; 12:581–91.
6. O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. Annu Rev Immunol **2013**; 31:475–527.
7. Cobat A, Gallant CJ, Simkin L, et al. Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J Exp Med **2009**; 206:2583–91.
8. Cobat A, Gallant CJ, Simkin L, et al. High heritability of antimycobacterial immunity in an area of hyperendemicity for tuberculosis disease. J Infect Dis **2010**; 201:15–19.

9. Cobat A, Barrera LF, Henao H, et al. Tuberculin skin test reactivity is dependent on host genetic background in Colombian tuberculosis household contacts. Clin Infect Dis Off Publ Infect Dis Soc Am **2012**; 54:968–71.
10. Boisson-Dupuis S, El Baghdadi J, Parvaneh N, et al. IL-12Rβ1 deficiency in two of fifty children with severe tuberculosis from Iran, Morocco, and Turkey. PLoS One **2011**; 6:e18524.
11. Azad AK, Sadee W, Schlesinger LS. Innate immune gene polymorphisms in tuberculosis. Infect Immun **2012**; 80:3343–59.
12. Puffer RR. Familial susceptibility to tuberculosis: Its importance as a public health problem. Harvard University Press, Cambridge, MA, **1944**.
13. Möller M, Hoal EG. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberc Edinb Scotl **2010**; 90:71–83.
14. Bellamy R, Ruwende C, Corrah T, McAdam KP, Whittle HC, Hill AV. Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med **1998**; 338:640–4.
15. Li X, Yang Y, Zhou F, et al. SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PLoS One **2011**; 6:e15831.
16. Thye T, Vannberg FO, Wong SH, et al. Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet **2010**; 42:739–41.
17. Thye T, Owusu-Dabo E, Vannberg FO, et al. Common variants at 11p13 are associated with susceptibility to tuberculosis. Nat Genet **2012**; 44:257–9.
18. Chimusa ER, Zaitlen N, Daya M, et al. Genome-wide association study of ancestry-specific TB risk in the South African Coloured population. Hum Mol Genet **2014**; 23:796–809.
19. Mahasirimongkol S, Yanai H, Mushiroda T, et al. Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. J Hum Genet **2012**; 57:363–7.
20. Baghdadi JE, Orlova M, Alter A, et al. An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med **2006**; 203:1679–84.
21. Grant AV, El Baghdadi J, Sabri A, et al. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. Am J Hum Genet **2013**; 92:407–14.
22. Filipe-Santos O, Bustamante J, Chapgier A, et al. Inborn errors of IL-12/23- and IFN-γ-mediated immunity: molecular, cellular, and clinical features. Semin Immunol **2006**; 18:347–61.
23. Bogunovic D, Byun M, Durfee LA, et al. Mycobacterial disease and impaired IFN-γ immunity in humans with inherited ISG15 deficiency. Science **2012**; 337:1684–8.
24. Bustamante J, Arias AA, Vogt G, et al. Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. Nat Immunol **2011**; 12:213–21.
25. De Beaucoudrey L, Samarina A, Bustamante J, et al. Revisiting human IL-12Rβ1 deficiency: a survey of 141 patients from 30 countries. Medicine (Baltimore) **2010**; 89:381–402.
26. Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype–phenotype associations. Eur J Hum Genet EJHG **2001**; 9:301–6.
27. Cordell HJ, Barratt BJ, Clayton DG. Case/pseudocontrol analysis in genetic association studies: A unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. Genet Epidemiol **2004**; 26:167–85.
28. Cochran WG. The combination of estimates from different experiments. Biometrics **1954**; 10:101–29.
29. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics **2010**; 11:288.
30. Stark GR, Darnell JE Jr. The JAK-STAT pathway at twenty. Immunity **2012**; 36:503–14.
31. Casanova J-L, Holland SM, Notarangelo LD. Inborn errors of human JAKs and STATs. Immunity **2012**; 36:515–28.
32. Hijikata M, Shojima J, Matsushita I, et al. Association of IFNGR2 gene polymorphisms with pulmonary tuberculosis among the Vietnamese. Hum Genet **2012**; 131:675–82.
33. Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. N Engl J Med **2011**; 365:1612–23.
34. Remmers EF, Cosan F, Kirino Y, et al. Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behçet's disease. Nat Genet **2010**; 42:698–702.
35. Hou S, Yang Z, Du L, et al. Identification of a susceptibility locus in STAT4 for Behçet's disease in Han Chinese in a genome-wide association study. Arthritis Rheum **2012**; 64:4104–13.
36. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res **2012**; 22:1790–7.
37. Messeguer X, Escudero R, Farré D, Núñez O, Martínez J, Albà MM. PROMO: detection of known transcription regulatory elements using species-tailored searches. Bioinforma Oxf Engl **2002**; 18:333–4.
38. Veyrieras J-B, Kudaravalli S, Kim SY, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet **2008**; 4:e1000214.
39. Stern JNH, Keskin DB, Romero V, et al. Molecular signatures distinguishing active from latent tuberculosis in peripheral blood mononuclear cells, after in vitro antigenic stimulation with purified protein derivative of tuberculin (PPD) or Candida: a preliminary report. Immunol Res **2009**; 45:1–12.
40. Berry MPR, Graham CM, McNab FW, et al. An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature **2010**; 466:973–7.
41. Alcaïs A, Quintana-Murci L, Thaler DS, Schurr E, Abel L, Casanova J-L. Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann N Y Acad Sci **2010**; 1214:18–33.
42. Alcaïs A, Alter A, Antoni G, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. Nat Genet **2007**; 39:517–22.