REPORT

Age-Dependent Association between Pulmonary Tuberculosis and Common *TOX* Variants in the 8q12–13 Linkage Region

Audrey V. Grant,[1,2,14] Jamila El Baghdadi,[3,14,*] Ayoub Sabri,[3] Safa El Azbaoui,[3] Kebir Alaoui-Tahiri,[4] Ismail Abderrahmani Rhorfi,[4] Yasser Gharbaoui,[4] Ahmed Abid,[4] Majid Benkirane,[5] Vaomalala Raharimanga,[6] Vincent Richard,[6] Marianna Orlova,[7] Anne Boland,[8] Mélanie Migaud,[1,2] Satoshi Okada,[9] Daniel K. Nolan,[1,2] Jacinta Bustamante,[1,2,10] Luis B. Barreiro,[11] Erwin Schurr,[7] Stephanie Boisson-Dupuis,[1,2,9] Voahangy Rasolofo,[12] Jean-Laurent Casanova,[1,2,9,13,15] and Laurent Abel[1,2,9,15,*]

Only a small fraction of individuals infected with *Mycobacterium tuberculosis* develop clinical tuberculosis (TB) in their lifetime. Genetic epidemiological evidence suggests a genetic determinism of pulmonary TB (PTB), but the molecular basis of genetic predisposition to PTB remains largely unknown. We used a positional-cloning approach to carry out ultrafine linkage-disequilibrium mapping of a previously identified susceptibility locus in chromosomal region 8q12–13 by genotyping 3,216 SNPs in a family-based Moroccan sample including 286 offspring with PTB. We observed 44 PTB-associated SNPs ($p < 0.01$), which were genotyped in an independent set of 317 cases and 650 controls from Morocco. A single signal, consisting of two correlated SNPs close to *TOX*, rs1568952 and rs2726600 (combined $p = 1.1 \times 10^{-5}$ and 9.2×10^{-5}, respectively), was replicated. Stronger evidence of association was found in individuals who developed PTB before the age of 25 years (combined p for rs1568952 $= 4.4 \times 10^{-8}$; odds ratio of PTB for AA versus AG/GG = 3.09 [1.99–4.78]). The association with rs2726600 ($p = 0.04$) was subsequently replicated in PTB-affected subjects under 25 years in a study of 243 nuclear families from Madagascar. Stronger evidence of replication in Madagascar was obtained for additional SNPs in strong linkage disequilibrium with the two initial SNPs ($p = 0.003$ for rs2726597), further confirming the signal. We thus identified around rs1568952 and rs2726600 a cluster of SNPs strongly associated with early-onset PTB in Morocco and Madagascar. SNP rs2726600 is located in a transcription-factor binding site in the 3′ region of *TOX*, and further functional explorations will focus on CD4 T lymphocytes.

Tuberculosis (TB), caused by *Mycobacterium tuberculosis*, results in 8.8 million new cases and 1.1 million deaths each year[1] and remains a major public-health problem worldwide. One-third of the world's population is exposed to *M. tuberculosis*, and after exposure, most, but not all, individuals become infected. About 5% of infected individuals develop clinical TB within 2 years of infection and experience a minimal latency phase or none at all. This "primary" TB is particularly common in children and is often associated with aggressive miliary and/or extrapulmonary disease. Other individuals remain infected in a state of immunological equilibrium of variable duration, and most (~95%) maintain a latency of infection of indefinite duration.[2,3] The remaining 5% of infected individuals develop clinical TB later in life, typically pulmonary TB (PTB) due to reactivation of the original infection or, in some cases, reinfection.[2,3] The natural history of TB is therefore characterized by great interindividual variability at the initial infection step and at the junctures leading to clinical symptoms of either primary TB or reactivation PTB.[2] In addition to environmental (e.g., microbial) and nongenetic (e.g., acquired immunodeficiency) host factors, there is considerable evidence to suggest that host genetic control of the response to *M. tuberculosis* contributes to this variability.[2,3] This is the case for the control of TB infection (e.g., as measured by the tuberculin skin test),[4–7] as shown by the recent identification of two major loci associated with TB by genome-wide linkage analysis.[8] Evidence has also been obtained for a genetic basis for the development of clinical disease for both childhood TB and adult PTB.[2,9–11]

The study of Mendelian susceptibility to mycobacterial disease (MSMD [MIM 209950]), a clinical disease caused by weakly virulent mycobacteria, was instrumental to

[1]Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Institut National de la Santé et de la Recherche Médicale U980, 75015 Paris, France; [2]Paris Descartes University, Sorbonne Paris Cité, Imagine Institute, 75015 Paris, France; [3]Genetics Unit, Military Hospital Mohamed V, Hay Riad, 10100 Rabat, Morocco; [4]Department of Pneumology, Military Hospital Mohamed V, Hay Riad, 10100 Rabat, Morocco; [5]Blood Transfusion Center, Military Hospital Mohamed V, Hay Riad, 10100 Rabat, Morocco; [6]Epidemiology Unit, Institut Pasteur de Madagascar, 101 Antananarivo, Madagascar; [7]McGill Centre for the Study of Host Resistance, The Research Institute of the McGill University Health Centre, Montreal, QC H3G 1A4, Canada; [8]CEA, Institut de Génomique, Centre National de Génotypage, 91000 Evry, France; [9]St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY 10065, USA; [10]Center for the Study of Primary Immunodeficiencies, Necker Hospital, Assistance Publique-Hôpitaux de Paris, 75015 Paris, France; [11]Sainte-Justine Hospital Research Centre, Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal, QC H3T 1C5, Canada; [12]Mycobacteria Unit, Institut Pasteur de Madagascar, 101 Antananarivo, Madagascar; [13]Pediatric Hematology-Immunology Unit, Necker Hospital, Assistance Publique-Hôpitaux de Paris, 75015 Paris, France
[14]These authors contributed equally to this work
[15]These authors contributed equally to this work
*Correspondence: baghdadijamila@yahoo.fr (J.E.B.), laurent.abel@inserm.fr (L.A.)
http://dx.doi.org/10.1016/j.ajhg.2013.01.013.

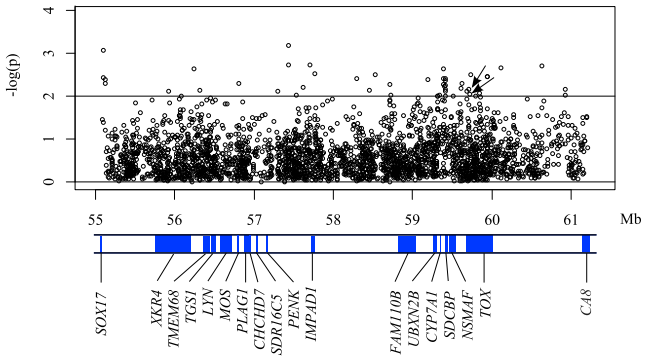

Figure 1. Fine-Mapping Association Results for 203 Moroccan Families with at Least One Child with PTB

Statistical significance, expressed as the $-\log_{10}$ of the most significant FBAT p value from the additive, dominant, and recessive tests, is shown against the position (in Mb) on chromosome 8 for each of 2,865 quality-control-filtered SNPs. Locations and symbols of genes (from start to stop codons) are shown below the chromosomal-position axis. The statistical-significance threshold for SNP selection for genotyping in the case-control population is indicated by a horizontal black line at $p = 0.01$. The two SNPs (rs1568952 and rs2726600) found to be significant after replication in the Moroccan case-control study are indicated by arrows.

initial genetic research on severe childhood TB. Mutations resulting in impaired IFN-γ immunity have been identified in nine genes in individuals with MSMD.[2,12–15] Several children with severe TB have been found to carry biallelic mutations of one of these genes, *IL12RB1*.[10,16–22] In an attempt to estimate the proportion of severe pediatric TB cases that are Mendelian, Boisson-Dupuis et al. found that two out of a total of 50 cases bore biallelic mutations in *IL12RB1* after sequencing.[10] We recently discovered autosomal-dominant IL-12Rβ2 deficiency in eight individuals with disseminated TB (unpublished data). More common genetic factors of susceptibility to pediatric TB were also identified in an association study focusing on variants of the natural-resistance-associated macrophage protein 1 (*NRAMP1* or *SLC11A1*) and taking gene-environment interactions into account.[23] These findings for subjects with childhood TB provide proof of principle for the genetic basis of TB susceptibility.

The study of genetic susceptibility to PTB has proven more difficult, although there is considerable evidence, particularly from twin studies, to support a major role for human genetic factors in the development of PTB.[24–27] Most classical genetic association studies investigating PTB have focused on candidate genes, and a number of common risk variants have been reported.[9] However, few have been satisfactorily replicated, and alleles of human leukocyte antigen (HLA) class II[28,29] and *NRAMP1*[30–32] have provided the most consistent results. Recent transcriptomic analyses of peripheral-blood cells identified a neutrophil-driven interferon-inducible transcript signature in individuals with active PTB,[33] providing interesting new candidate genes and pathways to be tested. In a recent genome-wide association study (GWAS), only two intergenic SNPs with modest risk effects were identified in samples from Ghana and The Gambia.[34,35] Fewer GWASs have been published for TB than for other common diseases, but current data suggest an underlying heterogeneity in the genetic variants affecting PTB. In particular, it seems likely that different subgroups of PTB-affected individuals have different genetic risk factors. These subgroups can be defined on the basis of individual factors, such as clinical characteristics, or extrinsic factors, such as pathogen variability.[36]

By genome-wide linkage analysis, we previously identified in chromosomal region 8q12–13 a locus linked to PTB in a Moroccan study population.[37] In a similar situation, the positional-cloning approach has proven successful in the study of genetic susceptibility to leprosy, the second most prevalent human mycobacterial infection, leading to the identification of susceptibility variants in *PARK2/PACRG*,[38] *LTA*,[39] and the HLA class I region.[40] Here, we conducted a TB association study involving two independent study populations from Morocco and one from Madagascar. The sampling of all the cohorts was approved by the appropriate ethical committees and/or institutional review boards, and written informed consent was obtained from all adults and from the parents of minors. We first conducted a dense fine-mapping study of 8q12–13 by selecting and genotyping 3,216 SNPs covering the linkage region from 55.1–61.2 Mb in human assembly GRCh37.p5 (Figure S1, available online) in a primary Moroccan family-based sample including 203 nuclear families with 286 PTB-affected offspring (Tables S1 and S2). In total, we tested 2,865 quality-control-filtered SNPs for association in family-based association tests (FBATs) (Figure S1).

The results of this first screen are shown in Figure 1. A p value below 0.01 was obtained for 44 SNPs (Table S3), which were selected for replication and genotyped in a second independent sample of 317 cases and 650 controls from Morocco (Table S2). Forty of these 44 SNPs passed quality-control filters. A SNP was replicated if the association analysis yielded a one-tailed p value < 0.01 under the same genetic model of inheritance for the same risk allele as in the primary study. Under the assumption of a recessive model for the minor allele, only two SNPs, rs1568952 (G/A) and rs2726600 (A/G), satisfied this condition with p values of 0.0003 and 0.0046, respectively (Table 1). We observed that the odds ratios (ORs) were slightly lower in the replication sample than in the family-based sample (Table S4), which might be the result of the "winner's curse" phenomenon,[41] although minor misclassification of controls cannot be excluded (Table S2). The minor alleles of rs1568952 (A) and rs2726600 (G) had frequencies of 0.33 and 0.36, respectively, in the control population. With a pairwise R^2 of 0.77, these SNPs were in strong linkage disequilibrium (LD), indicative of a single association signal. SNP rs2726600 is located within the eighth intron of *TOX*, whereas rs1568952 is located 6 kb downstream of the last *TOX* exon (Figure 2).

Table 1. Genetic Association Results for Replicated SNPs rs1568952 and rs2726600 in the Primary Moroccan Family-Based Study, the Moroccan Case-Control Replication Study, and Combined Analyses under the Recessive Model for the Minor Allele

					Family-Based Study		Case-Control Study		Combined	
Stratum	SNP	Minor Allele	Major Allele	MAF[a]	OR (95% CI)	p Value[b]	OR (95% CI)	p Value[b]	OR (95% CI)	p Value[b]
Full	rs1568952	A	G	0.36	3.21 (1.41–7.35)	0.007	1.98 (1.33–2.94)	6×10^{-4}	2.18 (1.53–3.10)	1.1×10^{-5}
	rs2726600	G	A	0.40	2.65 (1.27–5.56)	0.0093	1.61 (1.12–2.31)	0.0092	1.81 (1.34–2.43)	9.2×10^{-5}
<25 Years	rs1568952	A	G	-	5.54 (1.97–15.53)	0.0003	2.86 (1.72–4.77)	2.9×10^{-5}	3.09 (1.99–4.78)	4.4×10^{-8}
	rs2726600	G	A	-	2.56 (1.37–4.80)	0.0025	2.00 (1.24–3.23)	0.0039	2.19 (1.52–3.14)	3.2×10^{-5}
≥25 Years	rs1568952	A	G	-	0.65 (0.12–3.66)	0.62	1.52 (0.93–2.47)	0.094	1.42 (0.88–2.27)	0.15
	rs2726600	G	A	-	1.73 (0.56–5.33)	0.33	1.38 (0.89–2.14)	0.15	1.42 (0.95–2.13)	0.09

The following abbreviations are used: MAF, minor allele frequency; OR, odds ratio; and CI, confidence interval.
[a]MAF was estimated from among 316 founders.
[b]All p values are two sided.

We then estimated the overall effect of these two SNPs in the combined Moroccan samples. The combined sample of 603 PTB-affected subjects consisting of affected offspring from the family-based study and cases from the case-control replication study had a sex ratio (male/female) of 1.71 and a mean (±SD) age at the diagnosis of PTB of 25.3 (±10.4) years (Table S2). The distributions of age by study and affection status are shown in Figure S2. In a conditional-logistic-regression analysis (Table S4), the association-test p values for the whole sample were 1.1×10^{-5} for rs1568952 and 9.2×10^{-5} for rs2726600. The overall OR of developing PTB at rs1568952 for AA homozygous subjects versus those with an AG/GG genotype was 2.18 (1.53–3.10). Similar results were obtained in a classical meta-analysis, further validating the conditional-logistic-regression approach. We were also able to show the absence of significant heterogeneity between the ORs obtained in our family-based and case-control samples (Table S5). Two classical factors known to influence the occurrence of PTB are sex and age, and the association between PTB and SNPs rs1568952 and rs2726600 was thus examined after stratification for these two factors. This stratification had no significant effect on the results obtained in analyses stratified for sex (data not shown).

We conducted age-stratified analyses by considering two groups of subjects of approximately equal size on the basis of age at PTB diagnosis (<25 years or ≥25 years) and found evidence of heterogeneity (Cochran's Q p = 0.019 for rs1568952). In the combined sample, the evidence of association with rs1568952 was markedly stronger in younger PTB cases (p value = 4.4×10^{-8}), giving an OR for developing PTB of 3.09 (1.99–4.78) (Table 1). By contrast, this association was not significant in older PTB-affected subjects. A similar pattern of strong association in younger PTB-affected subjects was observed for rs2726600 and gave an OR of developing PTB of 2.19 (1.52–3.14). We also estimated the p values and ORs for younger affected individuals by classical meta-analysis, which gave very similar results (Table S5). Finally, we examined the impact of lowering the age cutoff from 25 to 18 years on OR estimates for the younger stratum. Interestingly, we observed that the effect of both SNPs clearly increased among younger affected individuals when the cutoff was decreased to 20 years or below (Figure S3). We conducted additional analyses to make use of the between-family information provided by our primary family sample (Table S6). The expected frequency of the risk allele calculated from our families with PTB-affected offspring under the age of 25 years was significantly higher than that observed in controls for both rs1568952 (0.39 versus 0.33, p = 0.04) and rs2726600 (0.43 versus 0.36, p = 0.02) (Table S6). This between-family information thus provides additional support for the association between early-onset PTB and SNPs rs1568952 and rs2726600.

We then investigated the LD patterns of our two associated SNPs with HapMap Release 28 Phases II and III (August 2010) (Table S7). We identified five SNPs with R^2 values between 0.44 and 0.96 for rs1568952 and between 0.54 and 0.76 for rs2726600 (Table S7). These five SNPs had been genotyped in our primary family-based sample and provided p values between 0.057 and 0.87 for association with PTB (Table S4). We successfully imputed these SNPs in our case-control replication study (Table S4) and performed a combined analysis of the family-based and case-control samples. The signal was stronger in younger subjects for all SNPs. rs2726596 and rs2726598, the two SNPs in strongest LD with rs1568952 in the Moroccan population, yielded p values between those for rs1568952 and rs2726600 ($p = 3 \times 10^{-6}$ and 8×10^{-7}, respectively), whereas the SNP for which the lowest level of LD was observed (rs2252970) was not significant in any of these analyses. However, none of the associations tested was more significant than that for rs1568952. We also carried out multivariable regression analysis (Table S4) with these seven SNPs. When the most significant SNP (rs1568952) was introduced into the model, the other SNPs ceased to be significantly associated with early-onset PTB, confirming that this association consisted of a single signal. Overall, these analyses identified a cluster of variants associated with

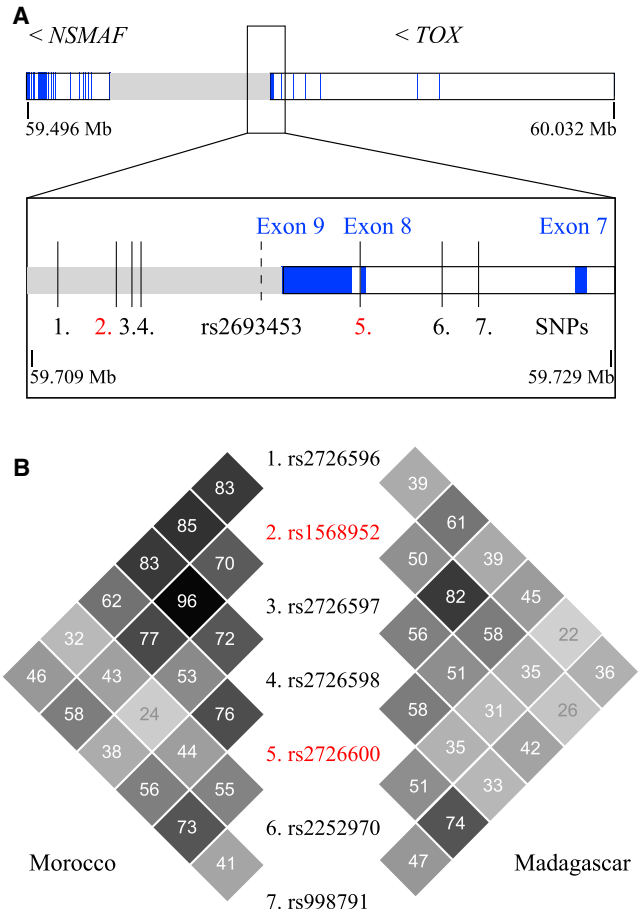

Figure 2. Association between PTB and a Cluster of SNPs in the 3′ Region of *TOX*

(A) Chromosomal location of the SNPs in the 8q region, from 59.496 to 60.032 Mb. *NSMAF* and *TOX* are outlined in black, exons are in blue, and the intergenic region is in gray. The direction of transcription is indicated by the "<" sign. Locations of the cluster of seven SNPs (numbered 1–7), including rs1568952 and rs2726600, are indicated by black vertical lines in the magnified region. The location of rs2693453, which was found to be in strong LD with this cluster in the 1000 Genomes Project but was not genotyped in the two study populations, is indicated by a black dashed vertical line.

(B) LD patterns for the cluster of seven SNPs genotyped in the Moroccan family-based study (left) and the Madagascan family-based study (right). Pairwise R^2 values for all pairs of SNPs are given as percentages, and shading from white to black indicates intensity (R^2 ranges from 0 to 1). The two SNPs (rs1568952 and rs2726600) replicated in the Moroccan studies are numbered in red.

early-onset PTB in Morocco, and rs1568952 displayed the strongest association.

Finally, we tested this cluster of replicated SNPs in a family-based sample consisting of 243 nuclear families from Madagascar (Table S8). These families included 244 offspring with PTB, a sex ratio of 1.54, and a mean (±SD) age at TB diagnosis of 26.5 (±10.0) years (Table S2); age distributions by affection status are illustrated in Figure S2. A single SNP (rs998791) failed but was successfully imputed (Table 2). The frequencies of the alleles associated with a risk of PTB in the Moroccan studies (frequencies ~0.38) were lower in the Madagascan study population (frequencies ~0.15) than in the Moroccan population, decreasing the number of families informative for FBAT analyses (Table 2). However, rs2726600 was found to be significantly associated with PTB in the subsample of Madagascan affected individuals with an age at diagnosis under 25 years ($p = 0.042$), whereas the number of informative families was too low for testing at rs1568952. Interestingly, SNP rs2726598, which had the highest level of LD with rs1568952 in Morocco ($R^2 = 0.96$) and Madagascar ($R^2 = 0.81$), also tended to be associated with PTB in younger age groups in the Madagascan sample, although this trend was not significant ($p = 0.065$). Two additional SNPs, rs2726597 and rs2726596, in strong LD with rs1568952 in Madagascar ($R^2 = 0.52$ and 0.40, respectively) had a minor allele frequency (MAF) (0.22) higher than that of rs1568952 and therefore yielded a larger number of families informative for association. These two SNPs were significantly associated (p values of 0.003 for rs2726597 and 0.029 for rs2726596) with PTB in the full Madagascan sample, and this association was, again, stronger in the younger PTB-affected individuals (Table 2). For rs2726597, the OR of developing PTB before the age of 25 years was estimated at 5.61 (1.82–17.29). Finally, no evidence was found for an association between rs2252970 and PTB, consistent with the results obtained in Morocco. These results for the Madagascan sample strongly support the association between a cluster of SNPs around rs1568952 and rs2726600 and the development of early-onset PTB, validating the initial findings we obtained in Morocco.

Analyses in the two study populations from Morocco identified rs1568952 as the SNP with the strongest evidence for association from this cluster. However, the low MAF and correspondingly small number of informative families, particularly for the age group under 25 years, made it difficult to assess the effect of rs1568952 in the Madagascan population, but the effect of rs2726600 was replicated. The SNPs rs2726596 and rs2726597 were the most strongly associated with PTB in the Madagascan study, and these markers were in strong LD with rs1568952 in both the Moroccan and Madagascan populations (Table 2) and with rs2726600 in the CEU (Utah residents with ancestry from northern and western Europe from the CEPH collection) population (Table S7). Refined analysis of the cluster, including pairwise interaction studies in the whole Moroccan sample, and haplotype analysis in the two family-based samples from Morocco and Madagascar did not identify any combination of SNPs for which there was substantially stronger evidence of association than for the best single SNP. A search for proxy SNPs in the 1000 Genomes Project with the web-based application SNAP[42] led to the identification of one other SNP, rs2693453, in strong LD ($R^2 > 0.6$) with the six associated SNPs in the CEU population ($R^2 = 0.84$ with rs1568952 and 0.80 with rs2726600). SNP rs2693453 was not released in the HapMap project and

Table 2. Genetic Association Results at rs1568952, rs2726600, and Five SNPs in Linkage Disequilibrium with rs1568952 in 243 Nuclear Families from Madagascar under the Recessive Model for the Minor Allele

SNP[a]	Position[b]	Minor Allele	Major Allele	MAF[c]	R^{2}[d]	Full Population			Under 25 Years		
						n	p Value	OR (95% CI)[e]	n	p Value	OR (95% CI)[e]
rs1568952	59,712,363	A	G	0.14 (0.36)	-	4[f]	-	-	-	-	-
rs2726600	59,720,604	G	A	0.15 (0.40)	0.57 (0.77)	10	0.065	1.79 (0.69–4.62)	8	0.042	2.68 (0.89–8.05)
rs2726598	59,713,199	A	G	0.14 (0.37)	0.81 (0.96)	6	0.19	2.1 (0.52–8.4)	5	0.065	3.71 (0.80–17.14)
rs2726597	59,712,891	T	G	0.22 (0.44)	0.52 (0.71)	21	0.009	2.85 (1.34–6.05)	11	0.0028	5.61 (1.82–17.29)
rs2726596	59,710,386	T	C	0.22 (0.40)	0.40 (0.84)	21	0.045	1.86 (0.91–3.78)	15	0.029	2.62 (1.11–6.16)
rs2252970	59,723,371	C	A	0.23 (0.46)	0.37 (0.44)	18	>0.5	-	11	>0.5	-
rs998791[g]	59,724,602	C	T	0.12 (0.47)	0.26 (0.58)	3[f]	-	-	-	-	-

The following abbreviations are used: MAF, minor allele frequency; OR, odds ratio; CI, confidence interval; and n, number of informative families for FBAT.
[a]Selected SNPs for the validation population from Madagascar were genotyped with custom-designed Illumina VeraCode GoldenGate Assays. Departures from Hardy-Weinberg equilibrium at each marker and MAFs were calculated among founders. Replicated SNPs were tested with the use of FBATs and conditional-logistic-regression analyses with a one-sided test and a type I error rate of 0.05.
[b]Position on chromosome 8 according to human assembly GRCh37.p5 coordinates.
[c]Values in Morocco are in parentheses.
[d]Pairwise R^2 values with rs1568952 in Madagascar. Values in Morocco are in parentheses.
[e]ORs and one-sided 95% CIs were derived from the one-sided family-based association test under the recessive model.
[f]Too few informative families to be tested.
[g]SNP rs99871 was imputed in the study population from Madagascar. See Table S4.

is present within the span of the chromosome 8 region defined by our cluster of associated SNPs (Figure 2). We did not find any additional highly correlated SNPs in the sub-Saharan African or the two Asian HapMap populations. Given the discovery of the association with PTB in a Moroccan population very similar to the European population in terms of MAF and LD patterns, the panel of variants representative of our signal is probably restricted to those identified in our cluster of six SNPs (with the addition of rs2693453).

We then used data from the ENCODE project to investigate the potential involvement of the region defined by our SNP cluster in transcriptional regulation. The cluster of associated SNPs overlaps the 3′ region of *TOX* and the downstream neighbor *NSMAF* (140 kb from the cluster) (Figure 2). Our analyses strongly suggest that rs2726600 is located in an important regulatory region of ~200 bp. This region is characterized by the presence of the enhancer-associated histone mark (H3K4Me1) and DNase I hypersensitivity sites in almost all cell types tested. These sites are the strongest known predictors of transcriptional activity[43–45] and are bound by a large number of transcription factors (e.g., NF-κB and STAT1), as shown by chromatin-immunoprecipitation sequencing[44] (Figure S4). We therefore explored the possible involvement of the PTB-associated SNPs in changes to *TOX* and *NSMAF* expression by using available data from a previous study exploring expression quantitative trait loci in primary dendritic cells (DCs) from 65 individuals of European descent before and after infection with *M. tuberculosis*.[46] Genotype data were available for SNP rs2726600, and expression levels for DCs showed *TOX* expression to be significantly weaker ($p = 0.037$) in the group with the at-risk homozygous genotype at baseline (Figure S5), whereas *NSMAF* expression was similar across genotypes (data not shown). These initial findings require confirmation—particularly in other cell types, such as T cells—but they suggest that because the SNP cluster resides in the transcription-factor binding site, it might be involved in the modulation of *TOX* expression.

We report the discovery of susceptibility variants for PTB on the basis of a refined LD-mapping study of chromosomal region 8q12–13, which we previously identified as linked with PTB in a positional-cloning strategy applied successively to three independent samples from Morocco and Madagascar. A cluster of six correlated SNPs was associated with PTB, and this association was particularly strong in subjects who were diagnosed with TB before the age of 25 years. Some of the multiplex families from the original linkage study[37] were included in the present study, and we found some support for the hypothesis that the associated SNPs might in part explain the linkage peak, although the sample size was clearly too small for any definitive conclusions to be drawn (Table S1). The frequencies of the PTB risk alleles for these SNPs were lower in Madagascar than in Morocco, but we were nevertheless able to replicate the signal supported by the two Moroccan populations in a highly admixed Madagascan population with source populations from Asia and Africa[47] through the use of a family-based study design robust to population stratification. Accounting for different LD patterns across ethnicities was shown to be critical for the validation of SNPs in the *PARKIN* regulatory region associated with leprosy in an Indian population after their discovery in a Vietnamese population.[38,48] Our results further demonstrate the importance of carefully considering LD patterns in the process of validating association findings across ethnicities.

One of the major findings of our study was the demonstration that the effect of this cluster of SNPs was clearly age dependent: it was almost entirely restricted to subjects who developed PTB before the age of 25 years. As an example, the OR of developing PTB at rs1568952 for homozygotes for the risk allele was 2.18 in the full population and 3.09 in the group under 25 years old in Morocco. In addition, when we decreased the age cutoff to 18 years, the OR increased to 5.03, which provides further evidence of the inverse relationship between age at diagnosis and strength of the genetic effect (Figure S3). A similar effect of age has already been reported for leprosy: the estimated risk of the *LTA*+80 variant is much higher in early-onset than in late-onset leprosy subjects.[39] Molecular, immunological, and clinical data, as well as other theoretical and population-genetics considerations, have led to the development of a theory explaining the relationship between age at disease onset and the underlying genetic factors.[49] According to this model, the genetic architecture of infectious diseases is a continuous spectrum extending from the role of single genes with strong individual effects for Mendelian disorders of childhood to the role of polygenic factors with modest individual effects for complex diseases in adults. The general idea is that in areas in which a particular microbe is endemic, stronger individual genetic effects are associated with an earlier onset of the corresponding disease.[49] The results reported here are consistent with this hypothesis and demonstrate the strong age dependence of at least some of the genetic factors underlying susceptibility to PTB.

The cluster of associated SNPs defines a 15 kb region overlapping the 3′ region of *TOX* (Figure 2). SNPs rs998791 and rs2726600 are located in introns 7 and 8 of *TOX*, respectively, whereas the remaining four SNPs are in the 3′ UTR of the gene. *TOX* encodes the nuclear factor thymocyte selection-associated high-mobility group box, which is known for its role in the generation of the immune system, including the development of T cells and natural killer cells, and lymph node organogenesis.[50] In particular, *TOX* is required for the development of the CD4 T lineage.[51] The high incidence of PTB in HIV^+ individuals bears witness to the critical role of CD4 T cells in maintaining immunological equilibrium after infection with *M. tuberculosis*,[52] and the role of T cells in mycobacterial infection and disease is an area of active research.[53] The neighboring downstream gene, located 140 kb from the SNP cluster, is *NSMAF*, which encodes the neutral sphingomyelinase (N-SMase)-activation-associated factor, also known as FAN (factor associated with N-SMase activation). FAN is an adaptor protein that binds constitutively to TNF-R1 and has been implicated in TNF-induced gene expression and leukocyte recruitment to inflammatory sites.[54] Treatments targeting TNF have been shown to disrupt the immunological equilibrium of infected individuals and subsequently lead to PTB. *NSMAF* might therefore also be of interest for functional exploration, although FAN-deficient mice have been shown to be no more susceptible than wild-type mice to infection with several intracellular pathogens.[54]

The classification of PTB-affected individuals encompasses considerable phenotypic heterogeneity and summarizes a number of events—from infection to the diagnosis of clinical symptoms of variable nature—arising through different mechanisms. To some extent, this might account for the difficulties in replicating findings of genetic association with PTB in candidate-gene studies and for the limited insight provided by the largest GWAS reported to date.[34,35] The present study highlights the merits of alternative strategies, such as positional cloning, and the importance of refining the PTB phenotype by accounting for age at first PTB diagnosis in particular. In areas where *M. tuberculosis* is endemic and where migration is negligible, age at first PTB onset is clearly correlated with the duration of the latency period. Identified here, the signal close to *TOX* therefore appears to play an important role in prompting the critical transition from infection to active pulmonary disease. It is not yet possible to determine whether this effect influences short (i.e., <2 years in the context of a primary infection or reinfection) or long (i.e., >2 years in the context of a reactivation) latency periods or both. Our results highlight the importance of reducing phenotypic heterogeneity so that it becomes easier to decipher the complex mechanisms involved in the development of clinical TB.

Supplemental Data

Supplemental Data include five figures and eight tables and can be found with this article online at http://www.cell.com/AJHG.

Acknowledgments

We thank all pulmonary-tuberculosis (PTB)-affected study participants and family members from Morocco and Madagascar and all healthy Moroccan subjects for their participation in this study. We thank the late M. Chentoufi (Centre de Diagnostic et Traitement des Maladies Respiratoires [CDTMR] of Hay Mohammadi), I. Sentissi and F. Bourkadi (CDTMR of Salé), and F. El Bordi for recruitment of PTB-affected subjects. We thank N. El Amraoui and M. Hakam (National Blood Transfusion Center) for recruiting healthy donors. We thank B. Rakotondrasoa and D. Rakotomanajanahary for recruiting PTB-affected subjects and family members in Madagascar. We thank E. Ranaivoson, J. Razananirinomenjanahary, B. Ravalison, D. Ramarinarisoa, and the staff at the Dispensaire Anti-tuberculeux de Tananarive, as well as Professors H. Raobijaona and N. Ravelomanana and the staff of the pediatric units of Antananarivo hospitals, for assistance. We thank the Centre National de Génotypage for carrying out the genotyping for the initial fine-mapping screen. This work was supported by grants from the French National Agency for Research, European Commission (HEALTH-F3-2008-200732), European Research Council (ERC-2010-AdG-268777), Pasteur Institute of Paris (Progamme Transversal de Recherche PTR202), Bill and Melinda Gates Foundation, St. Giles Foundation, Jeffrey Modell Foundation and Talecris Biotherapeutics, Rockefeller University Center for Clinical and Translational Science (5UL1RR024143-04),

Rockefeller University, and National Institute of Allergy and Infectious Diseases (1R01AI089970-01 and U01AI088685). A.V.G. was partly supported by the Fondation pour la Recherche Médicale and Fondation BNP Paribas. L.B.B. is a scholar of the Fonds de la Recherche en Santé du Québec.

Received: September 11, 2012
Revised: October 12, 2012
Accepted: January 22, 2013
Published: February 14, 2013

Web Resources

The URLs for the data presented herein are as follows:

1000 Genomes Project, http://www.1000genomes.org/
ENCODE, http://genome.ucsc.edu/ENCODE/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org
SNP Annotation and Proxy Search (SNAP), http://www.broadinstitute.org/mpg/snap/

References

1. World Health Organization. (2011). Global tuberculosis control: WHO report 2011 (France: World Health Organization).
2. Casanova, J.-L., and Abel, L. (2002). Genetic dissection of immunity to mycobacteria: The human model. Annu. Rev. Immunol. *20*, 581–620.
3. Ernst, J.D. (2012). The immunological life cycle of tuberculosis. Nat. Rev. Immunol. *12*, 581–591.
4. Cobat, A., Gallant, C.J., Simkin, L., Black, G.F., Stanley, K., Hughes, J., Doherty, T.M., Hanekom, W.A., Eley, B., Beyers, N., et al. (2010). High heritability of antimycobacterial immunity in an area of hyperendemicity for tuberculosis disease. J. Infect. Dis. *201*, 15–19.
5. Cobat, A., Barrera, L.F., Henao, H., Arbeláez, P., Abel, L., García, L.F., Schurr, E., and Alcaïs, A. (2012). Tuberculin skin test reactivity is dependent on host genetic background in Colombian tuberculosis household contacts. Clin. Infect. Dis. *54*, 968–971.
6. Jepson, A., Fowler, A., Banya, W., Singh, M., Bennett, S., Whittle, H., and Hill, A.V. (2001). Genetic regulation of acquired immune responses to antigens of Mycobacterium tuberculosis: A study of twins in West Africa. Infect. Immun. *69*, 3989–3994.
7. Sepulveda, R.L., Heiba, I.M., Navarrete, C., Elston, R.C., Gonzalez, B., and Sorensen, R.U. (1994). Tuberculin reactivity after newborn BCG immunization in mono- and dizygotic twins. Tuber. Lung Dis. *75*, 138–143.
8. Cobat, A., Gallant, C.J., Simkin, L., Black, G.F., Stanley, K., Hughes, J., Doherty, T.M., Hanekom, W.A., Eley, B., Jaïs, J.-P., et al. (2009). Two loci control tuberculin skin test reactivity in an area hyperendemic for tuberculosis. J. Exp. Med. *206*, 2583–2591.
9. Möller, M., and Hoal, E.G. (2010). Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. Tuberculosis (Edinb.) *90*, 71–83.
10. Boisson-Dupuis, S., El Baghdadi, J., Parvaneh, N., Bousfiha, A., Bustamante, J., Feinberg, J., Samarina, A., Grant, A.V., Janniere, L., El Hafidi, N., et al. (2011). IL-12Rβ1 deficiency in two of fifty children with severe tuberculosis from Iran, Morocco, and Turkey. PLoS ONE *6*, e18524.
11. Alcaïs, A., Fieschi, C., Abel, L., and Casanova, J.-L. (2005). Tuberculosis in children and adults: Two distinct genetic diseases. J. Exp. Med. *202*, 1617–1621.
12. Filipe-Santos, O., Bustamante, J., Chapgier, A., Vogt, G., de Beaucoudrey, L., Feinberg, J., Jouanguy, E., Boisson-Dupuis, S., Fieschi, C., Picard, C., and Casanova, J.L. (2006). Inborn errors of IL-12/23- and IFN-gamma-mediated immunity: molecular, cellular, and clinical features. Semin. Immunol. *18*, 347–361.
13. Bustamante, J., Arias, A.A., Vogt, G., Picard, C., Galicia, L.B., Prando, C., Grant, A.V., Marchal, C.C., Hubeau, M., Chapgier, A., et al. (2011). Germline CYBB mutations that selectively affect macrophages in kindreds with X-linked predisposition to tuberculous mycobacterial disease. Nat. Immunol. *12*, 213–221.
14. Hambleton, S., Salem, S., Bustamante, J., Bigley, V., Boisson-Dupuis, S., Azevedo, J., Fortin, A., Haniffa, M., Ceron-Gutierrez, L., Bacon, C.M., et al. (2011). IRF8 mutations and human dendritic-cell immunodeficiency. N. Engl. J. Med. *365*, 127–138.
15. Bogunovic, D., Byun, M., Durfee, L.A., Abhyankar, A., Sanal, O., Mansouri, D., Salem, S., Radovanovic, I., Grant, A.V., Adimi, P., et al. (2012). Mycobacterial disease and impaired IFN-γ immunity in humans with inherited ISG15 deficiency. Science *337*, 1684–1688.
16. 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.
17. Fieschi, C., Dupuis, S., Catherinot, E., Feinberg, J., Bustamante, J., Breiman, A., Altare, F., Baretto, R., Le Deist, F., Kayal, S., et al. (2003). Low penetrance, broad resistance, and favorable outcome of interleukin 12 receptor beta1 deficiency: Medical and immunological implications. J. Exp. Med. *197*, 527–535.
18. de Beaucoudrey, L., Samarina, A., Bustamante, J., Cobat, A., Boisson-Dupuis, S., Feinberg, J., Al-Muhsen, S., Jannière, L., Rose, Y., de Suremain, M., et al. (2010). Revisiting human IL-12Rβ1 deficiency: A survey of 141 patients from 30 countries. Medicine (Baltimore) *89*, 381–402.
19. Tabarsi, P., Marjani, M., Mansouri, N., Farnia, P., Boisson-Dupuis, S., Bustamante, J., Abel, L., Adimi, P., Casanova, J.-L., and Mansouri, D. (2011). Lethal tuberculosis in a previously healthy adult with IL-12 receptor deficiency. J. Clin. Immunol. *31*, 537–539.
20. Altare, F., Ensser, A., Breiman, A., Reichenbach, J., Baghdadi, J.E., Fischer, A., Emile, J.F., Gaillard, J.L., Meinl, E., and Casanova, J.L. (2001). Interleukin-12 receptor beta1 deficiency in a patient with abdominal tuberculosis. J. Infect. Dis. *184*, 231–236.
21. Caragol, I., Raspall, M., Fieschi, C., Feinberg, J., Larrosa, M.N., Hernández, M., Figueras, C., Bertrán, J.-M., Casanova, J.-L., and Español, T. (2003). Clinical tuberculosis in 2 of 3 siblings with interleukin-12 receptor beta1 deficiency. Clin. Infect. Dis. *37*, 302–306.
22. Ozbek, N., Fieschi, C., Yilmaz, B.T., de Beaucoudrey, L., Demirhan, B., Feinberg, J., Bikmaz, Y.E., and Casanova, J.-L. (2005). Interleukin-12 receptor beta 1 chain deficiency in a child with disseminated tuberculosis. Clin. Infect. Dis. *40*, e55–e58.
23. Malik, S., Abel, L., Tooker, H., Poon, A., Simkin, L., Girard, M., Adams, G.J., Starke, J.R., Smith, K.C., Graviss, E.A., et al. (2005).

Alleles of the NRAMP1 gene are risk factors for pediatric tuberculosis disease. Proc. Natl. Acad. Sci. USA *102*, 12183–12188.
24. Comstock, G.W. (1978). Tuberculosis in twins: A re-analysis of the Prophit survey. Am. Rev. Respir. Dis. *117*, 621–624.
25. van der Eijk, E.A., van de Vosse, E., Vandenbroucke, J.P., and van Dissel, J.T. (2007). Heredity versus environment in tuberculosis in twins: the 1950s United Kingdom Prophit Survey Simonds and Comstock revisited. Am. J. Respir. Crit. Care Med. *176*, 1281–1288.
26. Puffer, R.R. (1944). Familial Susceptibility to Tuberculosis: Its Importance as a Public Health Problem (Cambridge, MA: Harvard University Press).
27. Dubos, R.J., and Dubos, J. (1987). The White Plague: Tuberculosis, Man, and Society (New Brunswick, NJ: Rutgers University Press).
28. Goldfeld, A.E., Delgado, J.C., Thim, S., Bozon, M.V., Uglialoro, A.M., Turbay, D., Cohen, C., and Yunis, E.J. (1998). Association of an HLA-DQ allele with clinical tuberculosis. JAMA *279*, 226–228.
29. Delgado, J.C., Baena, A., Thim, S., and Goldfeld, A.E. (2006). Aspartic acid homozygosity at codon 57 of HLA-DQ beta is associated with susceptibility to pulmonary tuberculosis in Cambodia. J. Immunol. *176*, 1090–1097.
30. Bellamy, R., Ruwende, C., Corrah, T., McAdam, K.P., Whittle, H.C., and Hill, A.V. (1998). Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N. Engl. J. Med. *338*, 640–644.
31. Greenwood, C.M., Fujiwara, T.M., Boothroyd, L.J., Miller, M.A., Frappier, D., Fanning, E.A., Schurr, E., and Morgan, K. (2000). Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am. J. Hum. Genet. *67*, 405–416.
32. Zhang, W., Shao, L., Weng, X., Hu, Z., Jin, A., Chen, S., Pang, M., and Chen, Z.W. (2005). Variants of the natural resistance-associated macrophage protein 1 gene (NRAMP1) are associated with severe forms of pulmonary tuberculosis. Clin. Infect. Dis. *40*, 1232–1236.
33. Berry, M.P.R., Graham, C.M., McNab, F.W., Xu, Z., Bloch, S.A.A., Oni, T., Wilkinson, K.A., Banchereau, R., Skinner, J., Wilkinson, R.J., et al. (2010). An interferon-inducible neutrophil-driven blood transcriptional signature in human tuberculosis. Nature *466*, 973–977.
34. 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.
35. 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.
36. Comas, I., Chakravartti, J., Small, P.M., Galagan, J., Niemann, S., Kremer, K., Ernst, J.D., and Gagneux, S. (2010). Human T cell epitopes of Mycobacterium tuberculosis are evolutionarily hyperconserved. Nat. Genet. *42*, 498–503.
37. Baghdadi, J.E., Orlova, M., Alter, A., Ranque, B., Chentoufi, M., Lazrak, F., Archane, M.I., Casanova, J.-L., Benslimane, A., Schurr, E., and Abel, L. (2006). An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J. Exp. Med. *203*, 1679–1684.
38. Mira, M.T., Alcaïs, A., Nguyen, V.T., Moraes, M.O., Di Flumeri, C., Vu, H.T., Mai, C.P., Nguyen, T.H., Nguyen, N.B., Pham, X.K., et al. (2004). Susceptibility to leprosy is associated with PARK2 and PACRG. Nature *427*, 636–640.
39. Alcaïs, A., Alter, A., Antoni, G., Orlova, M., Nguyen, V.T., Singh, M., Vanderborght, P.R., Katoch, K., Mira, M.T., Vu, H.T., et al. (2007). Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. Nat. Genet. *39*, 517–522.
40. Alter, A., Huong, N.T., Singh, M., Orlova, M., Van Thuc, N., Katoch, K., Gao, X., Thai, V.H., Ba, N.N., Carrington, M., et al. (2011). Human leukocyte antigen class I region single-nucleotide polymorphisms are associated with leprosy susceptibility in Vietnam and India. J. Infect. Dis. *203*, 1274–1281.
41. Kraft, P. (2008). Curses—winner's and otherwise—in genetic epidemiology. Epidemiology *19*, 649–651, discussion 657–658.
42. Johnson, A.D., Handsaker, R.E., Pulit, S.L., Nizzari, M.M., O'Donnell, C.J., and de Bakker, P.I.W. (2008). SNAP: A web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics *24*, 2938–2939.
43. Degner, J.F., Pai, A.A., Pique-Regi, R., Veyrieras, J.-B., Gaffney, D.J., Pickrell, J.K., De Leon, S., Michelini, K., Lewellen, N., Crawford, G.E., et al. (2012). DNase I sensitivity QTLs are a major determinant of human expression variation. Nature *482*, 390–394.
44. ENCODE Project Consortium. (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.
45. Ernst, J., Kheradpour, P., Mikkelsen, T.S., Shoresh, N., Ward, L.D., Epstein, C.B., Zhang, X., Wang, L., Issner, R., Coyne, M., et al. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.
46. Barreiro, L.B., Tailleux, L., Pai, A.A., Gicquel, B., Marioni, J.C., and Gilad, Y. (2012). Deciphering the genetic architecture of variation in the immune response to Mycobacterium tuberculosis infection. Proc. Natl. Acad. Sci. USA *109*, 1204–1209.
47. Tofanelli, S., Bertoncini, S., Castrì, L., Luiselli, D., Calafell, F., Donati, G., and Paoli, G. (2009). On the origins and admixture of Malagasy: New evidence from high-resolution analyses of paternal and maternal lineages. Mol. Biol. Evol. *26*, 2109–2124.
48. Alter, A., Fava, V.M., Huong, N.T., Singh, M., Orlova, M., Van Thuc, N., Katoch, K., Thai, V.H., Ba, N.N., Abel, L., et al. (2013). Linkage disequilibrium pattern and age-at-diagnosis are critical for replicating genetic associations across ethnic groups in leprosy. Hum. Genet. *132*, 107–116.
49. Alcaïs, A., Quintana-Murci, L., Thaler, D.S., Schurr, E., Abel, L., and Casanova, J.-L. (2010). Life-threatening infectious diseases of childhood: single-gene inborn errors of immunity? Ann. N Y Acad. Sci. *1214*, 18–33.
50. Aliahmad, P., Seksenyan, A., and Kaye, J. (2012). The many roles of TOX in the immune system. Curr. Opin. Immunol. *24*, 173–177.
51. Aliahmad, P., and Kaye, J. (2008). Development of all CD4 T lineages requires nuclear factor TOX. J. Exp. Med. *205*, 245–256.
52. Rom, W.N., and Garay, S.M. (2004). Tuberculosis (Philadelphia: Lippincott Williams & Wilkins).
53. Cooper, A.M. (2009). T cells in mycobacterial infection and disease. Curr. Opin. Immunol. *21*, 378–384.
54. Montfort, A., de Badts, B., Douin-Echinard, V., Martin, P.G.P., Iacovoni, J., Nevoit, C., Therville, N., Garcia, V., Bertrand, M.-A., Bessières, M.-H., et al. (2009). FAN stimulates TNF(alpha)-induced gene expression, leukocyte recruitment, and humoral response. J. Immunol. *183*, 5369–5378.