Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy

Furen Zhang1–4, Hong Liu1,3, Shumin Chen1,3, Huiqi Low5, Liangdan Sun6,7, Yong Cui6,7, Tongsheng Chu1,3, Yi Li5, Xi’an Fu1,3, Yongxiang Yu1,3, Gongqi Yu1,3, Benqin Shi2–4, Hongqing Tian2,4, Dianchang Liu1,3, Xiulu Yu1,3, Jinhu Li1,3, Nan Lu1,3, Fangfang Bao1,3, Chunying Yuan1,3, Jian Liu1,3, Huaxu Liu1,3, Lin Zhang1,3, Yonghu Sun1,3, Mingfei Chen1,3, Qing Yang1,3, Haitao Yang8, Rongde Yang9, Lianhua Zhang8, Qiang Wang10, Hong Liu1,3, Fuguang Zuo1,3, Haizhen Zhang1,3, Chiea Chuen Khor1,11, Martin L Hibberd12,15, Sen Yang6,7, Jianjun Liu1,5,11,13 & Xuejun Zhang5,7

We performed a genome-wide association study with 706 individuals with leprosy and 5,581 control individuals and replicated the top 24 SNPs in three independent replication samples, including a total of 3,301 individuals with leprosy and 5,299 control individuals from China. Two loci not previously associated with the disease identified with genome-wide significance: rs2275606 (combined P = 3.94 × 10−14, OR = 1.30) on 6q24.3 and rs3762318 (combined P = 3.27 × 10−11, OR = 0.69) on 1p31.3. These associations implicate IL23R and RAB32 as new susceptibility genes for leprosy. Furthermore, we identified evidence of interaction between the NOD2 and RIPK2 loci, which is consistent with the biological association of the proteins encoded by these genes (NOD2-RIPK2 complex) in activating the NF-κB pathway as a part of the host defense response to infection. Our findings have expanded the biological functions of IL23R by uncovering its involvement in infectious disease susceptibility and suggest a potential involvement of autophagocytosis in leprosy pathogenesis. The IL23R association supports previous observations of the marked overlap of susceptibility genes for leprosy and Crohn’s disease, implying common pathogenesis mechanisms.

Infectious diseases represent major health problems worldwide, with the vast majority of the disease burden falling on developing countries. Pathogen, host genetic and environmental factors interact to determine both the susceptibility to a particular microbial infection and the course of infection. Although exposure to the pathogen is critical for the initiation of pathogenesis, recent breakthroughs using genome-wide association studies (GWAS) have firmly established roles for host genetic factors in human susceptibility to infection and in the progression of infectious diseases.1,2

Leprosy is a chronic granulomatous infectious disease caused by Mycobacterium leprae that affects both the skin and peripheral nerves. Although the prevalence of leprosy has declined dramatically since the introduction of multidrug therapy in the 1980s, more than 200,000 new cases are reported globally each year, and leprosy remains a major public health problem, especially in China and India3. Family studies and population epidemiological surveys have clearly demonstrated a substantial contribution of host genetics to the susceptibility of individuals to leprosy, with estimated heritability of up to 57% (ref. 4). However, there is an incomplete understanding of the genetic basis of leprosy, which is compounded by the lack of suitable animal hosts for M. leprae and the difficulty of culturing it in vitro; both limitations have greatly hindered research on the mechanisms underlying leprosy.

In 2009, we performed a GWAS of leprosy and identified six susceptibility loci (CCDC122, LACCI (C1orf31), NOD2, TNFSF15, RIPK2 and the HLA-DR–HL-DQ locus) in the Chinese population, which indicated the importance of NOD2-mediated innate immunity in protection against infection by M. leprae5. More recently, Wong et al. conducted a genome-wide gene-centric association study and found that TLR1, HLA-DRB1 and HLA-DQA1 associated with leprosy in the Indian population6. Although these two studies have provided valuable insights into the mechanism of leprosy progression and revealed remarkable similarity between the susceptibility genes for leprosy and Crohn’s disease, it is clear that many more genetic susceptibility loci remain to be discovered.

In this study, we performed an expanded GWAS of leprosy by combining our published GWAS data set of 706 subjects with leprosy and 1,225 control subjects with an additional 4,367 control subjects of Chinese Han

---

1Shandong Provincial Institute of Dermatology and Venereology, Provincial Academy of Medical Science, Jinan, China. 2Shandong Provincial Hospital for Skin Diseases, Shandong University, Jinan, China. 3Shandong Provincial Key Lab for Dermatovenerology, Jinan, China. 4Shandong Provincial Medical Center for Dermatovenerology, Jinan, China. 5Human Genetics, Genome Institute of Singapore, A*STAR, Singapore. 6Institute of Dermatology and Department of Dermatology at No. 1 hospital, Anhui Medical University, Hefei, China. 7State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Hefei, China. 8Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China. 9Wenshan Institute of Dermatology, Wenshan, China. 10Anhui Provincial Institute of Dermatology, Hefei, China. 11Center for Molecular Epidemiology, National University of Singapore, Singapore. 12Department of Epidemiology and Public Health, National University of Singapore, Singapore. 13School of Life Sciences, Anhui Medical University, Hefei, China. Correspondence should be addressed to F.Z. (zhangfuren@hotmail.com).

Received 9 May; accepted 19 September; published online 23 October 2011; doi:10.1038/ng.973
To minimize the adverse impact of population stratification, association analysis was performed using the top five principal components as covariates via logistic regression. This reduced the extent of genome-wide inflation to 1.04, a level considered acceptable by conventional GWAS standards. Consistent with our previous GWAS, we observed strong associations within the HLA-DR and HLA-DQA1 loci (data not shown) and with the previously identified genes outside of the major histocompatibility complex (MHC) loci (Supplementary Table 1). After excluding the SNPs within these previously identified regions, a substantial presence of extremely small P values remained, thus suggesting the existence of additional associations beyond the ones already identified (Supplementary Fig. 2a). In this expanded analysis, 22 SNPs in 21 new loci were found with \( P < 2.0 \times 10^{-7} \) (Supplementary Fig. 2b).

We then performed a replication study using three independent leprosy samples from populations of Chinese descent: (i) 2,307 individuals with leprosy and 4,585 control subjects of northern Chinese Han descent; (ii) 273 individuals with leprosy and 214 control subjects of southern Chinese Han descent and (iii) 721 individuals with leprosy and 500 control subjects from southern Chinese minority populations (Supplementary Table 2). We selected the most significant SNP from each of the 21 distinct loci mentioned above (with the exception that two SNPs were chosen from the 1p31.3 locus). In addition, rs5743618 (I602S) and rs17616475 in CYLD were also included. Altogether, 24 SNPs within 22 loci were genotyped in the three replication samples, for a total of 3,301 subjects with leprosy and 5,299 control subjects.

The replication analysis identified five SNPs within four loci (the neighboring C1orf141 and IL23R genes, RAB32, RPS6KA4, and CYLD) that showed consistent association in the GWAS and replication samples, and two of these reached genome-wide significance in the combined replication samples alone: rs3762318 at the C1orf141-IL23R locus \((P = 2.67 \times 10^{-9})\) and rs2275606 in RAB32 \((P = 5.51 \times 10^{-12})\). We also performed association analysis of the combined GWAS and replication samples (excluding the additional 4,367 controls from the GWAS discovery analysis), which revealed an additional SNP (rs16948876 in CYLD) that reached genome-wide significance \((P = 1.64 \times 10^{-10})\) (Table 1). Because the additional 4,367 samples used in the genome-wide discovery analysis came from individuals with other diseases, they were excluded from this joint analysis. It has been shown that optimal statistical power for detection can be achieved for a given number of cases by increasing the control to case ratio\(^9\). In light of this observation, we included an additional 4,367 population controls in the GWAS analysis solely to maximize statistical power for the discovery of novel loci, a method that has previously been employed to identify novel susceptibility loci for psoriasis and SLE beyond original GWAS analyses\(^{16,17}\).

In total, three novel associations have been identified with genome-wide significance: rs2275606 in RAB32 on 6q24.3 \((P = 3.94 \times 10^{-14})\), OR = 1.30), rs3762318 in the C1orf141-IL23R locus on 1p31.3 \((P = 3.27 \times 10^{-11})\), OR = 0.69) and rs16948876 in CYLD on 16q12.1.

**Table 1. Summary of association results for the three confirmed loci in the GWAS, replications and combined samples.**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr position</th>
<th>MAF</th>
<th>Gene</th>
<th>Allele</th>
<th>P</th>
<th>OR</th>
<th>Chr position</th>
<th>MAF</th>
<th>Gene</th>
<th>Allele</th>
<th>P</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3762318</td>
<td>Chr:1:67869707</td>
<td>0.04</td>
<td>I602S</td>
<td>AG</td>
<td>0.06</td>
<td>OR</td>
<td>rs3762318</td>
<td>Chr:1:67869707</td>
<td>0.04</td>
<td>I602S</td>
<td>AG</td>
<td>0.06</td>
</tr>
<tr>
<td>rs2275606</td>
<td>Chr:6:146960643</td>
<td>0.02</td>
<td>RAB32</td>
<td>AG</td>
<td>0.02</td>
<td>OR</td>
<td>rs2275606</td>
<td>Chr:6:146960643</td>
<td>0.02</td>
<td>RAB32</td>
<td>AG</td>
<td>0.02</td>
</tr>
<tr>
<td>rs16948876</td>
<td>Chr:16:49411919</td>
<td>0.01</td>
<td>CYLD</td>
<td>AG</td>
<td>0.01</td>
<td>OR</td>
<td>rs16948876</td>
<td>Chr:16:49411919</td>
<td>0.01</td>
<td>CYLD</td>
<td>AG</td>
<td>0.01</td>
</tr>
</tbody>
</table>
Table 2  Estimated genotype ORs of rs40457 stratified by the genotype at rs9302752

<table>
<thead>
<tr>
<th>rs9302752</th>
<th>rs40457</th>
<th>Number (frequency)</th>
<th>Stratified association analysis of rs40457</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>OR</td>
<td>95% CI</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CC</td>
<td>GG</td>
<td>173 (59%)</td>
<td>147 (48%)</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>116 (40%)</td>
<td>127 (43%)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>17 (6%)</td>
<td>19 (6%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>306</td>
<td>293</td>
</tr>
<tr>
<td>CT</td>
<td>GG</td>
<td>495 (52%)</td>
<td>670 (52%)</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>376 (39%)</td>
<td>515 (40%)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>84 (9%)</td>
<td>112 (9%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>955</td>
<td>1,297</td>
</tr>
<tr>
<td>TT</td>
<td>GG</td>
<td>545 (58%)</td>
<td>806 (47%)</td>
</tr>
<tr>
<td></td>
<td>GA</td>
<td>321 (34%)</td>
<td>725 (42%)</td>
</tr>
<tr>
<td></td>
<td>AA</td>
<td>75 (8%)</td>
<td>202 (12%)</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>941</td>
<td>1,733</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>2,202</td>
<td>3,323</td>
</tr>
</tbody>
</table>

*P* values were acquired by genotype test using logistic regression. **P** values were acquired by the 1-degree-of-freedom score test modeled using logistic regression.

(P = 1.64 × 10<sup>-10</sup>, OR = 1.56) (Table 1). Whereas the associations at rs2275606 and rs3762318 did not show any heterogeneity among the four independent GWAS and replication samples, the association at rs16948876 showed moderate heterogeneity (P = 0.03) (Supplementary Table 3), which was largely due to the substantially weaker association in individuals from Chinese minority groups (OR = 1.01) relative to the three samples groups from individuals of Chinese Han descent (OR = 2.55 in the GWAS sample, 1.57 in the replication sample of northern Chinese Han individuals and 1.51 in the replication sample of southern Chinese Han individuals; P<sub>heterogeneity</sub> = 0.16 for samples from the three Chinese Han groups combined). Nonetheless, the association of rs16948876 in CYLD reached genome-wide significance regardless of whether the individuals from Chinese minority groups were included in the analysis (P = 1.64 × 10<sup>-10</sup>, OR = 1.56) or not (P = 8.00 × 10<sup>-12</sup>, OR = 1.66) (Table 1). All these associations retained genome-wide significance after adjusting for age and gender (Supplementary Table 4).

We also investigated the interaction among the identified susceptibility loci. In a subset of subjects that included the 2,202 individuals with leprosy and 3,323 control individuals in which all the susceptibility SNPs in nine loci were genotyped, we performed pairwise interaction analysis by choosing the top SNP from each locus. We saw evidence for an interaction between the SNPs in NOD2 (rs9302752) and RIPK2 (rs40457) (P = 0.011 before correction, P = 0.036 after correction for multiple testing). The stratified analysis showed that the protective association at rs40457 is only significant in subjects with the TT genotype of rs9302752 (P = 3.03 × 10<sup>-14</sup>) but not in subjects with either the CT (P = 0.14) or CC (P = 0.11) genotype (Table 2). Although intriguing, the finding of an interaction between the SNPs in NOD2 (rs9302752) and RIPK2 (rs40457) needs to be validated by independent studies.

We also identified suggestive associations for rs6588248 within IL23R on 1p31.3 (P = 7.93 × 10<sup>-6</sup>, OR = 0.87) and rs538147 within RPS6KA4 on 11q11-13 (P = 2.70 × 10<sup>-6</sup>, OR = 1.17) (Supplementary Table 3). For both SNPs, consistent association was observed in the independent GWAS and the three replication samples, but association in the combined analysis of the GWAS and replication samples failed to reach genome-wide significance (P < 5.0 × 10<sup>-8</sup>). We did not observe evidence of the previously reported association of TLR1 with leprosy<sup>a</sup>. We tested the rs5743618 SNP (I602S) identified in that study and rs17616475, the top SNP in TLR1 in our GWAS and replication sample of 3,301 individuals with leprosy and 5,299 control individuals, but we did not observe evidence of association (rs5743618, P = 0.67 and rs17616475, P = 0.96) (Supplementary Table 3 and Supplementary Fig. 3a). However, we did notice that there was a
substantial difference in the frequency of rs5743618 between Chinese (MAF = 0.01 in Chinese Han) and Indian (MAF = 0.13 in New Delhi) populations. Despite the low allele frequency of 0.01, the Chinese replication sample of 3,301 individuals with leprosy and 5,299 control individuals should have a sufficient power of 99% to detect the association at rs5743618 with OR = 0.31 (as observed in the Indian population) and a substantial power of 70% to detect even weaker association (OR = 0.5) at a significance of 0.001. The clear disparity of the association at TLR1 between the Chinese and Indian populations suggests a possible population-specific effect of TLR1 variation and potential heterogeneity of leprosy susceptibility based on ancestry. This heterogeneity may not be surprising, given the common belief that the human population has been subjected to strong selection by infection, and therefore susceptibility to pathogens is likely to vary across different human populations. The association results for the other 19 SNPs from the replication analysis are summarized (Supplementary Table 3).

The rs3762318 SNP is located within a linkage-disequilibrium block of 150 kb where IL23R and C1orf41 are present (Fig. 1a). This SNP was correlated with the rs6588248 SNP within IL23R (D’ = 0.95). Conditional analysis showed that the association at rs6588248 is not independent from rs3762318 (Pconditional = 0.08), suggesting that the association signal observed on 1p31.3 is likely to be localized to IL23R. Although searches of the SNP and CNV annotation (SCAN) database did not identify any expression quantitative trait locus (eQTL) effect (data not shown), analysis of data from the Sanger Institute Genevar database revealed a moderate eQTL effect of rs3762318 on IL23R expression in the lymphoblastoid cell lines of the 195 subjects of the HapMap project (P = 0.017) (Supplementary Table 5). However, the UCSC browser did not reveal any regulatory element at rs3762318 (data not shown). Whereas the previously reported SNPs in IL23R that associated with Behcet’s disease and ulcerative colitis are in a different linkage-disequilibrium block than rs3762318, the other previously reported SNPs associated with psoriasis, Crohn’s disease and ulcerative colitis are located within the same linkage-disequilibrium block (Supplementary Fig. 3c) but did not show association with leprosy (data not shown)18–21. The protein encoded by IL23R forms a receptor for the interleukin (IL)-23 cytokine and is part of a signaling pathway involving the gene product of another leprosy-associated locus, TNFSF15 (ref. 22). Together with the β1 subunit of IL-12 (encoded by IL12RB1 (ref. 23)), IL23R is part of the IL-12–IL-23 and IFN-γ cascades, which have essential roles in immunity to mycobacteria24.

The rs2275606 SNP is located within a 250-kb linkage-disequilibrium block on 6q24.3 where C6orf103 and RAB32 reside (Fig. 1b). C6orf103 is a hypothetical protein-coding locus, and the function of the corresponding gene product has not been annotated. RAB32 is a member of the Rab superfamily of low molecular weight G proteins and is required for the formation of autophagic vacuoles and the regulation and clearance of aggregated proteins by autophagy24. A recent study reported that the Rab32 protein participates in controlling the recruitment of cathepsin D to phagosomes containing Mycobacterium tuberculosis25, suggesting that Rab32 may have a similar role in the pathogenesis of leprosy.

The rs16948876 SNP was mapped just 19 kb downstream of CYLD and about 140 kb away from the previously identified association within NOD2 (Supplementary Fig. 3b). The two associations are in different linkage-disequilibrium blocks and are separated by a recombination hotspot, and the top SNPs from the two loci, rs16948876 and rs9302752, have moderate linkage disequilibrium (D’ = 0.51, r2 = 0.03 in the HapMap CHB and JPT samples). We genotyped the two SNPs in 3,386 individuals with leprosy and 6,241 control individuals of Chinese Han descent and performed conditional analysis. Both SNPs showed strong association (rs9302752: P = 1.84 × 10−52, OR = 1.64; rs16948876: P = 1.68 × 10−12, OR = 1.64), but conditioning on rs9302752 dramatically weakened the association at rs16948876 (Pconditional = 3.64 × 10−4, ORconditional = 1.30). Therefore, further study is needed to confirm whether rs16948876 (in CYLD) is truly an independent association locus.

Whereas the identification of IL23R as a new susceptibility gene has firmly established the involvement of innate immunity in the pathogenesis of leprosy, the identification of RAB32 has provided new biological insight into the mechanism of leprosy development. This finding highlights the potential involvement of autophagy in host defense against M. leprae infection25, which has parallels with Crohn’s disease26. Furthermore, our discovery has expanded the known biological functions of IL23R by demonstrating, for the first time, its involvement in infectious disease. The observed interaction between the NOD2 and RIPK2 loci is also interesting because the NOD2 and RIRK2 proteins form a NOD2-RIRK2 complex in the NOD2-mediated signaling pathway that activates nuclear factor NF-κB as a part of the host immune response to infection.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

ACKNOWLEDGMENTS

We thank the individuals who participated in this project. This work was funded by grants from the National Natural Science Foundation of China (8107288, 81072391), the Project of Research Foundation of the Shandong Provincial Institute of Dermatology and Venereology (2008.07–2011.07), the Shandong Provincial Leprosy Control Special Financial Support (2009–2011), the Anhui Provincial Special Scientific Program (2007.07–2011.07) and the Agency for Science, Technology, and Research of Singapore. Grant support was also provided by the Project of Taishan scholar (2008–) and the Project of Medical leading scholar of Shandong Province (2010–). F. Zhang was supported by funding from the Project of Taishan scholar award (2008–) and the Project of Medical Leading Scholar of Shandong Province award (2010–).

AUTHOR CONTRIBUTIONS


COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/


Note: Supplementary information is available on the Nature Genetics website.
ONLINE METHODS

Study subjects. To increase the power of discovery analysis by GWAS, we combined samples from 706 individuals with leprosy and 1,225 control individuals used in our previously published GWAS with an additional 4,367 population control subjects that were genotyped through a series of GWAS of the Chinese population, including 10 newly genotyped control subjects, 1,012 individuals with atopic dermatitis, 1,139 individuals with psoriasis, 1,082 individuals with SLE and 1,124 individuals with vitiligo. All individuals were of Chinese Han descent and were genotyped using the same Illumina Human 610-Quad Bead chips. Three independent replication samples were used in the validation study, including 2,307 individuals with leprosy and 4,585 control individuals of Chinese Han descent recruited from northern China (Anhui, Shandong and Jiangsu provinces), 273 individuals with leprosy and 214 control individuals of Chinese Han descent from southern China (Yunnan province) and 721 individuals with leprosy and 500 control individuals from Chinese minority populations from Yunnan province, including subjects of Chuang (314 individuals with leprosy and 221 control individuals), Miaoas (189 individuals with leprosy and 157 control individuals) and Yizu descent (154 individuals with leprosy and 113 control individuals) and from several other smaller Chinese minority populations (64 individuals with leprosy and 9 control individuals in total). All subjects with leprosy and control subjects were recruited using uniform criteria and matched according to ancestry and geographic region. All the replication controls were individuals without leprosy, autoimmune or systemic disorders, or family history of leprosy (including first-, second- and third-degree relatives). All individuals with leprosy and control subjects were recruited using uniform criteria and matched according to ancestry and geographic region.

The clinical diagnoses of all individuals with leprosy were based on medical records stored in local leprosy-control institutions and clinical assessments at the time that blood was taken (looking for evidence of leprosy, such as claw hand, lagothalomas or foot drop). Written informed consent was obtained from all individuals with leprosy and control individuals. This study was approved by the institutional review boards of the Shandong Provincial Institute of Dermatology and Venerology and the Shandong Academy of Medical Science.

Genotype imputation and quality control in expanded GWAS analyses. We imputed 6,298 samples (706 individuals with leprosy and 5,592 control subjects) by using genotyped SNPs whose genotypes all passed the quality control criteria of call rate >90%, MAF >1%, deviation from Hardy-Weinberg equilibrium of $P < 1.0 \times 10^{-5}$ in control subjects. Imputation was performed by using IMPUTE version 1 and HapMap reference data (HapMap phase II, CHB+JPT data). Individual genotypes with probability <90% were excluded from further analysis, and SNPs with imputation certainty of <80%, MAF <1% and missing rate >10% for genotypes were excluded from further analysis. In total, 1,209,790 imputed SNPs passed quality control and remained in association analysis.

SNP selection for replication. A locus was chosen for replication when it met the following criteria: at least one SNP with a P value $< 2.0 \times 10^{-5}$ and more than one SNP showing evidence of association (P value $< 1.0 \times 10^{-4}$). For validation, we selected the most significantly associated SNP within each of the 21 newly suggested loci, with the exception that two SNPs were selected within the locus on 1p31.3. In addition, the previously reported SNP rs5743618 and our top GWAS SNP rs17616475 within the TLR1 gene were also selected for replication analysis.

Genotyping analysis of the replication study. Genotyping analyses of the three replication samples were conducted using approximately 15 ng of genomic DNA on the Sequenom MassARRAY system. Sample DNA was amplified by multiplex PCR and the PCR products were then used for locus-specific single-base extension reactions. The resulting products were desalted and transferred to the 384-element SpectroCHIP array. Allele detection was performed using MALDI-TOF mass spectrometry. The mass spectrometers were analyzed using Sequenom MassARRAY TYPER software. In each replication sample, we excluded SNPs with a call rate <95%, low minor allele frequency (<0.01) or deviation from Hardy-Weinberg equilibrium proportions ($P < 0.01$) in the control subjects.

Statistical analysis. Quantile-quantile plots and the calculations of genomic control values were done using the statistical analysis program R, which evaluated the overall significance of genome-wide association results and the potential impact of population stratification. All samples were assessed for population outliers and stratification by using a principal component analysis (PCA)-based approach. First, 6,298 samples (706 individuals with leprosy and 5,592 control individuals) were analyzed together with the 206 reference samples from the International HapMap Project, which includes samples from 57 individuals of Yoruba ancestry from Ibadan, Nigeria (YRI), 44 individuals of Japanese descent from Tokyo, Japan (JPT), 45 individuals of Han Chinese descent from Beijing, China (CHB) and 60 individuals of Northern and Western European descent from Utah, USA (CEU). Eleven population outliers were detected. After removing these outliers, we carried out a second PCA using the remaining case and control samples. In total, 1,701,673 SNPs (both genotyped and imputed) in 706 individuals with leprosy and 5,581 control individuals were used in analyzing genotype-phenotype association.

Genome-wide association testing was performed in PLINK by using logistic regression, with the first five principal components from the second PCA included as covariates to adjust for population stratification. Cochran-Armitage trend tests were used to test genotype-phenotype association in each replication sample of the validation study.

Cochran-Mantel-Haenszel tests were used to test genotype-phenotype association in the combined replication samples by treating three individual replication samples as independent studies. The final joint analysis of the combined GWAS and replication samples was performed by using the Cochran-Mantel-Haenszel test in the GWAS sample of 706 individuals with leprosy and 514 genetically matched control subjects and three independent validation samples. Breslow-day tests and Q tests were performed to evaluate the significance of heterogeneity among individual studies. In the current study, a P value $< 0.05$ was considered to indicate significant heterogeneity. If the P value was $> 0.05$, the fixed-effect model (Mantel-Haenszel) was used to combine the results of different cohorts; otherwise, the random-effect model (DerSimonian-Laird) was used.

The recombination plot of each susceptibility locus was generated in R using information from the HapMap project (Han Chinese individuals from Beijing, China (CHB)) and Japanese individuals from Tokyo, Japan (JPT) samples) (http://hapmap.ncbi.nlm.nih.gov/).

Interaction analysis. The pairwise interaction among nine susceptibility loci was investigated in the subset cohort of 2,202 individuals with leprosy and 3,323 control individuals. This cohort included the 1,220 individuals from the GWAS and a subset of the replication subjects that were analyzed in both our previous (reference) and current study, where the top SNP for each locus (rs3762318 (IL23R), rs602875 (HLA-DR) and HLA-DQ), rs2275606 (RAB32), rs40457 (RIPK2), rs6478108 (TNSF15), rs538147 (RPS6KA4), rs1873613 (LRK2), rs3764147 (C13orf31) and rs9302752 (NOD2)) were genotyped in all the 2,202 individuals with leprosy and 3,323 control individuals. Thirty-six pairwise interaction tests among the nine SNPs were performed using logistic regression and likelihood ratio tests. P values for the interactions were calculated by likelihood ratio tests to compare the two models with and without the interaction term, where SNP1, SNP2 and study (northern Chinese Han, southern Chinese Han and southern Chinese minority populations) were included in the models as covariates. We further performed stratified association analysis of rs40457 by the genotype at rs9302752. All samples were stratified into three strata according to the genotypes at rs9302752, and then the association between the genotype at rs40457 and leprosy was tested within each stratum by using logistic regression analysis. In one of the logistic regression analyses, the genotype at rs40457 was included as a factor, with the most common genotype as a reference category. In the second analysis, we applied logistic regression assuming a log-additive model where rs40457 was coded as 0, 1 or 2 for the number of minor alleles (1-degree-of-freedom score test).

eQTL analysis of rs3762318. eQTL analysis was first performed by searching the SNP and CNV annotation (SCAN) database (http://www.scandb.org/newinterface/about.html), but no eQTL effect was found for rs3762318

Interaction analysis. The pairwise interaction among nine susceptibility loci was investigated in the subset cohort of 2,202 individuals with leprosy and 3,323 control individuals. This cohort included the 1,220 individuals from the GWAS and a subset of the replication subjects that were analyzed in both our previous (reference) and current study, where the top SNP for each locus (rs3762318 (IL23R), rs602875 (HLA-DR) and HLA-DQ), rs2275606 (RAB32), rs40457 (RIPK2), rs6478108 (TNSF15), rs538147 (RPS6KA4), rs1873613 (LRK2), rs3764147 (C13orf31) and rs9302752 (NOD2)) were genotyped in all the 2,202 individuals with leprosy and 3,323 control individuals. Thirty-six pairwise interaction tests among the nine SNPs were performed using logistic regression and likelihood ratio tests. P values for the interactions were calculated by likelihood ratio tests to compare the two models with and without the interaction term, where SNP1, SNP2 and study (northern Chinese Han, southern Chinese Han and southern Chinese minority populations) were included in the models as covariates. We further performed stratified association analysis of rs40457 by the genotype at rs9302752. All samples were stratified into three strata according to the genotypes at rs9302752, and then the association between the genotype at rs40457 and leprosy was tested within each stratum by using logistic regression analysis. In one of the logistic regression analyses, the genotype at rs40457 was included as a factor, with the most common genotype as a reference category. In the second analysis, we applied logistic regression assuming a log-additive model where rs40457 was coded as 0, 1 or 2 for the number of minor alleles (1-degree-of-freedom score test).

eQTL analysis of rs3762318. eQTL analysis was first performed by searching the SNP and CNV annotation (SCAN) database (http://www.scandb.org/newinterface/about.html), but no eQTL effect was found for rs3762318
(data not shown). eQTL analysis was then performed by using the genotype and gene expression data from the 195 HapMap II samples, including the 55 CEU, 42 CHB, 42 JPT and 56 YRI individuals. The genotype data for these individuals were downloaded from the HapMap project, and expression data for the genes surrounding rs3762318 (C1orf141, IL23R, LOC100130497 and IL12RB2) were downloaded from the Genevar database\(^3\). Correlation between the genotype at rs3762318 and the expression levels at the surrounding genes was tested using ANOVA analysis. Searching for regulatory elements associated with rs3762318 was performed on the UCSC Genome Browser\(^3\).