Zhang, FR; Huang, W; Chen, SM; Sun, LD; Liu, H; Li, Y; Cui, Y; Yan, XX; Yang, HT; Yang, RD; Chu, TS; Zhang, C; Zhang, L; Han, JW; Yu, GQ; Quan, C; Yu, YX; Zhang, Z; Shi, BQ; Zhang, LH; Cheng, H; Wang, CY; Lin, Y; Zheng, HF; Fu, XA; Zuo, XB; Wang, Q; Long, H; Sun, YP; Cheng, YL; Tian, HQ; Zhou, FS; Liu, HX; Lu, WS; He, SM; Du, WL; Shen, M; Jin, QY; Wang, Y; Low, HQ; Erwin, T; Yang, NH; Li, JY; Zhao, X; Jiao, YL; Mao, LG; Yin, G; Jiang, ZX; Wang, XD; Yu, JP; Hu, ZH; Gong, CH; Liu, YQ; Liu, RY; Wang, DM; Wei, D; Liu, JX; Cao, WK; Cao, HZ; Li, YP; Yan, WG; Wei, SY; Wang, KJ; Hibberd, ML; Yang, S; Zhang, XJ; Liu, JJ (2009) Genomewide association study of leprosy. The New England journal of medicine, 361 (27). pp. 2609-18. ISSN 0028-4793 DOI: 10.1056/NEJMoa0903753

Downloaded from: http://researchonline.lshtm.ac.uk/2875457/

DOI: 10.1056/NEJMoa0903753

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by-nc-nd/2.5/
Genomewide Association Study of Leprosy

Fu-Ren Zhang, M.D., Ph.D., Wei Huang, Ph.D., Shu-Min Chen, M.D., Ph.D., Liang-Dan Sun, M.D., Ph.D., Hong Liu, M.D., Yi Li, Ph.D., Yong Cui, M.D., Ph.D., Xiao-Xiao Yan, M.D., Hai-Tao Yang, M.D., Rong-De Yang, M.D., Tong-Sheng Chu, M.D., Chi Zhang, M.D., Lin Zhang, M.D., Jian-Wen Han, M.D., Gong-Qi Yu, B.S., Cheng Quan, M.D., Yong-Xiang Yu, B.S., Zheng Zhang, M.D., Ben-Qing Shi, M.D., Jian-Hua Zhang, M.D., Hui Cheng, M.D., Chang-Yuan Wang, M.D., Yan Lin, M.D., Hou-Feng Zheng, M.D., Xi-An Fu, M.D., Xian-Bo Zuo, M.S., Qiang Wang, M.D., Heng Long, M.D., Yi-Ping Sun, M.D., Yi-Lin Cheng, M.S., Hong-Qing Tian, M.D., Fu-Sheng Zhou, B.S., Hua-Xu Liu, M.D., Ph.D., Wen-Sheng Lu, M.D., Su-Min He, M.D., Wen-Li Du, B.S., Min Shen, B.S., Qi-Yi Jin, B.S., Ying Wang, Ph.D., Hui-Qi Low, B.S., Tantoso Erwin, B.S., Ning-Han Yang, B.S., Jin-Yong Li, M.D., Xin Zhao, M.D., Yue-Lin Jiao, M.D., Li-Guo Mao, M.D., Gang Yin, M.D., Zhen-Xia Jiang, M.D., Xiao-Dong Wang, M.D., Jing-Ping Yu, M.D., Zong-Hou Hu, M.D., Cui-Hua Gong, M.D., Yu-Qiang Liu, M.D., Rui-Yu Liu, M.D., De-Min Wang, M.D., Dong Wei, M.D., Jin-Xian Liu, M.D., Wei-Kun Cao, M.D., Hong-Zhong Cao, M.D., Yong-Ping Li, M.D., Wei-Guo Yan, M.D., Shi-Yu Wei, M.D., Kui-Jun Wang, M.D., Martin L. Hibberd, Ph.D., Sen Yang, M.D., Ph.D., Xue-Jun Zhang, M.D., Ph.D., and Jian-Jun Liu, Ph.D.

ABSTRACT

BACKGROUND
The narrow host range of Mycobacterium leprae and the fact that it is refractory to growth in culture has limited research on and the biologic understanding of leprosy. Host genetic factors are thought to influence susceptibility to infection as well as disease progression.

METHODS
We performed a two-stage genomewide association study by genotyping 706 patients and 1225 controls using the Human610-Quad BeadChip (Illumina). We then tested three independent replication sets for an association between the presence of leprosy and 93 single-nucleotide polymorphisms (SNPs) that were most strongly associated with the disease in the genomewide association study. Together, these replication sets comprised 3254 patients and 5955 controls. We also carried out tests of heterogeneity of the associations (or lack thereof) between these 93 SNPs and disease, stratified according to clinical subtype (multibacillary vs. paucibacillary).

RESULTS
We observed a significant association (P<1.00×10−10) between SNPs in the genes CCDC122, C13orf31, NOD2, TNFSF15, HLA-DR, and RIPK2 and a trend toward an association (P=5.10×10−5) with a SNP in LRRK2. The associations between the SNPs in C13orf31, LRRK2, NOD2, and RIPK2 and multibacillary leprosy were stronger than the associations between these SNPs and paucibacillary leprosy.

CONCLUSIONS
Variants of genes in the NOD2-mediated signaling pathway (which regulates the innate immune response) are associated with susceptibility to infection with M. leprae.
leprosy is a chronic infectious disease caused by Mycobacterium leprae. It affects the skin and peripheral nerves and can cause irreversible impairment of nerve function and consequent chronic disabilities. Despite a dramatic decrease in its prevalence over the past two decades (largely due to the worldwide introduction of multidrug therapy in 1982), leprosy remains a major public health problem and one of the most important preventable disabilities in many developing countries. It is therefore particularly unfortunate that research into the mechanisms underlying infection and clinical sequelae has been limited by the fact that M. leprae infects only humans and cannot be cultured in vitro.

The clinical disease of leprosy develops in a minority of infected persons, and it manifests as a spectrum of disease symptoms that result from interactions between the host’s immune response and the bacterium. Tuberculoid and lepromatous leprosy are at opposite ends of the spectrum, each being associated with a relatively stable immune status of the host. “Borderline” categories of the disease, characterized by a variety of clinical manifestations, are associated with an unstable immune response to the bacilli.

The unusually low diversity of genomic sequences among M. leprae strains makes it unlikely that differences in susceptibility or clinical manifestation are governed by the strain of M. leprae or variation within each strain. Therefore, the immunologic response of the host is thought to play a critical role; multibacillary infection is associated with a type 2 helper T (Th2) cell response, whereas paucibacillary infection is associated with an immune response mediated by type 1 helper T (Th1) cells.

Host genetic factors have been implicated in susceptibility to leprosy in studies of familial clustering, studies of twins, complex segregation analyses, and tests of association with the HLA genes. Markers in several genes and genomic regions (e.g., HLA-DR [the gene encoding major histocompatibility complex class II DR], PARK2–PACRG [genes encoding proteins related to Parkinson’s disease], LTA [the gene encoding lymphotxin alpha], and chromosome 10p13) have been reported to be associated with susceptibility to leprosy or the development of a particular clinical form of the disease, but few of these associations have been replicated. We performed a genomewide association study involving large numbers of patients with leprosy and unaffected persons (controls).

**Methods**

We carried out a genomewide association study of leprosy in a “discovery” set of 706 affected patients and 1225 unaffected controls, all of whom were Han Chinese from eastern China. The first replication set consisted of Han Chinese from eastern China, and the second and third replication sets were made up of Han Chinese as well as persons from minority, non-Han ethnic groups (including the Chung, Miao, Yízú, and other smaller groups) from southern China.

Leprosy was diagnosed on the basis of consensus by at least two dermatologists. From medical records, we determined the clinical subtype of the disease, whether there was a family history of leprosy, and the age at onset of disease. The controls did not have a history of leprosy, autoimmune, or systemic disorders or a family history of leprosy (among first-, second-, or third-degree relatives). Patients and controls self-reported their age, sex, and ethnic group on a questionnaire. All participants reported that they were free of infection by M. tuberculosis and chronic infection by other agents (with the exception of M. leprae in the case patients). Patients and controls were matched according to ethnic origin and geographic region of recruitment. All participants provided written informed consent, and the study was approved by local institutional ethics committees (see the Supplementary Appendix, available with the full text of this article at NEJM.org).

We carried out the genomewide association study using Human610-Quad BeadChip (Illumina) and the follow-up genotyping using the iPLEX system (Sequenom) and the TaqMan assay (Applied Biosystems). We tested for population stratification in the discovery set using a method based on principal-components analysis and tested for the presence of genotype–phenotype associations using the Cochran–Armitage trend test with and without correction for population stratification. We also carried out heterogeneity analyses of the 93 single-nucleotide polymorphisms (SNPs) with the strongest associations with disease susceptibility in the genomewide association study to determine whether these associations were disro-
Genomewide Association Study of Leprosy

RESULTS

Genomewide Association Analysis
After filtering the data obtained by genomewide association study, for purposes of quality control, a total of 491,883 SNPs from 706 case patients and 1225 controls remained and were subjected to statistical analysis (see the Supplementary Appendix). Principal-components analysis, using the 206 HapMap reference samples, confirmed that all participants were of Chinese ancestry (Fig. 1 in the Supplementary Appendix), although the case patients and controls showed some genetic stratification (Fig. 2 in the Supplementary Appendix). To minimize the effect of population stratification, we tested for the presence of genotype-phenotype associations using two approaches. First, we analyzed the genomewide genotypes of the 706 case patients and 1225 controls using the Cochran–Armitage trend test with correction for population stratification based on principal-components analysis.15 Second, we tested the genotypes for an association with affected status, without correction for population stratification, after removing the 711 genetically unmatched controls (Fig. 3 in the Supplementary Appendix). (The summary statistics of the full data set obtained by means of the genomewide association analysis can be obtained from the National Center for Biotechnology Information’s database of genotypes and phenotypes [dbGaP; www.ncbi.nlm.nih.gov/gap], accession number phs000217.v1.p1.)

The results of these two analyses indicated that there was no overall inflation of the associations with leprosy because of population stratification (Fig. 4 in the Supplementary Appendix). Moreover, the results of the two analyses were generally consistent, suggesting a strong association within the major histocompatibility complex (MHC) region (on chromosome 6p21) and additional associations at chromosome 16q12 (rs9302752; \( P = 1.42 \times 10^{-17} \); odds ratio for leprosy, 2.12) and chromosome 13q14 (rs3764147; \( P = 4.06 \times 10^{-7} \); odds ratio, 1.97) (Fig. 1 and Table 1). The \( P \) values yielded by both analyses showed a deviation from the null distribution of no association after the SNPs within the MHC region were removed from the analyses, suggesting that the observed \( P \) values within the tail of the distribution are smaller than those expected on the basis of chance and therefore probably reflect true genetic associations (Fig. 4 in the Supplementary Appendix).

We observed two associations with leprosy within the MHC region. One was within the HLA-B–HLA-C locus (encoding MHC, classes I, B and C), at which the most strongly associated SNP was rs9264868 (\( P = 1.96 \times 10^{-18} \); odds ratio, 2.12), and the other was within the HLA-DR–DQ locus (encoding MHC, class II, DR and DQ), at which the most strongly associated SNP was rs9271366 (\( P = 1.94 \times 10^{-17} \); odds ratio, 2.35) (Fig. 5 in the Supplementary Appendix). After controlling for the genetic effect of rs9271366, the association within the HLA-B–HLA-C locus remained significant (Table 2 in the Supplementary Appendix), suggesting that these two associations are independent of each other.

Tests of Replication
We genotyped 93 SNPs — those that showed the strongest association with leprosy in the genomewide association study — in samples from three replication sets: two consisting of Han Chinese and one of Chinese minority groups — collectively, 3254 case patients and 5955 controls (Table 1). In addition to these tests of association carried out using each of the three replication sets, we carried out a combined analysis of the results obtained by means of targeted genotyping of the samples in the replication sets and the genomewide genotyping of the samples in the discovery set. (With respect to the discovery set, we used results from the second analysis, in which we used the smaller group of matched control samples.)

With respect to evaluating the MHC region in the replication sets, we genotyped two SNPs: rs602875 at the HLA-DR–DQ locus (\( P = 3.47 \times 10^{-4} \); odds ratio, 0.58) (since rs9271366, also at this locus and with a stronger association, was refractory to genotyping) and rs9264868 at the HLA-B–C locus (\( P = 1.96 \times 10^{-7} \); odds ratio, 2.12) (Fig. 5 in the Supplementary Appendix). The results of the combined analysis strongly support an association between rs602875 and susceptibility to leprosy.
Figure 1. Results of Genomewide Association Analysis.

The $-\log_{10}$ genomewide $P$ values, as calculated with the use of the Cochran–Armitage trend test, are shown. Panel A shows the $P$ values, after EIGENSTRAT correction, calculated on the basis of data from 491,883 polymorphic single-nucleotide polymorphisms (SNPs) identified in the 706 patients with leprosy and 1225 controls (total, 1931 samples). Panel B shows the $P$ values calculated on the basis of data from 492,109 polymorphic SNPs in the 706 patients and the 514 genetically matched controls (total, 1220 samples). The red horizontal lines indicate $P = 1.00 \times 10^{-5}$, the threshold used for a trend toward genomewide significance.
The associations with susceptibility to disease were replicated for two SNPs (rs42490 and rs40457) within \textit{RIPK2} (the gene encoding receptor-interacting serine–threonine kinase 2, on chromosome 8q21), three SNPs (rs4574921, rs10114470, and rs6478108) within \textit{TNFSF15} (the gene encoding tumor necrosis factor [ligand] superfamily member 15, on chromosome 9q32), two SNPs (rs3764147 and rs10507522) within \textit{C13orf31} (the gene encoding chromosome 13 open reading frame 31, on chromosome 13q14), two SNPs (rs9533634 and rs3088362) within \textit{CCDC122} (the gene encoding coiled-coil domain containing 122, on chromosome 13q14), and two SNPs (rs9302752 and rs7194886) within \textit{NOD2} (the gene encoding nucleotide-binding oligomerization domain containing 2, on chromosome 16q12) (Table 2). At least two SNPs in each of these five genes showed significant association (P<1.00×10^{-10} for all analyses combined) with affected status. To investigate the independence of the multiple associations observed within each of the five genes, we performed conditional association analyses, in which the genetic effect of the most strongly associated SNP at each locus was controlled. These analyses revealed at least two independently associated SNPs, located in different blocks of linkage disequilibria.

\begin{table}
\centering
\begin{tabular}{|l|c|c|c|c|c|}
\hline
\textbf{Characteristic} & \textbf{Genomewide Association Study} & \textbf{Replication Study 1} & \textbf{Replication Study 2} & \textbf{Replication Study 3} & \textbf{Total} \\
\hline
\textbf{Case patients} & & & & & \\
\hline
No. & 706 & 2164 & 304 & 786 & 3960 \\
Mean age (yr) & 65.5 & 66.5 & 58.7 & 54.9 & 62.8 \\
Mean age at onset of leprosy (yr) & 21.8 & 21.4 & 26.5 & 26.2 & 23.3 \\
Sex (no.) & & & & & \\
Male & 562 & 1742 & 221 & 535 & 3060 \\
Female & 144 & 422 & 83 & 251 & 900 \\
Clinical subtype of leprosy (no.)† & & & & & \\
 Multibacillary & 305 & 918 & 166 & 379 & 1768 \\
 Paucibacillary & 397 & 1081 & 124 & 357 & 1959 \\
Disabled from leprosy (no.)‡ & & & & & \\
 Yes & 570 & 947 & 149 & 333 & 2019 \\
 No & 90 & 654 & 150 & 340 & 1234 \\
Family history of leprosy (no.)§ & & & & & \\
 Familial & 185 & 165 & 62 & 80 & 492 \\
 Sporadic & 521 & 1799 & 193 & 574 & 3087 \\
\hline
\textbf{Controls} & & & & & \\
No. & 1225 & 4373 & 709 & 873 & 7180 \\
Mean age (yr) & 34.9 & 63.0 & 40.4 & 43.9 & 48.0 \\
Sex (no.) & & & & & \\
Male & 670 & 2913 & 300 & 590 & 4473 \\
Female & 555 & 1460 & 409 & 283 & 2707 \\
\hline
\end{tabular}
\caption{Baseline Characteristics of the Case Patients and Controls, According to Cohort.}
\end{table}

\(^{*}\) Samples for the genomewide association study and replication study 1 were collected from Han Chinese in Shandong, Anhui, and Jiangsu provinces, eastern China; samples for replication study 2, from Han Chinese in Yunnan province, southern China; and samples for the replication study 3, from the Chung, Miao, Yizú, and other smaller minority groups in Yunnan province, southern China.

\(^{†}\) Data on clinical subtype were missing for 233 patients.

\(^{‡}\) Data on disability status were missing for 707 patients.

\(^{§}\) Data on family history of leprosy were missing for 381 patients.

(P=5.33×10^{-27}; odds ratio, 0.67) but not between rs9264868 and the disease (P=2.33×10^{-3}; odds ratio, 1.14).

The associations with susceptibility to disease were replicated for two SNPs (rs42490 and rs40457) within \textit{RIPK2} (the gene encoding receptor-interacting serine–threonine kinase 2, on chromosome 8q21), three SNPs (rs4574921, rs10114470, and rs6478108) within \textit{TNFSF15} (the gene encoding tumor necrosis factor [ligand] superfamily member 15, on chromosome 9q32), two SNPs (rs3764147 and rs10507522) within \textit{C13orf31} (the gene encoding chromosome 13 open reading frame 31, on chromosome 13q14), two SNPs (rs9533634 and rs3088362) within \textit{CCDC122} (the gene encoding coiled-coil domain containing 122, on chromosome 13q14), and two SNPs (rs9302752 and rs7194886) within \textit{NOD2} (the gene encoding nucleotide-binding oligomerization domain containing 2, on chromosome 16q12) (Table 2). At least two SNPs in each of these five genes showed significant association (P<1.00×10^{-10} for all analyses combined) with affected status. To investigate the independence of the multiple associations observed within each of the five genes, we performed conditional association analyses, in which the genetic effect of the most strongly associated SNP at each locus was controlled. These analyses revealed at least two independently associated SNPs, located in different blocks of linkage disequilibria.

\(^{a}\) Samples for the genomewide association study and replication study 1 were collected from Han Chinese in Shandong, Anhui, and Jiangsu provinces, eastern China; samples for replication study 2, from Han Chinese in Yunnan province, southern China; and samples for the replication study 3, from the Chung, Miao, Yizú, and other smaller minority groups in Yunnan province, southern China.
The odds ratios for leprosy and P values were calculated with the use of the Cochran–Armitage trend test. The P values were calculated after adjustment for age and sex.

---

Table 2. Associations with Leprosy for 16 Single-Nucleotide Polymorphisms (SNPs) within the Seven Susceptibility Genes, According to Analysis.*

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>Major Allele/Minor Allele</th>
<th>Gene</th>
<th>Minor-Allele Frequency†</th>
<th>Genomewide Association Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs602875</td>
<td>6</td>
<td>32681607</td>
<td>A/G</td>
<td>HLA-DR–DQ</td>
<td>0.32</td>
<td>P Value</td>
</tr>
<tr>
<td>rs42490</td>
<td>8</td>
<td>90847650</td>
<td>G/A</td>
<td>RIPK2</td>
<td>0.42</td>
<td>3.47×10⁻⁴ 0.58 (0.43–0.79)</td>
</tr>
<tr>
<td>rs40457</td>
<td>8</td>
<td>90892832</td>
<td>A/G</td>
<td>LRRK2</td>
<td>0.28</td>
<td>1.43×10⁻² 0.73 (0.53–0.99)</td>
</tr>
<tr>
<td>rs10982385</td>
<td>9</td>
<td>116532838</td>
<td>A/C</td>
<td>TNFSF15</td>
<td>0.44</td>
<td>6.09×10⁻² 1.28 (0.98–1.68)</td>
</tr>
<tr>
<td>rs4574921</td>
<td>9</td>
<td>116578155</td>
<td>A/G</td>
<td>TNFSF15</td>
<td>0.32</td>
<td>2.38×10⁻¹ 1.46 (1.10–1.94)</td>
</tr>
<tr>
<td>rs10114470</td>
<td>9</td>
<td>116587593</td>
<td>A/G</td>
<td>TNFSF15</td>
<td>0.47</td>
<td>1.47×10⁻⁴ 1.60 (1.22–2.10)</td>
</tr>
<tr>
<td>rs6478108</td>
<td>9</td>
<td>116598524</td>
<td>G/A</td>
<td>TNFSF15</td>
<td>0.46</td>
<td>4.55×10⁻⁴ 1.54 (1.18–2.01)</td>
</tr>
<tr>
<td>rs1873613</td>
<td>12</td>
<td>38838684</td>
<td>A/G</td>
<td>LRRK2</td>
<td>0.25</td>
<td>9.37×10⁻³ 0.67 (0.49–0.91)</td>
</tr>
<tr>
<td>rs9533634</td>
<td>13</td>
<td>43295815</td>
<td>A/G</td>
<td>CCDC122</td>
<td>0.24</td>
<td>1.43×10⁻³ 0.85 (0.62–1.17)</td>
</tr>
<tr>
<td>rs3088362</td>
<td>13</td>
<td>43331630</td>
<td>C/A</td>
<td>CCDC122</td>
<td>0.26</td>
<td>2.00×10⁻² 1.87 (1.38–2.53)</td>
</tr>
<tr>
<td>rs3764147</td>
<td>13</td>
<td>43355925</td>
<td>A/G</td>
<td>C13orf31</td>
<td>0.31</td>
<td>4.06×10⁻⁷ 1.97 (1.49–2.62)</td>
</tr>
<tr>
<td>rs10507522</td>
<td>13</td>
<td>43377000</td>
<td>A/G</td>
<td>C13orf31</td>
<td>0.31</td>
<td>4.17×10⁻⁴ 0.55 (0.40–0.75)</td>
</tr>
<tr>
<td>rs9302752</td>
<td>16</td>
<td>49276604</td>
<td>A/G</td>
<td>NOD2</td>
<td>0.29</td>
<td>1.42×10⁻⁹ 2.28 (1.70–3.06)</td>
</tr>
<tr>
<td>rs7194886</td>
<td>16</td>
<td>49282694</td>
<td>G/A</td>
<td>NOD2</td>
<td>0.14</td>
<td>4.43×10⁻⁷ 2.25 (1.58–3.21)</td>
</tr>
<tr>
<td>rs8057341</td>
<td>16</td>
<td>49295481</td>
<td>A/G</td>
<td>NOD2</td>
<td>0.22</td>
<td>5.22×10⁻² 1.33 (0.96–1.84)</td>
</tr>
<tr>
<td>rs3135499</td>
<td>16</td>
<td>49323628</td>
<td>A/C</td>
<td>NOD2</td>
<td>0.21</td>
<td>9.21×10⁻² 1.26 (0.91–1.74)</td>
</tr>
</tbody>
</table>

* The odds ratios for leprosy and P values were calculated with the use of the Cochran–Armitage trend test. The P values were calculated after adjustment for age and sex. C13orf31 denotes the gene encoding chromosome 13 open reading frame 31, CCDC122 the gene encoding coiled-coil domain containing 122, CI confidence interval, HLA-DR the gene encoding major histocompatibility complex class II DR, LRRK2 the gene encoding leucine-rich repeat kinase 2, NOD2 the gene encoding nucleotide-binding oligomerization domain containing 2, RIPK2 the gene encoding receptor-interacting serine–threonine kinase 2, and TNFSF15 the gene encoding tumor necrosis factor (ligand) superfamily member 15.

† The minor-allele frequency is based on the controls.

---

The results indicate a trend toward an association between the SNP rs1873613 in LRRK2 (the gene encoding leucine-rich repeat kinase 2, on chromosome 12q12) and susceptibility to leprosy (Table 2). Inclusion of the replication samples strengthened the evidence for the association for this SNP (P=5.10×10⁻⁶; odds ratio, 0.82). The subgroup analysis of the multibacillary and paucibacillary clinical subtypes of leprosy revealed significant evidence for heterogeneity at five SNPs (rs3764147, rs10507522, rs9302752, rs42490, and rs1491938) within four genes (C13orf31, LRRK2, NOD2, and RIPK2). The associations of these SNPs were stronger with the multibacillary form of leprosy than with the paucibacillary form, and the difference in the strength of association was significant (defined as P<0.05 after correction for multiple testing for the 16 SNPs listed in Table 2) (Table 3). The rs1491938 variant (in LRRK2) showed a significant association with the multibacillary form (P=2.26×10⁻⁶; odds ratio, 0.81) but not the paucibacillary form (P=2.96×10⁻⁶; odds ratio, 0.96).

ANALYSIS OF SUBGROUPS OF PATIENTS

The subgroup analysis of the multibacillary and paucibacillary clinical subtypes of leprosy revealed significant evidence for heterogeneity at five SNPs (rs3764147, rs10507522, rs9302752, rs42490, and rs1491938) within four genes (C13orf31, LRRK2, NOD2, and RIPK2). The associations of these SNPs were stronger with the multibacillary form of leprosy than with the paucibacillary form, and the difference in the strength of association was significant (defined as P<0.05 after correction for multiple testing for the 16 SNPs listed in Table 2) (Table 3). The rs1491938 variant (in LRRK2) showed a significant association with the multibacillary form (P=2.26×10⁻⁶; odds ratio, 0.81) but not the paucibacillary form (P=2.96×10⁻⁶; odds ratio, 0.96).
We did not detect effects of heterogeneity in other subgroups of patients (data not shown).

**PATHWAY ANALYSIS**

We explored possible functional relationships between the seven identified susceptibility genes using the Ingenuity Pathways Analysis knowledge database (Ingenuity Systems). A single network of 35 genes, including 5 of the 7 susceptibility genes, was identified through unsupervised network analysis, resulting in a highly significant score (one-sided $P=1.00 \times 10^{-15}$ by Fisher’s exact test). Figure 2 illustrates the functional relationship between the five susceptibility genes (together with five other genes), creating a plausible biologic network underlying susceptibility to leprosy.

<table>
<thead>
<tr>
<th>Replication Study 1</th>
<th>Replication Study 2</th>
<th>Replication Study 3</th>
<th>All Studies Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$ Value</td>
<td>Odds Ratio (95% CI)</td>
<td>$P$ Value</td>
<td>Odds Ratio (95% CI)</td>
</tr>
<tr>
<td>$8.81 \times 10^{-22}$</td>
<td>$0.64$ (0.59–0.71)</td>
<td>$3.29 \times 10^{-1}$</td>
<td>$0.85$ (0.65–1.12)</td>
</tr>
<tr>
<td>$2.45 \times 10^{-13}$</td>
<td>$0.74$ (0.68–0.80)</td>
<td>$2.40 \times 10^{-1}$</td>
<td>$0.88$ (0.69–1.11)</td>
</tr>
<tr>
<td>$2.45 \times 10^{-10}$</td>
<td>$0.74$ (0.68–0.81)</td>
<td>$6.79 \times 10^{-1}$</td>
<td>$0.97$ (0.74–1.26)</td>
</tr>
<tr>
<td>$4.94 \times 10^{-6}$</td>
<td>$1.18$ (1.09–1.27)</td>
<td>$1.54 \times 10^{-1}$</td>
<td>$1.16$ (0.92–1.47)</td>
</tr>
<tr>
<td>$9.23 \times 10^{-9}$</td>
<td>$1.27$ (1.17–1.37)</td>
<td>$2.29 \times 10^{-2}$</td>
<td>$1.29$ (1.02–1.63)</td>
</tr>
<tr>
<td>$6.63 \times 10^{-8}$</td>
<td>$1.25$ (1.15–1.35)</td>
<td>$4.95 \times 10^{-1}$</td>
<td>$1.08$ (0.85–1.37)</td>
</tr>
<tr>
<td>$1.80 \times 10^{-11}$</td>
<td>$1.32$ (1.21–1.43)</td>
<td>$3.95 \times 10^{-2}$</td>
<td>$1.30$ (1.02–1.66)</td>
</tr>
<tr>
<td>$3.62 \times 10^{-3}$</td>
<td>$0.87$ (0.79–0.95)</td>
<td>$1.39 \times 10^{-3}$</td>
<td>$0.65$ (0.50–0.85)</td>
</tr>
<tr>
<td>$9.48 \times 10^{-8}$</td>
<td>$0.77$ (0.70–0.85)</td>
<td>$2.45 \times 10^{-1}$</td>
<td>$0.85$ (0.64–1.15)</td>
</tr>
<tr>
<td>$6.64 \times 10^{-29}$</td>
<td>$1.53$ (1.40–1.67)</td>
<td>$4.69 \times 10^{-4}$</td>
<td>$1.60$ (1.22–2.09)</td>
</tr>
<tr>
<td>$1.46 \times 10^{-37}$</td>
<td>$1.70$ (1.57–1.85)</td>
<td>$7.29 \times 10^{-6}$</td>
<td>$1.74$ (1.36–2.22)</td>
</tr>
<tr>
<td>$3.32 \times 10^{-18}$</td>
<td>$0.66$ (0.60–0.72)</td>
<td>$3.07 \times 10^{-2}$</td>
<td>$0.75$ (0.57–0.98)</td>
</tr>
<tr>
<td>$3.83 \times 10^{-28}$</td>
<td>$1.59$ (1.47–1.73)</td>
<td>$7.11 \times 10^{-2}$</td>
<td>$1.26$ (0.97–1.63)</td>
</tr>
<tr>
<td>$5.26 \times 10^{-18}$</td>
<td>$1.56$ (1.41–1.73)</td>
<td>$1.31 \times 10^{-2}$</td>
<td>$1.51$ (1.09–2.10)</td>
</tr>
<tr>
<td>$2.23 \times 10^{-5}$</td>
<td>$1.22$ (1.11–1.33)</td>
<td>$4.51 \times 10^{-1}$</td>
<td>$1.10$ (0.82–1.47)</td>
</tr>
<tr>
<td>$8.16 \times 10^{-5}$</td>
<td>$1.20$ (1.10–1.32)</td>
<td>$5.59 \times 10^{-1}$</td>
<td>$1.07$ (0.79–1.44)</td>
</tr>
</tbody>
</table>

**DISCUSSION**

Through a genomewide association study of susceptibility to leprosy, we have implicated genetic variants in six genes that show a significant association with disease and a seventh gene that shows a trend toward an association with disease. The controls, particularly the 1225 with samples analyzed in the genomewide association study, were younger than the case patients. It is possible that the controls were too young for the disease to have developed clinically but the effect of their comparative age on the strength of the association is probably minimal, given the low incidence of leprosy in the general population.

The genomewide association study showed that, of the four SNPs at the NOD2 locus associated with leprosy, two SNPs (rs9302752 and rs7194886) lie between NOD2 and its 5′ neighboring gene, SNX20 (which encodes the sorting nexin 20 protein); these two intergenic SNPs are more strongly associated with leprosy than the two linked SNPs located within NOD2. We believe that the associations of leprosy with rs9302752 and rs7194886 probably reflect the effects of regulatory variants on NOD2 expression.

All the implicated gene variants, with the possible exception of those of LRRK2, seem to confer susceptibility to both multibacillary and paucibacillary forms of leprosy, indicating shared mechanisms underlying the development of these two clinical forms of the disease. However, several gene variants seem to be more strongly associated
with the multibacillary form of the disease than the paucibacillary form, and our data indicate an association between variants at the LRRK2 locus and the multibacillary form only. Our results are consistent with those of the two-step model for the development of leprosy, in which successful infection of *M. leprae* is first established in genetically predisposed persons, and the subsequent clinical manifestation of disease is influenced by other host factors and environmental factors. Genomewide association studies that directly test for a genetic association with the multibacillary or the paucibacillary form may uncover additional host genetic factors involved in the second step of disease development.

Variants of HLA genes, HLA-DRB1 in particular, have been associated with leprosy; both protective and risk alleles have been described. We too observed an association with leprosy within the MHC region (SNP rs602875, next to HLA-DRB1), although we did not observe an association of the disease with other previously reported “risk” loci: PARK2–PACRG, LTA, and a locus on chromosome 10p13 (Tables 3 and 4 in the Supplementary Appendix). The association of disease with the HLA-DR–DQ locus observed in this study is consistent with the previously identified association between leprosy and HLA-DRB1 and the fact that there is extensive linkage disequilibrium within the MHC region.

HLA-DR molecules present *M. leprae* peptide antigens to CD4+ T cells, which allows the T cells to be activated. In leprosy, this process is thought to lead to the generation of Th1 cells, which produce interferon-γ, resulting in macrophage maturation and the production of antimycobacterial molecules. Failure of this process is thought to be critical for susceptibility to leprosy and infection by other mycobacteria. Although HLA-DR is a well-established initiator of this process, the theoretical biologic network (as generated with the use of an unsupervised Ingenuity Pathways Analysis) (Fig. 2) suggests that interferon-γ may also be regulated by genes implicated in our analysis — NOD2, RIPK2, and TNFSF15 — and is consistent with the finding that persons with mutant interferon-γ are susceptible to mycobacterial infection.

TNFSF15 is a tumor necrosis factor (TNF)–like molecule expressed in macrophages and T cells; it binds a TNF-family receptor (expressed primarily on T cells) that mediates the switch from Th1 cells to Th2 cells. NOD2 is an intracellular sensing molecule that recognizes the bacterial-cell-wall peptidoglycan and the muramyl dipeptides motif. It is expressed by macrophages and epithelial cells. Ligand bound to NOD2 initiates signaling, which is mediated by RIPK2 through a ubiquination process that involves the recruitment of TAK1 (transforming growth factor β–activated kinase 1) and NEMO (nuclear factor-κB essential modulator) to the NOD2–RIPK2 complex; IKK proteins (encoded by NFKBIA and NFKBIB [nuclear factor of kappa light polypeptide gene enhancer in B cells inhibitor alpha and beta, respectively]) becomes degraded, leading to the

### Table 3. Heterogeneity Analysis of the Five Single-Nucleotide Polymorphisms (SNPs) Found to Be Differentially Associated with Clinical Subtypes.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome</th>
<th>Position</th>
<th>Gene</th>
<th>Multibacillary Leprosy (N = 1768)</th>
<th>Paucibacillary Leprosy (N = 1959)</th>
<th>P Value for Heterogeneity†</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3764147</td>
<td>13</td>
<td>4335925</td>
<td>C13orf31</td>
<td>P = 7.47×10^-6</td>
<td>OR = 1.93 (1.79–2.09)</td>
<td>P = 5.11×10^-41</td>
</tr>
<tr>
<td>rs10507522</td>
<td>13</td>
<td>4337000</td>
<td>C13orf31</td>
<td>P = 2.99×10^-28</td>
<td>OR = 0.60 (0.54–0.65)</td>
<td>P = 9.40×10^-10</td>
</tr>
<tr>
<td>rs9302752</td>
<td>16</td>
<td>4927604</td>
<td>NOD2</td>
<td>P = 8.87×10^-41</td>
<td>OR = 1.73 (1.60–1.88)</td>
<td>P = 1.44×10^-19</td>
</tr>
<tr>
<td>rs42490</td>
<td>8</td>
<td>9084750</td>
<td>RIPK2</td>
<td>P = 1.35×10^-19</td>
<td>OR = 0.69 (0.64–0.75)</td>
<td>P = 9.54×10^-8</td>
</tr>
<tr>
<td>rs1491938</td>
<td>12</td>
<td>3893197</td>
<td>LRRK2</td>
<td>P = 2.26×10^-6</td>
<td>OR = 0.81 (0.75–0.89)</td>
<td>P = 2.96×10^-1</td>
</tr>
</tbody>
</table>

† C13orf31 denotes the gene encoding chromosome 13 open reading frame 31, LRRK2 the gene encoding leucine-rich repeat kinase 2, NOD2 the gene encoding nucleotide-binding oligomerization domain containing 2, RIPK2 the gene encoding receptor-interacting serine–threonine kinase 2, and SNP single-nucleotide polymorphism.

‡ P values were calculated on the basis of the difference in the odds ratios for the development of multibacillary leprosy and for the development of paucibacillary leprosy.
movement of NF-κB to the nucleus and the subsequent activation of NF-κB target genes,\textsuperscript{28} such as TNFSF15. Consistent with our data are the phenotypes of mice deficient in Nod2 and Ripk2. These mice are highly susceptible to infection with \textit{M. tuberculosis}\textsuperscript{29} and \textit{Chlamydia pneumoniae},\textsuperscript{30} respectively, owing to a failure to produce inflammatory cytokines known to initiate the Th1-cell responses.\textsuperscript{31}

PARK2 is implicated through our network analysis, and variants of PARK2 are reported to be associated with susceptibility to leprosy, although we did not observe any such association in our analyses.\textsuperscript{15} PARK2 is an E3 ligase, thought to regulate innate immunity.\textsuperscript{32} We therefore hypothesize that PARK2 participates in ubiquination-mediated NOD2 signaling. Variants of both PARK2 and LRRK2 are associated with susceptibility to Parkinson's disease and interact directly. LRRK2 is thought to regulate the ligase activity of PARK2.\textsuperscript{33}

Taken together, it seems that five of the genes directly implicated in our study feature in the NOD2-mediated regulatory node of innate immunity. The functions of the other two implicated genes, CDC122 and CI3orf31, are as yet unknown.

Variants of \textit{NOD2} and \textit{TNFSF15} are associated with Crohn's disease and are linked to altered production of interleukin-10 and altered Th1-Th2 switching.\textsuperscript{34-36} It is therefore all the more notable that leprosy and Crohn's disease have common immunologic features, including a Th1-cell response with granuloma formation. Moreover, mycobacterial infection has been described as a risk factor for Crohn's disease.\textsuperscript{37,38}

In summary, our genomewide association study highlights variants of genes encoding proteins involved in the innate immune response as risk factors for developing leprosy.

Supported by grants from the Research Foundation of Shandong Provincial Institute of Dermatology and Venereology (2008-7), the Research Foundation of Shandong Academy of Medical Science (2005), the Shandong Provincial Leprosy Control Special Financial Support (2007), the Shandong Provincial Research Fund of Science and Technology (2006GG2302029), the Outstanding Scholarship 1020 Project of Shandong Provincial Health System (2008), the Shandong Provincial Taishan Scholar Construction Project (2009), the National Natural Science Foundation (30771943/C030116), the Anhui Provincial Special Scientific Program (2007-7), and the Agency for Science, Technology, and Research of Singapore.

No potential conflict of interest relevant to this article was reported.

We thank all the participants involved in this research.

\textbf{REFERENCES}

6. Ridley DS, Jopling WH. Classification of leprosy according to immunity: a five-

Copyright © 2009 Massachusetts Medical Society.