ORIGINAL ARTICLE

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., Lian-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 ($P<1.00\times10^{-10}$) between SNPs in the genes *CCDC122*, *C13orf31*, *NOD2*, *TNFSF15*, *HLA-DR*, and *RIPK2* and a trend toward an association ($P=5.10\times10^{-5}$) 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*.

The authors' affiliations are listed in the Supplementary Appendix, available with the full text of this article at NEJM.org. Address reprint requests to Dr. Fu-Ren Zhang at Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Science, 57, Jiyan Lu, Jinan, Shandong 250022, China, or at zhangfuren@hotmail.com; to Dr. Xue-Jun Zhang at the Institute of Dermatology and Department of Dermatology at No. 1 Hospital, Anhui Medical University, Hefei, Anhui 230022, China, or at ayzxj@vip.sina.com; or to Dr. Jian-Jun Liu at the Department of Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore, or at liuj3@gis.a-star.edu.sg.

This article (10.1056/NEJMoa0903753) was published on December 16, 2009, at NEJM.org.

N Engl J Med 2009;361:2609-18.

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.[1] Despite a dramatic decrease in its prevalence over the past two decades (largely due to the worldwide introduction of multidrug therapy in 1982),[2] leprosy remains a major public health problem and one of the most important preventable disabilities in many developing countries.[3] 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.[4]

The clinical disease of leprosy develops in a minority of infected persons,[5] 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.[6]

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.[7] 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.[8]

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.[9-13] 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 lymphotoxin 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.[14-17] 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 dispro-

portionately driven by the presence or absence of family history of leprosy, presence or absence of disability from leprosy, the age at onset of the disease, or its clinical subtype. More information on the samples, genotyping, quality control, and statistical analyses is provided in the Supplementary Appendix.

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.[18] 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; $P=1.42\times10^{-9}$; odds ratio for leprosy, 2.28) and chromosome 13q14 (rs3764147; $P=4.06\times10^{-7}$; 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 ($P=1.96\times10^{-4}$; 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 ($P=1.94\times10^{-17}$; 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 ($P=3.47\times10^{-4}$; 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 ($P=1.96\times10^{-4}$; 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

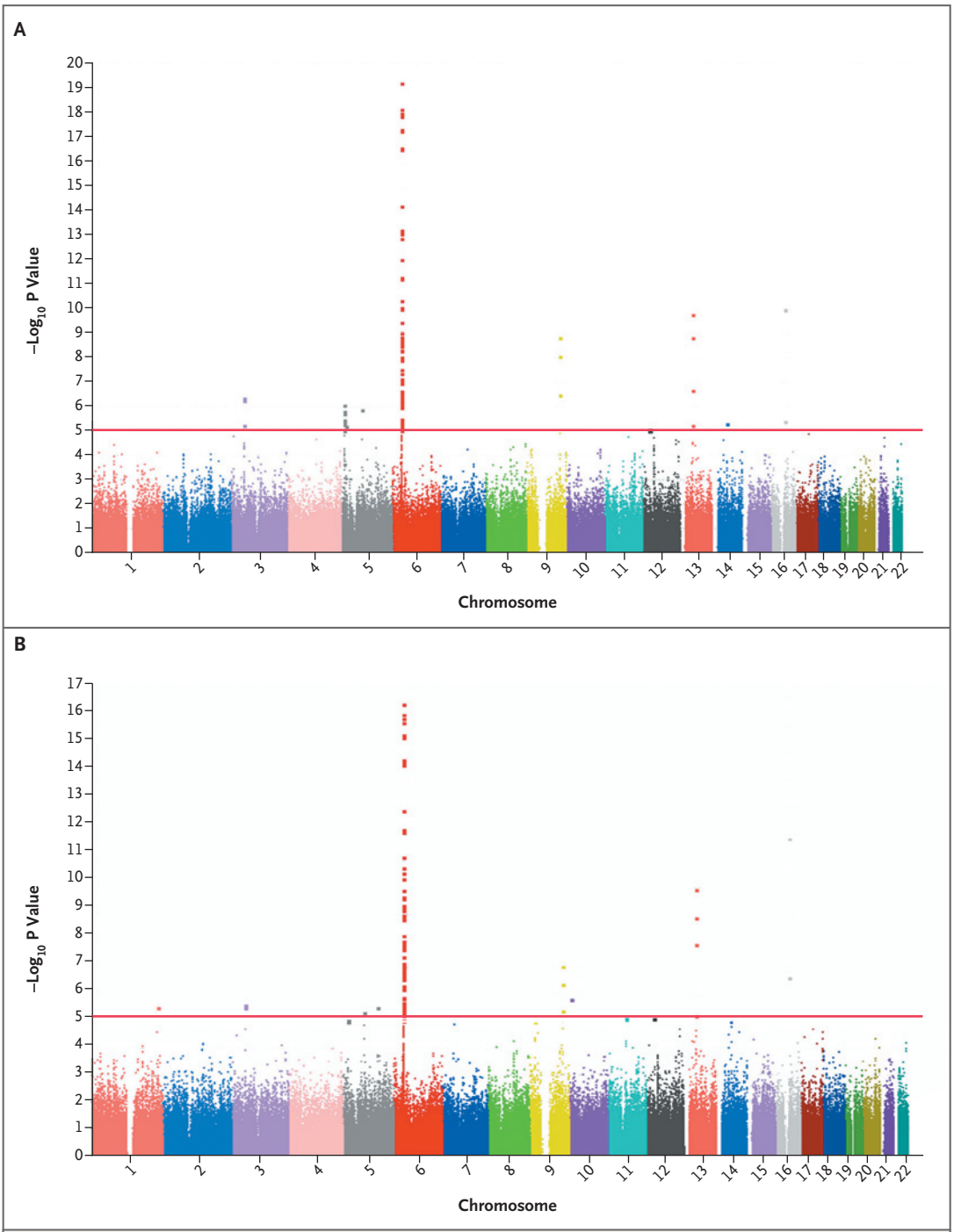

Figure 1. Results of Genomewide Association Analysis.

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 $P=1.00\times10^{-5}$, the threshold used for a trend toward genomewide significance.

Table 1. Baseline Characteristics of the Case Patients and Controls, According to Cohort.*

Characteristic	Genomewide Association Study	Replication Study 1	Replication Study 2	Replication Study 3	Total
Case patients					
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	353	2019
No	90	654	150	340	1234
Family history of leprosy (no.)§					
Familial	185	165	62	80	492
Sporadic	521	1799	193	574	3087
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

* 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.

($P=5.33\times10^{-27}$; odds ratio, 0.67) but not between rs9264868 and the disease ($P=2.33\times10^{-3}$; odds ratio, 1.14).

The associations with susceptibility to disease were replicated for two SNPs (rs42490 and rs40457) within *RIPK2* (the gene encoding receptor-interacting serine–threonine kinase 2, on chromosome 8q21), three SNPs (rs4574921, rs10114470, and rs6478108) within *TNFSF15* (the gene encoding tumor necrosis factor [ligand] superfamily member 15, on chromosome 9q32), two SNPs (rs3764147 and rs10507522) within *C13orf31* (the gene encoding chromosome 13 open reading frame 31, on chromosome 13q14), two SNPs (rs9533634 and rs3088362) within *CCDC122* (the gene encoding coiled-coil domain containing 122, on chromosome 13q14), and two SNPs (rs9302752 and rs7194886) within *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 ($P<1.00\times10^{-10}$ 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 disequilib-

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

SNP	Chromosome	Position	Major Allele/ Minor Allele	Gene	Minor-Allele Frequency†	Genomewide Association Study	
						P Value	Odds Ratio (95% CI)
rs602875	6	32681607	A/G	*HLA-DR–DQ*	0.32	3.47×10^{-4}	0.58 (0.43–0.79)
rs42490	8	90847650	G/A	*RIPK2*	0.42	1.23×10^{-3}	0.66 (0.51–0.87)
rs40457	8	90892832	A/G	*RIPK2*	0.28	1.43×10^{-2}	0.73 (0.53–0.99)
rs10982385	9	116532838	A/C	*TNFSF15*	0.44	6.09×10^{-2}	1.28 (0.98–1.68)
rs4574921	9	116578155	A/G	*TNFSF15*	0.32	2.38×10^{-3}	1.46 (1.10–1.94)
rs10114470	9	116587593	A/G	*TNFSF15*	0.47	1.47×10^{-4}	1.60 (1.22–2.10)
rs6478108	9	116598524	G/A	*TNFSF15*	0.46	4.55×10^{-4}	1.54 (1.18–2.01)
rs1873613	12	38838684	A/G	*LRRK2*	0.25	9.37×10^{-3}	0.67 (0.49–0.91)
rs9533634	13	43295815	A/G	*CCDC122*	0.24	1.43×10^{-1}	0.85 (0.62–1.17)
rs3088362	13	43331630	C/A	*CCDC122*	0.26	2.00×10^{-6}	1.87 (1.38–2.53)
rs3764147	13	43355925	A/G	*C13orf31*	0.31	4.06×10^{-7}	1.97 (1.49–2.62)
rs10507522	13	43377000	A/G	*C13orf31*	0.31	4.17×10^{-5}	0.55 (0.40–0.75)
rs9302752	16	49276604	A/G	*NOD2*	0.29	1.42×10^{-9}	2.28 (1.70–3.06)
rs7194886	16	49282694	G/A	*NOD2*	0.14	4.43×10^{-7}	2.25 (1.58–3.21)
rs8057341	16	49295481	A/G	*NOD2*	0.22	5.22×10^{-2}	1.33 (0.96–1.84)
rs3135499	16	49323628	A/C	*NOD2*	0.21	9.21×10^{-2}	1.26 (0.91–1.74)

* 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.

rium (Table 2 in the Supplementary Appendix) and with low pairwise r^2 values (<0.3) at each locus (Fig. 6 in the Supplementary Appendix).

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 an association for this SNP ($P=5.10\times10^{-5}$ for all analyses combined; odds ratio, 0.86). Joint analysis of the 1931 samples (including the 711 unmatched controls) in the genomewide association study and those in all three replication sets also supported an association ($P=3.68\times10^{-5}$; odds ratio, 0.86), with an even stronger association from joint analysis of all the Han samples (from 3174 case patients and 6307 controls) ($P=5.68\times10^{-6}$; odds ratio, 0.82).

The results for the other 77 SNPs included in the replication analyses are summarized in Table 1 in the Supplementary Appendix.

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 ($P=2.26\times10^{-6}$; odds ratio, 0.81) but not the paucibacillary form ($P=2.96\times10^{-1}$; odds ratio, 0.96).

Replication Study 1		Replication Study 2		Replication Study 3		All Studies Combined	
P Value	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)
8.81×10^{-22}	0.64 (0.59–0.71)	3.29×10^{-1}	0.85 (0.65–1.12)	6.32×10^{-3}	0.77 (0.64–0.93)	5.35×10^{-27}	0.67 (0.62–0.72)
2.45×10^{-13}	0.74 (0.68–0.80)	2.40×10^{-1}	0.88 (0.69–1.11)	1.21×10^{-2}	0.83 (0.71–0.96)	1.38×10^{-16}	0.76 (0.71–0.81)
2.45×10^{-10}	0.74 (0.68–0.81)	6.79×10^{-1}	0.97 (0.74–1.26)	2.03×10^{-2}	0.84 (0.72–0.98)	1.34×10^{-12}	0.77 (0.72–0.83)
4.94×10^{-5}	1.18 (1.09–1.27)	1.54×10^{-1}	1.16 (0.92–1.47)	5.66×10^{-2}	1.16 (1.00–1.34)	8.73×10^{-8}	1.19 (1.12–1.27)
9.23×10^{-9}	1.27 (1.17–1.37)	2.29×10^{-2}	1.29 (1.02–1.63)	1.79×10^{-5}	1.39 (1.19–1.61)	2.17×10^{-16}	1.31 (1.23–1.40)
6.63×10^{-8}	1.25 (1.15–1.35)	4.95×10^{-1}	1.08 (0.85–1.37)	9.87×10^{-5}	1.35 (1.16–1.57)	5.42×10^{-14}	1.28 (1.20–1.37)
1.80×10^{-11}	1.32 (1.21–1.43)	3.95×10^{-2}	1.30 (1.02–1.66)	8.20×10^{-7}	1.47 (1.26–1.71)	3.39×10^{-21}	1.37 (1.28–1.46)
3.62×10^{-3}	0.87 (0.79–0.95)	1.39×10^{-3}	0.65 (0.50–0.85)	8.92×10^{-1}	1.01 (0.86–1.18)	5.10×10^{-5}	0.86 (0.80–0.92)
9.48×10^{-8}	0.77 (0.70–0.85)	2.45×10^{-1}	0.85 (0.64–1.15)	7.31×10^{-5}	0.69 (0.57–0.83)	4.77×10^{-12}	0.76 (0.70–0.82)
6.64×10^{-23}	1.53 (1.40–1.67)	4.69×10^{-4}	1.60 (1.22–2.09)	1.11×10^{-2}	1.26 (1.05–1.51)	1.36×10^{-31}	1.52 (1.41–1.63)
1.46×10^{-37}	1.70 (1.57–1.85)	7.29×10^{-6}	1.74 (1.36–2.22)	4.54×10^{-8}	1.55 (1.32–1.82)	3.72×10^{-54}	1.68 (1.57–1.80)
3.32×10^{-18}	0.66 (0.60–0.72)	3.07×10^{-2}	0.75 (0.57–0.98)	1.05×10^{-2}	0.81 (0.69–0.95)	4.64×10^{-24}	0.68 (0.63–0.74)
3.83×10^{-28}	1.59 (1.47–1.73)	7.11×10^{-2}	1.26 (0.97–1.63)	9.21×10^{-5}	1.44 (1.20–1.72)	3.77×10^{-40}	1.59 (1.49–1.71)
5.26×10^{-18}	1.56 (1.41–1.73)	1.31×10^{-2}	1.51 (1.09–2.10)	1.86×10^{-7}	1.77 (1.42–2.19)	1.77×10^{-30}	1.63 (1.50–1.77)
2.23×10^{-5}	1.22 (1.11–1.33)	4.51×10^{-1}	1.10 (0.82–1.47)	1.49×10^{-1}	0.86 (0.70–1.06)	5.53×10^{-5}	1.17 (1.09–1.26)
8.16×10^{-5}	1.20 (1.10–1.32)	5.59×10^{-1}	1.07 (0.79–1.44)	1.13×10^{-1}	0.84 (0.68–1.04)	2.52×10^{-4}	1.16 (1.07–1.25)

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 $P=1.00\times10^{-15}$ 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.

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 associations 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

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

SNP	Chromosome	Position	Gene	Multibacillary Leprosy (N=1768)		Paucibacillary Leprosy (N=1959)		P Value for Heterogeneity†
				P Value	Odds Ratio (95% CI)	P Value	Odds Ratio (95% CI)	
rs3764147	13	43355925	*C13orf31*	7.47×10^{-62}	1.93 (1.79–2.09)	5.11×10^{-21}	1.44 (1.33–1.55)	5.44×10^{-10}
rs10507522	13	43377000	*C13orf31*	2.99×10^{-28}	0.60 (0.54–0.65)	9.40×10^{-10}	0.77 (0.71–0.84)	2.57×10^{-6}
rs9302752	16	49276604	*NOD2*	8.87×10^{-41}	1.73 (1.60–1.88)	1.44×10^{-19}	1.44 (1.33–1.55)	3.98×10^{-5}
rs42490	8	90847650	*RIPK2*	1.35×10^{-19}	0.69 (0.64–0.75)	9.54×10^{-8}	0.82 (0.76–0.88)	8.23×10^{-4}
rs1491938	12	38931897	*LRRK2*	2.26×10^{-6}	0.81 (0.75–0.89)	2.96×10^{-1}	0.96 (0.88–1.04)	8.55×10^{-4}

* *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.

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.[19] 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[20]; both protective and risk alleles have been described.[21] 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).[14-16] 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.[22] 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.[19]

TNFSF15 is a tumor necrosis factor (TNF)–like molecule expressed in macrophages and T cells[23]; it binds a TNF-family receptor (expressed primarily on T cells) that mediates the switch from Th1 cells to Th2 cells.[24,25] NOD2 is an intracellular sensing molecule that recognizes the bacterial-cell-wall peptidoglycan and the muramyl dipeptides motif.[26] 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 [NF-κB] essential modulator) to the NOD2–RIPK2 complex[27]; IKb 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

movement of NF-κB to the nucleus and the subsequent activation of NF-κB target genes,[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 *M. tuberculosis*[29] and *Chlamydophila pneumoniae*,[30] respectively, owing to a failure to produce inflammatory cytokines known to initiate the Th1-cell responses.[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.[15] PARK2 is an E3 ligase, thought to regulate innate immunity.[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.[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, *CCDC122* and *C13orf31*, are as yet unknown.

Variants of *NOD2* and *TNFSF15* are associated with Crohn's disease and are linked to altered production of interleukin-10 and altered Th1–Th2 switching.[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.[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.

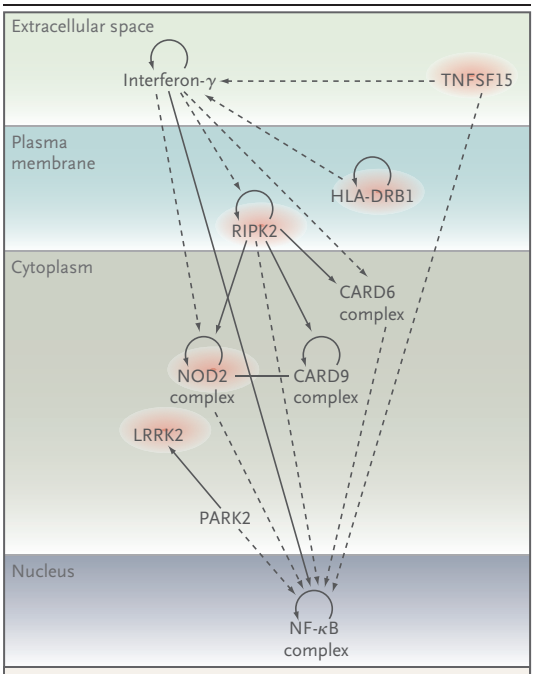

Figure 2. A Gene-Interaction Network of Five Genes Conferring Susceptibility to Leprosy and Five Other Genes.

An Ingenuity Pathways Analysis of the seven susceptibility genes identified a single, closely connected network of interacting genes, including 5 of the 7 genes implicated in the development of leprosy in the genomewide association study (highlighted in red) and 30 additional genes. The network was unlikely to have been identified by chance (one-sided $P=1.00\times10^{-15}$ by Fisher's exact test). Solid lines between genes represent known direct molecular interactions, and dashed lines, known indirect molecular interactions. CARD6 denotes caspase recruitment domain family member 6, CARD9 caspase recruitment domain family member 9, HLA-DRB1 major histocompatibility complex class II DR beta 1, LRRK2 leucine-rich repeat kinase 2, NF-κB nuclear factor κB, NOD2 nucleotide-binding oligomerization domain containing 2, RIPK2 receptor-interacting serine–threonine kinase 2, and TNFSF15 tumor necrosis factor (ligand) superfamily member 15.

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.

REFERENCES

1. Hastings RC, Opromolla DVA. Leprosy. 2nd ed. Edinburgh: Churchill Livingstone, 1994:291.
2. Chemotherapy of leprosy for control programmes. World Health Organ Tech Rep Ser 1982;675:1-33.
3. Britton WJ, Lockwood DN. Leprosy. Lancet 2004;363:1209-19.
4. Fine PE. Natural history of leprosy — aspects relevant to a leprosy vaccine. Int J Lepr Other Mycobact Dis 1983;51:553-5.
5. Quintana-Murci L, Alcaïs A, Abel L, Casanova JL. Immunology in natura: clinical, epidemiological and evolutionary genetics of infectious diseases. Nat Immunol 2007;8:1165-71.
6. Ridley DS, Jopling WH. Classification of leprosy according to immunity: a five-

group system. Int J Lepr Other Mycobact Dis 1966;34:255-73.
7. Monot M, Honoré N, Garnier T, et al. On the origin of leprosy. Science 2005; 308:1040-2.
8. Modlin RL. Th1-Th2 paradigm: insights from leprosy. J Invest Dermatol 1994;102:828-32.
9. Shields ED, Russell DA, Pericak-Vance MA. Genetic epidemiology of the susceptibility to leprosy. J Clin Invest 1987;79: 1139-43.
10. Chakravartti MR, Vogel F. A twin study on leprosy. Vol. 1 of Topics in human genetics. Stuttgart, Germany: Georg Thieme Verlag, 1973.
11. Abel L, Demenais F. Detection of major genes for susceptibility to leprosy and its subtypes in a Caribbean island: Desirade island. Am J Hum Genet 1988; 42:256-66.
12. Abel L, Vu DL, Oberti J, et al. Complex segregation analysis of leprosy in southern Vietnam. Genet Epidemiol 1995;12:63-82.
13. Todd JR, West BC, McDonald JC. Human leukocyte antigen and leprosy: study in northern Louisiana and review. Rev Infect Dis 1990;12:63-74.
14. Siddiqui MR, Meisner S, Tosh K, et al. A major susceptibility locus for leprosy in India maps to chromosome 10p13. Nat Genet 2001;27:439-41.
15. Mira MT, Alcaïs A, Nguyen VT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 2004;427: 636-40.
16. Alcaïs A, Alter A, Antoni G, et al. Stepwise replication identifies a low-producing lymphotoxin-alpha allele as a major risk factor for early-onset leprosy. Nat Genet 2007;39:517-22.
17. Moraes MO, Cardoso CC, Vanderborght PR, Pacheco AG. Genetics of host response in leprosy. Lepr Rev 2006;77:189-202.
18. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.
19. Casanova JL, Abel L. Genetic dissection of immunity to mycobacteria: the human model. Annu Rev Immunol 2002; 20:581-620.
20. Ottenhoff TH, de Vries RR. HLA class II immune response and suppression genes in leprosy. Int J Lepr Other Mycobact Dis 1987;55:521-34.
21. Geluk A, Ottenhoff TH. HLA and leprosy in the pre and postgenomic eras. Hum Immunol 2006;67:439-45.
22. Ottenhoff TH, Verreck FA, Hoeve MA, van de Vosse E. Control of human host immunity to mycobacteria. Tuberculosis (Edinb) 2005;85:53-64.
23. Croft M. The role of TNF superfamily members in T-cell function and diseases. Nat Rev Immunol 2009;9:271-85.
24. Meylan F, Davidson TS, Kahle E, et al. The TNF-family receptor DR3 is essential for diverse T cell-mediated inflammatory diseases. Immunity 2008;29:79-89.
25. Murphy KM, Reiner SL. The lineage decisions of helper T cells. Nat Rev Immunol 2002;2:933-44.
26. Franchi L, Warner N, Viani K, Nuñez G. Function of Nod-like receptors in microbial recognition and host defense. Immunol Rev 2009;227:106-28.
27. Hitotsumatsu O, Ahmad RC, Tavares R, et al. The ubiquitin-editing enzyme A20 restricts nucleotide-binding oligomerization domain containing 2-triggered signals. Immunity 2008;28:381-90.
28. Hasegawa M, Fujimoto Y, Lucas PC, et al. A critical role of RICK/RIP2 polyubiquitination in Nod-induced NF-kappaB activation. EMBO J 2008;27:373-83.
29. Divangahi M, Mostowy S, Coulombe F, et al. NOD2-deficient mice have impaired resistance to Mycobacterium tuberculosis infection through defective innate and adaptive immunity. J Immunol 2008;181:7157-65.
30. Shimada K, Chen S, Dempsey PW, et al. The NOD/RIP2 pathway is essential for host defenses against Chlamydophila pneumoniae lung infection. PLoS Pathog 2009;5(4):e1000379.
31. Moreira LO, El Kasmi KC, Smith AM, et al. The TLR2-MyD88-NOD2-RIPK2 signalling axis regulates a balanced pro-inflammatory and IL-10-mediated anti-inflammatory cytokine response to Gram-positive cell walls. Cell Microbiol 2008;10:2067-77.
32. Schurr E, Alcaïs A, de Léséleuc L, Abel L. Genetic predisposition to leprosy: a major gene reveals novel pathways of immunity to Mycobacterium leprae. Semin Immunol 2006;18:404-10.
33. Smith WW, Pei Z, Jiang H, et al. Leucine-rich repeat kinase 2 (LRRK2) interacts with parkin, and mutant LRRK2 induces neuronal degeneration. Proc Natl Acad Sci U S A 2005;102:18676-81.
34. Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. Nat Rev Immunol 2008;8:458-66.
35. Noguchi E, Homma Y, Kang X, Netea MG, Ma XA. Crohn's disease-associated NOD2 mutation suppresses transcription of human IL10 by inhibiting activity of the nuclear ribonucleoprotein hnRNP-A1. Nat Immunol 2009;10:471-9.
36. Thiébaut R, Kotti S, Jung C, et al. TNFSF15 polymorphisms are associated with susceptibility to inflammatory bowel disease in a new European cohort. Am J Gastroenterol 2009;104:384-91.
37. Behr MA, Schurr E. Mycobacteria in Crohn's disease: a persistent hypothesis. Inflamm Bowel Dis 2006;12:1000-4.
38. Pierce ES. Where are all the Mycobacterium avium subspecies paratuberculosis in patients with Crohn's disease? PLoS Pathog 2009;5(3):e1000234.