nature genetics

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

Furen Zhang[1–4], Hong Liu[1,3], Shumin Chen[1,3], Huiqi Low[5], Liangdan Sun[6,7], Yong Cui[6,7], Tongsheng Chu[1,3], Yi Li[5], Xi'an Fu[1,3], Yongxiang Yu[1,3], Gongqi Yu[1,3], Benqing Shi[2,4], Hongqing Tian[2,4], Dianchang Liu[1,3], Xiulu Yu[1,3], Jinghui Li[1,3], Nan Lu[1,3], Fangfang Bao[1,3], Chunying Yuan[1,3], Jian Liu[1,3], Huaxu Liu[1,3], Lin Zhang[1,3], Yonghu Sun[1,3], Mingfei Chen[1,3], Qing Yang[1,3], Haitao Yang[8], Rongde Yang[9], Lianhua Zhang[8], Qiang Wang[10], Hong Liu[1,3], Fuguang Zuo[1,3], Haizhen Zhang[1,3], Chiea Chuen Khor[5,11], Martin L Hibberd[5,12], Sen Yang[6,7], Jianjun Liu[1,5,11,13] & Xuejun Zhang[6,7]

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 were identified with genome-wide significance: rs2275606 (combined $P = 3.94 \times 10^{-14}$, OR = 1.30) on 6q24.3 and rs3762318 (combined $P = 3.27 \times 10^{-11}$, 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 India[3]. 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*, *LACC1* (*C13orf31*), *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. leprae*[5]. 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 population[6]. 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

[1]Shandong Provincial Institute of Dermatology and Venereology, Provincial Academy of Medical Science, Jinan, China. [2]Shandong Provincial Hospital for Skin Diseases, Shandong University, Jinan, China. [3]Shandong Provincial Key Lab for Dermatovenereology, Jinan, China. [4]Shandong Provincial Medical Center for Dermatovenereology, Jinan, China. [5]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore. [6]Institute of Dermatology and Department of Dermatology at No.1 hospital, Anhui Medical University, Hefei, China. [7]State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Hefei, China. [8]Jiangsu Provincial Center for Disease Control and Prevention, Nanjing, China. [9]Wenshan Institute of Dermatology, Wenshan, China. [10]Anhui Provincial Institute of Dermatology, Hefei, China. [11]Center for Molecular Epidemiology, National University of Singapore, Singapore. [12]Department of Epidemiology and Public Health, National University of Singapore, Singapore. [13]School 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

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

SNP	Chr. position	Allele[a]	Gene	MAF[b]	Expanded GWAS		Original GWAS[c]		Replication 1		Replication 2		Replication 3		Combined replication		All Han samples[d]		All samples[e]	
					P	OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR	*P*	OR
rs3762318	Chr. 1: 67369707	G/A	*IL23R*	0.10	2.20×10^{-9}	0.48	2.96×10^{-3}	0.64	2.03×10^{-7}	0.71	0.618	0.89	1.28×10^{-3}	0.55	2.67×10^{-9}	0.70	2.64×10^{-9}	0.71	3.27×10^{-11}	0.69
rs2275606	Chr. 6: 146960643	A/G	*RAB32*	0.21	1.12×10^{-6}	1.43	1.51×10^{-3}	1.37	1.11×10^{-10}	1.32	0.731	1.05	6.62×10^{-3}	1.27	5.51×10^{-12}	1.29	1.65×10^{-12}	1.30	3.94×10^{-14}	1.30
rs16948876	Chr. 16: 49411919	A/G	*CYLD*	0.03	4.07×10^{-9}	2.45	6.88×10^{-5}	2.55	3.29×10^{-8}	1.57	0.218	1.51	0.969	1.01	1.26×10^{-7}	1.47	8.00×10^{-12}	1.66	1.64×10^{-10}	1.56

[a]Minor/major allele. ORs were calculated according to the minor allele. [b]Minor allele frequency in the control individuals of the GWAS. [c]The original GWAS included 706 individuals with leprosy and 514 well-matched control subjects. [d]All Han samples included individuals from the original GWAS and from replications 1 and 2 and excluded the 4,367 control subjects used in the expanded GWAS discovery analysis. [e]All samples included individuals from the original GWAS and the three replication groups and excluded the 4,367 control subjects used in the expanded GWAS discovery analysis.

descent (consisting of 10 control individuals, 1,012 individuals with atopic dermatitis, 1,139 individuals with psoriasis, 1,082 individuals with systemic lupus erythematosus (SLE) and 1,124 individuals with vitiligo) from several previous GWAS of the Chinese population[7–10]. After imputation and quality control filtering (see Online Methods), a total of 1,701,673 single-nucleotide polymorphisms (SNPs) (including both genotyped and imputed ones) in 706 individuals with leprosy and 5,581 control individuals were used for genome-wide discovery analysis. Principal-component analysis confirmed all the subjects to be of Chinese ancestry and further indicated a moderate degree of genetic stratification between the infected individuals and controls ($\lambda_{GC} = 1.44$, **Supplementary Fig. 1**). To minimize the adverse impact of population stratification, association analysis was performed using the top five principal components as covariates via logistic regression[11]. This reduced the extent of genome-wide inflation to 1.04, a level considered acceptable by conventional GWAS standards[12–14]. 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 $P < 2.0 \times 10^{-5}$ (**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 *TLR1* (ref. 6) 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[15]. 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 2 Estimated genotype ORs of rs40457 stratified by the genotype at rs9302752

		Number (frequency)		Stratified association analysis of rs40457							
rs9302752	rs40457	Case	Control	OR	95% CI		$P > \|z\|$[a]	OR	95% CI		$P > \|z\|$[b]
CC	GG	173 (59%)	147 (48%)	1.00	–	–	–				
	GA	116 (40%)	127 (43%)	0.76	0.54	1.07	0.118	0.80	0.61	1.05	0.104
	AA	17 (6%)	19 (6%)	0.71	0.35	1.44	0.343				
	Subtotal	306	293								
CT	GG	495 (52%)	670 (52%)	1.00	–	–	–				
	GA	376 (39%)	515 (40%)	0.91	0.76	1.09	0.305	0.90	0.79	1.03	0.143
	AA	84 (9%)	112 (9%)	0.81	0.59	1.12	0.202				
	Subtotal	955	1,297								
TT	GG	545 (58%)	806 (47%)	1.00	–	–	–				
	GA	321 (34%)	725 (42%)	0.58	0.48	0.69	2.20×10^{-9}	0.60	0.53	0.69	3.03×10^{-14}
	AA	75 (8%)	202 (12%)	0.38	0.28	0.52	6.36×10^{-10}				
	Subtotal	941	1,733								
	Total	2,202	3,323								

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

($P = 1.64 \times 10^{-10}$, 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; $P_{heterogeneity} = 0.16$ 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 ($P = 1.64 \times 10^{-10}$, OR = 1.56) or not ($P = 8.00 \times 10^{-12}$, 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) ($P = 0.0011$ 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 ($P = 3.03 \times 10^{-14}$) 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 ($P = 7.93 \times 10^{-6}$, OR = 0.87) and rs538147 within *RPS6KA4* on 11q11-13 ($P = 7.20 \times 10^{-6}$, 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 ($P < 5.0 \times 10^{-8}$).

We did not observe evidence of the previously reported association of *TLR1* with leprosy[6]. 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

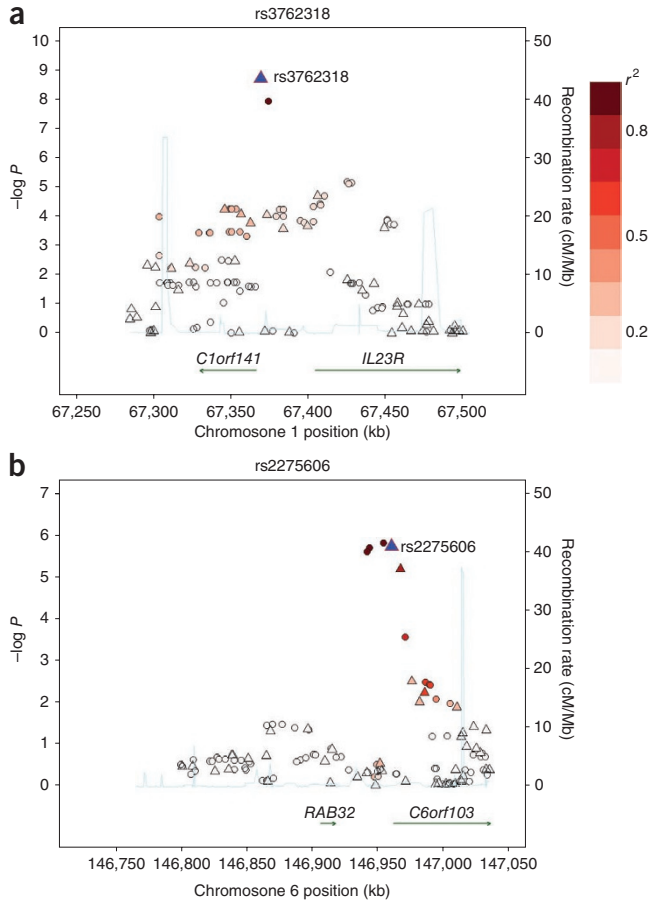

Figure 1 Regional association plots of rs3762318 and rs2275606. The *P* values for the SNPs (shown as $-\log_{10} P$ values on the *y* axis) were plotted against their map positions (*x* axis). The color of each SNP spot reflects its r^2 value, with the top SNP (large blue triangle) indicated for each locus. Estimated recombination rates (based on the combined CHB and JPT samples from the HapMap project) were plotted in light blue. Gene annotations were adapted from the UCSC Genome Browser. (**a**) The rs3762318 SNP on 1p31.3. (**b**) The rs2275606 SNP on 6q24.3.

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 *C1orf141* 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 ($P_{conditional}$ = 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 been reported to have essential roles in immunity to mycobacteria[24].

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 Ras superfamily of low molecular weight G proteins and is required for the formation of autophagic vacuoles and the regulation of the clearance of aggregated proteins by autophagy[24]. A recent study reported that the Rab32 protein participates in controlling the recruitment of cathepsin D to phagosomes containing *Mycobacterium tuberculosis*[25], 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, r^2 = 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: $P = 1.84 \times 10^{-52}$, OR = 1.64; rs16948876: $P = 1.68 \times 10^{-12}$, OR = 1.64), but conditioning on rs9302752 dramatically weakened the association at rs16948876 ($P_{conditional} = 3.64 \times 10^{-4}$, $OR_{conditional}$ = 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 autophagocytosis in host defense against *M. leprae* infection[25], which has parallels with Crohn's disease[26]. 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.

URLs. R, http://www.r-project.org/; International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; SCAN database, http://www.scandb.org/newinterface/about.html; UCSC genome browser, http://genome.ucsc.edu/.

METHODS

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

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

ACKNOWLEDGMENTS

We thank the individuals who participated in this project. This work was funded by grants from the National Natural Science Foundation of China (81071288, 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

F. Zhang, X.Z. and Jianjun Liu conceived of and designed the study. S.C., T.C., X.Y., Lin Zhang, D.L., R.Y., H.Y., Lianhua Zhang and Q.W. undertook recruitment and collected phenotype data. Hong Liu, X.F., G.Y., Y.Y., Q.L., F.B., N.L., C.Y., Y.S., M.C., Hong Liu, H.Z., F. Zuo and Q.Y. conducted sample selection and performed the genotyping of the validation study. Hong Liu, X.F., Jian Liu, B.S., H.T. and Huaxu Liu collected phenotype data, undertook related data handling and calculation, managed recruitment and obtained biological samples. Jianjun Liu, H. Low, Hong Liu and Y.L. undertook data checking, statistical analysis and bioinformatics analyses. S.Y., Hong Liu, L.S. and Y.C. were responsible for sample selection, genotyping and project management. C.C.K. and M.L.H. helped to revise the manuscript. All authors contributed to the final manuscript, with F. Zhang, Jianjun Liu, X.Z. and Hong Liu having key roles.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Davila, S. *et al.* Genome-wide association study identifies variants in the *CFH* region associated with host susceptibility to meningococcal disease. *Nat. Genet.* **42**, 772–776 (2010).
2. Thye, T. *et al.* Genome-wide association analyses identifies a susceptibility locus for tuberculosis chromosome 18q11.2. *Nat. Genet.* **42**, 739–741 (2010).

3. World Health Organization. Global leprosy situation, 2010. *Wkly. Epidemiol. Rec.* **85**, 337–348 (2010).
4. Shields, E.D., Russell, D.A. & Pericak-Vance, M.A. Genetic epidemiology of the susceptibility to leprosy. *J. Clin. Invest.* **79**, 1139–1143 (1987).
5. Zhang, F.R. *et al.* Genomewide association study of leprosy. *N. Engl. J. Med.* **361**, 2609–2618 (2009).
6. Wong, S.H. *et al.* Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.* **6**, e1000979 (2010).
7. Zhang, X.J. *et al.* Psoriasis genome-wide association study identifies susceptibility variants within *LCE* gene cluster at 1q21. *Nat. Genet.* **41**, 205–210 (2009).
8. Quan, C. *et al.* Genome-wide association study for vitiligo identifies susceptibility loci at 6q27 and the MHC. *Nat. Genet.* **42**, 614–618 (2010).
9. Han, J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234–1237 (2009).
10. Sun, L.D. *et al.* Genome-wide association study identifies two new susceptibility loci for atopic dermatitis in the Chinese Han population. *Nat. Genet.* **43**, 690–694 (2011).
11. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
12. Plenge, R.M. *et al.* Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. *Nat. Genet.* **39**, 1477–1482 (2007).
13. Mells, G.F. *et al.* Genome-wide association study identifies 12 new susceptibility loci for primary biliary cirrhosis. *Nat. Genet.* **43**, 329–332 (2011).
14. Franke, A. *et al.* Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat. Genet.* **42**, 1118–1125 (2010).
15. Hennessy, S. *et al.* Factors influencing the optimal control-to-case ratio in matched case-control studies. *Am. J. Epidemiol.* **149**, 195–197 (1999).
16. Sun, L.D. *et al.* Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat. Genet.* **42**, 1005–1009 (2010).
17. Sheng, Y.J. *et al.* Follow-up study identifies two novel susceptibility loci PRKCB and 8p11.21 for systemic lupus erythematosus. *Rheumatology (Oxford)* **50**, 682–688 (2011).
18. Strange, A. *et al.* A genome-wide association study identifies new psoriasis susceptibility loci and an interaction between *HLA-C* and *ERAP1*. *Nat. Genet.* **42**, 985–990 (2010).
19. Mizuki, N. *et al.* Genome-wide association studies identify *IL23R–IL12RB2* and *IL10* as Behçet's disease susceptibility loci. *Nat. Genet.* **42**, 703–706 (2010).
20. McGovern, D.P. *et al.* Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat. Genet.* **42**, 332–337 (2010).
21. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
22. Brand, S. Crohn's disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. *Gut* **58**, 1152–1167 (2009).
23. Parham, C. *et al.* A receptor for the heterodimeric cytokine IL-23 is composed of IL-12Rβ1 and a novel cytokine receptor subunit, IL-23R. *J. Immunol.* **168**, 5699–5708 (2002).
24. Ottenhoff, T.H., Verreck, F.A., Hoeve, M.A. & van de Vosse, E. Control of human host immunity to mycobacteria. *Tuberculosis (Edinb.)* **85**, 53–64 (2005).
25. Seto, S., Tsujimura, K. & Koide, Y. Rab GTPases regulating phagosome maturation are differentially recruited to mycobacterial phagosomes. *Traffic* **12**, 407–420 (2011).
26. Deretic, V. Autophagy in infection. *Curr. Opin. Cell Biol.* **22**, 252–262 (2010).

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[5] 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[7–10]. 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), Miaos (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 Venereology 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 spectrograms 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[27]. 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 of <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[28]; otherwise, the random-effect model (DerSimonian-Laird) was used[29].

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 (*TNFSF15*), rs538147 (*RPS6KA4*), rs1873613 (*LRRK2*), 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

doi:10.1038/ng.973

(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[30]. 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[31].

27. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
28. Mantel, N. & Haenszel, W. Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.* **22**, 719–748 (1959).
29. DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. *Control. Clin. Trials* **7**, 177–188 (1986).
30. Yang, T.P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474–2476 (2010).
31. Fujita, P.A. *et al.* The UCSC Genome Browser database: update 2010. *Nucleic Acids Res.* **38**, D613–D619 (2010).