Human Molecular Genetics, 2013, Vol. 22, No. 21 **4430–4437**
doi:10.1093/hmg/ddt286
Advance Access published on June 19, 2013

An association study of TOLL and CARD with leprosy susceptibility in Chinese population

Hong Liu[1,2,3,4,†], Fangfang Bao[1,3,†], Astrid Irwanto[5,6,†], Xi'an Fu[1,3], Nan Lu[1,3], Gongqi Yu[1,3], Yongxiang Yu[1,3], Yonghu Sun[1,3], Huiqi Low[5], Yi Li[5], Herty Liany[5], Chunying Yuan[1,3], Jinghui Li[1,3], Jian Liu[1,3], Mingfei Chen[1,3], Huaxu Liu[1,3], Na Wang[1,3], Jiabao You[1,3], Shanshan Ma[1,3], Guiye Niu[1,3], Yan Zhou[1,3], Tongsheng Chu[1,3], Hongqing Tian[2,4], Shumin Chen[1,3], Xuejun Zhang[8,10], Jianjun Liu[1,5,6,7] and Furen Zhang[1,2,3,4,9,*]

[1]Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Sciences, Shandong, China, [2]Shandong Provincial Hospital for Skin Diseases, Shandong University, Shandong, China, [3]Shandong Provincial Key Lab for Dermatovenereology, Shandong, China, [4]Shandong Provincial Medical Center for Dermatovenereology, Shandong, China, [5]Human Genetics, Genome Institute of Singapore, A*STAR, Singapore, [6]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore [7]School of Life Sciences, Anhui Medical University, Anhui, China, [8]Institute of Dermatology and Department of Dermatology at No.1 hospital, Anhui Medical University, Anhui, China and [9]Shandong Clinical College, Anhui Medical University, Anhui, China and [10]State Key Laboratory Incubation Base of Dermatology, Ministry of National Science and Technology, Anhui Medical University, Anhui, China

Received August 1, 2012; Revised May 29, 2013; Accepted June 12, 2013

Previous genome-wide association studies (GWASs) identified multiple susceptibility loci that have highlighted the important role of *TLR* (Toll-like receptor) and *CARD* (caspase recruitment domain) genes in leprosy. A large three-stage candidate gene-based association study of 30 *TLR* and 47 *CARD* genes was performed in the leprosy samples of Chinese Han. Of 4363 SNPs investigated, eight SNPs showed suggestive association ($P < 0.01$) in our previously published GWAS datasets (Stage 1). Of the eight SNPs, rs2735591 and rs4889841 showed significant association ($P < 0.001$) in an independent series of 1504 cases and 1502 controls (Stage 2), but only rs2735591 (next to *BCL10*) showed significant association in the second independent series of 938 cases and 5827 controls (Stage 3). Rs2735591 showed consistent association across the three stages ($P > 0.05$ for heterogeneity test), significant association in the combined validation samples ($P_{corrected} = 5.54 \times 10^{-4}$ after correction for 4363 SNPs tested) and genome-wide significance in the whole GWAS and validation samples ($P = 1.03 \times 10^{-9}$, OR = 1.24). In addition, we demonstrated the lower expression of *BCL10* in leprosy lesions than normal skins and a significant gene connection between *BCL10* and the eight previously identified leprosy loci that are associated with NFκB, a major regulator of downstream inflammatory responses, which provides further biological evidence for the association. We have discovered a novel susceptibility locus on 1p22, which implicates *BCL10* as a new susceptibility gene for leprosy. Our finding highlights the important role of both innate and adaptive immune responses in leprosy.

INTRODUCTION

Leprosy is a chronic human infectious disease caused by the intracellular pathogen *Mycobacterium leprae* (*M. leprae*), which remains a serious health problem in the developing countries, with more than 200 000 new cases being detected each year (1). It affects both the skin and peripheral nerves and can cause irreversible impairment of nerve function and chronic disabilities (2). It has been proven that both environmental factors and host genetics play a critical role in the prevalence

*To whom correspondence should be addressed. Tel: +86 53187298801; Fax: +86 53187984734; Email: zhangfuren@hotmail.com
†These authors contributed equally to this work.

of leprosy with estimated hereditable fractions of up to 57% (3,4).

Clinical manifestations of leprosy are classified according to World Health Organization (paucibacillary and multibacillary) schemes, reflective of a Th1 (cell-mediated) or Th2 (humoral) host immune response, respectively. Its divergent clinical forms reflect two distinct immune responses to the same pathogen, which maintains the importance of host genetics in susceptibility to leprosy. The results of genetic studies suggest the involvement of host genetics in both susceptibility to leprosy *per se* and development of clinical subtypes (5).

In 2009, we performed the first GWAS of leprosy and identified six susceptibility loci (*CCDC122, LACC1, NOD2, TNFSF15, HLA-DR* and *RIPK2*) and one suggestive locus (*LRRK2*) in the Chinese Han population, which has indicated the importance of the *NOD2*-mediated innate immunity against the infection by *M leprae* (6). NOD2 is a member of the CARD-containing protein family (CARD15), which plays an important role in the immune response by recognizing the bacterial molecules and activating the NF-κB protein (7–10). We also found association around *RIPK2*, also has a CARD domain in its C-terminal portion which can interact with NOD2 and induce activation of NF-κB (11). Other CARD-carrying proteins, including CARD9, CARD10, CARD11, CARD14, NOD1 and BCL10, have also been shown to facilitate NF-κB activation (12). Our subsequent GWAS has further identified *IL23R* and *RAB32* as susceptibility loci for leprosy (13), which has not only expanded the biological functions of *IL23R* by uncovering its involvement in infectious disease susceptibility, but also suggested a potential involvement of autophagocytosis (*RAB32*) in leprosy development.

Toll-like receptors (*TLR*s) recognize particular molecular patterns of diverse microorganisms to induce innate immune responses. The contribution of the *TLR* variations to the susceptibility for leprosy has been investigated in different populations, and *TLR1*, *TLR2* and *TLR4* variants have been reported to be associated with leprosy by previous studies in Indian, Nepalese and African populations (14–17). With the exception of *TLR1*, the association evidence of other *TLR* variants are, however, not conclusive.

Although we did not find a definitive evidence of TOLL/CARD enrichment through an unbiased gene-enrichment analysis of our previously published GWAS dataset (13) using MAGENTA (18) (Supplementary Material, Table S1), the evidence of association around *NOD2* and *RIPK2* in our previous findings and the roles of CARD-containing molecules, as well as TOLL receptors in the innate immune response against pathogens, led us to hypothesize that other TOLL/CARD genes might also play a role in the susceptibility to leprosy. Therefore, to further investigate the role of *TLRs* and *CARDs* in leprosy, we performed a large three-stage candidate gene-based association study of *TLR*s and *CARD*s in leprosy samples of Chinese Han population.

RESULTS

The current candidate gene-based genetic association study was performed in three stages using the previously published GWAS datasets and two new independent sample series of Chinese Han population. The GWAS datasets includes the matched 1220 samples (706 cases and 514 controls) used in our initial GWAS analysis (6) and additional controls of 5067 individuals used in our subsequent expanded GWAS analysis (13). The first independent validation series consists of 1504 cases and 1502 controls, and the second independent validation series includes 938 cases and 5827 controls. The samples of the GWAS datasets were described in our previous publications (6,13). For the two independent validation series, the cases and the controls were matched in terms of age, gender and geographic residence. The general characteristics of all the patients and control subjects are summarized in Table 1. In total, 3148 leprosy cases and 12 910 controls of Chinese Han were investigated in the current study.

In the current study, we performed a candidate gene-based association study on 30 *TLR* and 47 *CARD*-related genes that were identified through an UCSC Genome Browser. We first examined the association evidence within the critical regions of these candidate genes in the two published GWAS datasets of matched 1220 samples (6) and expanded 6287 samples (13). In total, 4363 SNPs within these 77 candidate genes were investigated in the two GWAS datasets (Supplementary Material, Table S2). After excluding the SNPs located in known susceptibility loci, SNPs with suggestive evidence of association were identified using the following criteria: (i) the *P*-value of association was $<1.0 \times 10^{-2}$ in the expanded GWAS analysis, and (ii) the *P*-value of association showed improvement between the first GWAS analysis of 706 cases and 514 matched controls and the expanded GWAS analysis of 706 cases and 5581 population controls. A moderate threshold of suggestive association was employed here to maximize the power and the chance to discover novel associations within these candidate genes. In total, eight suggestive SNPs within two *TLR*-related (*ZCCHC11* and *FREM1*) and six *CARD*-related (*BCL10*, *KCNQ1*, *MAP2K1*, *SGSH*, *NOD1* and *VISA*) candidate genes were brought forward for further validation.

In the first validation analysis (Stage 2), six of the eight selected SNPs were successfully genotyped in 1504 leprosy cases and 1502 healthy controls, while the other two SNPs failed in either the assay design (rs5743369) or genotyping analysis (rs6084506). The association results of the six SNPs are summarized in the Supplementary Material, Table S3. Of the six SNPs, two showed significant association ($P < 0.005$ after correction for six SNPs tested): rs2735591 ($P = 2.18 \times 10^{-6}$, OR = 1.32) and rs4889841 ($P = 1.94 \times 10^{-4}$, OR = 1.23). In the second validation (Stage 3), the two SNPs were further genotyped in additional 938 cases and 5827 controls. SNP rs2735591 showed significant association in the second validation sample series ($P = 4.58 \times 10^{-3}$, OR = 1.17), but rs4889841 failed to show consistent association ($P = 0.415$, OR = 1.04) (Supplementary Material, Table S4). Rs2735591 showed consistent association among the two independent validation and previously published GWAS (1220 samples) without evidence for genetic heterogeneity ($P > 0.05$) (Table 2). The association at rs2735591 is highly significant in the combined validation samples ($P = 1.27 \times 10^{-7}$, OR = 1.27), even after correction for all the 4363 SNPs tested in the current study ($P_{corrected} = 5.54 \times 10^{-4}$) and surpassed the genome-wide significance threshold ($P = 1.03 \times 10^{-9}$, OR = 1.24) in the combined GWAS and two independent validation

Table 1. Sample summary of 2442 cases and 8082 controls

	Cases				Controls		
	Sample size	Mean age	Mean age at onset	Males[a] (%)	Sample size	Mean age	Males[a] (%)
Validation set 1	1504	66.22	20.96	83%	1502	65.25	84%
Validation set 2	938	65.05	24.14	81%	5827	49.82	65%
Total	2442	65.78	22.2	82%	8082	52.42	68%

[a]Gender Information was missing for 63 patients and 136 control subjects in all the samples.

series, consisting of a total of 3148 cases and 7843 controls (Table 2).

We further investigated the association evidence within the surrounding region of rs2735591 on 1p22 by imputing our expanded GWAS dataset (13) using the genetic variation data from 1000 Genomes Project (version February 2012) as a reference panel to maximize the genetic variation coverage of the region in association analysis. In total, we tested the association at 1500 imputed SNPs and 296 typed SNPs within this region. Rs2735591 is located within a linkage-disequilibrium (LD) block of 300 Kb on 1p22, where *BCL10* as well as several other genes or transcripts reside (Fig. 1). Within the LD block, there are several other SNPs showing consistent association evidence and in strong LD with rs2735591, but none of them are coding variants. The top SNP within this region was rs233100 ($P = 3.73 \times 10^{-4}$ and OR = 1.26) that is in LD with rs2735591 ($r^2 = 0.33$, $D' = 0.90$) (Fig. 1). Conditioning on rs233100 could eliminate the association at rs2735591 ($P = 0.389$ and OR = 1.07), while conditioning on rs2735591 reduces the association at rs233100 ($P = 0.015$ and OR = 1.21) (Supplementary Material, Fig. S1), suggesting that the associations at the two SNPs are not fully independent, and there seems to be one association within the region.

To investigate the potential association of rs2735591 with any transcriptional regulation activity, we searched Genevar expression quantitative trait loci (eQTL) database (19) and found moderate evidence for the association of rs2735591 with *BCL10* expression level in HapMap3 CHB (Han Chinese in Beijing, China) ($P = 0.071$) and JPT (Japanese in Tokyo, Japan) ($P = 0.032$) samples. Moderate evidence for eQTL effect was also observed for rs233100 in HapMap3 LWK (Luhya in Webuye, Kenya) ($P = 0.030$) and GIH (Gujarati Indian from Houston, Texas) (0.047) samples (Supplementary Material, Table S5). Furthermore, by annotating SNPs that are highly correlated ($r^2 > 0.8$) with rs2735591 and rs233100 using the information from the databases HaploReg (20) and RegulomeDB (21), we discovered that rs2735592, a highly correlated SNP with rs2735591 ($r^2 = 1$, $D' = 1$) is located at the 3′ UTR of *BCL10*, which may affect the binding of FOXA1 and FOXA2 transcription factors by damaging the TEF binding motif. Moreover, similar to rs2735591, a moderate evidence was observed for rs2735592 to affect the expression of *BCL10* in HapMap CHB ($P = 0.071$) and JPT ($P = 0.043$) samples.

We further investigated and compared the expression of *BCL10* between the leprosy lesions from 22 patients and the normal skins from 33 healthy controls. Significantly lower expression of *BCL10* was observed in the lesions of patients than the normal skin tissues ($P = 3.83 \times 10^{-5}$) (Fig. 2).

We also assessed the connections between *BCL10* and the eight previously identified susceptibility genes for leprosy, including *IL23R*, *RAB32*, *RIPK2*, *TNFSF15*, *CYLD, LACC1, CCDC122* and *NOD2*. *BCL10* was among the top 10 most significantly correlated gene with other leprosy known loci ($P = 6.07 \times 10^{-4}$, Supplementary Material, Table S6), connected mostly through *TNFSF15*, *RIPK2*, *CYLD* and *NOD2* (Supplementary Material, Fig. S2). The most frequently connecting terms include kappab, caspase, metabotropic, inflammatory and death. The involvement of NF-κB has been known to be a major regulator of downstream inflammatory responses for mycobacteria infection (22,23). This provides further supporting evidence for the involvement of BCL10 in the development of leprosy.

DISCUSSION

Through a three-stage association study in three independent case–control series, we have identified a novel association at rs2735591 ($P = 1.03 \times 10^{-9}$, OR = 1.24), locating 889 bp upstream of *BCL10*. Further fine mapping analysis of the surrounding region of this SNP revealed a stronger, but correlated association at rs233100, a noncoding polymorphism located 28 Kb to *BCL10*. Besides *BCL10*, there are several other genes or transcripts within the LD region of the novel association locus (Fig. 1). *C1orf52* and *DDAH1* locate next to *BCL10*. *C1orf52* is a hypothetical protein-coding transcript without functional annotation. *DDAH1* belongs to the dimethylarginine dimethylaminohydrolase (*DDAH*) gene family and encodes an enzyme involved in nitric oxide generation by regulating cellular concentrations of methylarginine, which in turn inhibits nitric oxide synthase activity (24).

Given that all the significant SNPs and the top SNP (rs233100) within the locus are noncoding variants, it is not unreasonable to hypothesize that the observed association at 1p22 might be due to the effect of the causal variant of the association on the transcriptional regulation of *BCL10* expression. This is supported by evidence from eQTL databases that our associated SNP might affect *BCL10* expression. Moreover, gene regulatory databases support this notion by pointing evidences that the associated SNP is likely to affect the binding of FOXA1 and FOXA2 transcription factors, which might in turn affect the expression of *BCL10*. We also investigated the expression of *BCL10* gene in skin tissues and found that the level of *BCL10* mRNA was down regulated in the lesions of leprosy compared with the normal skin tissues. It has been shown *in vivo* that BCL10 deficiency increases susceptibility to bacterial infections (25).

Table 2. Summary of the association results of rs2735591

SNP	rs2735591				
Chromosome	1				
Position	85 517 060				
Minor/Major Allele	T/C				
Test allele[a]	T				
Gene	*BCL10*				
	MAF cases	MAF controls	*P*	OR (95% CI)	Q-test[c]
GWAS 1220 samples (Stage 1)[b]	0.33	0.29	1.66×10^{-2}	1.24 (1.04–1.48)	NA
Validation stage 2	0.36	0.30	2.18×10^{-6}	1.32 (1.17–1.48)	NA
Validation stage 3	0.32	0.29	4.58×10^{-3}	1.17 (1.05–1.3)	NA
Combined Validation stages 2 and 3	0.34	0.30	1.27×10^{-7}	1.27 (1.14–1.33)	0.127
Combined stage 1–3	0.34	0.29	1.03×10^{-9}	1.24 (1.16–1.33)	0.954

MAF, minor allele frequency; *P*, *P*-value; OR, odds ratio; CI, confidence interval.
[a]Test allele: the allele that was used for estimating the OR.
[b]Data from the previously published GWAS (6).
[c]*P*-value from Cochran's Q-test of heterogeneity.

In addition, *BCL10* was the eighth most significantly connected gene to our previously identified leprosy susceptibility genes based on the GRAIL analysis, indicating that there is a strong biological relationship between *BCL10* and previously implicated susceptibility genes. These results suggest that *BCL10* is likely the susceptibility gene of this novel association, although further fine mapping and functional investigations are needed to confirm the biological mechanism underlying the association of *BCL10*.

The protein encoded by *BCL10* contains a caspase recruitment domain (CARD) and plays an important role in the activation of NF-κB, an important regulator of immune response against the infection by *M. leprae* (26,27). BCL10 regulates the activation of an NF-κB signaling pathway by forming two complexes *CARMA1-Bcl10-MALT1* and *CARD9-Bcl10-MALT1* in lymphoid (L-CBM) and myeloid (M-CBM) cells, respectively (28–30). As a regulator of innate immunity, an M-CBM complex acts on NF-κB activation downstream of immunoreceptor tyrosine-based activation motif (ITAM) and hemITAM-coupled receptors. M-CBM can also regulate the activation of mitogen-activated protein kinases downstream of TLRs and the intracellular bacterial sensor *NOD2* through *RIPK2* as a part of immune response for microbe infection (28). As a regulator of adaptive immunity, L-CBM is crucial for the activation of NF-κB in numerous immune receptor signaling pathways, including the T-cell receptor (TCR) and B-cell receptor (BCR) signaling pathways. It has been suggested that L-CBM might also be involved in the *TLR4*-mediated signaling for NF-κB activation in B lymphocytes. Paul *et al*. investigated the TCR-to-NF-κB signaling and indicated that antigen signaling through the TCR triggers both activation of NF-κB and the selective autophagy of BCL10, inhibiting NF-κB activation. This study demonstrates that selective autophagy of BCL10 is a pathway-intrinsic homeostatic mechanism that modulates TCR signaling to NF-κB in effector T cells and may protect T cells from adverse consequences of unrestrained NF-κB activation (31). Our discovery of *BCL10* as a new susceptibility locus may suggest that adaptive immunity, in addition to the *NOD2*-mediated innate immunity, also play an important role in leprosy.

We did not observe supporting evidence for the association of TLRs with leprosy susceptibility in Chinese population. In our previous study (13), we analyzed the reported SNP rs5743618 (I602S) of *TLR1* in Indian and African populations (17) and another SNP rs17616475 of *TLR1* showing suggestive association in our extended GWAS analysis, but did not observe any evidence of association in further validation, despite the fact that our sample of 3301 cases and 5299 controls should have sufficient power (99%) to detect the association of rs5743618 with OR = 0.31 (assuming the same effect size in Indians) at $P = 0.05$. In the current study, a total of 1599 SNPs within the 30 candidate genes of the TLR family, including 25 SNPs within the *TLR1* locus (chromosome 4: 38.4–38.6 Mb) were investigated by *in silico* association analysis in 706 cases and 5581 controls. Of the 1599 SNPs, only two TLR SNPs showed suggestive associations with $P < 0.01$, but both were not confirmed by further validation study. Power analysis indicates that the sample of 706 cases and 5581 controls should provide sufficient power (>85%) to detect associations in SNPs with a moderate effect (OR = 1.50) and frequencies as low as 5% at a significance threshold of 0.01. Taken together, our studies did not reveal any evidence for the associations of *TLRs* in Chinese population. The absence of association evidence for TLRs in Chinese population is intriguing, given that the TLRs are known to be the critical mediators of innate immune recognition of microbial pathogens, including *M. leprae* (32), and the association of *TLR1* with leprosy was replicated in Indian and African populations (17). Although the LD structure of the *TLR1* region is similar between the Chinese (CHB) and Indian (GIH) samples of Hapmap3 data (data not shown), we could not rule out the possibility that there could still be differences in some SNP's allele frequency or LD, and consequently the associations identified in Indian population may not be well captured in our study. Further studies are needed to understand whether the disparity of association evidence for *TLRs* between Chinese and Indian/African population may suggest the genetic heterogeneity of leprosy susceptibility between ethnic populations.

In summary, we have performed a comprehensive candidate gene association study of TLR and CARD by deep *in silico* association analysis in our previously published GWAS datasets and

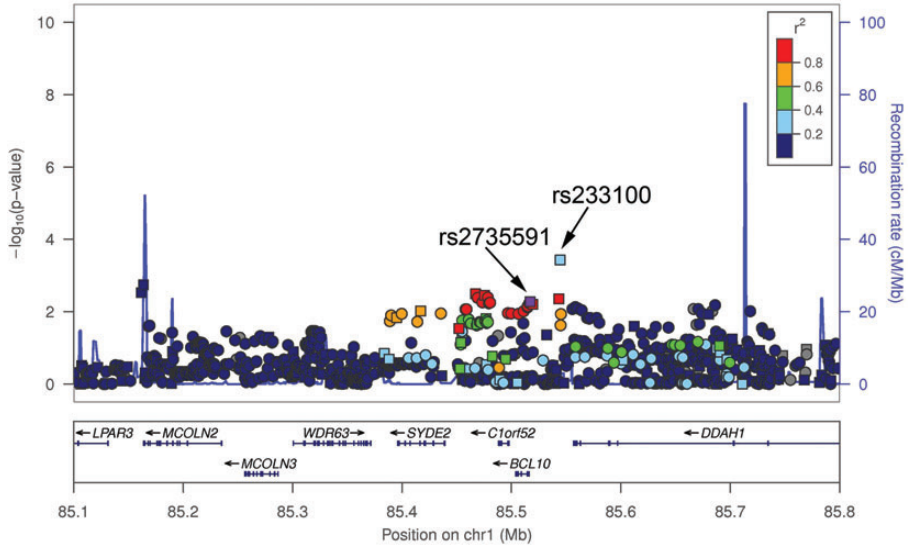

Figure 1. Regional association plots of 1p22 based on the imputation analysis in the expanded GWAS dataset of 706 cases and 5581 controls (13) (See Methods). The *P*-values for the SNPs (shown as $-\log_{10}$ *P*-values on the *Y*-axis) were plotted against their mapping positions (*X*-axis). Imputed SNPs are denoted as circles, whereas typed SNPs as squares. The color of each SNP reflects its r^2 value with the confirmed SNP rs2735591. Estimated recombination rates [based on the 1000 Genomes CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan)] were plotted in light blue. Plots were generated using LocusZoom (39). Top SNP (rs233100) as well as the confirmed SNP (rs2735591) within the LD block of 300 Kb on 1p22.3 region were labeled out.

further validation analysis in two large independent sample series of leprosy. By integrating the hypothesis-driven study with genome-wide analysis, we are able to identify novel associations beyond the original GWAS analysis. Our discovery of the novel association at 1p22 implicates *BCL10* as the new susceptibility gene for leprosy, highlighting the important role of both innate and adaptive immune responses in leprosy.

MATERIALS AND METHODS

Ethical statement

The study was approved by the institutional IRB committees at the Shandong Provincial Institute of Dermatology and Venereology, Shandong Academy of Medical Science. All the cases and the controls were recruited with written informed consent.

Subjects

Stage 1 discovery analysis was performed *in silico* using our previously published GWAS datasets (6,13). The validation analysis of Stage 2 was performed in an independent series of 1504 cases and 1502 controls, and further validation analysis of Stage 3 was performed in the second independent series of 938 cases and 5827 controls. All the patients and controls of the three stages were recruited from northern China and matched in terms of age, gender and residence, which minimizes the possibility of confounding due to population stratification. High correlation between subpopulation genetic and geographic structures of Han Chinese has been demonstrated to be a good proxy for genetic matching (33).

Table 1 shows the clinical information of all the samples. Leprosy was diagnosed on the basis of consensus by at least two dermatologists. The clinical diagnoses of all the leprosy cases were based on medical records stored in local leprosy control institutions and clinical assessments at the time of blood taken (looking for evidence of leprosy such as claw hand, lagothalomas or foot drop and pathology of skin lesions, etc.). The controls were recruited without history of leprosy or a family history of leprosy or other autoimmune diseases.

Candidate gene and SNP selection

By searching in the UCSC Genome Browser (http://genome.ucsc.edu/) using keyword '*TLR*' or '*CARD*', we found 37 candidate genes of *TLR* family and 53 candidate genes of *CARD* family. After removing overlapping transcripts, a total of 77 candidate genes of *TLR* (30 genes) and *CARD* (47 genes) were obtained. We then determined the critical region of each candidate gene (coding region $\pm$ 20 Kb) and extracted all the association results within these critical regions from our previously published GWAS analyses in genetically matched 1220 samples (706 cases and 514 controls) (6) and the expanded GWAS sample of 706 cases and 5581 population controls (13). There were a total of 4363 SNPs within 77 candidate genes extracted from the GWAS dataset, 2764 SNPs from CARD-related genes and 1599 SNPs from TLR-related genes. SNPs with suggestive association evidence were identified by using the following criteria: (i) the *P*-value of association was <0.01 in the expanded GWAS analysis; (ii) the *P*-value of association showed improvement between the initial GWAS analysis of 706 cases and 514 matched controls and the expanded GWAS analysis of 706 cases and 5581 population controls and (iii) SNPs located in known susceptibility loci, NOD2, RIPK2, were excluded. Finally, eight suggestive SNPs within two *TLR* genes (*ZCCHC11, FREM1*) and six *CARD* genes (*BCL10*, *KCNQ1*, *MAP2K1*, *CARD14*, *NOD1* and *VISA*) were brought forward for further validation analysis.

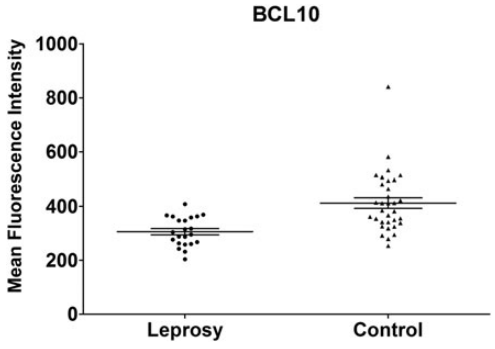

Figure 2. Expression analysis of *BCL10* in skin tissues. A significantly lower expression of *BCL10* was observed in the lesions of patients than the normal skin tissues of healthy controls ($P = 3.83 \times 10^{-5}$).

Gene-set enrichment analysis of leprosy genetic associations

MAGENTA (18) was employed to test for the enrichment of genetic associations within pathways by using the genome-wide association results from our published GWAS dataset (13) as input. This dataset includes 1 701 673 genotyped and imputed SNPs.

Evidence of significance was taken from the *q*-values as presented in the 'FDR_95PERC_CUTOFF' of MAGENTA's output. By using a less stringent level of 20% FDR (False Discovery Rate), we observed four pathways (Supplementary Material, Table S1), containing 141 unique genes, which is 0.7% of all the known genes in human genome. Among these 141 unique genes, 2 are *TLR*/*CARD* genes, which comprises 2.5% of the total *TLR*/*CARD* genes within the genome, but Pearson's chi-square test with Yates' continuity correction (as implemented in R statistical package under the function chisq.test) did not reach statistical significance and only showed suggestive evidence of enrichment (one-sided $P = 0.0995$). We further took 10 000 random gene sets comprising the same number of genes as the actual set at a level of 20% FDR (141 genes). Among these 10 000 random sets, there were 884 sets containing two or more *TLR*/*CARD* genes, giving us a *P*-value of 0.0884. As this exceeds the 5% conventional cut-off for the type-I error rate, this indicates that there is no significant evidence of enrichment for *TLR*/*CARD* genes.

Genotyping analysis

Genotyping analyses of all the validation samples were conducted by using the Sequenom MassArray system. Approximately 15 ng of genomic DNA was used to genotype each sample using the Sequenom MassArray system (San Diego, USA). The sample DNA was amplified by a multiplex PCR, and the PCR products were then used for a locus-specific single-base extension reaction. The resulting products were desalted and transferred to a 384-element SpectroCHIP array. Allele detection was performed using MALDI-TOF MS. The mass spectrograms were analyzed using the Sequenom MassARRAY TYPER software (San Diego, USA). In each validation series, we excluded SNPs with a call rate of $<95\%$, low minor allele frequency (<0.01) or deviation from Hardy–Weinberg equilibrium proportions ($P < 0.01$) in the controls. Out of eight suggestive SNPs validated in Stage 2, one SNP was unsuccessfully designed by Sequenom and the other one had poor clustering (Supplementary Material, Fig. S3), hence cannot be analyzed (rs6084506 and rs5743369, respectively, both SNPs are in CARD genes).

Imputation of ungenotyped SNPs within 1p22 locus in the expanded GWAS dataset

Imputation of ungenotyped SNPs within 1p22 locus was performed in the GWAS dataset (13) of 706 cases and 5581 controls by using IMPUTE version 2.2.2 and the genetic variation data from the 1000 Genomes Project (version February 2012) as a reference panel, which includes 1092 individuals from Africa (246 samples), North America (181 samples), Asia (286 samples) and Europe (379 samples). Imputed genotypes with a probability of $<90\%$, as well as SNPs with imputation certainty $<80\%$, minor allele frequency $<1\%$ and a missing rate of $>1\%$ were excluded from further analysis. In total, there were 1500 SNPs that were successfully imputed within this region. Together with 296 genotyped SNPs, a total of 1796 SNPs within 1p22 were tested for association in the 706 cases and 5581 controls.

Statistical analysis

The Cochran–Armitage trend test was used to analyze the genotype–phenotype association in each of the two validation samples using Plink v1.07 software (34). The Cochran–Mantel–Haenszel test was used to test genotype–phenotype association in the combined GWAS and validation samples by treating the GWAS (1220 samples) and two validation samples as independent studies. An expanded GWAS dataset described in our previous study (13) (706 cases and 5581 controls) were only used in the initial *in silico* association analysis to select SNPs for further validation analysis. The final association analysis of the combined samples was performed by using the matched GWAS dataset of 706 cases and 514 controls and the two independent validation series, consisting a total of 3148 cases and 7843 controls (this number excludes those extended samples used as controls with reasons as previously explained (13)). Cochran's Q test was performed to evaluate the significance of heterogeneity among the three studies, and the *P*-value of <0.05 after correction for multiple SNP testing was considered as significant heterogeneity. If the corrected *P*-value was >0.05, the fixed-effect model (Mantel–Haenszel) was used to combine the results of the three independent samples (35); otherwise, the random-effects model (DerSimonian-Laird) was used (36). Multiple SNP testing correction was performed using Bonferroni correction (37).

Gene relationship investigation with GRAIL

The web-based software called Gene Relationships Across Implicated Loci (GRAIL) (38) was used to investigate the biological relationship between the eight known susceptibility genes (including *IL23R*, *RAB32*, *RIPK2*, *TNFSF15*, *CYLD*, *CCDC122*, *LACC1* and *NOD2*) (6,13) and 30 and 47 candidate genes in the *TLR* and *CARD* gene family, respectively. GRAIL provides a relatedness score of two genes by measuring the text-based similarity based on the text in PubMed abstracts. Relatedness of the genes tested is not biased with respect to phenotype as

GRAIL does not take any information on the phenotype used. We selected the December 2006 PubMed data as a basis for text mining and CHB and JPT HapMap population as a basis for LD calculations in GRAIL.

Multiplex-branched DNA assay

FFPE skin tissue samples were obtained from the Shandong Provincial Institute of Dermatology and Venereology between 1989 and 2007 (33 cases and 34 controls), with the cases being collected generally earlier (older batch) than the controls. Samples were all of Northern Han Chinese descent with the mean age in the cases being 46.5 and the mean age in the controls being 42.6 (Student's *t*-test *P*-value = 0.33). Modified probe design software was used to design oligonucleotide probe sets for *BCL10*. The bDNA assays were performed according to the recommended procedure of QuantiGene Reagent System. Relative levels of gene expression were normalized to the expression of *PPIB* and *HPRT1* to control for variability in the preparation of the tissue samples as well as input material. Eleven cases and one control had low fluorescence intensity and, therefore, were not included in the analysis. Hence, the remaining samples to be analyzed include 22 cases and 33 controls. Since the expression data did not follow a normal distribution, statistical significance was tested using a non-parametric test, Wilcoxon's rank sum test implemented in R package 'exactRankTests'. Significance was defined at $P < 0.01$.

URLS

Genevar, http://www.sanger.ac.uk/resources/software/genevar/java/genevar.jnlp.

RegulomeDB, http://regulome.stanford.edu/.

HaploReg, http://www.broadinstitute.org/mammals/haploreg/haploreg.php.

MAGENTA, http://www.broadinstitute.org/mpg/magenta/.

GRAIL, http://www.broadinstitute.org/mpg/grail/.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank the individuals who participated in this project.

Conflict of Interest statement. None declared.

FUNDING

This work was funded by grants from the National Natural Science Foundation of China (81071288, 81072391, and 81101187), the National 973 program (2011CB512105), the Project of Taishan scholar (2008–present), the Project of Medical leading scholar of Shandong Province (2010–present), the Natural Science Foundation of Shandong Province (ZR2011HQ003), the Innovative Research Team of the Ministry of Education of China (IRT-1046) and the Agency for Science, Technology and Research (A*STAR) of Singapore. A.I. was supported by the Singapore International Graduate Award. Funding to pay the Open Access publication charges for this article was provided by special foundation of Shandong Medical Leading Scholar(20121212).

REFERENCES

1. 2012) Global leprosy situation, 2012. *Wkly Epidemiol. Rec.*, **87**, 317–328.
2. Misch, E.A., Berrington, W.R., Vary, J.C. and Hawn, T.R. (2010) Leprosy and the human genome. *Microbiol. Mol. Biol. Rev.*, **74**, 589–620.
3. Shields, E.D., Russell, D.A. and Pericak-Vance, M.A. (1987) Genetic epidemiology of the susceptibility to leprosy. *J. Clin. Invest.*, **79**, 1139–1143.
4. Bakker, M.I., May, L., Hatta, M., Kwenang, A., Klatser, P.R., Oskam, L. and Houwing-Duistermaat, J.J. (2005) Genetic, household and spatial clustering of leprosy on an island in Indonesia: a population-based study. *BMC Med. Genet.*, **6**, 40.
5. Alter, A., Grant, A., Abel, L., Alcaïs, A. and Schurr, E. (2011) Leprosy as a genetic disease. *Mamm. Genome*, **22**, 19–31.
6. Zhang, F.R., Huang, W., Chen, S.M., Sun, L.D., Liu, H., Li, Y., Cui, Y., Yan, X.X., Yang, H.T., Yang, R.D. *et al.* (2009) Genomewide association study of leprosy. *N. Engl. J. Med.*, **361**, 2609–2618.
7. Ferwerda, G., Kullberg, B.J., de Jong, D.J., Girardin, S.E., Langenberg, D.M., van Crevel, R., Ottenhoff, T.H., Van der Meer, J.W. and Netea, M.G. (2007) Mycobacterium paratuberculosis is recognized by Toll-like receptors and NOD2. *J. Leukoc. Biol.*, **82**, 1011–1018.
8. Mahapatra, S., Crick, D.C., McNeil, M.R. and Brennan, P.J. (2008) Unique structural features of the peptidoglycan of Mycobacterium leprae. *J. Bacteriol.*, **190**, 655–661.
9. Al-Sayeqh, A.F., Loughlin, M.F., Dillon, E., Mellits, K.H. and Connerton, I.F. (2010) Campylobacter jejuni activates NF-kappaB independently of TLR2, TLR4, Nod1 and Nod2 receptors. *Microb. Pathog.*, **49**, 294–304.
10. Shaw, M.H., Reimer, T., Kim, Y.G. and Nuñez, G. (2008) NOD-like receptors (NLRs): bona fide intracellular microbial sensors. *Curr. Opin. Immunol.*, **20**, 377–382.
11. McCarthy, J.V., Ni, J. and Dixit, V.M. (1998) RIP2 is a novel NF-kappaB-activating and cell death-inducing kinase. *J. Biol. Chem.*, **273**, 16968–16975.
12. Blonska, M. and Lin, X. (2011) NF-êB signaling pathways regulated by CARMA family of scaffold proteins. *Cell Res.*, **21**, 55–70.
13. Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y. *et al.* (2011) Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.*, **43**, 1247–1251.
14. Misch, E.A., Macdonald, M., Ranjit, C., Sapkota, B.R., Wells, R.D., Siddiqui, M.R., Kaplan, G. and Hawn, T.R. (2008) Human TLR1 deficiency is associated with impaired mycobacterial signaling and protection from leprosy reversal reaction. *PLoS Negl. Trop. Dis.*, **2**, e231.
15. Bochud, P.Y., Sinsimer, D., Aderem, A., Siddiqui, M.R., Saunderson, P., Britton, S., Abraham, I., Tadesse Argaw, A., Janer, M., Hawn, T.R. *et al.* (2009) Polymorphisms in Toll-like receptor 4 (TLR4) are associated with protection against leprosy. *Eur. J. Clin. Microbiol. Infect. Dis.*, **28**, 1055–1065.
16. Bochud, P.Y., Hawn, T.R., Siddiqui, M.R., Saunderson, P., Britton, S., Abraham, I., Argaw, A.T., Janer, M., Zhao, L.P., Kaplan, G. *et al.* (2008) Toll-like receptor 2 (TLR2) polymorphisms are associated with reversal reaction in leprosy. *J. Infect. Dis.*, **197**, 253–261.
17. Wong, S.H., Gochhait, S., Malhotra, D., Pettersson, F.H., Teo, Y.Y., Khor, C.C., Rautanen, A., Chapman, S.J., Mills, T.C., Srivastava, A. *et al.*. (2010) Leprosy and the adaptation of human toll-like receptor 1. *PLoS Pathog.*, **6**, e1000979.
18. Segrè, A.V., Groop, L., Mootha, V.K., Daly, M.J., Altshuler, D., Consortium, D. and investigators, M. (2010) Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS. Genet.*, **6**, e1001058.
19. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P. and Dermitzakis, E.T. (2010) Genevar: a database and Java application for the analysis and

visualization of SNP-gene associations in eQTL studies. *Bioinformatics*, **26**, 2474–2476.
20. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
21. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S. *et al.* (2012) Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.*, **22**, 1790–1797.
22. Zhang, Y., Broser, M. and Rom, W.N. (1994) Activation of the interleukin 6 gene by Mycobacterium tuberculosis or lipopolysaccharide is mediated by nuclear factors NF-IL6 and NF-kappa B. *Proc. Natl Acad. Sci. USA.*, **91**, 2225–2229.
23. Gutierrez, M.G., Gonzalez, A.P., Anes, E. and Griffiths, G. (2009), In *Cell Microbiol.*, England, Vol. 11, pp. 406–420.
24. Ellger, B., Richir, M.C., van Leeuwen, P.A., Debaveye, Y., Langouche, L., Vanhorebeek, I., Teerlink, T. and Van den Berghe, G. (2008) Glycemic control modulates arginine and asymmetrical-dimethylarginine levels during critical illness by preserving dimethylarginine–dimethylaminohydrolase activity. *Endocrinology*, **149**, 3148–3157.
25. Xue, L., Morris, S.W., Orihuela, C., Tuomanen, E., Cui, X., Wen, R. and Wang, D. (2003), In *Nat. Immunol.*, United States, Vol. 4, pp. 857–865.
26. Bhattacharyya, S., Borthakur, A., Tyagi, S., Gill, R., Chen, M.L., Dudeja, P.K. and Tobacman, J.K. (2010) B-cell CLL/lymphoma 10 (BCL10) is required for NF-kappaB production by both canonical and noncanonical pathways and for NF-kappaB-inducing kinase (NIK) phosphorylation. *J. Biol. Chem.*, **285**, 522–530.
27. Kang, T.J., Lee, G.S., Kim, S.K., Jin, S.H. and Chae, G.T. (2010) Comparison of two mice strains, A/J and C57BL/6, in caspase-1 activity and IL-1beta secretion of macrophage to Mycobacterium leprae infection. *Mediators Inflamm.*, **2010**, 708713.
28. Hara, H. and Saito, T. (2009) CARD9 versus CARMA1 in innate and adaptive immunity. *Trends Immunol.*, **30**, 234–242.
29. Wegener, E. and Krappmann, D. (2007) CARD-Bcl10-Malt1 signalosomes: missing link to NF-kappaB. *Sci. STKE*, **2007**, pe21.
30. Wegener, E., Oeckinghaus, A., Papadopoulou, N., Lavitas, L., Schmidt-Supprian, M., Ferch, U., Mak, T.W., Ruland, J., Heissmeyer, V. and Krappmann, D. (2006) Essential role for IkappaB kinase beta in remodeling Carma1-Bcl10-Malt1 complexes upon T cell activation. *Mol. Cell*, **23**, 13–23.
31. Paul, S. and Schaefer, B.C. (2012) Selective autophagy regulates T cell activation. *Autophagy*, **8**, 1690–1692.
32. Krutzik, S.R., Ochoa, M.T., Sieling, P.A., Uematsu, S., Ng, Y.W., Legaspi, A., Liu, P.T., Cole, S.T., Godowski, P.J., Maeda, Y. *et al.* (2003) Activation and regulation of Toll-like receptors 2 and 1 in human leprosy. *Nat. Med.*, **9**, 525–532.
33. Chen, J., Zheng, H., Bei, J.X., Sun, L., Jia, W.H., Li, T., Zhang, F., Seielstad, M., Zeng, Y.X., Zhang, X. *et al.* (2009) Genetic structure of the Han Chinese population revealed by genome-wide SNP variation. *Am. J. Hum. Genet.*, **85**, 775–785.
34. Shi, H., Medway, C., Brown, K., Kalsheker, N. and Morgan, K. (2011) Using Fisher's method with PLINK 'LD clumped' output to compare SNP effects across Genome-wide Association Study (GWAS) datasets. *Int. J. Mol. Epidemiol. Genet.*, **2**, 30–35.
35. Mantel, N. and Haenszel, W. (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. Natl. Cancer Inst.*, **22**, 719–748.
36. DerSimonian, R. and Laird, N. (1986) Meta-analysis in clinical trials. *Control Clin. Trials*, **7**, 177–188.
37. Bonferroni, C.E. (1935), In *Studi in Onore del Professore Salvatore Ortu Carboni*, pp. 13–60.
38. Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D. and Daly, M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.*, **5**, e1000534.
39. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.