nature genetics

Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus

Jian-Wen Han[1,2,18], Hou-Feng Zheng[1,2,18], Yong Cui[1,2,18], Liang-Dan Sun[1,2], Dong-Qing Ye[1,2], Zhi Hu[1,2], Jin-Hua Xu[3], Zhi-Ming Cai[4], Wei Huang[5], Guo-Ping Zhao[5], Hong-Fu Xie[6], Hong Fang[7], Qian-Jin Lu[8], Jian-Hua Xu[9], Xiang-Pei Li[10], Yun-Feng Pan[11], Dan-Qi Deng[12], Fan-Qin Zeng[13], Zhi-Zhong Ye[14], Xiao-Yan Zhang[15], Qing-Wen Wang[4], Fei Hao[16], Li Ma[15], Xian-Bo Zuo[1,2], Fu-Sheng Zhou[1,2], Wen-Hui Du[1,2], Yi-Lin Cheng[1,2], Jian-Qiang Yang[1,2], Song-Ke Shen[1,2], Jian Li[1,2], Yu-Jun Sheng[1,2], Xiao-Xia Zuo[6], Wei-Fang Zhu[7], Fei Gao[8], Pei-Lian Zhang[12], Qing Guo[13], Bo Li[14], Min Gao[1,2], Feng-Li Xiao[1,2], Cheng Quan[1,2], Chi Zhang[1,2], Zheng Zhang[1,2], Kun-Ju Zhu[1,2], Yang Li[1,2], Da-Yan Hu[1,2], Wen-Sheng Lu[1,2], Jian-Lin Huang[11], Sheng-Xiu Liu[1,2], Hui Li[1,2], Yun-Qing Ren[1,2], Zai-Xing Wang[1,2], Chun-Jun Yang[1,2], Pei-Guang Wang[1,2], Wen-Ming Zhou[1,2], Yong-Mei Lv[1,2], An-Ping Zhang[1,2], Sheng-Quan Zhang[1,2], Da Lin[1,2], Yi Li[17], Hui Qi Low[17], Min Shen[5], Zhi-Fang Zhai[16], Ying Wang[5], Feng-Yu Zhang[1,2], Sen Yang[1,2], Jian-Jun Liu[1,2] & Xue-Jun Zhang[1–3]

We performed a genome-wide association study (GWAS) of systemic lupus erythematosus (SLE) in a Chinese Han population by genotyping 1,047 cases and 1,205 controls using Illumina Human610-Quad BeadChips and replicating 78 SNPs in two additional cohorts (3,152 cases and 7,050 controls). We identified nine new susceptibility loci (*ETS1*, *IKZF1*, *RASGRP3*, *SLC15A4*, *TNIP1*, 7q11.23, 10q11.22, 11q23.3 and 16p11.2; $1.77 \times 10^{-25} \leq P_{combined} \leq 2.77 \times 10^{-8}$) and confirmed seven previously reported loci (*BLK*, *IRF5*, *STAT4*, *TNFAIP3*, *TNFSF4*, 6q21 and 22q11.21; $5.17 \times 10^{-42} \leq P_{combined} \leq 5.18 \times 10^{-12}$). Comparison with previous GWAS findings highlighted the genetic heterogeneity of SLE susceptibility between Chinese Han and European populations. This study not only advances our understanding of the genetic basis of SLE but also highlights the value of performing GWAS in diverse ancestral populations.

Systemic lupus erythematosus (SLE) is a prototypic systemic autoimmune disease that is characterized by a diverse array of autoantibody production, complement activation, immune complex deposition and tissue and organ damage[1] and that is influenced by both genetic and environmental factors[2]. SLE affects predominantly women (prevalence ratio of women to men is 9:1), particularly during the childbearing years. There are marked disparities in SLE incidence and prevalence worldwide; SLE prevalence varies among different ethnic and geographical populations[3]. The prevalence of SLE ranges from 31 to 70 cases per 100,000 persons among Chinese populations[4] and from 7 to 71 cases per 100,000 persons in European populations[5], and lupus nephritis is more prevalent in Chinese populations than in European populations[3]. The ethnic and genetic heterogeneity of SLE may contribute to the complexity of its clinical manifestation.

Over the past two decades, numerous studies have identified multiple genetic factors related to SLE. In particular, four recent genome-wide association studies (GWAS) of SLE in European populations have identified more than 20 robust susceptibility genes and/or loci[6–9]. The genetic heterogeneity between ethnic populations has

[1]Institute of Dermatology and Department of Dermatology at NO. 1 Hospital, Anhui Medical University, Hefei, Anhui, China. [2]Key Laboratory of Dermatology, Anhui Medical University, Ministry of Education, China, Hefei, Anhui, China. [3]Department of Dermatology, Huashan Hospital of Fudan University, Shanghai, China. [4]Department of Rheumatology, Peking University Shenzhen Hospital, Shenzhen, Guangdong, China. [5]Chinese National Human Genome Center at Shanghai, Shanghai, China. [6]Department of Dermatology, Xiangya Hospital, Central South University, Changsha, Hunan, China. [7]Department of Dermatology, First Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China. [8]Department of Dermatology, Epigenetic Research Center, Second Xiangya Hospital, Central South University, Changsha, Hunan, China. [9]Department of Rheumatology at NO. 1 Hospital, Anhui Medical University, Hefei, Anhui, China. [10]Department of Rheumatology, Affiliated Provincial Hospital of Anhui Medical University, Hefei, Anhui, China. [11]Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China. [12]Department of Dermatology, The Second Affiliated Hospital of Kunming Medical University, Kunming, Yunnan, China. [13]Department of Dermatology, The Second Affiliated Hospital of Sun Yat-sen University, Guangzhou, Guangdong, China. [14]Department of Rheumatology, Shenzhen Fourth People's Hospital, Shenzhen, Guangdong, China. [15]Department of Dermatology, China-Japan Friendship Hospital, Beijing, China. [16]Department of Dermatology, Southwest Hospital, Third Military Medical University, Chongqing, China. [17]Human Genetics, Genome Institute of Singapore, Singapore, Singapore. [18]These authors contributed equally to this work. Correspondence should be addressed to X.-J.Z. (ayzxj@vip.sina.com), J.-J.L. (liuj3@gis.a-star.edu.sg) or S.Y. (yangsen@medmail.com.cn).

Received 13 July; accepted 9 September; published online 18 October 2009; doi:10.1038/ng.472

Table 1 Summary of samples used in GWAS and replication studies

	Cases					Controls			
	Sample size	Mean age (s.d.)	Mean age of onset (s.d.)	Male/female	Female (%)	Sample size	Mean age (s.d.)	Male/female	Female (%)
GWAS[a]	1,047	34.02 (11.53)	29.81 (10.08)	63/984	93.98	1,205	34.75 (12.97)	673/532	44.15
Replication 1[b]	1,643	35.36 (12.10)	30.89 (11.21)	136/1,507	91.72	5,930	29.64 (11.31)	2,729/3,201	53.98
Replication 2[c]	1,509	32.85 (11.16)	28.18 (10.42)	113/1,396	92.51	1,120	32.23 (14.48)	415/705	62.95
Total	4,199	34.12 (11.68)	29.65 (10.96)	312/3,887	92.57	8,255	30.74 (12.18)	3,817/4,438	53.76

[a]GWAS samples are from central China. [b]Replication 1 samples are from central China. [c]Replication 2 sample are from southern China.

been suggested to be important in SLE risk[10], showing the need for further GWAS in non-European populations.

We conducted a GWAS of SLE in a Chinese Han population by analyzing 493,955 autosomal SNPs in 1,047 affected individuals (cases) and 1,205 controls (**Table 1**). Principal component analysis and genomic control ($\lambda_{gc} = 1.045$ for all 493,955 SNPs and 1.039 after excluding 4,840 SNPs within the major histocompatibility complex (MHC) region) indicated minimal inflation of the GWA results due to population stratification (**Supplementary Fig. 1**).

The analysis revealed association at three loci—6p21 (MHC)[6,7,9], 2q32.3 (*STAT4*)[6,9] and 8p23.1 (*BLK*)[6]—with genome-wide significance ($P < 5 \times 10^{-8}$) (**Fig. 1a**). Furthermore, a quantile-quantile plot of the observed *P* values showed a clear deviation at the tail of the distribution from the null distribution (the distribution expected if there were no association) even after 4,840 SNPs were removed from the MHC region. This suggests that the observed *P* values, particularly the ones within the tail of the distribution, are smaller than those expected by chance and probably reflect true genetic association (**Fig. 1b**).

In the MHC region, 13 SNPs, all located within the HLA class II region, showed genome-wide significant association ($P < 5 \times 10^{-8}$); the most significant association was identified at rs9271100 ($P = 1.42 \times 10^{-12}$, odds ratio (OR) = 1.9, 95% CI = 1.59–2.27) (**Supplementary Fig. 2**). Further conditional association analysis of the 13 SNPs confirmed that there were two independent associations within the HLA class II region at rs9271100 near *HLA-DRB1* and at rs3997854 near *HLA-DQA2*. Our association results within the MHC region are largely consistent with findings (including association at *HLA-DRB1*) identified in previous GWAS of SLE in European populations[6].

We performed a replication study by genotyping 78 non-MHC SNPs (from 67 loci) in two additional cohorts of Chinese Han individuals (replication 1: 1,643 cases and 5,930 controls; replication 2: 1,509 cases and 1,120 controls; **Table 1** and **Supplementary Table 1**). In the replication study, the association analysis was performed in the two individual replication samples separately as well as in the combined GWAS and replication sample. Twenty-one SNPs within 16 loci were validated, with independent evidence supporting association ($P < 0.03$) for the two replication samples and highly significant evidence in the combined sample that surpassed genome-wide significance ($P_{combined} < 5 \times 10^{-8}$). The 16 confirmed susceptibility loci were located at 1q25.1, 2p22.3, 2q32.3, 5q33.1, 6q21, 6q23.3, 7p12.2, 7q11.23, 7q32.1, 8p23.1, 10q11.22, 11q23.3, 11q24.3, 12q24.32, 16p11.2 and 22q11.21 ($5.17 \times 10^{-42} \leq P_{combined} \leq 2.77 \times 10^{-8}$) (**Table 2**). The associations at these 16 loci were independent of the associations within the MHC region because the associations at these loci remained similar even after controlling for the genetic effect of rs9271100 and rs3997854 within the MHC region (**Supplementary Table 2**). Pairwise interaction analysis was also performed among the top SNPs from the 16 non-MHC loci and the 2 MHC loci (rs9271100 and rs3997854), but no interaction was identified ($P > 0.05$ after correction for multiple testing; data not shown).

The associations at 1q25.1, 2q32.3, 6q21, 6q23.3, 7q32.1, 8p23.1 and 22q11.21 were reported by previous GWAS in European populations,

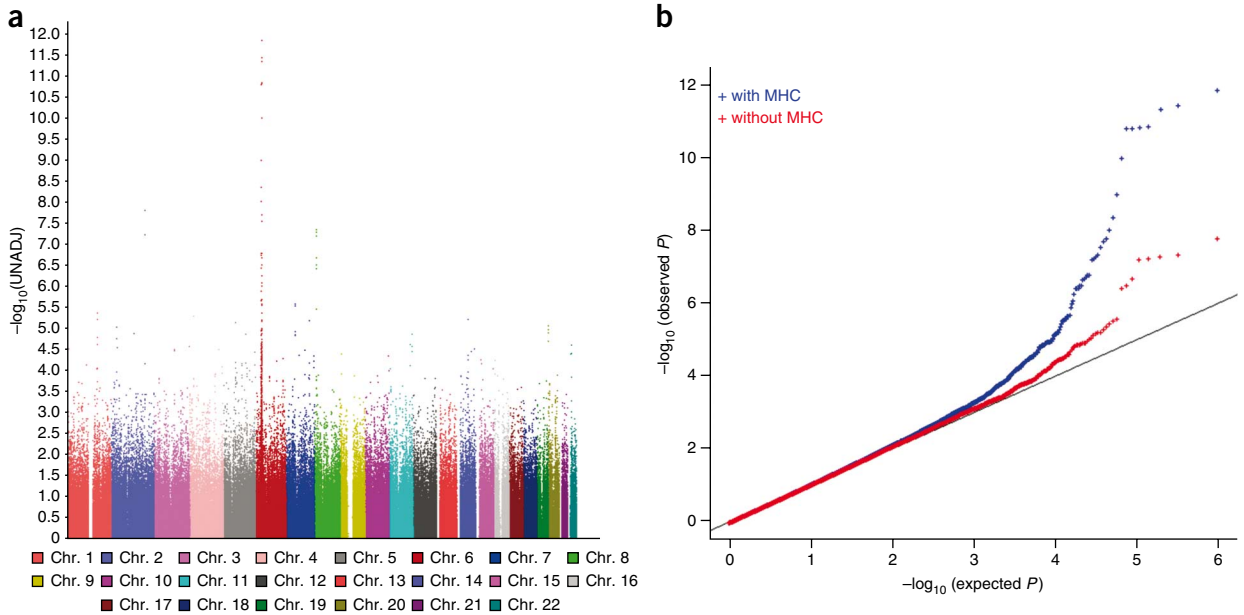

Figure 1 Summary of genome-wide association results for 1,047 cases and 1,205 controls. (**a**) The genome-wide *P* values ($-\log_{10} P$) of the logistic regression analysis adjusted for gender (493,955 SNPs) plotted against position on each chromosome. Each chromosome is depicted in a different color. (**b**) Quantile-quantile plots of the observed *P* values versus the expected values from *P* value of association. The plot in blue was based on the entire set of 493,955 SNPs, whereas the plot in red was obtained after removing 4,840 SNPs within MHC region (Chr. 6: 25–37 Mb).

Table 2 Association evidence for 21 SNPs at 16 loci in GWAS, replication and combined studies

SNP	Chr	Gene	Allele[a]	GWAS				Replication 1 (central)				Replication 2 (Southern)				Combined			
				1,047 cases, 1,205 controls				1,643 cases, 5,930 controls				1,509 cases, 1,120 controls				4,199 cases, 8,255 controls			
				MAF		OR	P^b	MAF		OR	P^b	MAF		OR	P^b	MAF		OR	P^c
				Case	Control			Case	Control			Case	Control			Case	Control	(95% CI)	
rs1234315	1q25.1	*TNFSF4*	A/G	0.5000	0.4253	1.32	8.92×10^{-05}	0.4987	0.4164	1.35	1.93×10^{-12}	0.5107	0.4385	1.36	5.48×10^{-07}	0.5033	0.4208	1.37 (1.29–1.45)	2.34×10^{-26}
rs2205960	1q25.1	*TNFSF4*	A/C	0.3395	0.2602	1.43	4.17×10^{-06}	0.3566	0.2698	1.43	1.58×10^{-15}	0.3411	0.2692	1.39	8.07×10^{-07}	0.3468	0.2682	1.46 (1.37–1.56)	2.53×10^{-32}
rs13385731	2p22.3	*RASGRP3*	G/A	0.1108	0.1480	0.64	2.82×10^{-05}	0.1107	0.1489	0.69	9.47×10^{-09}	0.1149	0.1535	0.73	6.30×10^{-04}	0.1122	0.1494	0.70 (0.64–0.76)	1.25×10^{-15}
rs7574865	2q32.3	*STAT4*	A/C	0.4193	0.3249	1.53	1.57×10^{-08}	0.4150	0.3248	1.47	3.75×10^{-19}	0.4334	0.3431	1.52	6.51×10^{-11}	0.4227	0.3273	1.51 (1.43–1.61)	5.17×10^{-42}
rs10036748	5q33.1	*TNIP1*	G/A	0.2053	0.2539	0.80	5.70×10^{-03}	0.2063	0.2353	0.84	9.79×10^{-04}	0.1965	0.2460	0.75	7.87×10^{-05}	0.2026	0.2395	0.81 (0.75–0.87)	1.67×10^{-09}
rs548234	6q21	*PRDM1-ATG5*	G/A	0.3176	0.2672	1.27	2.51×10^{-03}	0.3014	0.2628	1.19	1.48×10^{-04}	0.3120	0.2685	1.27	4.25×10^{-04}	0.3093	0.2642	1.25 (1.17–1.33)	5.18×10^{-12}
rs2230926	6q23.3	*TNFAIP3*	C/A	0.0755	0.0456	1.57	2.49×10^{-03}	0.0755	0.0429	1.95	4.35×10^{-14}	0.0623	0.0335	1.83	5.34×10^{-05}	0.0708	0.0420	1.72 (1.52–1.94)	1.37×10^{-17}
rs4917014	7p12.2	*IKZF1*	C/A	0.2560	0.3154	0.70	2.93×10^{-06}	0.2596	0.3272	0.72	1.45×10^{-12}	0.2438	0.2969	0.77	1.48×10^{-04}	0.2530	0.3214	0.72 (0.68–0.77)	2.75×10^{-23}
rs1167796	7q11.23	Multiple genes	A/G	0.2311	0.2832	0.80	4.74×10^{-03}	0.2555	0.2932	0.82	3.13×10^{-05}	0.2638	0.3052	0.83	4.02×10^{-03}	0.2523	0.2934	0.83 (0.78–0.89)	2.12×10^{-08}
rs4728142	7q32.1	*IRF5*	A/G	0.1796	0.1361	1.37	1.57×10^{-03}	0.1766	0.1348	1.44	2.48×10^{-10}	0.1819	0.1186	1.66	7.09×10^{-09}	0.1792	0.1328	1.43 (1.32–1.54)	8.14×10^{-19}
rs7812879	8p23.1	*BLK*	A/G	0.1734	0.2456	0.62	4.44×10^{-08}	0.1808	0.2366	0.71	3.17×10^{-10}	0.1800	0.2444	0.68	9.03×10^{-08}	0.1786	0.2389	0.69 (0.64–0.74)	2.09×10^{-24}
rs2618479	8p23.1	*BLK*	A/G	0.1939	0.2610	0.65	2.06×10^{-07}	0.2084	0.2611	0.76	5.63×10^{-08}	0.2036	0.2641	0.70	8.29×10^{-07}	0.2030	0.2615	0.72 (0.67–0.77)	5.26×10^{-21}
rs2248932	8p23.1	*BLK*	G/A	0.2001	0.2556	0.73	1.67×10^{-04}	0.2120	0.2556	0.79	5.06×10^{-06}	0.2051	0.2543	0.76	1.23×10^{-04}	0.2066	0.2554	0.76 (0.71–0.82)	1.63×10^{-14}
rs1913517	10q11.22	*LRRC18-WDFY4*	A/G	0.3338	0.2871	1.28	1.28×10^{-03}	0.3431	0.2893	1.27	7.72×10^{-08}	0.3166	0.2807	1.17	1.43×10^{-02}	0.3312	0.2878	1.24 (1.17–1.32)	7.22×10^{-12}
rs4639966	11q23.3	Multiple genes	G/A	0.3386	0.2816	1.39	2.39×10^{-05}	0.3327	0.2981	1.16	1.22×10^{-03}	0.4033	0.3393	1.30	2.56×10^{-05}	0.3594	0.3013	1.29 (1.22–1.37)	1.25×10^{-16}
rs6590330	11q24.3	*ETS1*	A/G	0.4078	0.3436	1.37	1.33×10^{-05}	0.4081	0.3431	1.33	9.31×10^{-11}	0.4249	0.3351	1.44	1.08×10^{-08}	0.4141	0.3421	1.37 (1.29–1.45)	1.77×10^{-25}
rs10847697	12q24.32	*SLC15A4*	A/G	0.2622	0.2058	1.31	1.24×10^{-03}	0.2511	0.2059	1.25	7.38×10^{-06}	0.2336	0.2030	1.19	1.56×10^{-02}	0.2475	0.2055	1.26 (1.17–1.34)	3.54×10^{-11}
rs1385374	12q24.32	*SLC15A4*	A/G	0.2617	0.2046	1.32	9.16×10^{-04}	0.2502	0.2054	1.24	1.15×10^{-05}	0.2344	0.2009	1.22	6.64×10^{-03}	0.2474	0.2046	1.26 (1.18–1.35)	1.77×10^{-11}
rs7197475	16p11.2	Multiple genes	A/G	0.0969	0.0660	1.58	6.52×10^{-04}	0.0976	0.0848	1.18	2.18×10^{-02}	0.1128	0.0854	1.44	4.80×10^{-04}	0.1029	0.0821	1.31 (1.20–1.46)	2.77×10^{-08}
rs463426	22q11.21	*HIC2-UBE2L3*	G/A	0.4647	0.5261	0.83	9.12×10^{-03}	0.4790	0.5268	0.81	9.54×10^{-07}	0.4206	0.4771	0.82	1.17×10^{-03}	0.4545	0.5199	0.78 (0.74–0.83)	1.48×10^{-16}
rs131654	22q11.21	*HIC2-UBE2L3*	C/A	0.4131	0.4768	0.78	5.11×10^{-04}	0.4264	0.4795	0.79	2.14×10^{-08}	0.3769	0.4230	0.81	7.52×10^{-04}	0.4054	0.4714	0.78 (0.74–0.83)	2.99×10^{-16}

[a]Minor allele/major allele. [b]Association statistic adjusted for gender. [c]Association statistic adjusted for gender and study.

implicating *TNFSF4* (ref. 11), *STAT4* (refs. 6,9), *PRDM1-ATG5* (ref. 7), *TNFAIP3* (ref. 9), *IRF5* (refs. 6,7,9), *BLK*[6] and *HIC2-UBE2L3* (ref. 7) as SLE susceptibility genes. The replication results of the remaining 57 SNPs are summarized in **Supplementary Table 3**.

To identify susceptibility genes underlying each of the nine newly discovered associations, we investigated patterns of recombination and linkage disequilibrium (LD) around the risk-associated SNPs and the gene(s) located within each region harboring the association (**Supplementary Fig. 3**). Only one gene was implicated by the associations at 2p22.3 (*RASGRP3*), 7p12.2 (*IKZF1*), 11q24.3 (*ETS1*) and 12q24.32 (*SLC15A4*), where a single gene was found within the LD block harboring the association. At 5q33.1, two genes (*TNIP1* and *GPX3*) were found within this locus, but *TNIP1* is more plausible as a susceptibility gene due to its important role in modulating cell activation, cytokine signaling and apoptosis[12]. For the associations at 7q11.23, 10q11.22, 11q23.3 and 16p11.2, there were more than one gene or transcript within the each region. Further fine mapping analyses are required to determine the susceptibility genes underlying these newly discovered SLE risk loci. Of note, the association at 16p11.2 is located close to the *ITGAM-ITGAX* susceptibility locus identified in European populations[6,7,13]. However, the association identified in our study is within an LD block that is distinct from the one in which *ITGAM* and *ITGAX* are located. Furthermore, the minor allele frequencies (MAFs) of five previously reported SNPs within the *ITGAM-ITGAX* locus were very low (<0.01) in the Chinese Han population (**Supplementary Table 4a**). Further studies will be needed to confirm the susceptibility gene within this region.

To investigate how these loci might influence SLE susceptibility, we further explored the biological function of the five newly discovered susceptibility genes (*RASGRP3*, *IKZF1*, *ETS1*, *SLC15A4* and *TNIP1*). *ETS1* encodes a member of the ETS family of transcription factors that regulate the development, senescence and death of many immune cells and that play a role both in innate and adaptive immune response[14]. *IKZF1* encodes a lymphoid-restricted zinc finger transcription factor that regulates lymphocyte differentiation and proliferation[15] as well as self-tolerance through regulation of B-cell receptor signaling[16]. RasGRP3 is responsible for Ras-ERK signaling mediated by B-cell receptor ligation in B cells and might be involved in immunoglobulin production and B-cell proliferation[17]. TNIP1 interacts with TNFAIP3 and could act as a negative regulator of the NF-κB pathway downstream of TNF-α, which has a key role in modulating cell activation, cytokine signaling and apoptosis[12]. SLC15A4 is a transporter for Nod1 ligands in early endosomes and is involved in Nod1-dependent NF-κB signaling, with potential participation in the antigen presentation in immune response[18]. These newly discovered susceptibility genes provide further supporting evidence for the three biological processes that were highlighted by findings from recent GWAS of SLE in European populations[19,20]: immune complex processing (*SLC15A4*); Toll-like receptor function and type I interferon production (*TNIP1*); and immune signal transduction in lymphocytes (*ETS1*, *RASGRP3* and *IKZF1*). It is worth pointing out that *TNIP1* and *TNFAIP3* are also associated with psoriasis[21], suggesting that the genetic basis of SLE and psoriasis might be partially shared, possibly through regulation of inflammation.

This first GWAS of SLE in a Chinese Han population allowed us to perform a direct comparison with GWAS findings in European populations. We first investigated the association evidence from our study for 48 SLE-associated SNPs (from 22 loci) identified by previous GWAS in European populations, including seven SNPs within the MHC region (**Supplementary Table 4a**). In addition to the three SNPs within three loci that were confirmed by our GWAS and replication studies, the

associations at 10 SNPs within *BLK*, *IRF5*, *NMNAT2*, *STAT4*, *UBE2L3*, 1q25.1, 5q33.3 and 6q21 were also supported by our GWAS analysis with significant or suggestive evidence ($P < 0.01$) (**Supplementary Table 4a**, Group 1). However, our study did not provide evidence for the associations at the remaining 35 SNPs within *ATG5*, *BANK1*, *BLK*, *FCGR2A*, the MHC region, *ICA1*, *IRF5*, *ITGAM*, *PHRF1* (also known as *KIAA1542*), *LYN*, *PTPN22*, *PXK*, *SCUBE1*, *TNFAIP3* and *XKR6* ($P > 0.05$). Of these 35 SNPs, 24 SNPs (including 5 within the MHC region) are very rare (20 SNPs) or have low population frequency (4 SNPs) in the Chinese Han population (MAF < 5%) (**Supplementary Table 4a**, Group 2 and 3). The rareness or low frequency of these 24 SNPs may be responsible for the failure to detect their associations in the Chinese Han population. Of note, the remaining 11 SNPs, including 2 SNPs within the HLA class II region identified by the International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) study[7], have a similar MAF in both the Chinese Han and European populations but failed to show any associations in the Chinese Han population (**Supplementary Table 4a**, Group 4). Conversely, none of the top 13 MHC SNPs identified in our GWAS analysis showed significant association in a GWAS of European populations[6] (**Supplementary Table 4b**). Furthermore, of the 10 newly discovered non-MHC SNPs identified by this study, only rs10036748 within *TNIP1* showed significant association ($P = 10^{-4}$~10^{-3}) in European populations; this SNP also shows a much higher frequency (77%) in European populations than in Chinese populations (26%) (**Supplementary Table 4b**). None of the ten SNPs is unique to the Chinese populations, but some SNPs, such as rs13385731 (*RASGRP3*) and rs6590330 (*ETS1*), did show different allele frequencies between the two populations (**Supplementary Table 4b**).

The disparities of the GWAS findings between Chinese Han and European populations might suggest genetic heterogeneity of SLE between the two populations, especially in instances where risk alleles show a similar allele frequency in the two populations but association in only one population. Furthermore, although our study has >90% power to detect a SNP with MAF >0.2 and an OR >1.3, it may not have sufficient power to detect risk SNPs that are less frequent and/or associated with weaker genetic effects, for example, several risk SNPs identified in European populations that were found to be very rare in Chinese populations. The genetic effects (ORs) of some associations may also be overestimated by previous studies in European populations. Consequently, some of the association disparities may simply be due to insufficient power and/or chance. In addition, some of the disparities could also be due to different LD and haplotype patterns between the two populations. It is possible that, although existing in both the Chinese Han and European populations, some of the causal variants are not sufficiently tagged by any SNPs screened in our GWAS analysis and thus failed to be detected in our study of the Chinese Han population. Furthermore, differences in clinical sub-phenotypes and/or environment risk factors between the two populations might also explain these disparities. Additional studies using large samples of diverse ethnic populations will help clarify the basis for disparities of SLE association between different populations and provide a better understanding of the genetic basis of SLE.

In summary, we performed a large GWAS of SLE in a Chinese Han population and identified nine new susceptibility loci, as well as confirming seven previously reported risk loci. Our study not only advances our understanding of the genetic basis of SLE susceptibility but also highlights the genetic heterogeneity of disease susceptibility between different ethnic populations.

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 all study participants. This study was funded by the National Natural Science Foundation (30972727), the Key Project of National Natural Science Foundation (30530670 and 30830089), the Anhui Provincial Special Scientific Program (2007-7) and the Anhui Skin Genetic Study Innovative Research Team Program (TD200701).

AUTHOR CONTRIBUTIONS

X.-J.Z. conceived of this study and obtained financial support. X.-J.Z., J.-J.L. and S.Y. designed the study. J.-W.H., H.-F.Z., Y.C. and L.-D.S. participated in the design and were responsible for sample selection, genotyping and project management. D.-Q.Y., Z.H., Jin-Hua Xu, Z.-M.C., H.-F.X., H.F., Q.-J.L., Jian-Hua Xu, X.-P.L., Y.-F.P., D.-Q.D., F.-Q.Z., Z.-Z.Y., X.-Y.Z., Q.-W.W., F.H., L.M., X.-X.Z., W.-F.Z., F.G., P.-L.Z., Q.G., B.L., M.G., F.-L.X., J.-L.H., S.-X.L., H.L., Z.-X.W., C.-J.Y., P.-G.W., W.-M.Z., A.-P.Z., D.L. and Z.-F.Z. conducted sample selection and data management. W.-H.D., J.-Q.Y., S.-K.S., J.L., Y.-J.S., K.-J.Z., Yang Li, D.-Y.H. and Y.-M.L. undertook recruitment, collected phenotype data, undertook related data handling and calculation, managed recruitment and obtained biological samples. W.H., G.-P.Z., F.-S.Z., Y.-L.C., C.Q., C.Z., Z.Z., W.-S.L., Y.-Q.R., S.-Q.Z., M.S. and Y.W. performed genotyping analysis. H.-F.Z., X.-B.Z., Yi Li, H.Q.L. and F.-Y.Z. undertook data processing, statistical analysis and bioinformatics investigations. All the authors contributed to the final paper, with X.-J.Z., J.-J.L., S.Y., J.-W.H., H.-F.Z., Y.C. and L.-D.S playing key roles.

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

1. Hopkinson, N.D., Doherty, M. & Powell, R.J. Clinical features and race-specific incidence/prevalence rates of systemic lupus erythematosus in a geographically complete cohort of patients. *Ann. Rheum. Dis.* **53**, 675–680 (1994).
2. Wang, J. *et al.* Systemic lupus erythematosus: a genetic epidemiology study of 695 patients from China. *Arch. Dermatol. Res.* **298**, 485–491 (2007).
3. Lau, C.S., Yin, G. & Mok, M.Y. Ethnic and geographical differences in systemic lupus erythematosus: an overview. *Lupus* **15**, 715–719 (2006).
4. Zeng, Q.Y. *et al.* Rheumatic diseases in China. *Arthritis Res. Ther.* **10**, R17 (2008).
5. Danchenko, N., Satia, J.A. & Anthony, M.S. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* **15**, 308–318 (2006).
6. Hom, G. *et al.* Association of systemic lupus erythematosus with *C8orf13-BLK* and *ITGAM-ITGAX*. *N. Engl. J. Med.* **358**, 900–909 (2008).
7. Harley, J.B. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in *ITGAM, PXK, KIAA1542* and other loci. *Nat. Genet.* **40**, 204–210 (2008).
8. Kozyrev, S.V. *et al.* Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 211–216 (2008).
9. Graham, R.R. *et al.* Genetic variants near *TNFAIP3* on 6q23 are associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 1059–1061 (2008).
10. Yang, W. *et al.* Population differences in SLE susceptibility genes: *STAT4* and *BLK*, but not *PXK*, are associated with systemic lupus erythematosus in Hong Kong Chinese. *Genes Immun.* **10**, 219–226 (2009).
11. Cunninghame Graham, D.S. *et al.* Polymorphism at the TNF superfamily gene *TNFSF4* confers susceptibility to systemic lupus erythematosus. *Nat. Genet.* **40**, 83–89 (2008).
12. Beyaert, R., Heyninck, K. & Van Huffel, S. A20 and A20-binding proteins as cellular inhibitors of nuclear factor-kappa B-dependent gene expression and apoptosis. *Biochem. Pharmacol.* **60**, 1143–1151 (2000).
13. Nath, S.K. *et al.* A nonsynonymous functional variant in integrin-alpha(M) (encoded by *ITGAM*) is associated with systemic lupus erythematosus. *Nat. Genet.* **40**, 152–154 (2008).
14. Gallant, S. & Gilkeson, G. ETS transcription factors and regulation of immunity. *Arch. Immunol. Ther. Exp. (Warsz.)* **54**, 149–163 (2006).
15. Georgopoulos, K. *et al.* The Ikaros gene is required for the development of all lymphoid lineages. *Cell* **79**, 143–156 (1994).
16. Wojcik, H., Griffiths, E., Staggs, S., Hagman, J. & Winandy, S. Expression of a non-DNA-binding Ikaros isoform exclusively in B cells leads to autoimmunity but not leukemogenesis. *Eur. J. Immunol.* **37**, 1022–1032 (2007).
17. Coughlin, J.J., Stang, S.L., Dower, N.A. & Stone, J.C. RasGRP1 and RasGRP3 regulate B cell proliferation by facilitating B cell receptor-Ras signaling. *J. Immunol.* **175**, 7179–7184 (2005).
18. Lee, J. *et al.* pH-dependent internalization of muramyl peptides from early endosomes enables Nod1 and Nod2 signaling. *J. Biol. Chem.* **284**, 23818–23829 (2009).
19. Harley, I.T., Kaufman, K.M., Langefeld, C.D., Harley, J.B. & Kelly, J.A. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. *Nat. Rev. Genet.* **10**, 285–290 (2009).
20. Graham, R.R., Hom, G., Ortmann, W. & Behrens, T.W. Review of recent genome-wide association scans in lupus. *J. Intern. Med.* **265**, 680–688 (2009).
21. Nair, R.P. *et al.* Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. *Nat. Genet.* **41**, 199–204 (2009).

ONLINE METHODS

Subjects. All Chinese Han samples used in this study were obtained from doctors through collaboration with multiple hospitals in two geographic regions (central and southern China)[22–24]. Samples in the initial stage (1,099 SLE cases and 1,254 controls) were recruited from central China; samples in the replication studies were recruited from central (replication 1: 1,643 cases and 5,930 controls) and southern (replication 2: 1,509 cases and 1,120 controls) China. All affected individuals defined as cases were diagnosed using the revised criteria for the classification of SLE from the American College of Rheumatology[25]. Clinical information was collected from the affected individuals through a full clinical checkup by physician specialists. Additional demographic information was collected from both cases and controls through a structured questionnaire. All controls were clinically assessed to be without SLE, other autoimmune disorders, systemic disorders or family history of autoimmune disorders (including first-, second- and third-degree relatives). All participants provided written informed consent. The study was approved by the Institutional Ethical Committee of each hospital and was conducted according to Declaration of Helsinki principles.

EDTA anticoagulated venous blood samples were collected from all participants. Genomic DNA was extracted from peripheral blood lymphocytes by standard procedures using Flexi Gene DNA kits (Qiagen) and was diluted to working concentrations of 50 ng/μl for genome-wide genotyping and 15–20 ng/μl for the validation study.

Genotyping and quality controls in GWAS. The genome-wide genotyping analysis was conducted using Illumina Human610-Quad BeadChips at the Chinese National Human Genome Center at Shanghai, China. Genotyping was performed as previously described[26] according to the Infinium HD protocol from Illumina. Briefly, 200 ng of genomic DNA of each sample was whole-genome amplified, fragmented, precipitated and resuspended in the appropriate hybridization buffer. Denatured samples were hybridized on prepared Illumina Human610-Quad BeadChips. After hybridization, the BeadChip oligonucleotides were extended by a single labeled base, which was detected by fluorescence imaging with an Illumina Bead Array Reader. Normalized bead intensity data obtained for each sample were loaded into the Illumina BeadStudio 3.2 software, which converted the fluorescence intensities into SNP genotypes. The clustering of genotypes was carried out with the GenCall software version 6.2.0.4, which assigns a quality score to each locus and an individual genotype confidence score that is based on the distance of a genotype from the center of the nearest cluster.

All SNPs on the X, Y and mitochondrial chromosomes as well as copy number variation–related SNPs were excluded. SNPs were excluded if they showed either a call rate lower than 90% in cases or controls, a minor allele frequency <1% in the population or significant deviation from Hardy-Weinberg equilibrium in the controls ($P \leq 10^{-7}$). In the GWAS stage, 1,099 SLE cases and 1,254 controls were genotyped initially using the Illumina Human610-Quad BeadChip array. Fifty samples with genotyping call rate less than 98% were removed from analysis (33 cases and 17 controls). We then examined potential genetic relatedness based on pairwise identity by state for all the successfully genotyped samples using PLINK 1.06 software[27]. In total, 51 samples were removed due to sample duplications and genetic relatedness (19 cases and 32 controls). After quality control, the genotype data of 493,955 autosomal SNPs in 1,047 cases and 1,205 controls were used for the GWAS analysis. The remaining samples were subsequently assessed for population stratification using the software package EIGENSTRAT[28]. The original script from EIGENSTRAT was modified to extract the principal components for plotting (**Supplementary Fig. 1**).

SNP selection for replication. First, we chose SNPs with P value <0.01 after adjustment by gender and with P value $<5 \times 10^{-4}$ without adjustment by gender for further evaluation. Then, all the chosen SNPs were further analyzed by including another 4,001 samples as controls. These samples were genotyped in a series of GWAS of various diseases in the Chinese Han population, including psoriasis, leprosy, vitiligo and atoptic dermatitis. The improvement of association evidence by the inclusion of the 4,001 samples as controls was used as one of the criteria for selecting SNPs for validation analysis; however, because these additional controls had diseases other than SLE, these 4,001 samples were not included in the association analysis of the validation study or in the final joint association analysis of the combined samples. Additional criteria for SNPs selection were high MAF (>0.05 both in cases and controls), high P value of Hardy-Weinberg equilibrium ($P \geq 0.01$ in controls) and proximity to putative candidate genes (immune-related or involved in immune cell proliferation and differentiation) or known susceptibility loci for SLE. In total, 78 SNPs were selected for the validation analysis, including 12 SNPs within or close to known susceptibility loci for SLE identified by previous studies (including *STAT4*, *TNFSF4*, *BLK*, *IRF5*, *TNFAIP3*, 6q21 and 22q11.21) (**Table 2** and **Supplementary Table 3**).

Genotyping and quality controls in replication studies. Genotyping analysis of the SNPs selected for fast-track validation analysis was performed using the Sequenom MassArray system. Approximately 15 ng of genomic DNA was used to genotype each sample. Locus-specific PCR and detection primers were designed using the MassARRAY Assay Design 3.0 software (Sequenom) at the Key Laboratory of Dermatology at Anhui Medical University, Ministry of Education, China following the manufacturer's instructions. The DNA samples were amplified by multiplex PCR reactions, and the PCR products were then used for locus-specific single-base extension reactions. 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 by the MassARRAY Typer software (Sequenom). To evaluate the quality of the genotype data for the validation analysis, 100 randomly selected samples from the GWAS stage were re-genotyped in replication samples by using the Sequenom system. The concordance rate between the genotypes from the Illumina and the Sequenom analyses was >99%. We excluded SNPs with call rate lower than 90% in cases or controls and with Hardy-Weinberg equilibrium in the controls ($P \leq 0.001$). After quality control, 9 SNPs with low call rates were excluded, and 69 SNPs were left for final analysis. The remaining SNP cluster patterns of the genotyping data from the Illumina and Sequenom analyses were checked to confirm their good quality.

Statistical analysis. In GWAS stage, single-marker association analyses were performed using logistic regression with gender as a covariate. For the replication studies, 69 SNPs that passed quality control were analyzed in each replication study using logistic regression with gender as a covariate. The joint analysis of all combined samples was performed using logistic regression with gender and sample cohorts as covariates. The P values adjusted by gender were reported without correction for multiple testing. Multiple logistic regression analysis was used to test the independence of multiple associations of top 13 SNPs within the MHC region. We used logistic regression analysis to test the independence of association evidence between MHC and non-MHC SNPs by controlling for the effect of rs9271100 and rs3997854 (adjusted for gender). To test for any two-locus interactions, we computed a logistic regression model for each pair of the top SNPs of the 16 non-MHC loci and 2 MHC SNPs (rs9271100 and rs3997854). Recombination plots of each newly discovered susceptibility locus were generated using the information from the HapMap project (CHB and JPT samples).

22. Chu, J.Y. *et al.* Genetic relationship of populations in China. *Proc. Natl. Acad. Sci. USA* **95**, 11763–11768 (1998).
23. Ke, Y. *et al.* Y-chromosome haplotype distribution in Han Chinese populations and modern human origin in East Asians. *Sci. China C Life Sci.* **44**, 225–232 (2001).
24. Shi, Y. *et al.* Genetic structure adds power to detect schizophrenia susceptibility at *SLIT3* in the Chinese Han population. *Genome Res.* **14**, 1345–1349 (2004).
25. Hochberg, M.C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1725 (1997).
26. 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).
27. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
28. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).

doi:10.1038/ng.472