Chinese-European SLE GWAS meta-analysis findings include ten new loci and a genetic basis for increased non-European prevalence

David L Morris[1,19] , Yujun Sheng[2,3,4,19], Yan Zhang[5,19], Yong–Fei Wang[5,], Zhengwei Zhu[2,3] Philip Tombleson[1,], Lingyan Chen[1], Deborah S Cunninghame Graham[1], James Bentham[6], Amy L Roberts[1], Ruoyan Chen[5], Xianbo Zuo[2,3], Tingyou Wang[5], Leilei Wen[2,3], Chao Yang[2,3], Lu Liu[2,3], Lulu Yang[2,3], Feng Li[2,3], Yuanbo Huang[2,3], Xianyong Yin[2,3], Sen Yang[2,3], Lars Rönnblom[7], Barbara G Fürnrohr[8–11], Reinhard E Voll[8,9,12,13], Georg Schett[8,9], Nathalie Costedoat–Chalumeau[14], Patrick M Gaffney[15], Yu Lung Lau[5,16], Xuejun Zhang[2,3,17], Wanling Yang[5,20], Yong Cui[2,3,4,20], Timothy J Vyse[1,19,20]

[19] These authors contributed equally to this work
[20] These authors jointly supervised the work

[1] Division of Genetics and Molecular Medicine, King's College London, UK.
[2] Department of Dermatology, NO. 1 Hospital, Anhui Medical University, Hefei, Anhui, China.
[3] Key Laboratory of Dermatology, Ministry of Education, Anhui Medical University, Hefei, Anhui, China.
[4] Department of Dermatology, China–Japan Friendship Hospital, Beijing, China.
[5] Department of Paediatrics and Adolescent Medicine, LKS Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.
[6] Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, UK.
[7] Department of Medical Sciences, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[8] Department of Internal Medicine 3, Ulmenweg 18, University of Erlangen–Nuremberg, 91054 Erlangen, Germany
[9] Institute for Clinical Immunology, Ulmenweg 18, University of Erlangen–Nuremberg, 91054 Erlangen, Germany
[10] Division of Genetic Epidemiology, Innrain 80/IV, Medical University Innsbruck, 6020 Innsbruck, Austria
[11] Division of Biological Chemistry, Innrain 80/IV, Medical University Innsbruck, 6020 Innsbruck, Austria
[12] Department of Rheumatology, University Hospital Freiburg, Freiburg, Germany, Germany
[13] Clinical Immunology & Centre of Chronic Immunodeficiency, University Hospital Freiburg, Freiburg, Germany, Germany
[14] AP–HP, Hôpital Cochin, Centre de référence maladies auto–immunes et systémiques rares, Paris, France; Université Paris Descartes–Sorbonne Paris Cité, Paris, France.
[15] Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104.
[16] The University of Hong Kong Shenzhen Hospital, ShenZhen, China.
[17] Department of Dermatology, Huashan Hospital of Fudan University, Shanghai, China.
[18] Division of Immunology, Infection and Inflammatory Disease, King's College London, UK.

Correspondence should be addressed to TJV (timothy.vyse@kcl.ac.uk), YC (wuhucuiyong@vip.163.com) and WY (yangwl@hku.hk).

Systemic lupus erythematosus (SLE; OMIM 152700) is a genetically complex autoimmune disease. Over 50 loci have been found to be robustly associated by GWAS in single ethnicities. We combined three GWAS' datasets from two ethnicities: Chinese (1,659 cases and 3,398 controls) and European (4,036 cases and 6,959 controls). A meta–analysis of these studies found that over half of the published SLE genetic associations are present in both populations. A replication study in the Chinese (3,043 cases and 5,074 controls) and Europeans (2,643 cases and 9,032 controls) found 10 novel SLE loci. Our study reveals further evidence that the majority of genetic polymorphisms exerting risk for SLE are contained within the same regions across the Chinese and European populations. Furthermore, comparing risk allele frequencies and genetic risk scores suggests that the increased prevalence of SLE in non–Europeans (including Asians) has a genetic basis.

SLE is a highly complex disease, with occurrence heavily influenced by genetics (heritability=66%[1]). SLE incidence varies markedly across populations, with Europeans showing 3–4 fold lower prevalence compared with individuals of African or Asian ancestry[2]. In recent years, our understanding of SLE genetic aetiology has been transformed by GWAS, with the largest study in Europeans (4,036 cases and 6,959 controls)[3] finding evidence of association at 41 autosomal loci. Meanwhile, there have been two published GWAS[4,5] in Chinese populations and follow up studies in Asians[6-10] that found association at 31 loci, 11 of which are not published in Europeans. Thus 52 SLE disease susceptibility autosomal loci have been mapped by GWAS in these two populations.

While fine mapping of a selected number of known SLE associated loci[11-13] has been successfully undertaken by combining genetic results obtained from association mapping in different populations, to date transancestral approaches have not been employed at genome-wide level in SLE. Studies of other diseases[14] have also shown the benefit of comparing data from differing ancestries to exploit differences in LD.

Our initial objective was to compare observed genetic association signals across the genome in Chinese and Europeans. To provide additional power to identify potentially novel SLE associated loci we imputed each GWAS [A European study: 4,036 cases and 6,959 Controls[3] (λ_{GC}=1.16 with $\lambda_{1,000}$=1.02); a study from Anhui province in mainland China: 1,047 cases and 1,205 Controls[4] (λ_{GC}=1.05) and a study from Hong Kong: 612 cases and 2,193 Controls[5,7] (λ_{GC}=1.04)] to the density of the 1000 Genomes (1KG) data (see Online Methods). Analyses of association results in each population suggested that SLE susceptibility loci were shared extensively. Manhattan plots showing these similarities are presented in Fig. 1, where it can be seen that the association signals are mostly mirrored between populations. Details of the association data for individual SNPs are presented in Supplementary Table 1. Comparing the published genome-wide significant allelic associations in SLE, we see that many of the alleles hitherto thought to be associated with SLE in only one population have evidence for association in both European and Chinese SLE. Ranking genomic regions based on strength of association, we also find a significant correlation (P=2.7×10^{-9}, Kendall's Tau=0.08, see methods) between the two populations' GWAS. These observations suggested that combining GWAS data in a meta–analysis would likely yield novel association signals. Fig. 1b shows a Manhattan plot of the GWAS meta–analysis results, which included three associations in novel loci (rs17603856 6p23; rs1887428 [9p24]; rs669763 [16q13]) with genome wide level of significance (P<5×10^{-08}). In addition, it can be seen in this Figure that the Major Histocompatibility Complex (MHC) and to a lesser extent the *IRF5* locus on chromosome 7, exhibit significant trans–ancestral heterogeneity.

We then carried out a two–stage replication study, incorporating rs17603856, rs1887428 and rs669763. The 1KG-imputed data were scanned for association at loci independent of those previously published and excluding the MHC. A total of 66 SNPs at 56 loci (Online Methods describes SNP selection) were successfully genotyped in a further 3,043 cases and 5,074 controls of Chinese ancestry recruited from Anhui Province. Eighteen of these SNPs (at 17

independent loci) showed association in this replication study, passing a false discovery rate (FDR) of 0.01. These included rs17603856 and rs1887428 but not rs669763, which failed quality control. We then genotyped these 18 SNPs in a European replication cohort, comprising 1,478 cases and 6,925 controls[3]. Data from an additional European–American GWAS (1,165 independent cases and 2,107 controls) were also included in this final analysis[15] (Supplementary Table 2a). Of the 18 candidate SNPs, 11 passed a standard genome–wide level of significance ($P<5\times10^{-08}$) in the combined meta-analysis (11,381 cases and 24,463 controls) of all three main GWAS and the three replication studies (Table 1; forest plots are presented in Supplementary Fig. 1). The strongest association signal following this meta-analysis was rs1887428 (9p24). Additional statistically significant associations were found at rs34889541 (1q31.3), rs2297550 (1q32.1), rs6762714 (3q28), rs17603856 (6p23), rs597325 (6q15), rs73135369 (7q11.23), rs494003 (11q13.1) and rs1170426 (16q22.1), while two SNPs at 2p23.1 (rs1732199 and rs7579944) were replicated as being independently associated (see Online Methods and Table 1). The full set of results for the 18 candidate markers can be seen in Supplementary Table 2.

In order to highlight potential causal genes at the ten newly described susceptibility loci, the associated SNPs at each locus were tested for correlation with *cis*–acting gene expression in *ex vivo* naïve CD4+ T cells and CD14 monocytes in both Asian and European population data[16], and B cells, T cells and monocytes (stimulated and naïve) in Europeans only[17]. We calculated Regulatory Trait Concordance (RTC) scores[18] (see Online Methods) to test the relationship between eQTLs driven by disease-associated alleles, and other, potentially stronger eQTLs, which we identified at each locus. Supplementary Table 3 and Supplementary Fig. 2 present results for this analysis in all cell types in circumstances where eQTLs were found in at least one cell type/population. The eQTLs were consistent across cell type and population for *LBH* (rs19991732), *CTSW* (rs494003), *RNASEH2C* (rs494003) and *ZFP90* (rs1170426), with carriage of the SLE risk allele correlating with reduced expression (except in LPS stimulated monocytes for *RNASEH2C* where the eQTL results

were not significant and the RTC scores were very low). The SNP rs2297550 was found to be an eQTL for *IKBKE* with the SLE risk allele correlated with reduced expression in T cells, IFN stimulated monocytes, B cells and NK cells, but increased expression in monocytes.

We integrated the results of the eQTL analyses with an *in silico* survey of murine phenotype data resulting from knockouts of genes within the associated SLE loci (Table 2)[19-28]. These lines of evidence point to one likely causal gene at some loci, *IKBKE* and *JAK2* for example. In other instances, we found evidence that supports the role of multiple genes as candidates at a given locus; for example, *CTSW / RNASEH2C* and *CDH1 / ZFP90*. Locus Zoom[29] plots, using the European and meta-analysed Chinese data, for all 10 loci can be seen in Supplementary Fig. 3, which facilitate a comparison of the alignment of the association signals in the two populations. The potential roles of the putative causal genes at the loci mapped in this study are described in Supplementary Table 4.

The level of shared association we noted in our initial combination of the two ethnicities' GWAS was exploited further using fine mapping analyses of all published associated loci (Supplementary Table 1) and the loci we present as novel in this paper. We derived Bayesian credibility sets (C.S.) in each population for the most likely causal variants using a previously published approach[30-32]. We report the intersection of these sets (see methods) and Supplementary Fig. 4 displays the observed cumulative distribution for the number of SNPs in the intersection over a range of levels. Using the least stringent criterion (75% C.S.), 80% of the mapped loci had sets identifying 10 or less likely causal SNPs. Using a very rigorous criterion (99% C.S.), seven of the loci comprised less than 10 SNPs (Supplementary Table 5). *STAT4* is a good example of the co–localisation of signals from each ancestry, which we show in detail in Fig. 2. In contrast we show two examples in the Figure where the association arises in one population only: *IRF7* (European) and *ELF1* (Chinese). In each case it is evident that the likely explanation for the discrepant association signal is population-specific allele frequency differences within the credible SNP set. Supplementary Fig. 5 displays fine mapping data for the novel loci.

We downloaded epigenetic data covering each of the novel 10 associated loci identified by the meta-analysis (Table 1) from the RoadMap consortium for all blood cell types[33]. This was performed for all SNPs within the C.S. at each locus. Fig. 3 displays the results for SNPs at three loci, showing the level of RNA expression (RNA–seq), accessibility to DNAse, histone modification by acetylation (H3K27ac, H3K9ac) and histone modification by methylation (H3K27me3, H3K9me3). Supplementary Fig. 6 displays results for the other seven SNPs. The histone marks were selected to indicate the activation status of promoter and enhancer regions and regions of repression. This epigenetic annotation provides an interesting point of comparison with the eQTL results. Two intense histone acetylation peaks were observed around the associated SNPs rs2297550 (*IKBKE*) and rs1887428 (*JAK2*), yet only the variant in *IKBKE* showed a significant eQTL in the cells examined. Although we did find a significant eQTL for rs1887428 with *JAK2* in monocytes, the RTC scores were low (<0.4). At SNPs, rs34889541 (*CD45*) and rs597325 (*BACH2*), there was local evidence of histone acetylation in lymphocytes, but the two SNPs were not significant eQTLs. In contrast, rs1170426 (*ZFP90*) was a very significant eQTL, but the region around the associated SNP showed little evidence of regulatory function. However there was strong evidence of epigenetic effects at other SNPs contained in the *ZFP90* C.S. Some of the discrepancies between eQTL and epigenetic annotation likely represent the limited set of activation states (and perhaps samples sizes) of primary immune cells that have been subject to eQTL investigation.

The amount of shared risk effects between the Chinese and European populations was further investigated with a co-heritability analysis using LD score regression[34] (see methods) which showed a significant (P=4.0×10^{-03}, r_g=0.51) correlation between the two populations, with this correlation being stronger (P = 4.88 × 10^{-05}, r_g=0.62) after removing the MHC which emphasises its heterogeneity (Fig. 1b). These results beg the question: does the higher prevalence of SLE in Asians (compared with Europeans) have a genetic basis? We observed that on average the risk allele frequencies (RAF) in Chinese were significantly

higher than those in Europeans in the respective GWAS controls (paired *t*–test, P=0.02, Supplementary Fig. 7a) while the effect sizes (ORs) were not statistically different (P=0.47, Supplementary Fig. 7b). We also compared the genetic risk scores (GRS) – the joint effect of ORs and RAFs – between populations in data from 1KG (Phase III) (Fig. 4) and between the Chinese and European GWAS controls (Supplementary Fig. 8a). The GRS for SLE in East Asians (EAS) was significantly higher than that in Europeans (EUR) in the 1KG data [fold (EAS/EUR)=1.27, $P=4.99\times10^{-179}$; EUR=7.38 95%C.I. 7.31–7.45; EAS=9.35, 95%C.I. 9.27–9.43]. There was a similar difference in score between the GWAS controls [fold (Chinese/European) = 1.28, $P=1.00\times10^{-797}$; European=7.42, 95%C.I. 7.40–7.44; Chinese=9.51 95%C.I. 9.46–9.55]. With more associations to be identified in future studies, especially with increased power in non-European populations including East Asians, the difference in genetic predisposition between populations revealed by GWAS might further increase. We note that an analyses of chip heritability (using all genotyped SNPs to calculate heritability explained, see methods) in both the CHN and EUR data resulted in 28% (s.e.=2.6%) explained in CHN and 27% (s.e.=1.0%) explained in EUR.

Furthermore, we see a correlation between the GRS across all five major 1KG super–populations and rank of the prevalence[2] (see methods) of SLE (Fig. 4). A *t*–test on mean GRS between each pair of population data was highly significant ($P<10^{-16}$) for all pairs except AMR versus SAS (P=0.67) and a linear model with rank of prevalence predicting the GRS was significant ($P<10^{-16}$, r^2=0.39). We have excluded the MHC from this analysis due to the difficulty of defining the best model of association in this region, due to the extensive LD and limited genotyping of SNPs and classical HLA in both populations.

The increased genetic load in Chinese would help explain the continued increased prevalence in Asians following migration to Western locations[2]. We acknowledge that the trends we observe are a snapshot, as all available genotyped SNPs explained <30% disease heritability, and the comparison of GRS may not be a full reflection of genetic risk amongst the populations. A more detailed study of the increased prevalence of SLE in

Asians, and Africans, will require extensive comparisons of genetic and environmental data, including generation of DNA sequence data to exclude the European bias in genotyping arrays.

URLs. Department of Twin Research, King's College London**,** Twins–UK samples, http://www.twinsuk.ac.uk; Ingenuity Pathway Analysis, http://www.ingenuity.com/; Immunobase, http://www.immunobase.org. Systems Biology and Complex Disease Genetics, http://insidegen.com.

RoadMap data (http://egg2.wustl.edu/roadmap/data/byFileType/signal/consolidatedImputed/

Data Access

All 1KG imputed summary statistics are available at http://insidegen.com/insidegen–LUPUS–data.html

Acknowledgements

PT is employed by the Biomedical Research Centre. LC was funded by the China Scholarship Council, number 201406380127. The research was funded/supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London.

TJV was awarded funding to carry out genotyping and analysis from George Koukis and an Arthritis Research UK Special Strategic Award and by a Wellcome Trust grant (ref. 085492)

TJV was awarded funding by the MRC L002604/1 "Functional genomics of SLE: A transancestral approach"

For the replication study in Europeans, samples were provided by the Swedish SLE Network. Replication genotyping was performed by the SNP&SEQ Technology Platform in

Uppsala, which is part of the Swedish National Genomics Infrastructure (NGI) hosted by Science for Life Laboratory.

The controls for the European GWAS and replication were obtained from dbGaP under: accession phs000187.v1 [a study sponsored by the National Institute on Aging (grant numbers U01AG009740, RC2AG036495, and RC4AG039029) and was conducted by the University of Michigan]; a melanoma study data under accession number phs000187.v1.p1; a blood clotting study under accession number phs000304.v1.p1; a prostate cancer study data obtained from dbGaP under accession phs000207v1.

The French cases for the European replication study were provided by Felix Ackermann, Zahir Amoura, Bouchra Asli, Leonardo Astudillo, Olivier Aumaître, Cristina Belizna, Nadia Belmatoug, Olivier Benveniste, Audrey Benyamine, Holly Bezanahary, Patrick Blanco, Olivier Bletry, Pierre Bourgeois, Benoit Brihaye, Patrice Cacoub, Emmanuel Chatelus, Judith Cohen–Bittan, Richard Damade, Eric Daugas, Christian De–Gennes, Jean–François Delfraissy, Aurélien Delluc, Helene Desmurs–Clavel, Pierre Duhaut, Alain Dupuy, Isabelle Durieu, Hang–Korng Ea, Olivier Fain, Dominique Farge, Christian Funck–Brentano, Camille Frances, Lionel Galicier, Frédérique Gandjbakhch, Justine Gellen–Dautremer, Bertrand Godeau, Cécile Goujard, Catherine Grandpeix, Claire Grange, Gaëlle Guettrot, Loïc Guillevin, Eric Hachulla, Jean–Robert Harle, Julien Haroche, Pierre Hausfater, Jean–Sébastien Hulot, Moez Jallouli, Jean Jouquan, Gilles Kaplanski, Homa Keshtmand, Mehdi Khellaf, Olivier Lambotte, David Launay, Philippe Lechat, Du Le Thi Huong, Véronique Le–Guern, Jean–Emmanuel Kahn, Gaëlle Leroux, Hervé Levesque, Olivier Lidove, Nicolas Limal, Frédéric Lioté, Eric Liozon, Kim LY, Matthieu Mahevas, Kubéraka Mariampillai, Xavier Mariette, Alexis Mathian, Karin Mazodier, Marc Michel, Nathalie Morel, Luc Mouthon, Jacques Ninet, Eric Oksenhendler, Thomas Papo, Jean–Luc Pellegrin, Laurent Perard, Olivier Peyr, Anne–Marie Piette, Jean–Charles Piette, Vincent Poindron, Jacques Pourrat, Fabienne Roux, David Saadoun, Karim Sacre, Sabrinel Sahali, Laurent Sailler, Bernadette Saint–Marcoux, Françoise Sarrot–Reynauld, Yoland Schoindre, Jérémie Sellam, Damien

Sène, Jacques Serratrice, Pascal Seve, Jean Sibilia, Clause Simon, Amar Smail, Christelle Sordet, Jérome Stirnemann, Salim Trad, Jean–François Viallard, Elisabeth Vidal, Bertrand Wechsler, Pierre–Jean Weiller, Noël Zahr.

The Chinese GWAS data was funded through Key Basic Research Program of China (2014CB541901, 2012CB722404 and 2011CB512103), the National Natural Science Foundation of China (81402590, 81371722, 81320108016 and 81171505), the Research Project of Chinese Ministry of Education (No.213018A), the Program for New Century Excellent Talents in University (NCET–12–0600) and the Natural Science Fund of Anhui province (1408085MKL27).

YLL thank generous donations from Shun Tak District Min Yuen Tong of Hong Kong that partially supported the SLE GWAS in Hong Kong. YLL and WY thank the doctors who contributed SLE cases and colleagues in LKS Faculty of Medicine, University of Hong Kong who provided controls used in the GWAS. WY and YLL also thank support from Research Grant Council of the Hong Kong Government (GRF HKU783813M, HKU 784611M, 17125114 and HKU 770411M). YZ thanks the Health and Medical Research Fund (Ref:12133701) from the Food and Health Bureau, Hong Kong.

We thank Towfique Raj and Phil De Jager for contributing gene expression data (CD4 T cells and CD14/16 monocytes in Asian and European populations). These gene expression data are deposited in the National Center for Biotechnology Information Gene Expression Omnibus under accession no. GSE56035. We thank Ben Fairfax and Julian Knight for contributing the gene expression data on NK cells, naïve monocytes, monocytes stimulated by LPS (harvested after 2 hours and 24 hours), IFN and B cells. We thank Samuel Daffern for downloading the ChIP–seq data in contribution to the epigenetic analysis.

Author contributions

YFW, ZZ and PT contributed equally to this work.

TJV, XJZ, YC, YLL and WY supervised the study. Z‐WZ, L‐LW, CY, LL, L‐LY, FL, Y‐BH and SY performed sample selection and data management, undertook recruitment, collected phenotype data for the Anhui Chinese data. LR, BGF, BEV, NC–C and PMG performed sample selection and data management, undertook recruitment, collected phenotype data for the European data. ALR worked on both the Chinese and European replication studies' genotyping. DLM, YJS, YZ and YFW carried out statistical analysis of the GWAS data. DLM and PT carried out the 1000 genomes imputation in the European GWAS. RC and TW carried out the 1000 genomes imputation in the Anhui and Hong Kong Chinese GWAS. DLM, PT, XBZ, YFW and YZ carried out statistical analysis for the meta–analysis of the 1000 genomes imputed data. DLM, YJS and YZ designed the replication studies' chips. BGF and REV contributed data to the European replication cohort. DM and JB performed quality control on the European data for the replication study. DM analyzed the European replication data. DM, YJS and YZ analyzed the Anhui replication data. YFW and DM designed and performed genetic risk score comparison between the populations. YFW performed the LD score regression analysis. DM and LY carried out the eQTL analysis. DLM and DSCG carried out the epigenetic analysis. DLM, TJV, DSCG, XJZ, YC YJS, and WY wrote the manuscript. All authors have read and contributed to the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

1. Lawrence, J.S., Martins, C.L. & Drake, G.L. A Family Survey of Lupus-Erythematosus .1. Heritability. *Journal of Rheumatology* **14**, 913-921 (1987).
2. Danchenko, N., Satia, J. & Anthony, M. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* **15**, 308-318 (2006).
3. Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nature Genetics* **47**, 1457-1464 (2015).
4. Han, J.W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nature Genetics* **41**, 1234-1237 (2009).
5. Yang, W.L. *et al.* Genome-Wide Association Study in Asian Populations Identifies Variants in ETS1 and WDFY4 Associated with Systemic Lupus Erythematosus. *Plos Genetics* **6**, e1000841 (2010).
6. Sheng, Y.J. *et al.* Follow-up study identifies two novel susceptibility loci *PRKCB* and 8p11.21 for systemic lupus erythematosus. *Rheumatology* **50**, 682-688 (2011).
7. Yang, W. *et al.* Meta-analysis followed by replication identifies loci in or near *CDKN1B, TET3, CD80, DRAM1*, and *ARID5B* as associated with systemic lupus erythematosus in Asians. *American Journal of Human Genetics* **92**, 41-51 (2013).
8. Li, Y. *et al.* Association analyses identifying two common susceptibility loci shared by psoriasis and systemic lupus erythematosus in the Chinese Han population. *J Med Genet* **50**, 812-8 (2013).
9. Sheng, Y.J. *et al.* Association analyses confirm five susceptibility loci for systemic lupus erythematosus in the Han Chinese population. *Arthritis Res Ther* **17**, 85 (2015).
10. Yang, J. *et al. ELF1* is associated with systemic lupus erythematosus in Asian populations. *Human Molecular Genetics* **20**, 601-607 (2011).
11. Fernando, M.M.A. *et al.* Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at *MSH5, HLA-DPB1* and *HLA-G*. *Annals of the Rheumatic Diseases* **71**, 777-784 (2012).
12. Manku, H. *et al.* Trans-Ancestral Studies Fine Map the SLE-Susceptibility Locus *TNFSF4*. *Plos Genetics* **9**, e1003554 (2013).
13. Adrianto, I. *et al.* Association of a functional variant downstream of *TNFAIP3* with systemic lupus erythematosus. *Nature Genetics* **43**, 253-258 (2011).
14. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nature Genetics* **46**, 234-244 (2014).
15. Hom, G. *et al.* Association of systemic lupus erythematosus with C8orf13-*BLK* and *ITGAM-ITGAX*. *N Engl J Med* **358**, 900-909 (2008).
16. Raj, T. *et al.* Polarization of the Effects of Autoimmune and Neurodegenerative Risk Alleles in Leukocytes. *Science* **344**, 519-523 (2014).
17. Fairfax, B.P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nature Genetics* **44**, 502-510 (2012).
18. Nica, A.C. *et al.* Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. *Plos Genetics* **6**, e1000895 (2010).
19. Jury, E.C., Kabouridis, P.S., Flores-Borja, F., Mageed, R.A. & Isenberg, D.A. Altered lipid raft-associated signaling and ganglioside expression in T lymphocytes from patients with systemic lupus erythematosus. *Journal of Clinical Investigation* **113**, 1176-1187 (2004).
20. Wang, C. *et al.* Contribution of *IKBKE* and *IFIH1* gene variants to SLE susceptibility. *Genes and Immunity* **14**, 217-222 (2013).

21. Kim, K. *et al.* High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Annals of the Rheumatic Diseases* **74**, 10.1136/annrheumdis-2013-204749 (2015).
22. Plagnol, V. *et al.* Genome-Wide Association Analysis of Autoantibody Positivity in Type 1 Diabetes Cases. *Plos Genetics* **7**, e1002216 (2011).
23. Medici, M. *et al.* Identification of Novel Genetic Loci Associated with Thyroid Peroxidase Antibodies and Clinical Thyroid Disease. *Plos Genetics* **10**, e1004123 (2014).
24. Tantin, D., Tussie-Luna, M.I., Roy, A.L. & Sharp, P.A. Regulation of immunoglobulin promoter activity by TFII-I class transcription factors. *Journal of Biological Chemistry* **279**, 5460-5469 (2004).
25. Lu, L.D. *et al.* Depletion of Autoreactive Plasma Cells and Treatment of Lupus Nephritis in Mice Using CEP-33779, a Novel, Orally Active, Selective Inhibitor of JAK2. *Journal of Immunology* **187**, 3840-3853 (2011).
26. Crow, Y.J. *et al.* Characterization of Human Disease Phenotypes Associated with Mutations in *TREX1, RNASEH2A, RNASEH2B, RNASEH2C, SAMHD1, ADAR, and IFIH1*. *American Journal of Medical Genetics Part A* **167**, 296-312 (2015).
27. Gunther, C. *et al.* Defective removal of ribonucleotides from DNA promotes systemic autoimmunity. *Journal of Clinical Investigation* **125**, 413-424 (2015).
28. Huang, C. *et al.* Cutting Edge: a novel, human-specific interacting protein couples *FOXP3* to a chromatin-remodeling complex that contains *KAP1/TRIM28*. *J Immunol* **190**, 4470-3 (2013).
29. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
30. Beecham, A.H. *et al.* Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis. *Nature Genetics* **45**, 1353-1360 (2013).
31. Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nature Genetics* **44**, 1294-1301 (2012).
32. Gaulton, K.J. *et al.* Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci. *Nat Genet* **47**, 1415-1425 (2015).
33. Bernstein, B.E. *et al.* The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology* **28**, 1045-1048 (2010).
34. Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47**, 1236-1241 (2015).

Figure Legends

Figure 1 (a) Manhattan plot of the European and Chinese (meta-analysis of two Chinese GWAS) GWASs. The $-\log_{10}$ *P*-values for Europeans are shown in light blue with the $\log_{10}$ *P*–values for the Chinese in pink. The 52 loci with published evidence of association are highlighted in blue and red while the 10 novel loci indented as associated from this study are highlighted in black. (b) $-\text{Log}_{10}$ *P*-values for meta-analysis (Europeans combined with Chinese GWAS) in grey with the $\log_{10}$ *P*-values for a test of heterogeneity between the European and Chinese GWAS in brown. The 52 loci with published evidence of association are highlighted in black (meta *P*-values) and dark brown (heterogeneity test) while the 10 novel loci indented as associated from this study are highlighted in black.

Figure 2 Fine mapping examples for *STAT4*, *IRF7* and *ELF1*. The upper plots are LocusZoom plots showing association significance [$-\log_{10}$(*P*-value)] and local LD (colour coded). Circular points represent SNPs contained within the credibility sets and square points represent SNPs not contained in the sets. The lower plots display the minor allele frequencies (MAF) for all the SNPs in the intersection of the European (EUR) and Chinese (CHN) credibility sets. The MAF is plotted in red. The SNPs with highest posterior probability within the intersection of C.I.s are highlighted in BLUE (highest posterior probability in the EUR data), RED (highest posterior probability in the CHN data) and BLACK (highest posterior probability in the CHN-EUR Meta data). The C.S. coverage (99% for STAT4, 90% for *IRF7* and *ELF1*) was chosen as the maximum coverage that included a maximum of 30 SNPs.

Figure 3 3D enrichment plots depict epigenetic modifications +/–50bp overlapping all SNPs in the Credibility Sets for the 11 novel associated SNPs. The SNPs are shown as individual tracks on the x-axis with the SNP used in the replication study marked (*) and the SNP that shows the best evidence for co-localisation with the most prominent epigenetic mark (#).

Other SNP identities are listed in Supplementary Table 6. The z–axis represents $\log_{10}$ *P*-value against the null hypothesis that peak intensity arises from the control distribution. The z-axis is truncated at a lower level of ($P<10^{-04}$). Each novel associated locus has a separate panel with results for RNA expression (RNA-seq), accessibility to DNAse, histone modification by acetylation (H3K27ac, H3K9ac) and histone modification by methylation (H3K27me3, H3K9me3) over 27 immune cells. The data from the blood cell types are consistently ordered on the y-axis according to the annotation to the right of the figure: categories 1–9 innate response immune cells; categories 10–24 adaptive response immune cell types (categories 10–11 B-cells; categories 12–24 T-cells) and then categories 25–27 cell lines.

Figure 4 Box plots of genetic risk score (GRS) for across the five major population groups. These are standard box plots with medians, interquartile ranges and whiskers at 1.5 of the interquartile range (Tukey box plots) displayed. (EUR European N=498, AMR Amerindian N=347, SAS South Asian N=487, EAS East Asian N=503, AFR African N=657) in the 1,000 Genomes phase III release. The dotted line represented the increase in prevalence with the rank order presented (R1 representing the lowest prevalence and R4 the highest).

Table 1: Summary of statistical associations for new loci

SNP	Chr	Position	Risk allele[a]	Chinese MAF[b]		European MAF[b]		Chinese 4,702 cases 8,472 controls		European 6,679 cases 15,991 controls		Meta all 11,381 cases 24,463 controls		Gene[d]	Association with other Autoimmune diseases[f]
				Case	Control	Case	Control	OR (95% CI)[c]	*P*	OR (95% CI)	*P*	OR (95% CI)	*P*		
rs34889541	1q31.3	198,594,769	G	0.126	0.14	0.058	0.07	0.78 (0.72 – 0.84)	2.96E–10	0.86 (0.79 – 0.94)	5.34E–04	0.81 (0.76 – 0.86)	2.44E–12	*PTPRC (CD45)*	*MS, RA, T1D*
rs2297550	1q32.1	206,643,772	G	0.577	0.546	0.14	0.12	1.14 (1.08 – 1.20)	1.73E–07	1.18 (1.09 – 1.27)	1.43E–05	1.16 (1.11 – 1.21)	1.31E–11	*IKBKE*	
rs7579944	2p23.1	30,445,026	C	0.59	0.641	0.338	0.366	0.87 (0.82 – 0.92)	5.52E–06	0.92 (0.88 – 0.96)	3.96E–05	0.90 (0.87 – 0.93)	1.41E–09	*LBH*[e]	*RA, AA, IBD, NAR, PSC, SJO, SSC, VIT*
rs17321999	2p23.1	30,479,857	C	0.16	0.164	0.161	0.191	0.82 (0.77 – 0.88)	9.55E–09	0.84 (0.79 – 0.89)	2.26E–09	0.83 (0.79 – 0.87)	2.22E–16	*LBH*[e]	*RA, AA, IBD, NAR, PSC, SJO, SSC, VIT*
rs6762714	3q28	188,470,238	T	0.848	0.825	0.421	0.392	1.20 (1.12 – 1.29)	5.56E–07	1.14 (1.09 – 1.19)	7.97E–10	1.16 (1.12 – 1.20)	4.00E–15	*LPP, TPRG1–AS1*	*ATD, CEL VIT*
rs17603856	6p23	16,630,898	T	0.221	0.222	0.325	0.355	0.86 (0.80 – 0.92)	1.61E–05	0.89 (0.85 – 0.93)	3.34E–08	0.88 (0.85 – 0.91)	3.27E–12	*ATXN1*	
rs597325	6q15	91,002,494	G	0.485	0.52	0.357	0.385	0.84 (0.80 – 0.89)	1.05E–10	0.92 (0.88 – 0.96)	2.65E–04	0.89 (0.86 – 0.92)	4.03E–12	*BACH2*	*AS, ATD, CEL, CRO, MS, T1D IBD, PSC, VIT*
rs73135369	7q11.23	73,940,978	C	0.107	0.076	0.028	0.022	1.38 (1.26 – 1.51)	7.33E–13	1.20 (1.05 – 1.38)	9.00E–03	1.32 (1.23 – 1.42)	8.77E–14	*GTF2IRD1–GTF2I*	
rs1887428	9p24	4,984,530	G	0.372	0.346	0.398	0.373	1.24 (1.17 – 1.31)	4.49E–14	1.11 (1.06 – 1.16)	1.25E–06	1.16 (1.12 – 1.20)	2.19E–17	*JAK2*	*CRO, UC IBD, PSC, VIT*
rs494003	11q13.1	65,542,298	A	0.116	0.117	0.213	0.19	1.16 (1.06 – 1.27)	8.38E–04	1.13 (1.07 – 1.19)	1.68E–06	1.14 (1.09 – 1.19)	5.81E–09	*RNASEH2C*	*CRO, IBD*
rs1170426	16q22.1	68,603,798	C	0.198	0.176	0.252	0.235	1.20 (1.12 – 1.28)	4.36E–08	1.06 (1.02 – 1.10)	3.59E–03	1.12 (1.08 – 1.17)	2.24E–08	*ZFP90*	*MS, UC IBD, PSC, VIT*

"Chinese" comprises the two Chinese GWAS (1,659 cases and 3,398 controls) and the Chinese Replication (3,043 cases and 5,074 controls). "European" comprises the European GWAS (4,036 cases and 6,959 controls), the additional European GWAS (1,165 cases and 2,107 controls) and the European replication study (1,478 cases and 6,925 controls)

[a] The risk allele refers to the effect in the overall meta-analysis

[b] MAF refers to the frequency of allele that is minor in Europeans.

[c] The Odds Ratio (OR) is with respect to the minor allele.

[d] For the rationale for candidate gene selection at the associated loci see Table 2

[e] *C6orf1* is also known as *LBH*, but we chose *LBH* as our gene because there are two separate signals in *LBH*. rs7579944 and rs17321999 were found to be independently associated with SLE (see Online Methods): rs17321999 was significant (Chinese P = 2.62 x 10^{-11}; European P = 6.14 x 10^{-6}; Meta *P* = 3.33 x 10^{-15}) when using rs7579944 as a covariate in logistic regression, and rs7579944 was significant (Chinese *P* = 1.38 x10^{-8}; European *P* = 4.49 x 10^{-6}; Meta *P* = 4.16 x 10^{-13}) at meta-analysis when using rs17321999 as a covariate in logistic regression. The LD between these two SNPs was very weak in all studies (The r^2 was as follows in each data set: Anhui GWAS = 0.039; Hong Kong GWAS = 0.024; Anhui Replication study =0.030; European GWAS = 0.005; Hom et al GWAS = 0.007; European replication study =0.005)

[f] Association for the gene(s) implicated by each SNP in other autoimmune diseases (excluding SLE) in Immunobase (www.immunobase.org) – Type 1 diabetes (T1D), Celiac disease (CEL), Multiple Sclerosis (MS), Crohn's Disease (CRO), Primary Billiary Cirrhosis (PBC), Psoriasis (PSO), Rheumatoid Arthritis (RA), Ulcerative Colitis (UC), Ankylosing Spondylitis (AS), Autoimmune Thyroid Disease (ATD), Juvenile Idiopathic Arthritis (JIA), Alopecia Areata (AA), Inflammatory Bowel Disease (IBD), Narcolepsy (NAR), Primary Sclerosing Cholangitis (PSC), Sjögren's Syndrome (SJO), Systemic Scleroderma (SSc), Vitiligo (VIT).

Table 2: Candidate Genes at SLE Associated Loci in Meta–Analysis

Associated SNP	Chr	Genes within +/–200kb of SNP	Genes within same LD block as SNP[a]	Immune phenotype in murine model[b]	Cis eQTLs with SNP	Likely Causal Gene at Locus (Reference)
rs34889541	1	*ATP6V1G3, PTPRC, MIR181A1HG*	*PTPRC (CD45)*	*PTPRC (CD45)*		CD45[19]
rs2297550	1	*SRGAP2, SRGAP2D, IKBKE, RASSF5, EIF2D,DYRK3*	*IKBKE*	*IKBKE, RASSF5*	*IKBKE*	*IKBKE*[20]
rs17321999	2	*YPEL5, LBH, LOC285043, LCLAT1*	*LBH*		*LBH*	*LBH*[21]
rs6762714	3	*LPP, TPRG1–AS1*	*LPP*			
rs17603856	6	*ATXN1*	*ATXN1*			
rs597325	6	*BACH2*	*BACH2*	*BACH2*		*BACH2*[22,23]
rs73135369	7	*CLIP2, GTF2IRD1, GTF2I, LOC101926943*	*GTF2IRD1*			*GTF2IRD1/GTF2I*[24]
rs1887428	9	*RCL1, JAK2, INSL6*	*JAK2*	*JAK2*		*JAK2*[25]
rs494003	11	*EHBP1L1, KCNK7, MAP3K11, PCNXL3, SIPA1, RELA, KAT5, RNASEH2C, AP5B1, OVOL1, OVOL1–AS1, SNX32, CFL1, MUS81, EFEMP2, CTSW, FIBP, CCDC85B, FOSL1, C11orf68, DRAP1, TSGA10IP, SART1*	*AP5B1, OVOL1, OVOL1–AS1*	*CTSW, MUS81, RELA, SIPA1*	*CTSW, FIBP, MUS81, RNASEH2C*	*RNASEH2C*[26,27]
rs1170426	16	*SMPD3, ZFP90, CDH3, CDH1*	*ZFP90, CDH3*	*CDH1*	*ZFP90*	*ZFP90* (FIK)[28]

[a] The LD block is defined as SNPs showing a correlation (r^2) of 0.75 with the associated SNP

[b] The immune phenotype designation is taken from http://www.informatics.jax.org/phenotypes.shtml of genes within +/–200kb of associated SNP

ONLINE METHODS

Study design in brief

We combined summary genome wide association data from two Chinese GWAS[4,5] [Anhui province, mainland China: 1,047 cases (63 males) and 1,205 Controls (673 males), λ_{GC} = 1.05; Hong Kong: 612 cases (50 males) and 2,193 Controls (919 males),λ_{GC} = 1.04] and a European GWAS [4,036 cases (365 males) and 6,959 Controls (2,785 males), λ_{GC} = 1.16 with $\lambda_{1,000}$ = 1.02], after imputing all three studies to the 1000 Genome (1KG) data density, and performed a meta–analysis. As the European data comprise 70% of both total cases and total controls, and was therefore the driving force in this meta–analysis, we selected SNPs for replication in a further set of Chinese samples first. We identified a subset of SNPs in the Chinese replication that passed an FDR of 1% to take forward for replication in European samples. We then performed replication using a second European GWAS[15] independent of our main European GWAS and also *de novo* genotyping in a new data cohort of European ancestry.

Imputation

We pre-phased each of the three studies separately using SHAPEIT[35]. The studies were then separately imputed (IMPUTE[36]) with 1KG reference data (Phase-I integrated set March 2012 build 37). The three datasets were aligned and meta-analyzed using R[37] by the King's College London group and also independently by the groups at Anhui and Hong Kong using METAL[38]. SNPs with imputation INFO scores < 0.7 in any of the three studies were removed from further analysis. The number of SNPs available pre- and post-QC, per chromosome and per associated locus are displayed in Supplementary Tables 7a and 7f respectively. A

summary of INFO scores and imputation cross validation are in Supplementary Tables 7b–e for each chromosome and Supplementary Tables 7g–j for each associated locus.

See supplementary note 3 for a discussion of the limitation of using imputed data.

Statistical analysis

Association testing: Following Imputation, each GWAS dataset was analysed for association (SNPTEST[36]), fitting an additive model. We used the inverse variance method for meta–analysis, combining data from the three studies for SNPs with an imputation INFO score > 0.7 in all three studies.

Testing for heterogeneity: We tested for heterogeneity between the associations signals in the Chinese and European data using Cochran's Q statistic (1 degree of freedom in this case). The *P*-values on the $-\log_{10}$ scale are plotted in Fig. 1b. QQ-plots (one per chromosome) for the heterogeneity *P*-values can be seen in Supplementary Fig. 9a and Bland-Altman plots for differences in genetic effect (log odds-ratio) estimates are in Supplementary Fig. 9b.

Assessment of shared association between ancestries: To assess the extent to which genetic association with SLE was shared between the Chinese and European populations, we compared association results in the European GWAS[3] with a meta–analysis of both Chinese GWAS, for SNPs published as associated in Europeans[3] and/or Chinese studies[4,6-9]. Association signals were declared as "shared" between the Chinese and Europeans if the SNP passed any one of the following four tests:

1: the locus had a published association in both Chinese and European studies at a genome–wide level of significance ($P < 5 \times 10^{-08}$);

2: the SNP was only published in Europeans but the association *P*-value in the Chinese meta–analysis was significant (FDR < 0.01 across all SNPs in this group) and the direction of effect in all three GWAS was the same.

3: the SNP was only published in a Chinese study but the association *P*-value in the European GWAS was significant (FDR < 0.01) and the direction of effect in all three GWAS was the same.

4: If the SNP failed either of tests 2 or 3, we performed a gene–based test (applying the software KGG[39-41]) on genes within +/–1Mb of the published SNP. The locus was deemed shared if the gene-based *P*-value was significant at the 0.01 level after adjusting for multiple testing across all genes tested.

We also performed a meta-analysis (European GWAS + both Chinese GWASs) of all loci published in either Chinese or European studies (each published SNP +/–1Mb) and recorded the most associated SNP. For loci published in Europeans, we declared the loci shared if the *P*-value (adjusted for multiple testing over all SNPs tested within the 2Mb region) in the Chinese data passed an FDR at 0.01 across all the loci published only in Europeans. We performed the reverse test for all loci published only in Chinese. While this did not identify any further shared loci (Supplementary Table 1b), two loci showed suggestive evidence ($P < 0.05$ after multiple testing adjustment within loci but not after adjusting across loci.)

Consistency of association between ancestries: We tested the hypothesis that the genome–wide association signals were consistent between the two populations. Post 1KG imputed association data were used for SNPs with INFO > 0.7. These genome wide association signals were separated into 1Mb regions (moving 1MB windows across the genome, 2,698 in total). We removed the extended MHC with a conservative buffer zone (Chr–6, from 20Mb to 40Mb), leaving 2,678 regions. We also removed regions that had excessively (more than 2 standard deviations from the average) low (N < 1000) or high (N > 3000) density of SNPs. This removed only 10% of the regions, leaving 2,338 regions. The lowest *P*-value within each window was taken as the strength of association for that particular window. Each *P*-value within each region was adjusted for multiple testing using a

Bonferroni adjustment, to avoid bias in ranking agreement owing to the lowest *P*–value being correlated with the number of statistical tests. The 1Mb regions within each population's data were then ranked according to the *P*-value (lowest *P*-value having rank 1). We tested agreement in ranking using Kendall's Tau statistic. Supplementary Fig. 7c–i shows a heat map of the ranks [red for highest rank (lowest *P*-value) and blue for lowest rank (highest *P*-value)] for all 2,338 regions. The order in this heat map was determined by the sum of the ranks (the region at the top of the figure has the smallest rank sum across the two populations). European ranks were plotted next to the Chinese ranks. For comparison, a simulated ranked dataset is shown alongside; we permuted the numbers 1 to 2,338 in two separate datasets and produced a heat map ordered by the sum of the ranks. Supplementary Fig. 7c–ii shows the same data but only for the top 250 regions. Supplementary Fig. 7c–iii shows the top 50 regions.

Testing for independent effects within loci: We tested for independent effects of the two SNPs (rs17321999 and rs7579944) within the 2p23.1 locus by fitting a multiple regression model with both SNPs as explanatory variables (results for each SNP in this analysis are conditional on the other SNP as a covariate). We checked linkage disequilibrium between the two SNPs in all datasets. The conditional results were combined in meta-analysis in the same way as the single–marker analysis.

Selection of SNPs for replication study: SNPs were chosen for replication in the Chinese samples using a number of criteria. We only chose SNPs that were not within a 1Mb window of loci that had previously been published as associated with SLE. We selected SNPs that had *P*–value significance levels at meta–analysis < 10^{-04}. Three SNPs in loci not previously reported as associated with SLE had genome wide level of significance ($P < 5 \times 10^{-08}$) after meta–analysis. SNPs spanning a 1Mb window were considered as one region and we selected only independent SNPs within this region: using LD as a measure of independence. We performed a gene–based test on the meta–analyzed data, using only SNPs that passed INFO > 0.9, applying the software KGG[39-41]. One SNP from each of the loci that passed a

569 gene based test at the level of $P < 10^{-05}$ were chosen, some of which were already selected
570 as having $P < 10^{-04}$ in the meta–analysis as single markers. In total 105 SNPs were selected
571 for replication in the Chinese replication cohort. From these 66 passed QC and 18 SNPs,
572 that passed a FDR < 1%, were taken forward to a further replication in the European
573 replication.

574 **Genotyping of replication data**

575 Genotyping of 130 SNPs was performed in 3,614 cases and 5,924 controls forming the
576 Chinese replication set, using the Sequenom platform. This set of 130 SNPs included 105
577 SNPs in loci not previously reported as associated with SLE and 25 SNPs that were in loci
578 that had previously been published as associated with SLE. The 105 potentially novel SNPs
579 included, in some cases, multiple SNPs in the same loci where we had some evidence of
580 independence. Several quality control (QC) steps were performed. SNPs with >10% missing
581 data were removed (25 SNPs) followed by subjects with >5% missing data being removed.
582 Two SNPs were monomorphic. Of the remaining 103 SNPs, 77 were in regions of the
583 genome with potentially novel SLE associations. Thirteen SNPs were removed after
584 checking the genotyping allele intensity plots closely for clustering quality and testing for
585 Hardy Weinberg Equilibrium (HWE). SNPs were removed if HWE $P < 1.00 \times 10^{-04}$. Post-QC
586 the Chinese replication consisted of 3,043 cases, 5,074 controls with genotyping on 64
587 SNPs. The European replication data comprised 1,478 cases and 6,925 controls genotyped
588 for 18 SNPs that passed a False Discovery Rate of 1% in the Chinese replication study: the
589 cases were of European ancestry and were a subset of those used in the replication study in
590 the European GWAS[3], on which this current study performed new genotyping on these 18
591 SNPs, and the controls were the same as used in that study (these samples were checked
592 for European ancestry using a principal component analysis spiked with HapMap samples,
593 see original paper). One of the 18 SNPs typed in the European replication cohort for this
594 study (rs2297550) failed genotyping and the remaining 17 SNPs passed QC (< 3% missing

595 data, HWE $P > 1.00 \times 10^{-04}$). An additional European GWAS was also used for replication,
596 comprising 1,165 cases and 2,107 controls[15].

597 **Gene expression data**

598 Gene expression data were obtained from two sources: firstly, we obtained data from Fairfax
599 *et al*[17] and unpublished data from Fairfax and Knight for NK cells, naïve monocytes,
600 monocytes stimulated by LPS (harvested after 2 hours and 24 hours), monocytes stimulated
601 by IFN and B cells. The CD4 (CD4 T cells) and CD14 (CD14/16 monocytes) data were
602 obtained from a previous study of gene expression in immune related cells[16]. An adjustment
603 was made for multiple testing using a false discovery rate at 0.01. To test whether observed
604 associations between SNPs and expression levels of *cis*-acting genes were due to chance,
605 we calculated the RTC score[18].

606 **Fine mapping Bayesian credibility sets.**

607 For each of the associated loci in Supplementary Table 1 and Table 1, we calculated a
608 Bayes factor for each SNP within the 2Mb window. We used the approximate Bayes factor of
609 Wakefield[32]. We then calculated the posterior probability that each SNP was driving the
610 association, using the Bayes factors, and created credibility sets as recently described[32]. We
611 created credibility sets using the European data and the Chinese data separately and
612 overlaid these sets (presented in Supplementary Fig 5). We focused on the intersection of
613 these two sets and present the SNPs with highest posterior probability within this
614 intersection along with allele frequencies. We focus on the intersection of the two
615 populations' sets, as credibility sets calculated from the overall meta–analysis are driven by
616 the European data. This would also be true if we were to use Bayesian updating (where the
617 posterior probabilities from one population were used as priors in the other population). The
618 intersection of the sets gives a subset of each populations C.S. that more likely contain the
619 true casual SNP.

620 **RoadMap Data**

We downloaded the epigenetic data for SNPs within the credibility intervals (as defined in Supplementary Fig. 5) around each meta-analysis SNP (Table 1) from the RoadMap consortium for all blood cell types. We chose DNse, RNA-Seq, H3K27ac (distinguishing active enhancers/promoters), H3K27me3 (repressive domains), H3K9ac (promoters), H3K9me3 (constitutive heterochromatin). The files downloaded contained the consolidated imputed epigenetic data based on the *P*-value signals from each of the individual epigenetic marks in each of the cell types within whole blood. We used the UCSC genome browser (hg19) to subset each epigenetic track for regions containing each credibility SNP and then exported the signal data via Galaxy[42]. In selecting chromatin enrichments at each mark for each SNP within the credibility set, we ensured that no SNP was less than 10 bp away from the edge of the 25 base pair epigenetic interval containing it. For SNPs closer to the edge of the chromatin interval, we averaged the enrichment from two adjacent intervals. The "3D enrichment diagrams" were plotted for each chromatin mark in each cell type for each SNP within the credibility set (Fig. 3 and Supplementary Fig. 6). Fig. 3 and Supplementary Fig. 6 highlight SNPs contained within peaks of enrichment ($\log_{10} P < 1 \times 10^{-04}$) with tick marks, these SNPs are listed in Supplementary Table 6.

Genetic structure of SLE in European and Asian population

The genetic risk score was calculated according to the method described by Hughes et al.[43], taking the number of risk alleles (i.e., 0, 1 or 2) for a given SNP and multiplying it by the natural log of its odds ratio (OR). The cumulative risk score in each subject was calculated by summing the risk scores from the loci in Supplementary Table 1, excluding the MHC, plus the 11 novel SNPs reported in this paper, which robustly associated with SLE and passed quality control in each population:

$$Cumulative\ genetic\ risk\ score = \sum_{i=1}^{m} ln(OR_i)G_i$$

Where m represents number of SLE risk loci; OR_i indicates the OR of risk SNP_i and G is the number of risk alleles at a given SNP. Cumulative risk scores were calculated for 498 founders in EUR, 503 founders in EAS, 487 in SAS, 347 in AMR and 657 in AFR from the 1KG project phase III. We tested for differences in GRS using a *t*-test. A Q-Q plot for each data satisfied assumptions of normality and given the large sample sizes the central limit theorem will satisfy normality for the distribution of sample means. As there was evidence of differences in variances of the GRS between some pairs of populations (EUR vs AMR, $P = 9.97 \times 10^{-05}$; AMR vs SAS, $P = 5.37 \times 10^{-05}$ SAS vs EAS, $P = 4.50 \times 10^{-03}$), we used a Welch 2-sample *t*-test which does not assume equal variances. The variances in each group were as follows (Chinese controls = 0.75, European Controls = 0.69; 1KG EAS = 0.86, 1KG EUR = 0.67, 1KG SAS = 0.66, 1KG AMR = 0.99, 1KG AFR = 0.77). We used the SNPs from Supplementary Table 1a to calculate the GRS for each population. We used the estimated OR from the EUR GWAS for the calculation of the GRS in Europeans (EUR and GWAS controls) and the OR from the Chinese GWAS for the calculation of the GRS in the EAS and Chinese GWAS controls. The OR from the EUR-Chinese meta-analysis was used in calculating the GRS in the AMR, SAS and AFR populations.

See supplementary note 1 for an assessment of the robustness of our approach.

See supplementary note 2 for details on SLE prevalence.

Heritability explained

We calculated the heritability explained by all genotyped SNPs in the CHN and EUR populations using GCTA[44]. We assumed that the Chinese have approximately 3 fold increase in prevalence over the Europeans, so we set the prevalence at 0.0003 in EUR and 0.001 in CHN. We used a cut off for relatedness at 0.05 and we used sex as a covariate. The results were h^2=28.4% (SE = 2.6%) in CHN and h^2=27.0% (SE = 1.0%) in EUR for autosomal SNPs. We found that the results were robust to choice of relatedness for the autosomal SNPs [a cut-off of 0.125 resulted in h^2=28.4% (SE = 2.6%) in CHN and h^2=27%

(SE = 1.0%) in EUR] while not so for the X chromosome [a cut-off of 0.125 resulted in h^2=1.2% (SE = 0.5%) in CHN and h^2=1.1% (SE = 0.2%) in EUR] where a cut-off for relatedness at 0.05 resulted in h< 0.015 in both populations.

To compare both populations using the same SNP density we re-ran the analysis on the overlap of genotyped SNPs (267,005 SNPs with MAF > 1% in CHN and 264,833 with MAF > 1% in EUR) and find that the heritability explained was higher in the CHN data: h^2=30.2% (SE = 2.6%) in CHN and h^2=22.7% (SE = 0.9%) in EUR.

Genetic correlation between European and Chinese SLE GWAS

To estimate genetic correlation (r_g) we applied LD score regression[34] to the summary association data in the European GWAS and the meta-analysis of the Chinese data (the input data is all GWAS summary statistics not just the SLE risk loci discussed in this paper). While this methodology is designed to compare similarity of genetic risk across diseases in the same population it serves here only to illustrate similarity across populations for the same disease and to highlight the heterogeneity at the MHC. We performed this analysis using both Asian (r_g =0.49, P = 3.00 x 10^{-03}) and European (r_g=0.51, P = 4.00 x 10^{-03}) reference LD information. This analysis was performed using summary data on all the SLE risk loci presented in this paper and a further analysis after removing the MHC [Asian (r_g =0.63, P = 6.92 x 10^{-07}) and European (r_g =0.62, P = 4.88 x 10^{-05})]. The increase in r_g post removal of the MHC illustrates the major heterogeneity at this locus.

35. O'Connell, J. *et al.* A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *Plos Genetics* **10**, 10.1371/journal.pgen.1004234 (2014).
36. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**, 499-511 (2010).
37. R-Core-Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 2013).
38. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

39. Li, M.X., Kwan, J.S.H. & Sham, P.C. HYST: A Hybrid Set-Based Test for Genome-wide Association Studies, with Application to Protein-Protein Interaction-Based Association Analysis. *American Journal of Human Genetics* **91**, 478-488 (2012).
40. Li, M.X., Gui, H.S., Kwan, J.S.H. & Sham, P.C. GATES: A Rapid and Powerful Gene-Based Association Test Using Extended Simes Procedure. *American Journal of Human Genetics* **88**, 283-293 (2011).
41. Li, M.X., Sham, P.C., Cherny, S.S. & Song, Y.Q. A Knowledge-Based Weighting Framework to Boost the Power of Genome-Wide Association Studies. *Plos One* **5**, 10.1371/journal.pone.0014480 (2010).
42. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* **11**, 10.1186/gb-2010-11-8-r86 (2010).
43. Hughes, T. *et al.* Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. *Annals of the Rheumatic Diseases* **71**, 694-699 (2012).
44. Yang, J.A., Lee, S.H., Goddard, M.E. & Visscher, P.M. GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics* **88**, 76-82 (2011).

I declare that the authors have no competing interests as defined by Nature Publishing Group, or other interests that might be perceived to influence the results and/or discussion reported in this paper.