

1 Identification of 38 novel loci for systemic lupus erythematosus 2 and genetic heterogeneity that may underly population 3 disparities in this disease

4 Yong-Fei Wang^{1, #}, Yan Zhang^{2, #}, Zhiming Lin³, Huoru Zhang¹, Ting-You Wang^{1, 4}, Yujie
5 Cao¹, David L. Morris⁵, Yujun Sheng⁶, Xianrong Yin⁶, Shi-Long Zhong⁷, Xiaoqiong Gu⁸,
6 Yao Lei¹, Jing He², Qi Wu², Jiangshan Jane Shen¹, Jing Yang¹, Tai-Hing Lam⁹, Jia-
7 Huang Lin⁹, Zhi-Ming Mai^{9, 10}, Mengbiao Guo¹, Yuanjia Tang¹¹, Yanhui Chen¹², Qin
8 Song¹³, Bo Ban¹⁴, Chi Chiu Mok¹⁵, Yong Cui¹⁶, Liangjing Lu¹¹, Nan Shen¹¹, Pak C.
9 Sham¹⁷, Chak Sing Lau¹⁸, David K. Smith⁹, Timothy J. Vyse⁵, Xuejun Zhang⁶, Yu Lung
10 Lau^{1, *} & Wanling Yang^{1, *}

11

- 12 1. Department of Paediatrics and Adolescent Medicine, The University of Hong Kong;
- 13 2. Department of Pediatric Surgery, Guangzhou Institute of Pediatrics, Guangdong Provincial Key
14 Laboratory of Research in Structural Birth Defect Disease, Guangzhou Women and Children's
15 Medical Center, Guangzhou Medical University;
- 16 3. Department of Rheumatology, The Third Affiliated Hospital of Sun Yat-Sen University;
- 17 4. The Hormel Institute, University of Minnesota;
- 18 5. Division of Genetics and Molecular Medicine, King's College London;
- 19 6. Department of Dermatology, No.1 Hospital, Anhui Medical University;
- 20 7. Guangdong Provincial Key Laboratory of Coronary Heart Disease Prevention, Guangdong Provincial
21 People's Hospital;
- 22 8. Department of Clinical Biological Resource Bank, Guangzhou Institute of Pediatrics, Guangzhou
23 Women and Children's Medical Center;
- 24 9. School of Public Health, The University of Hong Kong;
- 25 10. Radiation Epidemiology Branch, Division of Epidemiology and Genetics, National Cancer Institute,
26 National Institutes of Health;
- 27 11. Shanghai Institute of Rheumatology, Renji Hospital, Shanghai Jiao Tong University School of
28 Medicine;
- 29 12. Department of Pediatrics, Union Hospital Affiliated to Fujian Medical University;
- 30 13. Department of Rheumatology, Affiliated Hospital of Jining Medical University;
- 31 14. Department of Endocrinology, Affiliated Hospital of Jining Medical University;
- 32 15. Department of Medicine, Tuen Mun Hospital;
- 33 16. Department of Dermatology, China-Japan Friendship Hospital;
- 34 17. Department of Psychiatry, The University of Hong Kong;
- 35 18. Department of Medicine, The University of Hong Kong;

36 [#]These authors contributed equally;

37 ^{*}Correspondence to: WY yangwl@hku.hk (W.Y.) or YLL lauylung@hku.hk (Y.L.L.)

38 **Abstract**

39 Systemic lupus erythematosus (SLE), a worldwide autoimmune disease with high
 40 heritability, shows differences in prevalence, severity and age of onset among different
 41 ancestral groups. Previous genetic studies have focused more on European
 42 populations, which appear to be the least affected. Consequently, the genetic
 43 variations that underly the commonalities, differences and treatment options in SLE
 44 among ancestral groups have not been well elucidated. To address this, we undertook
 45 a genome-wide association study, increasing the sample size of Chinese populations
 46 to the level of existing European studies. Thirty-eight novel SLE-associated loci and
 47 incomplete sharing of genetic architecture were identified. Nine disease loci showed
 48 clear ancestral group heterogeneity and implicated antibody production as a potential
 49 mechanism for differences in disease manifestation. Polygenic risk scores performed
 50 significantly better when trained on matched ancestral data sets. These analyses help
 51 to reveal the genetic bases for disparities in SLE among ancestral groups.

52 (146 words)

53

54 Introduction

55 Systemic lupus erythematosus (SLE; OMIM 152700) is an autoimmune disease
 56 characterized by production of autoantibodies and multiple organ damage. Genetic
 57 factors play a key role in the disease, with estimates of its heritability ranging from 43%
 58 to 66% across populations¹⁻³. Differences in the expression of the disease across
 59 ancestral groups have been reported with non-European populations showing an
 60 earlier age of onset, 2-4 fold higher prevalence and 2-8 fold higher risk of developing
 61 end-stage renal disease than European populations⁴⁻⁷. Responses to treatment of SLE
 62 with the novel monoclonal antibody against B-cell activating factor (BAFF), Belimumab,
 63 also show variation across ancestral groups^{8,9}. These findings highlight the
 64 heterogeneous nature of the disease, so closer examination of ancestral group
 65 differences is likely to improve disease risk prediction and lead to more precise
 66 treatment options.

67 More than 90 loci have been shown to be associated with SLE through genome-
 68 wide association studies (GWAS)¹⁰⁻¹². Trans-ancestral group studies conducted
 69 previously were primarily designed to increase power and to identify SLE susceptibility
 70 loci shared across ancestries^{13,14}. However, due to inadequate power in studies
 71 involving non-Europeans, current findings are biased towards loci associated with SLE
 72 in European populations. Some risk alleles reported from studies on European
 73 populations, such as those in or near *PTPN22*, *NCF2*, *SH2B3* and *TNFSF13B*, are
 74 absent in East Asian populations¹⁵ while a missense variant (rs2304256) in *TYK2*
 75 points to a European-specific disease association¹⁶⁻¹⁸.

76 The basis for ancestral group differences in the manifestation of SLE at the
 77 genome level remains poorly understood. Further studies on non-European
 78 populations will help define the genetic architecture underlying SLE and the
 79 consequences of patients' ancestral backgrounds. To this end, we genotyped 8,252
 80 participants of Han Chinese descent recruited from Hong Kong (HK), Guangzhou (GZ)
 81 and Central China (CC), and combined these data with previous datasets to give a
 82 total of ten SLE genetic cohorts consisting of 11,283 cases and 24,086 controls. The

increased sample size, particularly for those of Chinese ancestry, allowed identification of novel disease loci and comparative analyses of the genetic architectures of SLE between major ancestral groups. Functional annotation of these findings provided insights into ancestral group differences in SLE.

Results

Data set preparation

Han Chinese data: After removing individuals with a low genotyping rate or hidden relatedness, the 7,596 subjects of Han Chinese descent from HK, GZ and CC genotyped in this study and the 5,057 subjects from the existing Chinese GWAS¹³ gave a Chinese ancestry data set of 4,222 SLE cases and 8,431 controls (**Extended Data Table 2**). Ethnic European Data: Existing GWAS data from European populations¹⁹ were reanalyzed, based on principal components (PC) matching those for subjects from the 1,000 Genomes Project to minimize the potential influence of population substructures²⁰ (**Methods**) and grouped into three cohorts, EUR GWAS 1-3, (**Extended Data Fig. 2**). The recent SP GWAS²¹ data was included. After quality control, the European data included 4,576 cases and 8,039 controls. A further 2,485 SLE cases and 7,616 controls were included as summary statistics from an Immunochip study of East Asians²² (**Extended Data Table 2**).

Ancestral correlation of SLE

Genotype imputation and association analysis were performed independently for each GWAS cohort and as meta-analyses of each ancestral group (**Methods**). The trans-ancestral genetic-effect correlation, r_{ge} , between the Chinese and European GWAS was estimated to be 0.64 with a 95% confidence interval (CI) of 0.46 to 0.81 by Popcorn²³ (**Methods**), indicating a significant, but incomplete, correlation of the genetic factors for SLE between the two ancestries.

Novel SLE susceptibility loci

Meta-analyses, involving a total of 35,369 participants (**Methods**; **Extended Data Table 2**) were conducted. Of the 94 previously reported SLE associated variants

(**Extended Data Table 3**), 59 (62.8%) surpassed a genome-wide significance P -value threshold ($5.0E-08$) and 84 (89.4%) exceeded the threshold of $5E-05$ in our study. Thirty-four novel variants reached genome-wide significance and four variants had P -values approaching this threshold based on either ancestry-dependent or trans-ancestral meta-analyses. The newly identified loci included the immune checkpoint receptor *CTLA4*, the TNF receptor-associated factor *TRAF3* and the type I interferon gene cluster on 9p21 (**Table 1** and **Extended Data Table 4**). The new loci bring the total of SLE-associated loci to 132 and produce a 23.5% and 16.5% increase in heritability explained for East Asians and Europeans, respectively (**Methods**).

Annotation of SLE susceptibility loci

Functional annotations that might be enriched with SLE susceptibility loci were evaluated by the stratified LD score regression method²⁴ (**Methods**). For non-cell type-specific annotations, heritability was significantly enriched in transcription start sites ($P = 2.47E-05$), regions that are conserved in mammals ($P = 8.22E-03$) and ubiquitous enhancers that are marked with H3K27ac or H3K4me1 modifications ($P = 4.43E-02$, $P = 5.03E-02$, respectively; **Extended Data Fig. 5**). Based on H3K4me1 modifications (associated with active enhancers) across 127 cell types, enrichment of specific cell types was investigated (**Methods**). Cells that surpassed the false discovery threshold rate ($FDR < 0.05$) were mostly hematological cells, with B and T lymphocytes the most prominent cell types associated with SLE (**Extended Data Fig. 6a**). Similar results were observed based on H3K4me3 modifications (associated with promoters of active genes) (**Extended Data Fig. 6b**).

Identification of putative disease genes and pathways

Excluding the human leukocyte antigen (HLA) region, 179 putative disease genes were identified across the disease-associated loci reported before and those newly identified in this study (**Extended Data Table 5**; DEPICT²⁵, **Methods**). A significant level of protein-protein connectivity corresponding to genes found at the novel loci and known SLE-associated loci was observed ($P < 1E-16$; **Extended Data Fig. 7**). Forty-five pathways were significantly enriched with these putative SLE susceptibility genes

(ToppGene²⁶, FDR < 0.05; **Extended Data Table 6**). The pathways of cytokine signaling, IFN- α/β signaling, Toll-like receptor (TLR) signaling, and B and T cell receptor signaling showed greatest enrichment. The RIG-I-like receptor signaling ($P = 5.83\text{E-}10$) and TRAF6-mediated IRF7 activation ($P = 6.41\text{E-}10$) pathways were designated as SLE associated pathways primarily based on genes newly identified in this study.

Trans-ancestral fine-mapping of disease-associated loci

One hundred and eight SLE-associated loci tagged by SNPs having a minor allele frequency (MAF) greater than 1% in both Chinese and European populations were examined by PAINTOR²⁷, making use of the differences in LD between ancestries. The median number of putative causal variants in the 95% credible sets reduced from 57 per locus when using only the European GWAS to 16 per locus when using data from both ancestries (one-sided paired t-test $P = 9.79\text{E-}07$, **Fig. 2**). The number of disease-associated loci with five or fewer putative causal variants increased from four when using the European GWAS alone to 15 (**Extended Data Table 7**). A single putative causal variant was identified for the *WDFY4* and *TNFSF4* loci, the latter of which was functionally validated in a previous study²⁸ (**Extended Data Fig. 8**).

Ancestral group differences

Based on the analysis of Cochran's Q (CQ)-test that assesses heterogeneity of effect-size estimates from different ancestral groups, SLE-associated variants or loci were divided into four categories: 1) ancestry-shared disease loci tagged by variants with CQ-test $P \geq 0.05$; 2) putative ancestry-heterogeneous disease loci with CQ-tests of $P < 0.05$ but FDR adjusted CQ-test $P \geq 0.05$; 3) ancestry-heterogeneous disease loci with FDR adjusted CQ-test $P < 0.05$; and 4) disease loci tagged by associated variants with the risk allele absent in one of the two ancestries¹⁵ (**Extended Data Table 10**). Nine disease variants, other than those absent or rare (MAF < 0.01) in one of the two ancestries¹⁵, showed significant differences in effect-size estimates between the two ancestral groups and were considered ancestry-heterogeneous (FDR adjusted CQ-test $P < 0.05$, category 3; **Fig. 3a** and **Extended Data Table 8**). Within

this category, variants in the *HIP1*, *TNFRSF13B*, *PRKCB*, *PRRX1*, *DSE* and *PLD4* loci were associated with SLE only in East Asians and variants in *TYK2* and *NEURL4-ACAP1* only in Europeans ($P < 5.0E-08$ in one ancestry but $P > 0.01$ in the other, with non-overlapping of the 95% CIs of the ORs). These eight loci were thus considered ancestry-specific. SNP rs4917014, a variant near *IKZF1*, showed a significantly stronger effect in East Asians (OR = 1.33, $P = 5.18E-29$) than in Europeans (OR = 1.16, $P = 1.34E-06$; CQ-test $P = 4.02E-04$). These findings were supported by analyses in each cohort (**Extended Data Fig. 9-10**).

On reanalyzing data from association studies on rheumatoid arthritis (RA)²⁹, the variants in the *PRKCB* and *PLD4* loci were found to be associated with RA in East Asians but not in Europeans (CQ-test $P = 0.004$ and 0.010 for *PRKCB* and *PLD4*, respectively), while the variant in *TYK2* was found associated in Europeans only (CQ-test $P = 0.002$; **Fig. 3b**). This consistency with the differences found in the SLE study suggests that shared mechanisms could be responsible for ancestral group differences among autoimmune diseases.

Colocalization of loci across ancestral groups

Colocalization methods consider many SNPs, rather than only the leading variant in a locus, to compare association signals between ancestral groups. The ancestry-shared disease loci showed much higher posterior probabilities (PP) of colocalization (mean of PP = 0.42) than the ancestry-specific loci (mean of PP = 0.03; **Extended Data Fig 11a**). For example, the *MTF1*, *IKBKE* and *TNIP1* loci in category 1 showed strong posterior probabilities ($\geq 90\%$) of colocalization under a Bayesian test³⁰ (**Methods**), suggesting shared causal effects between the two ancestries. The eight ancestry-specific loci in category 3 showed low posterior probabilities of colocalization (1% - 10%), consistent with the CQ-test of the top variant of each locus (above and **Extended Data Fig. 10**).

Since LD differences between ancestries may affect the colocalization results, we compared the SLE association signals from the Chinese populations with those on 27 non-immune-related phenotypes studied in Europeans (**Extended Data Table 9**) to

serve as colocalization baseline values. While posterior probabilities for colocalization at the ancestry-shared *MTF1*, *IKBKE* and *TNIP1* loci were much greater than the baseline values, there were no differences for the six Asian-specific disease loci (**Extended Data Fig. 11b**), thereby excluding the potential influence of LD. The European-specific loci (*TYK2* and *NEURL4-ACAP1*) were not evaluated this way due to lack of public data.

Functional annotation of the ancestry-heterogeneous loci

The ancestry-heterogeneous loci appear to be enriched for functions related to antibody production. Two of the nine putative disease genes at ancestry-heterogeneous loci (category 3), *TNFRSF13B* and *IKZF1*, are causal genes for human primary immunodeficiency disorders (PID)³¹ presenting with primary antibody deficiencies (PADs), whereas none of the disease genes at the putative ancestry-heterogeneous loci (category 2, 0/22; Fisher exact test $P = 0.07$) or the ancestry-shared loci (category 1, 0/120; Fisher exact test $P = 0.004$) are known to cause PADs in humans. Two of the East Asian-specific disease variants, those in the *TNFRSF13B* and *PRKCB* loci, were associated with serum immunoglobulin levels in East Asian populations³²⁻³⁴, whereas none of the variants in loci belonging to category 1 and 2 were found to be associated.

TNFRSF13B, which encodes a BAFF receptor, TACI, plays a major role for immunoglobulin production³⁵⁻³⁷. In this study, a missense variant in *TNFRSF13B*, rs34562254, was specifically associated with SLE in Chinese populations (OR = 1.18, $P = 2.88E-08$ in Chinese; OR = 1.01, $P = 0.75$ in Europeans). In European populations, an SLE-associated variant in the 3'-UTR of *TNFRSF13B* (encoding BAFF), which is absent in Asian populations (category 4), was associated with serum levels of total IgG, IgG1, IgA and IgM³⁸.

Mice deficient in the orthologs of four of the nine putative disease genes (44.4%) at the ancestry-heterogeneous loci for SLE, *Tnfrsf13b*-, *Ikzf1*-, *Prkcb*- and *Tyk2*-, demonstrated abnormal IgG levels³⁹ (MP:0020174), while at putative ancestry-heterogeneous loci (4/22 or 18.2%; OR = 3.43, Fisher exact test $P = 0.185$) or ancestry-

shared loci (14/120 or 11.6%; OR = 5.92, Fisher exact test $P = 0.027$), proportionately fewer genes caused aberrant IgG levels in mice. Orthologs of four of the twelve (33.3%) putative disease genes from disease loci where the risk allele is monomorphic in one of the ancestries (*PTPN22*, *TNFSF13B*, *IKZF3* and *IGHG1*; category 4) also demonstrated abnormal immunoglobulin production in gene knockout mouse models^{35,38,40-42}.

Evolutionary signatures for the disease loci

Disease loci with heterogeneity between East Asians and Europeans might have undergone differential selection pressures in recent human history, as has been shown for the SLE risk variant in *TNFSF13B*³⁸. Frequency variances, as fixation indexes (F_{st}), for the variants of the first three categories were calculated using 3,324 controls from the HK cohort and 5,379 controls from EUR GWAS 2 cohort (**Methods**). Higher F_{st} would indicate a larger frequency difference between the two ancestries. Mean F_{st} values for the ancestry-shared, putative ancestry-heterogeneous and ancestry-heterogeneous variants were 0.054, 0.061 and 0.084, respectively. Although a small sample, three (*DSE*, *HIP1*, *TNFRSF13B*) of the nine ancestry-heterogeneous variants (33.3%) showed $F_{st} \geq 0.15$ (empirical $P < 0.03$), while only 10% of the putative ancestry-heterogeneous variants and 8.8% for the ancestry-shared disease variants had $F_{st} \geq 0.15$ (**Extended Data Fig. 12a**).

A significant correlation of standardized integrated Haplotype Scores (iHS)⁴³, estimated using control subjects from HK and EUR GWAS 2 cohorts (**Methods**), gave evidence of shared recent positive selection for the ancestry-shared disease variants (categories 1; $r = 0.28$, $P = 0.03$; **Extended Data Fig. 12b**). This is consistent with results using data from Southern Han Chinese (CHS) and Utah residents of European ancestry⁴⁴ (CEU; $r = 0.32$, $P = 0.008$). However, there was no evidence of such a correlation for disease variants that showed ancestry heterogeneity (category 2 and 3; **Extended Data Fig. 12b**). For example, in the BAFF system, the derived risk allele rs34562254-A in *TNFRSF13B* is much more prevalent in East Asians than in other populations (**Figure 4a**) and has a significantly longer haplotype for the derived risk

allele than the ancestral allele (more negative standardized iHS score) in East Asian populations than in Africans ($P = 3.2E-04$) or Europeans ($P = 4.4E-04$) (**Figure 4b**), suggesting recent positive selection for the risk allele in East Asians.

Polygenic risk scores for SLE and their accuracies across ancestries

Polygenic risk scores (PRS) have been used to estimate individual risk to complex diseases, such as coronary artery disease⁴⁵ and schizophrenia⁴⁶. However, as the majority of GWAS findings used to calculate these scores are based on European populations, their accuracy in other populations may be limited. PRS for SLE, trained by data on European populations, were tested on individuals of three Chinese cohorts using the lassosum algorithm⁴⁷ (**Methods**). The area under the receiver-operator curve (AUC) ranged from 0.62 to 0.64 for the three Chinese cohorts. Similar results were observed in the reverse case (**Extended Data Fig. 13a**). The LDpred⁴⁸ algorithm produced similar results (**Extended Data Fig. 13b**). These analyses suggest a partial transferability of PRS between the two ancestries.

Using samples from the GZ cohort as the validation dataset, performance of predictors trained using GWAS summary statistics from the HK and CC cohorts (2,618 cases and 7,446 controls) or from the European cohorts (4,576 cases and 8,039 controls) were evaluated. Ancestry-matched predictors significantly outperformed (AUC = 0.76, 95% CI: 0.74-0.78) ancestry mismatched predictors (AUC = 0.62, 95%CI: 0.60-0.64) (**Fig 5a**). When the analysis was repeated by randomly choosing the same number of samples (1,500 cases and 1,500 controls) from each of the Chinese and European GWAS as training data, a similar difference was observed (**Extended Data Fig. 14**). Ancestry-matched PRS for samples in the GZ cohort had a mean difference of 0.89 (standard deviation) between the SLE case and control groups (t -test $P = 9.01E-116$) and disease classification using the optimal threshold achieved 73.4% sensitivity and 65.4% specificity (**Fig 5b; Methods**). Disease risk increased with higher PRS, with individuals in the highest PRS decile having a much higher disease risk than those in the lowest decile (OR = 30.3, Chi-square test $P = 6.23E-54$; **Fig 5c**).

Discussion

Despite evidence of ancestral group differences in the disease manifestations of SLE, genetic studies to date have been unable to examine these differences in detail as most earlier data focused on European ancestral groups. As non-European groups appear to be more severely affected by SLE, greater genetic data on these groups are likely to be highly informative. By increasing the number of subjects of East-Asian ancestry to levels roughly equivalent to those of European subjects, and including previously published data, we have made cross ancestral group studies possible.

Through systematic comparisons between the two ancestral groups, it was found that there was a significant, but not complete, sharing of the genetic architecture of SLE between Chinese and Europeans. SLE showed a r_{ge} of 0.64. This represents a much lesser level of genetic sharing than seen for schizophrenia⁴⁹ between East Asian and European ancestries ($r_{ge} = 0.98$).

Through ancestry-dependent and trans-ancestral meta-analyses, we identified 38 novel loci associated with SLE, bringing the total number of SLE-associated loci to 132. High level functional annotation of these SLE associated loci implicated hematological cells, particularly B and T cells, cytokine signaling and other immune system pathways. Consistent with previous findings⁵⁰, we demonstrated the value of trans-ancestral data in significantly reducing the number of putative causal variants at each disease-associated locus, which may facilitate future functional and mechanistic studies.

There was strong evidence of heterogeneity on SLE associations between the two ancestries, which were not likely to be artefacts of study power. Eight variants (that are common in both ancestral groups) were associated with disease in only one of the ancestral groups. For three of these, a similar ancestral difference was found by re-analyzing association data on RA²⁹. This might suggest that common mechanisms account for ancestral differences in autoimmune diseases.

Genes at the ancestry-heterogeneous disease loci seemed more likely to be involved in regulation of immunoglobulins than genes at the ancestry-shared disease loci. Immunoglobulin levels are highly heritable⁵¹ and have been found to differ between ancestries, with African Americans and Asians having higher serum

immunoglobulin levels than people of European ancestry⁵²⁻⁵⁵. Higher antibody levels in non-European populations might have contributed to their higher prevalence of SLE and further study of intrinsic differences in immune function among ancestries may be informative.

That differential mechanisms may exist for antibody regulation between East Asian and European populations is supported by the association of SLE with the BAFF signaling system. This system, which may be part of the initial reaction to host-pathogen interactions³⁶, is under positive selection based on different environmental exposures^{36,38}. BAFF and its receptors, including TACI, play essential roles in B cell survival and differentiation³⁵⁻³⁷. The SLE risk allele in the gene encoding BAFF is completely absent in Chinese populations³⁸ and a missense variant in the gene encoding TACI was found to be specifically associated with SLE in East Asians in this study. Both these genes, *TNFSF13B*³⁸ in Europeans and *TNFRSF13B* in East Asians, were found to have undergone positive selection in recent human history. Adaptation of the host to resist pathogens may underlie some of the ancestral group-heterogeneity.

TACI is expressed at very low levels in human newborns and mice before exposure to pathogens^{56,57} and previous studies have shown that certain pathogens can ablate B cell responses by modulating the expression of TACI⁵⁸⁻⁶⁰. The BAFF risk allele was shown to significantly up-regulate humoral immunity³⁸, and whether this is the case for the risk allele in TACI should be investigated. TACI blockers, such as Atacicept⁶¹, might give better responses to SLE in patients of Asian ancestry and the variant found in *TNFRSF13B* may be a useful genetic marker for the prescription of Belimumab, a hypothesis that may warrant further study.

Our analyses have identified a substantial number of novel SLE-associated genetic loci and deepened our understanding of the genetic factors that may underly the differences in the manifestation of SLE between peoples of European and non-European ancestry. Like a recent PRS study in SLE⁶², but to a greater extent, we have shown that PRS achieved a far better performance when based on ancestry-matched populations. Our findings contribute new insights into precise treatments, and to risk

345 prediction and prevention of SLE.

346

Methods

Overview of samples

8,252 subjects of Han Chinese descent from Hong Kong (HK), Guangzhou (GZ) and Central China (CC) were genotyped in this study. The institutional review boards of the institutes collecting the samples (The University of Hong Kong, Hospital Authority Hong Kong West Cluster and Guangzhou Women and Children's Medical Center) approved the study and all subjects gave informed consent. These subjects were genotyped by the Infinium OmniZhongHua-8, the Infinium Global Screening Array-24 v2.0 (GSA) and the Infinium Asian Screening Array-24 v1.0 (ASA) platforms. Fourteen samples were randomly selected and genotyped by different platforms. High concordance rates (> 99.9%) were observed for genotypes derived from the different platforms (**Extended Data Table 1**). Principal component (PC) analysis was performed to examine potential batch effects and no significant differences were observed from the PCs for data genotyped by different platforms and from different batches (**Extended Data Fig. 1**). These new data were analyzed with data from our existing SLE GWAS¹³, which contained 5,057 Chinese subjects.

Quality control and association study

Genotype Harmonizer⁶³ was used to align the strands of variants of the Chinese GWAS to the reference of the 1,000 Genomes Project Phase 3 panel. Variants with a low call rate (< 90%), low minor allele frequency (< 0.5%) and violation of Hardy-Weinberg equilibrium (P -value < 1E-04) were removed. Quality control required the following criteria: i) missing genotypes were below 5%, ii) hidden relatedness (identity-by-descent) with other samples was $\leq 12.5\%$ iii) inbreeding coefficients with other samples ranged from -0.05 to 0.05 and iv) not having extreme PC values as computed for individuals using EIGENSTRAT embedded in PLINK^{64,65}. After quality control, pre-phasing used SHAPEIT⁶⁶ and individual-level genotype data were imputed to the density of the 1,000 Genomes Project Phase 3 reference using IMPUTE2⁶⁷. For association analysis SNPTTEST⁶⁸ was used to fit an additive model. Top PCs and the BeadChip types were included as covariates. The number of PCs to be adjusted for in

each analysis was determined using a scree plot with a cutoff when the plot levels off. Variants with imputed INFO scores < 0.7 were excluded. The genomic inflation factors (λ_{GC}) for the HK, GZ and CC GWAS were 1.04, 1.03 and 1.04, respectively, and the LD score regression (LDSC)⁶⁹ intercepts were 1.03, 1.02 and 1.03, respectively. Manhattan plots for each cohort are shown in **Extended Data Fig. 3**.

For the European SLE GWAS data, the λ_{GC} and LDSC intercept listed in LD hub seemed inflated ($\lambda_{GC} = 1.17$ and LDSC intercept = 1.10)²⁰. Thus, the data were reanalyzed to minimize the potential influence of sub-population stratification (and see below). PCA analysis showed that subjects from the existing European data were more diverse than the Chinese subjects used in this study (**Extended Data Fig. 2a**). The European individuals were grouped into three cohorts by their PCs relative to the subjects of the 1,000 Genomes Project. Subjects in the EUR GWAS 1, EUR GWAS 2, EUR GWAS 3 cohorts shared similar PCs with individuals of Spanish (IBS), northern and western European (CEU and GBR) and Italian (ITS) origins, respectively (**Extended Data Fig. 2b** and **Extended Data Table 2**). Quality control, imputation and association analyses were conducted, as for the Chinese datasets, in each cohort. λ_{GC} for the three European GWAS datasets were 1.05, 1.08 and 1.03, respectively, and the LDSC intercepts were 1.03, 1.04 and 1.00, respectively. Manhattan plots for each cohort are shown in **Extended Data Fig. 4**.

Genetic correlation between the two ancestries

Trans-ancestral genetic correlation from the meta-analysis results for Chinese and Europeans were estimated using the Popcorn algorithm²³ based on common SNPs in the autosomes. The disease prevalence in Chinese and European populations were set to be 1‰ and 0.3‰, respectively⁴. SNPs were removed from this analysis according the following criteria: 1) SNPs with strand-ambiguities (A/T or C/G alleles); 2) having MAF $< 5\%$; 3) having imputed INFO score < 0.9 . The cross-population LD scores were estimated using control subjects from the HK cohort ($n = 3,324$) and EUR GWAS 2 cohort ($n = 5,379$).

Meta-analyses of SLE association studies

Meta-analyses for the Chinese and European SLE GWAS were conducted independently. The summary association statistics from HK, GZ and CC GWAS (4,222 SLE cases and 8,431 controls) were combined in a meta-analysis using a fixed-effect model, weighted by the inverse-variance⁷⁰. The λ_{GC} for the Chinese meta-analysis was 1.09 and the LDSC intercept was 1.04. For the European data, the EUR GWAS 1-3 datasets were combined with a new GWAS on a Spanish population²¹ in the meta-analysis (4,576 cases and 8,039 controls). λ_{GC} and the LDSC intercept for the European SLE meta-analysis reduced to 1.11 and 1.03, respectively.

Trans-ancestral meta-analysis across the Chinese and European GWAS cohorts used the fixed-effect model. To increase statistical power, summary association statistics from an Immunochip study were included as an *in silico* replication, adding 2,485 SLE cases and 7,616 controls from Korea (KR), Han Chinese in Beijing (BJ) and Malaysian Chinese (MC)²². 11,283 SLE cases and 24,086 controls were involved in this study (**Extended Data Table 2**).

Heritability explained by the SLE-associated variants

The variance in liability explained by the SLE-associated variants was measured using VarExplained program⁷¹. Variants in the HLA region were excluded in the analysis. The disease prevalence was set to be 1‰ for East Asians and 0.3‰ for Europeans⁴. The novel loci increased the heritability explained from 0.10 to 0.13 for East Asians, and from 0.08 to 0.09 for Europeans.

Functional annotations of SLE associated SNPs

The stratified LD score regression method²⁴ was applied on the trans-ancestral meta-analysis result to partition SNP-heritability across functional annotations. Twenty-eight categories of annotations that are not cell type specific (**Extended Data Fig. 5**) provided by this source were studied. For cell type-specific analyses, H3K4me1 and H3K4me3 modifications across 127 cell types (**Extended Data Fig. 6**) were downloaded from the Roadmap Epigenomics Project⁷². The cell type-specific enrichment was performed under the “full baseline” model²⁴, which aimed to control for overlaps with annotations that are not cell type-specific.

Identification of putative SLE genes and gene-set enrichment analysis

Putative causal gene(s) across all the SLE-associated loci outside of the HLA region were identified using DEPICT²⁵. The default setting ($r^2 > 0.3$) was used to set boundaries for each SLE associated locus. Genes within (or overlapping) the boundaries were examined and those with a P -value < 0.05 were defined as putative causal genes. If no genes were selected at that locus, gene(s) identified from eQTL data from human whole blood^{73,74} were considered to be putatively causal. Gene-set enrichment analysis was performed using ToppGene²⁶, with the August, 2019 versions of the KEGG⁷⁵, Reactome⁷⁶ and mouse knockout phenotype³⁹ databases. The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies³¹ (PID) was used to obtain 320 human PID genes categorized into nine phenotypic classifications.

Trans-ancestral fine-mapping of the associated loci

The HLA region was excluded from this analysis, as extensive LD and limited genotyping of SNPs in both ancestries makes defining the best model of association difficult for this region. Disease loci with rare risk alleles ($MAF < 0.01$) or absent in one ancestry were also excluded, leaving 108 SLE-associated loci in the autosomes for this study. For each disease locus, all variants within the region were extracted for both ancestries. The genetic interval was determined by the closest recombination hotspots around a given disease-associated variant (defined as a recombination rate < 10 cM/Mb). A fine-mapping algorithm, PAINTOR (version 3.0)²⁷, was used to estimate the posterior probability of causality for each variant at a given locus based on the trans-ancestral model. For comparison, we also applied the fine-mapping algorithm on the Chinese and European SLE GWAS, separately. All analyses were run under the assumption of a single causal variant per locus. The LD matrix was calculated using control samples from HK ($n = 3,324$) and EUR GWAS 2 ($n = 5,379$) for Chinese and European populations. Variants with a cumulative posterior probability greater than 95% were defined as putative causal variants (95% credible set).

Identification of loci with differential effects between the two ancestries

Cochran's Q (CQ)-test⁷⁷ was used to examine effect-size differences between the

two ancestries for all the disease-associated variants in the autosomes. If the variants were also interrogated by the Immunochip²² system, the association results derived from the KR, BJ and MC cohorts were also included. CQ-test *P*-values were adjusted for a cutoff of 0.05 using the Benjamini-Hochberg method⁷⁸. For comparison, summary association statistics on RA were downloaded from a previous study²⁹ of 4,873 RA cases and 17,642 controls of Asian ancestry and 14,361 RA cases and 43,923 controls of European ancestry.

Colocalization analysis

Colocalization of association signals from the two ethnicities was determined using the R package coloc³⁰ on all variants with a MAF > 1% and imputed (IMPUTE2⁶⁷) INFO score > 0.9 within a given disease locus. To control for LD differences between ancestries, SLE association signals from Chinese populations were compared with those from 27 immune-unrelated phenotypes from European populations (LD hub²⁰; **Extended Data Table 9**) to generate baseline posterior probabilities of colocalization in the absence of a phenotypic relationship. Ancestry-shared causal effects for SLE were expected to be significantly greater than the baseline values.

Analysis of selection signatures for the associated variants

The fixation index (F_{st}) was used to test allele-frequency differences between the two ancestries. F_{st} was calculated based on the following formula⁷⁹:

$$F_{st} = \frac{H_t - H_s}{H_t},$$

where H_t is the expected proportion of heterozygosity in the pooled samples from all ethnicities based on Hardy-Weinberg equilibrium: $H_t = 2\bar{p}(1 - \bar{p})$, \bar{p} is the allele frequency in the overall pool. H_s , the expected proportion of heterozygosity in a subpopulation (either Chinese or Europeans), is estimated as $H_s = \frac{H_{p1} * N_{p1} + H_{p2} * N_{p2}}{N_{p1} + N_{p2}}$, where H_{pi} is the expected heterozygosity in the *i*th subpopulation estimated by the allele frequency in that subpopulation under Hardy-Weinberg equilibrium. N_{pi} is the sample size of the *i*th subpopulation.

Potential selective sweeps in respective ancestries were examined using the

Integrated Haplotype Score (iHS) method, which measures the extended haplotype homozygosity for the ancestral allele relative to the derived allele⁸⁰. Raw iHS scores were computed using the R package rehh⁸¹ and normalized by different frequency bins (50 bins over the range 0 to 1). Large negative standardized iHS values indicate long haplotypes carrying the derived allele, while large positive values suggest long haplotypes with the ancestral allele. The F_{st} and iHS values analyzed in this study were estimated using control subjects from HK (n = 3,324) and EUR GWAS 2 (n = 5,379). Standardized iHS scores based on the 1,000 Genomes Project were downloaded from a previous study⁴⁴ for comparison.

Calculation of polygenic risk scores

Polygenic risk scores (PRS) for individuals were computed using lassosum⁴⁷, a penalized regression framework. The meta-analysis results on Europeans were used to calculate PRS for individuals of Chinese ancestry, and vice versa. LD information among SNPs was calculated from the testing dataset. These analyses were repeated using LDpred⁴⁸.

The GZ SLE GWAS cohort was used as a test dataset to evaluate the influence of training data from different ancestries. Two predictors were constructed using lassosum based on meta-analysis results from: 1) HK and CC GWAS, 2,618 cases and 7,446 controls; 2) European GWAS, 4,576 cases and 8,039 controls. To control for influence from different sample sizes, 1,500 cases and 1,500 controls were randomly chosen from the Chinese and European populations to train two same-size predictors (repeated 3 times). PRS values generated from each test were scaled to a mean of 0 and a standard deviation of 1, and then evaluated based on the area under the ROC curve (AUC). The values and the 95% confidence intervals were calculated using the R package pROC⁸² and the optimal cut-off, the point that maximizes the sum of sensitivity and specificity, for case-control classification was estimated using the coords function.

519 **Acknowledgements**

520 We thank grant support from National Key Research and Development Program
521 of China (2017YFC0909001), Hong Kong PhD fellowship scheme (HKPF), HKU
522 Postgraduate Scholarships and the Edward & Yolanda Wong Fund for supporting
523 postgraduate students who participated in this work. We also thank the Hong Kong
524 Area of Excellence (AoE) NPC case-control study partially funded by the World Cancer
525 Research Fund (WCRF) for sharing their GWAS data. W. Yang and Y. L. Lau thank
526 Research Grant Council of Hong Kong for support on genetic studies of SLE (GRF
527 17146616, 17125114, HKU783813M). Y. Zhang thanks grant support from National
528 Natural Science Foundation of China (Grant No. 81601423). Y.-F. Wang thanks grant
529 support from National Natural Science Foundation of China (Grant No. 81801636).

530 **Author contributions**

531 W. Yang and Y.-F. Wang conceived the study. Y. L. Lau and Y. Zhang took the lead
532 in data collection and management. Z. Lin, H. Zhang, Y. Sheng, X. Yin, S.-L. Zhong, X.
533 Gu, J. He, Q. Wu, T.-H. Lam, J.-H. Lin, Z.-M. Mai, Y. Tang, Y. Chen, Q. Song, B. Ban,
534 C.C. Mok, C. Yong, L. Lu, N. Shen, P.C. Sham, C.C. Lau, J. Yang and X. Zhang
535 undertook subject recruitment and collected phenotype data. D.M. and T.J.V shared
536 SLE GWAS data on European populations. Y.-F. Wang, T.-Y. Wang, Y. Cao, M. Guo,
537 J.J. Shen carried out data analyses including quality control, genotype imputation,
538 association and meta-analyses. Y.-F. Wang, T.-Y. Wang and Y. Lei carried out fine-
539 mapping, selection signatures analyses and PRS comparison between the two
540 ancestral groups. Y.-F. Wang, W. Yang, Y. Zhang, Y. L. Lau, P.C. Sham and D.K. Smith
541 wrote the manuscript. All authors read and contributed to the manuscript.

542 **Competing interests**

543 None declared.

544

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 Wang, J. *et al.* Systemic lupus erythematosus: a genetic epidemiology study of 695 patients from China. *Arch Dermatol Res* **298**, 485-491, doi:10.1007/s00403-006-0719-4 (2007).
- 3 Kuo, C. F. *et al.* Familial Aggregation of Systemic Lupus Erythematosus and Coaggregation of Autoimmune Diseases in Affected Families. *JAMA Intern Med* **175**, 1518-1526, doi:10.1001/jamainternmed.2015.3528 (2015).
- 4 Johnson, A. E., Gordon, C., Palmer, R. G. & Bacon, P. A. The prevalence and incidence of systemic lupus erythematosus in Birmingham, England. Relationship to ethnicity and country of birth. *Arthritis and rheumatism* **38**, 551-558 (1995).
- 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 Costenbader, K. H. *et al.* Trends in the incidence, demographics, and outcomes of end-stage renal disease due to lupus nephritis in the US from 1995 to 2006. *Arthritis and rheumatism* **63**, 1681-1688, doi:10.1002/art.30293 (2011).
- 7 Ballou, S. P., Khan, M. A. & Kushner, I. Clinical features of systemic lupus erythematosus: differences related to race and age of onset. *Arthritis and rheumatism* **25**, 55-60, doi:10.1002/art.1780250109 (1982).
- 8 Stohl, W. *et al.* Efficacy and Safety of Subcutaneous Belimumab in Systemic Lupus Erythematosus: A Fifty-Two-Week Randomized, Double-Blind, Placebo-Controlled Study. *Arthritis Rheumatol* **69**, 1016-1027, doi:10.1002/art.40049 (2017).
- 9 Stohl, W. & Hilbert, D. M. The discovery and development of belimumab: the anti-BLyS-lupus connection. *Nature biotechnology* **30**, 69-77, doi:10.1038/nbt.2076 (2012).
- 10 Chen, L., Morris, D. L. & Vyse, T. J. Genetic advances in systemic lupus erythematosus: an update. *Current opinion in rheumatology*, doi:10.1097/BOR.0000000000000411 (2017).
- 11 Wen, L. *et al.* Exome-wide association study identifies four novel loci for systemic lupus erythematosus in Han Chinese population. *Annals of the rheumatic diseases*, doi:10.1136/annrheumdis-2017-211823 (2017).
- 12 Wang, Y. F. *et al.* Identification of ST3AGL4, MFHAS1, CSNK2A2 and CD226 as loci associated with systemic lupus erythematosus (SLE) and evaluation of SLE genetics in drug repositioning. *Annals of the rheumatic diseases*, doi:10.1136/annrheumdis-2018-213093 (2018).
- 13 Morris, D. L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nature genetics*, doi:10.1038/ng.3603 (2016).
- 14 Langefeld, C. D. *et al.* Transancestral mapping and genetic load in systemic lupus erythematosus. *Nature communications* **8**, 16021, doi:10.1038/ncomms16021 (2017).
- 15 Wang, Y. F., Lau, Y. L. & Yang, W. Genetic studies on systemic lupus erythematosus in East Asia point to population differences in disease susceptibility. *Am J Med Genet C Semin Med Genet*, doi:10.1002/ajmg.c.31696 (2019).
- 16 Cunninghame Graham, D. S. *et al.* Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS genetics* **7**, e1002341,

doi:10.1371/journal.pgen.1002341 (2011).

17 Li, P., Chang, Y. K., Shek, K. W. & Lau, Y. L. Lack of association of TYK2 gene polymorphisms in Chinese patients with systemic lupus erythematosus. *The Journal of rheumatology* **38**, 177-178, doi:10.3899/jrheum.100424 (2011).

18 Kyogoku, C. *et al.* Lack of association between tyrosine kinase 2 (TYK2) gene polymorphisms and susceptibility to SLE in a Japanese population. *Mod Rheumatol* **19**, 401-406, doi:10.1007/s10165-009-0173-1 (2009).

19 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, doi:10.1038/ng.3434 (2015).

20 Zheng, J. *et al.* LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics*, doi:10.1093/bioinformatics/btw613 (2016).

21 Julià, A. *et al.* Genome-wide association study meta-analysis identifies five new loci for systemic lupus erythematosus. *Arthritis research & therapy* **20**, 100, doi:10.1186/s13075-018-1604-1 (2018).

22 Sun, C. *et al.* High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nature genetics* **48**, 323-330, doi:10.1038/ng.3496 (2016).

23 Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes, C., Ye, C. J., Price, A. L. & Zaitlen, N. Transethnic Genetic-Correlation Estimates from Summary Statistics. *American journal of human genetics* **99**, 76-88, doi:10.1016/j.ajhg.2016.05.001 (2016).

24 Finucane, H. K. *et al.* Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature genetics* **47**, 1228-1235, doi:10.1038/ng.3404 (2015).

25 Pers, T. H. *et al.* Biological interpretation of genome-wide association studies using predicted gene functions. *Nature communications* **6**, 5890, doi:10.1038/ncomms6890 (2015).

26 Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic acids research* **37**, W305-311, doi:10.1093/nar/gkp427 (2009).

27 Kichaev, G. & Pasaniuc, B. Leveraging Functional-Annotation Data in Trans-ethnic Fine-Mapping Studies. *American journal of human genetics* **97**, 260-271, doi:10.1016/j.ajhg.2015.06.007 (2015).

28 Manku, H. *et al.* Trans-ancestral studies fine map the SLE-susceptibility locus TNFSF4. *PLoS genetics* **9**, e1003554, doi:10.1371/journal.pgen.1003554 (2013).

29 Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506**, 376-381, doi:10.1038/nature12873 (2014).

30 Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS genetics* **10**, e1004383, doi:10.1371/journal.pgen.1004383 (2014).

31 Bousfiha, A. *et al.* The 2017 IUIS Phenotypic Classification for Primary Immunodeficiencies. *J Clin Immunol* **38**, 129-143, doi:10.1007/s10875-017-0465-8 (2018).

633 32 Liao, M. *et al.* Genome-wide association study identifies common variants at TNFRSF13B
634 associated with IgG level in a healthy Chinese male population. *Genes and immunity* **13**,
635 509-513, doi:10.1038/gene.2012.26 (2012).

636 33 Osman, W. *et al.* Association of common variants in TNFRSF13B, TNFSF13, and ANXA3
637 with serum levels of non-albumin protein and immunoglobulin isotypes in Japanese. *PloS*
638 *one* **7**, e32683, doi:10.1371/journal.pone.0032683 (2012).

639 34 Yang, M. *et al.* Genome-wide scan identifies variant in TNFSF13 associated with serum
640 IgM in a healthy Chinese male population. *PloS one* **7**, e47990,
641 doi:10.1371/journal.pone.0047990 (2012).

642 35 Castigli, E. *et al.* TACI and BAFF-R mediate isotype switching in B cells. *The Journal of*
643 *experimental medicine* **201**, 35-39, doi:10.1084/jem.20032000 (2005).

644 36 Sakai, J. & Akkoyunlu, M. The Role of BAFF System Molecules in Host Response to
645 Pathogens. *Clin Microbiol Rev* **30**, 991-1014, doi:10.1128/CMR.00046-17 (2017).

646 37 Zhang, Y., Li, J., Zhang, Y. M., Zhang, X. M. & Tao, J. Effect of TACI signaling on humoral
647 immunity and autoimmune diseases. *Journal of immunology research* **2015**, 247426,
648 doi:10.1155/2015/247426 (2015).

649 38 Steri, M. *et al.* Overexpression of the Cytokine BAFF and Autoimmunity Risk. *The New*
650 *England journal of medicine* **376**, 1615-1626, doi:10.1056/NEJMoa1610528 (2017).

651 39 Eppig, J. T. *et al.* The Mouse Genome Database (MGD): comprehensive resource for
652 genetics and genomics of the laboratory mouse. *Nucleic acids research* **40**, D881-886,
653 doi:10.1093/nar/gkr974 (2012).

654 40 Dai, X. *et al.* A disease-associated PTPN22 variant promotes systemic autoimmunity in
655 murine models. *The Journal of clinical investigation* **123**, 2024-2036,
656 doi:10.1172/JCI66963 (2013).

657 41 Wang, J. H. *et al.* Aiolos regulates B cell activation and maturation to effector state.
658 *Immunity* **9**, 543-553 (1998).

659 42 Chen, X. *et al.* An autoimmune disease variant of IgG1 modulates B cell activation and
660 differentiation. *Science* (2018).

661 43 Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in
662 human populations. *Nature* **449**, 913-918, doi:10.1038/nature06250 (2007).

663 44 Johnson, K. E. & Voight, B. F. Patterns of shared signatures of recent positive selection
664 across human populations. *Nat Ecol Evol* **2**, 713-720, doi:10.1038/s41559-018-0478-6
665 (2018).

666 45 Khera, A. V. *et al.* Genome-wide polygenic scores for common diseases identify individuals
667 with risk equivalent to monogenic mutations. *Nature genetics*, doi:10.1038/s41588-018-
668 0183-z (2018).

669 46 Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108
670 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595
671 (2014).

672 47 Mak, T. S. H., Porsch, R. M., Choi, S. W., Zhou, X. & Sham, P. C. Polygenic scores via
673 penalized regression on summary statistics. *Genetic epidemiology* **41**, 469-480,
674 doi:10.1002/gepi.22050 (2017).

675 48 Vilhjalmsón, B. J. *et al.* Modeling Linkage Disequilibrium Increases Accuracy of Polygenic
676 Risk Scores. *American journal of human genetics* **97**, 576-592,

doi:10.1016/j.ajhg.2015.09.001 (2015).

49 Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *bioRxiv*, 445874, doi:10.1101/445874 (2018).

50 Wojcik, G. L. *et al.* Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514-518, doi:10.1038/s41586-019-1310-4 (2019).

51 Grundbacher, F. J. Heritability estimates and genetic and environmental correlations for the human immunoglobulins G, M, and A. *American journal of human genetics* **26**, 1-12 (1974).

52 Tollerud, D. J. *et al.* Racial differences in serum immunoglobulin levels: relationship to cigarette smoking, T-cell subsets, and soluble interleukin-2 receptors. *J Clin Lab Anal* **9**, 37-41, doi:10.1002/jcla.1860090107 (1995).

53 Mehta, A., Ramirez, G., Ye, G., McGeady, S. & Chang, C. Correlation Between IgG, IgA, IgM and BMI or Race in a Large Pediatric Population. *Journal of Allergy and Clinical Immunology* **129**, AB85, doi:10.1016/j.jaci.2011.12.643 (2012).

54 Roode, H. Serum immunoglobulin values in white and black South African pre-school children. Part I: Healthy children. *J Trop Pediatr* **26**, 104-107, doi:10.1093/tropej/26.3.104 (1980).

55 Lau, Y. L., Jones, B. M., Ng, K. W. & Yeung, C. Y. Percentile ranges for serum IgG subclass concentrations in healthy Chinese children. *Clin Exp Immunol* **91**, 337-341, doi:10.1111/j.1365-2249.1993.tb05905.x (1993).

56 Akkoyunlu, M. TACI expression is low both in human and mouse newborns. *Scand J Immunol* **75**, 368, doi:10.1111/j.1365-3083.2011.02651.x (2012).

57 Kanswal, S., Katsenelson, N., Selvapandian, A., Bram, R. J. & Akkoyunlu, M. Deficient TACI expression on B lymphocytes of newborn mice leads to defective Ig secretion in response to BAFF or APRIL. *Journal of immunology* **181**, 976-990, doi:10.4049/jimmunol.181.2.976 (2008).

58 Treml, L. S. *et al.* TLR stimulation modifies BlyS receptor expression in follicular and marginal zone B cells. *Journal of immunology* **178**, 7531-7539, doi:10.4049/jimmunol.178.12.7531 (2007).

59 Kanswal, S. *et al.* Suppressive effect of bacterial polysaccharides on BAFF system is responsible for their poor immunogenicity. *Journal of immunology* **186**, 2430-2443, doi:10.4049/jimmunol.1002976 (2011).

60 Moir, S. *et al.* Decreased survival of B cells of HIV-viremic patients mediated by altered expression of receptors of the TNF superfamily. *The Journal of experimental medicine* **200**, 587-599, doi:10.1084/jem.20032236 (2004).

61 Isenberg, D. *et al.* Efficacy and safety of atacicept for prevention of flares in patients with moderate-to-severe systemic lupus erythematosus (SLE): 52-week data (APRIL-SLE randomised trial). *Annals of the rheumatic diseases* **74**, 2006-2015, doi:10.1136/annrheumdis-2013-205067 (2015).

62 Chen, L. *et al.* Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. *Human molecular genetics*, doi:10.1093/hmg/ddaa030 (2020).

63 Deelen, P. *et al.* Genotype harmonizer: automatic strand alignment and format conversion for genotype data integration. *BMC Res Notes* **7**, 901, doi:10.1186/1756-0500-7-901 (2014).

721 64 Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide
722 association studies. *Nat Genet* **38**, 904-909, doi:10.1038/ng1847 (2006).

723 65 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer
724 datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).

725 66 Delaneau, O., Marchini, J., Genomes Project, C. & Genomes Project, C. Integrating
726 sequence and array data to create an improved 1000 Genomes Project haplotype
727 reference panel. *Nat Commun* **5**, 3934, doi:10.1038/ncomms4934 (2014).

728 67 Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G. R. Fast and accurate
729 genotype imputation in genome-wide association studies through pre-phasing. *Nature*
730 *genetics* **44**, 955-959, doi:10.1038/ng.2354 (2012).

731 68 Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for
732 genome-wide association studies by imputation of genotypes. *Nat Genet* **39**, 906-913
733 (2007).

734 69 Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity
735 in genome-wide association studies. *Nature genetics* **47**, 291-295, doi:10.1038/ng.3211
736 (2015).

737 70 Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide
738 association scans. *Bioinformatics* **26**, 2190-2191, doi:DOI 10.1093/bioinformatics/btq340
739 (2010).

740 71 So, H. C., Gui, A. H., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by
741 known susceptibility variants: a survey of ten complex diseases. *Genetic epidemiology* **35**,
742 310-317, doi:10.1002/gepi.20579 (2011).

743 72 Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human epigenomes.
744 *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

745 73 Westra, H. J. *et al.* Systematic identification of trans eQTLs as putative drivers of known
746 disease associations. *Nature genetics* **45**, 1238-1243, doi:10.1038/ng.2756 (2013).

747 74 Zhernakova, D. V. *et al.* Identification of context-dependent expression quantitative trait
748 loci in whole blood. *Nature genetics*, doi:10.1038/ng.3737 (2016).

749 75 Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference
750 resource for gene and protein annotation. *Nucleic acids research* **44**, D457-462,
751 doi:10.1093/nar/gkv1070 (2016).

752 76 Fabregat, A. *et al.* The Reactome Pathway Knowledgebase. *Nucleic acids research* **46**,
753 D649-D655, doi:10.1093/nar/gkx1132 (2018).

754 77 Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10**,
755 101-129, doi:10.2307/3001666 (1954).

756 78 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful
757 Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B*
758 *(Methodological)* **57**, 289-300 (1995).

759 79 Nei, M. Analysis of gene diversity in subdivided populations. *Proceedings of the National*
760 *Academy of Sciences of the United States of America* **70**, 3321-3323,
761 doi:10.1073/pnas.70.12.3321 (1973).

762 80 Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection
763 in the human genome. *PLoS biology* **4**, e72, doi:10.1371/journal.pbio.0040072 (2006).

764 81 Gautier, M. & Vitalis, R. rehh: an R package to detect footprints of selection in genome-

765 wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176-1177,
766 doi:10.1093/bioinformatics/bts115 (2012).
767 82 Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC
768 curves. *BMC bioinformatics* **12**, 77, doi:10.1186/1471-2105-12-77 (2011).
769

Table 1 Summary association statistics of newly identified SLE-associated variants

Order	RSid	Chr	Pos (hg19)	AA ^a	AB ^a	Gene context	Annotation ^d	EAS-specific meta-analysis				EUR-specific meta-analysis				Trans-ancestral meta-analysis			Cochran' s Q-test <i>P</i>
								6,707 cases and 16,047 controls				4,576 cases and 8,039 controls				11,283 cases and 24,086 controls			
								Freq ^b	OR ^c	SE	P	Freq ^b	OR ^c	SE	P	OR ^c	SE	P	
#1	rs12093154	1	1178925	G	A	<i>C1QTNF12</i>	missense	0.18	0.84	0.04	4.12E-06	0.10	0.82	0.05	1.38E-04	0.84	0.03	2.51E-09	0.648
#2	rs3795310	1	8431607	C	T	<i>RERE</i>	intronic	0.15	0.88	0.04	1.40E-03	0.46	0.88	0.03	8.03E-06	0.88	0.02	3.36E-08	0.920
#3	rs28411034	1	38276997	G	A	<i>MTF1</i>	UTR3	0.26	0.86	0.03	5.80E-06	0.29	0.87	0.03	5.93E-06	0.86	0.02	1.50E-10	0.860
#4	rs6702599	1	67825399	A	C	<i>IL12RB2</i>	intronic	0.94	0.88	0.06	3.56E-02	0.83	0.82	0.04	1.78E-08	0.84	0.03	3.18E-09	0.302
#5	rs1016140	1	117076547	G	T	<i>CD58</i>	intronic	0.56	0.87	0.02	2.95E-10	0.10	1.00	0.05	9.62E-01	0.89	0.02	1.26E-08	0.007
#6	rs11264750	1	157497160	A	G	<i>FCRL5</i>	intronic	0.08	0.75	0.04	1.95E-12	0.01	0.81	0.16	1.81E-01	0.75	0.04	8.70E-13	0.655
#7	rs12132445	1	170811799	G	A	<i>PRRX1, MROH9</i>	intergenic	0.62	1.17	0.03	6.75E-08	0.46	1.00	0.03	9.67E-01	1.08	0.02	1.74E-04	1.05E-04
#8	rs549669428	1	174895022	T	G	<i>RABGAP1L</i>	intronic	0.06	0.75	0.07	1.85E-05	0.27	0.87	0.04	1.67E-04	0.84	0.03	4.53E-08	0.062
#9	rs1547624	1	192543837	A	T	<i>RGS21, RGS1</i>	intergenic	0.82	1.17	0.03	4.55E-08	0.82	1.06	0.04	1.38E-01	1.13	0.02	1.95E-07	0.025
#10	rs2381401	2	144020974	C	T	<i>ARHGAP15</i>	intronic	0.30	1.18	0.03	1.11E-07	0.18	1.11	0.03	1.97E-03	1.15	0.02	1.73E-09	0.221
#11	rs9630991	2	191432139	G	A	<i>NEMP2, NAB1</i>	intergenic	0.23	0.85	0.03	3.34E-06	0.53	0.85	0.03	6.73E-09	0.85	0.02	1.08E-13	0.971
#12	rs11679484	2	198921604	C	A	<i>PLCL1</i>	intronic	0.18	1.15	0.03	5.59E-05	0.38	1.10	0.03	1.91E-04	1.12	0.02	6.26E-08	0.413
#13	rs3087243	2	204738919	G	A	<i>CTLA4</i>	downstream	0.23	0.88	0.03	6.44E-06	0.44	0.91	0.03	6.61E-04	0.89	0.02	2.29E-08	0.396
#14	rs438613	3	28072086	T	C	<i>LINC01980, CMC1</i>	intergenic	0.49	1.13	0.02	6.46E-08	0.47	1.07	0.03	1.98E-02	1.10	0.02	1.32E-08	0.201
#15	rs4690055	4	2748663	G	A	<i>TNIP2</i>	intronic	0.35	0.91	0.02	4.14E-05	0.47	0.91	0.03	4.73E-04	0.91	0.02	5.28E-08	0.776
#16	rs4697651	4	10721433	C	T	<i>CLNK, MIR572</i>	intergenic	0.16	0.85	0.04	4.49E-05	0.31	0.90	0.03	3.26E-04	0.88	0.02	6.36E-08	0.256
#17	rs6871748	5	35885982	T	C	<i>IL7R, CAPSL</i>	intergenic	0.16	0.88	0.03	1.44E-05	0.28	0.90	0.03	6.40E-04	0.89	0.02	3.96E-08	0.579
#18	rs6927090	6	252145	G	T	<i>DUSP22, IRF4</i>	intergenic	0.22	1.23	0.03	3.48E-10	\	\	\	\	\	\	\	\
#19	rs9387400	6	116694120	C	A	<i>DSE</i>	intronic	0.95	0.74	0.05	3.14E-08	0.37	0.94	0.03	2.96E-02	0.89	0.03	5.89E-06	1.17E-04
#20	rs13260060	8	71218360	G	A	<i>NCOA2</i>	intronic	0.41	0.89	0.02	1.92E-08	0.10	0.99	0.05	7.55E-01	0.90	0.02	7.94E-08	0.044
#21	rs2445610	8	128197088	A	G	<i>PRNCR1, CASC19</i>	intergenic	0.48	0.89	0.02	5.26E-08	0.36	0.96	0.03	1.82E-01	0.91	0.02	2.60E-07	0.027

#22	rs7815944	8	129427518	A	G	<i>LINC00824</i>	intronic	0.27	0.86	0.02	1.78E-09	0.03	0.94	0.08	4.27E-01	0.87	0.02	2.21E-09	0.303
#23	rs4978037	9	21228423	T	C	<i>IFNA17</i>	upstream	0.46	1.12	0.03	7.36E-05	0.78	1.12	0.03	3.92E-04	1.12	0.02	5.68E-08	0.902
#24	rs1405209	9	102585545	T	C	<i>NR4A3</i>	intronic	0.15	1.11	0.03	3.10E-04	0.36	1.12	0.03	8.55E-05	1.11	0.02	2.42E-08	0.816
#25	rs4745876	10	64425126	G	A	<i>ZNF365</i>	intronic	0.23	0.88	0.03	1.36E-06	0.14	0.90	0.04	1.40E-02	0.88	0.02	3.66E-08	0.569
#26	rs10999979	10	73501066	C	A	<i>CDH23</i>	intronic	0.18	1.16	0.03	1.01E-07	0.01	1.28	0.27	3.61E-01	1.17	0.03	4.65E-08	0.726
#27	rs7975703	12	121099302	C	T	<i>CABP1</i>	intronic	0.45	0.87	0.03	1.47E-06	0.21	0.91	0.03	3.89E-03	0.88	0.02	2.99E-08	0.367
#28	rs76725306	13	50177453	G	A	<i>RCBTB1, ARL11</i>	intergenic	0.32	1.15	0.03	1.38E-06	0.05	1.19	0.07	8.22E-03	1.16	0.03	3.60E-08	0.617
#29	rs1885889	13	100091300	A	G	<i>UBAC2, LINC01232</i>	intergenic	0.61	0.87	0.02	7.93E-10	0.84	0.86	0.04	5.69E-05	0.87	0.02	2.05E-13	0.870
#30	rs12148050	14	103263788	A	G	<i>TRAF3</i>	intronic	0.57	0.93	0.02	1.04E-03	0.65	0.87	0.03	1.11E-06	0.91	0.02	2.57E-08	0.062
#31	rs869310	15	77830306	T	G	<i>LINGO1</i>	intergenic	0.23	0.86	0.04	1.11E-05	0.36	0.90	0.03	2.54E-04	0.88	0.02	1.90E-08	0.294
#32	rs9899849	17	7234983	G	A	<i>NEURL4, ACAP1</i>	intergenic	0.29	0.97	0.03	3.32E-01	0.22	1.21	0.03	1.10E-08	1.08	0.02	1.08E-03	1.72E-06
#33	rs2384991	19	18386634	A	C	<i>KIAA1683, JUNB</i>	intergenic	0.34	1.10	0.02	2.88E-05	0.25	1.12	0.03	4.22E-04	1.11	0.02	4.95E-08	0.657
#34	rs13344313	19	18517767	G	A	<i>LRRC25, SSBP4</i>	intergenic	0.10	0.86	0.03	1.35E-05	0.28	0.86	0.03	2.33E-06	0.86	0.02	1.34E-10	0.959
#35	rs405858	19	33106621	C	T	<i>ANKRD27</i>	synonymous	0.31	0.87	0.02	1.70E-08	0.69	0.93	0.03	1.14E-02	0.89	0.02	2.38E-09	0.108
#36	rs3760667	19	49814832	C	T	<i>SLC6A16</i>	intronic	0.35	0.86	0.03	1.38E-06	0.21	0.88	0.03	3.73E-04	0.87	0.02	2.31E-09	0.595
#37	rs10419308	19	55739813	G	A	<i>TMEM86B</i>	intronic	0.06	0.89	0.06	3.60E-02	0.19	0.83	0.04	2.12E-07	0.84	0.03	3.60E-08	0.326
#38	rs6074813	20	1541752	G	T	<i>SIRPD, SIRPB1</i>	intergenic	0.80	1.12	0.03	8.49E-05	0.29	1.12	0.03	1.01E-04	1.12	0.02	3.23E-08	0.930

The EAS-specific meta-analysis was performed on association results from the three GWAS (HK, GZ and CC) and the three Immunochip (KR, BJ and MC) datasets. The EUR-specific meta-analysis involved the four European GWAS datasets described in Method. The trans-ancestral meta-analysis was performed on all ten datasets from East Asians and Europeans. Association results for each dataset are available in **Extended Data Table 4**. Cochran's Q test was used to assess differences in effect size estimates between East Asians and Europeans. SNP rs6927090 failed quality control in the European GWAS due to very low minor allele frequency and low quality on imputation.

^aAA: allele A, AB: allele B; ^bFreq refers to the frequency of allele B in HK GWAS and EUR GWAS 2, respectively. ^cThe odds ratio (OR) is with respect to allele B (allele B would be the risk allele when OR > 1). ^dThe genes and functional consequences of variants were annotated by ANNOVAR.

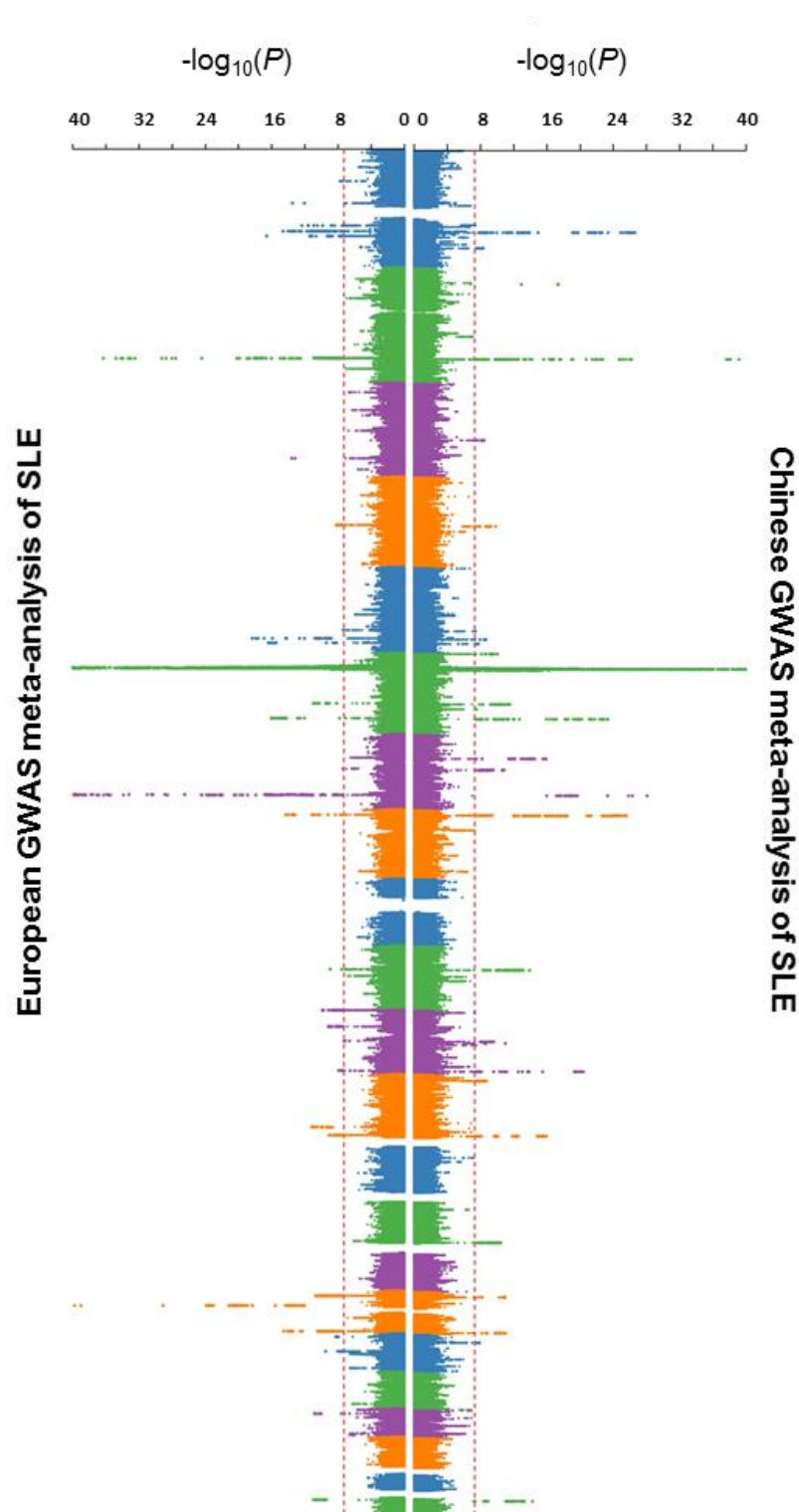


Fig. 1 Manhattan plots for the Chinese and European SLE GWAS meta-analyses results. The Chinese SLE GWAS comprised 4,222 cases and 8,431 controls and the European GWAS comprised 4,576 cases and 8,039 controls. The X-axis is the P -value of association, as $-\log_{10}(P)$, for the meta-analyses of the Chinese (**right**) and European (**left**) ancestries. Red dashed lines indicate the threshold of genome-wide statistical significance ($P = 5E-08$). SNPs with $P < 1E-40$ in an associated locus are not shown from the plot.

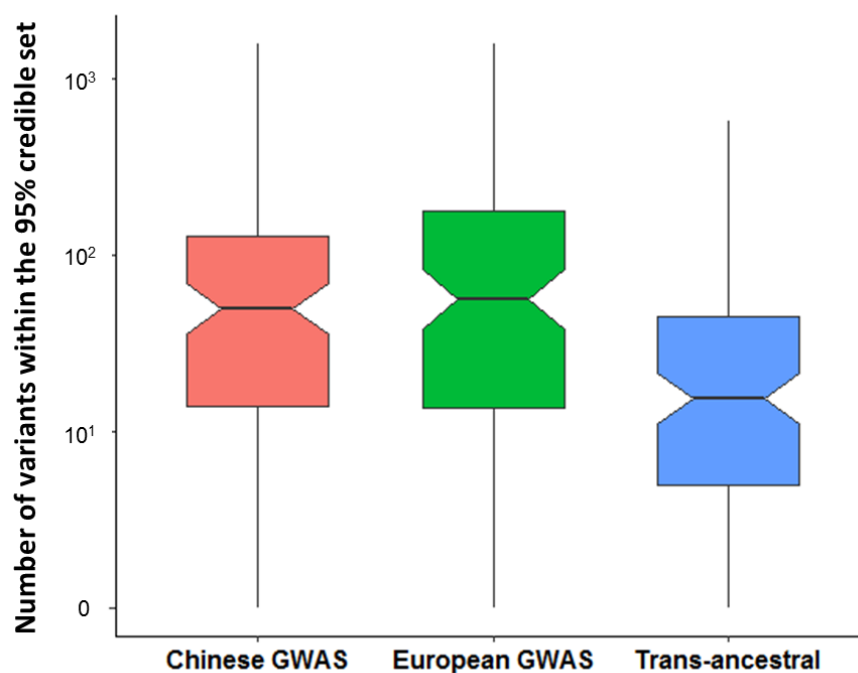
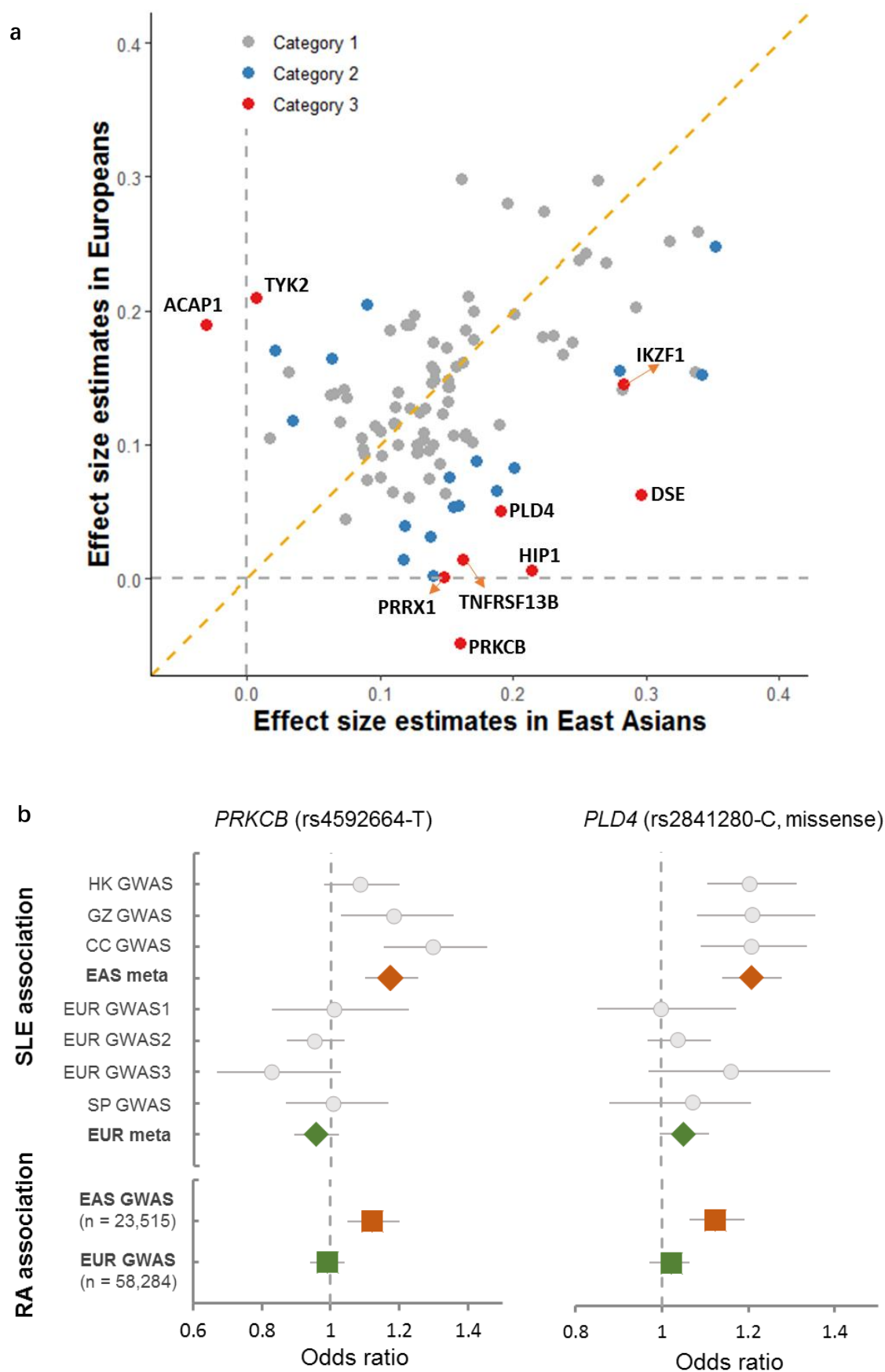


Fig. 2. Fine-mapping across SLE-associated loci based on the association results from the Chinese SLE GWAS, the European GWAS, and the trans-ancestral meta-analyses. The Y-axis indicates the number of potential causal variants at each locus based on the 95% credible sets of the association results from the Chinese SLE GWAS (**red**), the European GWAS (**green**) and trans-ancestral meta-analyses (**blue**).



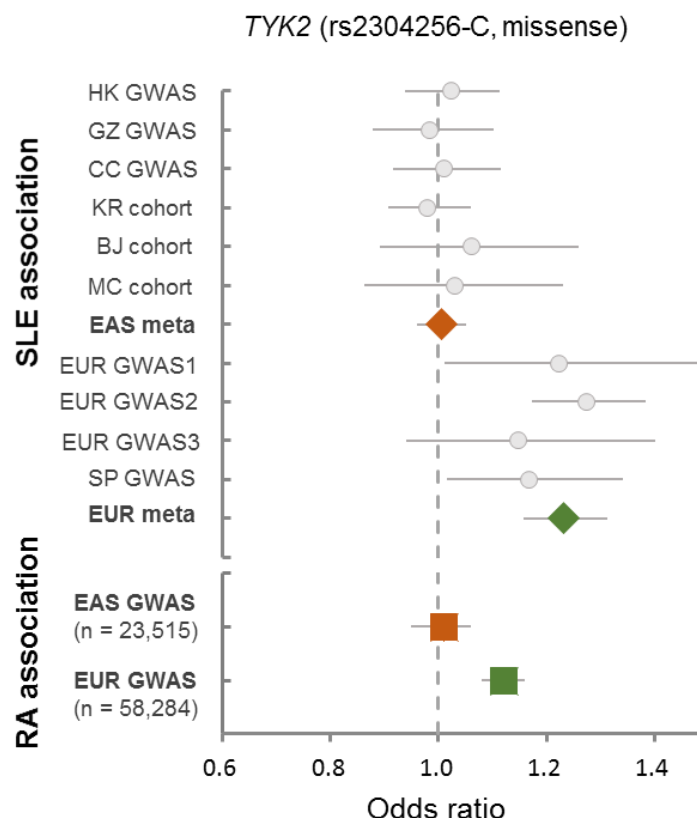


Fig. 3 Genetic loci showing significant ancestral differences in effect-size estimates for SLE and RA.

a. Correlation of effect-size estimates for SLE between East Asian (X-axis) and European (Y-axis) populations. Disease-associated variants with FDR adjusted CQ-test P -value < 0.05 (category 3) are labeled in **red**, and the variants with CQ-test P -value < 0.05 but FDR adjusted CQ-test P -value ≥ 0.05 (category 2) are labeled in **blue**. **b.** Forest plots of association from each cohort at the *PRKCB*, *PLD4* and *TYK2* loci. Diamonds represent the combined estimates of odds ratio (OR) in East Asians (EAS, **red**) and Europeans (EUR, **green**) for association with SLE. Error bars represent the 95% confidence intervals of the OR estimates. Squares represent OR estimates for rheumatoid arthritis (RA) in EAS (**red**) and EUR (**green**) populations. Regional plots for each locus are available in **Extended Data Fig. 10**. HK, Hong Kong; GZ, Guangzhou; CC, Central China; KR, Korean; BJ, Beijing; MC, Malaysian Chinese; SP, Spanish

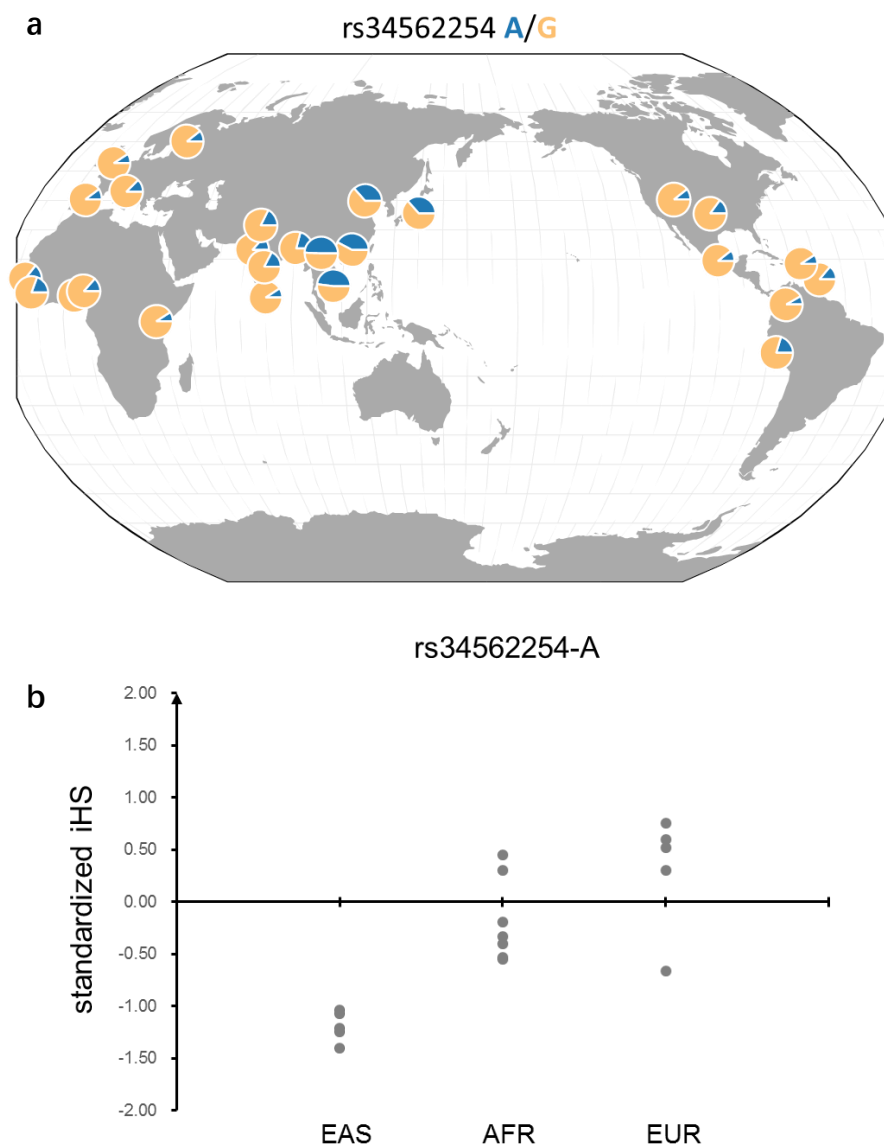
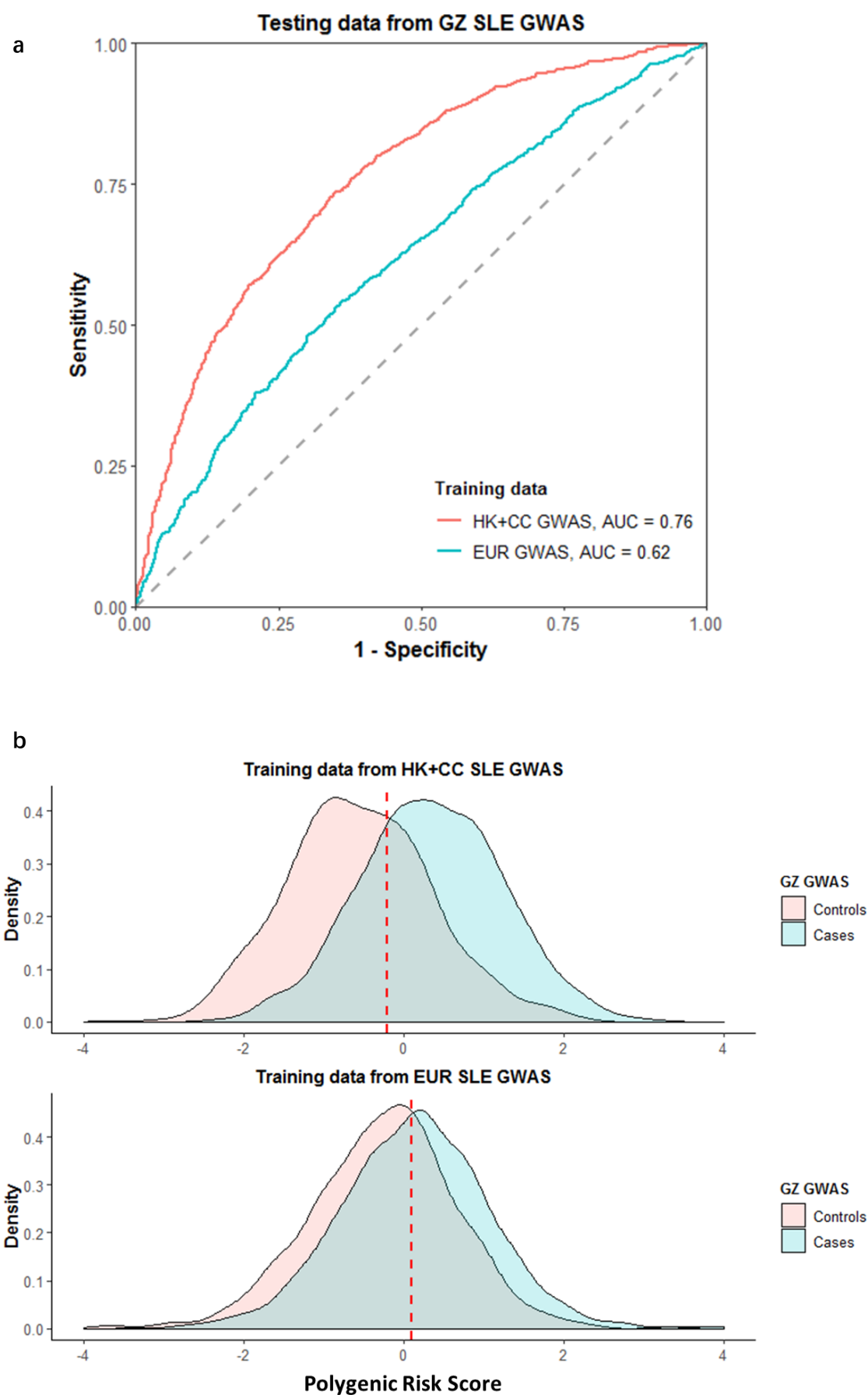


Fig. 4 Risk allele frequency and standardized iHS scores across different populations for the Asian-specific SLE variant at the *TNFRSF13B* locus. **a.** Frequency for the risk allele rs34562254-A across populations. **b.** Standardized iHS for the variants across different continental populations (EAS: East Asians; AFR: Africans; EUR: Europeans). The risk allele rs34562254-A is a derived allele, and negative iHS value indicates that the haplotypes carrying the derived allele are longer than the haplotypes carrying the ancestral allele. The frequency and iHS were calculated using data from the 1,000 Genomes Project.



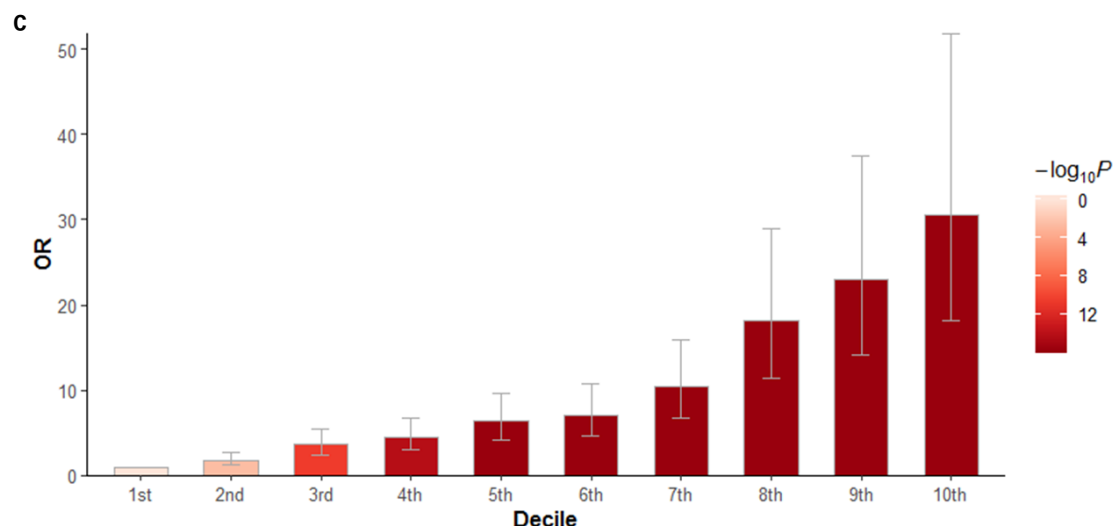


Fig. 5 Performance of polygenic risk scores (PRS) calculated by summary statistics from different ancestral groups. **a.** Performance of PRS evaluated by area under the ROC curve (AUC). PRS for individuals from the GZ cohort were calculated using summary statistics from ancestry-matched Chinese populations (HK and CC GWAS, 2,618 cases and 7,446 controls; **red**) and European populations (4,576 cases and 8,039 controls; **blue**). **b.** Distribution of PRS for SLE cases (**blue**) and controls (**pink**) from the GZ cohort. The PRS distribution was estimated using summary statistics from the ancestry-matched Chinese cohorts (upper panel) or from the mismatched European GWAS (lower panel). The optimal PRS threshold for disease risk prediction is indicated by the red vertical line. **c.** Odds ratios (ORs) of disease risk for individuals in the GZ GWAS across different deciles of PRS calculated by ancestry-matched data. ORs and 95% confidence intervals (bars) were calculated by reference to the 1st PRS decile.