

UCSF

UC San Francisco Previously Published Works

Title

Transancestral mapping and genetic load in systemic lupus erythematosus

Permalink

<https://escholarship.org/uc/item/8wj38401>

Journal

Nature Communications, 8(1)

ISSN

2041-1723

Authors

Langefeld, Carl D
Ainsworth, Hannah C
Graham, Deborah S Cunninghame
et al.

Publication Date

2017

DOI

10.1038/ncomms16021

Peer reviewed

ARTICLE

Received 1 Dec 2016 | Accepted 23 May 2017 | Published 17 Jul 2017

DOI: 10.1038/ncomms16021

OPEN

Transancestral mapping and genetic load in systemic lupus erythematosus

Carl D. Langefeld *et al.*[#]

Systemic lupus erythematosus (SLE) is an autoimmune disease with marked gender and ethnic disparities. We report a large transancestral association study of SLE using ImmunoChip genotype data from 27,574 individuals of European (EA), African (AA) and Hispanic Amerindian (HA) ancestry. We identify 58 distinct non-HLA regions in EA, 9 in AA and 16 in HA (~50% of these regions have multiple independent associations); these include 24 novel SLE regions ($P < 5 \times 10^{-8}$), refined association signals in established regions, extended associations to additional ancestries, and a disentangled complex HLA multigenic effect. The risk allele count (genetic load) exhibits an accelerating pattern of SLE risk, leading us to posit a cumulative hit hypothesis for autoimmune disease. Comparing results across the three ancestries identifies both ancestry-dependent and ancestry-independent contributions to SLE risk. Our results are consistent with the unique and complex histories of the populations sampled, and collectively help clarify the genetic architecture and ethnic disparities in SLE.

Correspondence and requests for materials should be addressed to C.D.L. (email: clangefe@wakehealth.edu) or to P.M.G. (email: gaffney@lupus.omrf.org) or to T.J.V. (email: tim.vyse@kcl.ac.uk).

[#]A full list of authors and their affiliations appears at the end of the paper.

Systemic lupus erythematosus (SLE) (OMIM 152,700) is a chronic autoimmune disease that affects multiple organs, and disproportionately affects women and individuals of non-European ancestry¹. Candidate gene and genome-wide association studies^{2–4} have successfully identified ~90 SLE risk loci that explain a significant proportion of SLE's heritability^{5–8}. These studies have been largely restricted to populations of European ancestry (EA). Yet, much of the heritability of SLE risk remains unexplained in EA populations, and is largely unknown in other ancestries. Here, we report the results of genotyping large samples of individuals of EA, African American (AA) and Hispanic (Amerindian) American ancestry (HA) on the Illumina Infinium ImmunoChip (196,524 polymorphisms: 718 small insertion deletions, 195,806 single nucleotide polymorphisms (SNPs)), a microarray designed to perform both deep replication and fine mapping of established major autoimmune and inflammatory disease loci⁹.

This study identifies 58 distinct non-HLA regions in EA, 9 in AA and 16 in HA. Approximately 50% of the associated regions have multiple independent associations. These 58 regions include 24 novel SLE regions reaching genome-wide significance ($P < 5 \times 10^{-8}$). Further, these results localize the association signals in established regions and extended associations to additional ancestries (for example, EA to AA or HA). Adjusting for the associated HLA alleles disentangles a complex multigenic effect just outside of the HLA region. The association between SLE and the risk allele genetic load (risk allele count) exhibits an accelerating nonlinear trend, greater than expected if the loci were acting independently on risk. This nonlinear risk relationship leads us to posit a cumulative hit hypothesis for autoimmune disease. Finally, we report both ancestry-dependent and ancestry-independent contributions to SLE risk.

Results

SLE genetic association study. In total, 27,574 SLE cases and controls from three ancestral groups were genotyped and passed quality control for the ImmunoChip (AA: 2,970 cases, 2,452 controls; EA: 6,748 cases, 11,516 controls; HA: 1,872 cases and 2,016 controls). Altogether, 146,111 SNPs passed quality control analyses in at least one ancestry (AA: 128,385, EA: 120,873, HA: 120,786). Restricting linkage disequilibrium (LD) to $r^2 < 0.2$ yielded 46,774 uncorrelated SNPs (union across ancestries) for an estimate of the number of independent tests. To minimize ancestry-specific inflation factors, 3, 4 and 2 admixture factors were included as covariates in the logistic regression model

for the EA, AA and HA association analyses, respectively (Supplementary Fig. 1). Inflation factors, scaled to 1,000 cases and 1,000 controls, were $\lambda_{AA,1000} = 1.03$, $\lambda_{EA,1000} = 1.03$ and $\lambda_{HA,1000} = 1.13$ (Supplementary Fig. 2). Power analyses are reported in Supplementary Fig. 3.

Single SNP association. Table 1 shows the number of distinct regions (see Methods) within each ancestry that reached three tiers of statistical significance (Tier 1: $P < 5 \times 10^{-8}$, Tier 2: $5 \times 10^{-8} < P < 1 \times 10^{-6}$ and Tier 3: $P > 1 \times 10^{-6}$ and $P_{FDR} < 0.05$) and lists the number of regions with novel SLE associations. The Tier 1 and Tier 2 thresholds are intentionally more stringent than even the conservative Bonferroni method to reduce the Type 1 error rate on this immune-centric genotyping platform. In total, 5, 38 and 7 distinct non-HLA regions met the Tier 1 threshold of significance for the AA, EA and HA cohorts, respectively; and of these Tier 1 associations, 2, 9 and 2 were novel to SLE regardless of ethnicity or to SLE for a specific ethnicity. An additional 4, 20 and 9 distinct non-HLA regions met the Tier 2 threshold (Fig. 1).

European ancestry. Statistically, EA had the most power and 58 regions met Tier 1 or Tier 2 thresholds (Supplementary Data 2). Many are novel SLE risk regions, and others are novel for EA (Table 2). More than 50% of these regions had multiple independent SNPs contributing to the association, based on regional stepwise analyses. In total, 223 distinct associations met $P_{FDR} < 0.01$ (Tables 1 and 2, Supplementary Table 2), which included both well-established and novel associations.

Novel Tier 1 regions of SLE association in EA and the proximal genes include 4p16 (*DGKQ*), 6p22 (*SLC17A4* and *LRRRC16A*), 6q23 (*OLIG3-LOC100130476*), 8p23 (*FAM86B3P*), 8q21 (*PKIA-ZC2HC1A*) and 17q25 (*GRB2*). Of the 20 EA Tier 2 associated regions, 16 appear novel to SLE.

African American ancestry. The AA sample was powered to detect $OR = 1.1$ to 1.2 at $\alpha = 1 \times 10^{-6}$. In addition to known regions in AA, novel AA regions identified include 5q33 (*PTTG1-MIR146A*), 6p21 (*UHRF1BP1-DEF6*) and 16q22 (*ZFP90*) (Tables 1 and 2; Supplementary Data 2). The 8p11 (*PLAT*) association is novel to SLE and was not observed in HA or EA as it was nearly monomorphic in both populations. The 1q25 region in AA is near the known anti-dsDNA-rs2205960 association between *TNFSF4* and *LOC100506023* in non-AA samples. The association at rs6681482 ($P = 8.11 \times 10^{-7}$, $OR = 0.73$) within *LOC100506023* appears independent and separated from the *TNFSF4* associations by a recombination hotspot. Three SNPs in this region met the stepwise significance threshold, but the strongest association in EA (rs2205960) was not genome-wide significant in AA ($OR = 1.35$, $P = 7.39 \times 10^{-4}$). The association with rs2431697 ($OR = 0.76$, $P = 1.27 \times 10^{-12}$) at 5q33 was previously associated with SLE and anti-dsDNA in EA, but not in AA (ref. 10).

Hispanic ancestry. HA samples had comparable power to the AA sample but exhibited more (nine versus four) novel associations at the Tier 1 and Tier 2 thresholds (Tables 1 and 2). Many regions had multiple independent associations, including cases of previously reported regions exhibiting additional novel loci. Novel Tier 1 regions include 14q31 (*GALC*) and 16p13 (*CLEC16A*). Novel Tier 2 regions include 3p11 (*EPHA3-PROS1*), 6p21 (*TCP11-SCUBE3*), 6q25 (*RSPH3*), 12q15 (*DYRK2-IFNG*), 12q21 (*SYT1*), 16q21 (*CSNK2A2-CCDC113*) and

Table 1 | Number of non-HLA independent regions per significance tier and ancestry (number of novel regions in parentheses*).

Tier and P value threshold	Ancestry		
	African American	European ancestry	Hispanic ancestry
Tier 1: P value $< 5 \times 10^{-8}$	5 (2)	38 (9)	7 (2)
Tier 2: P value $< 1 \times 10^{-6}$	4 (2)	20 (18)	9 (7)
Tier 3: FDR P value [†] < 0.01	18	165	66
Tier 3 Regions Only: FDR P values [‡] (not cumulative)			
0.01–0.05	55	312	154
0.001–0.01	17	119	57
0.0001–0.001	1	40	9
< 0.0001	0	6	0

*For Tier 1 and Tier 2 regions only; novel to SLE or first observed in specific ancestral cohort.

[†]Not cumulative.

[‡]FDR P value represents the Benjamini-Hochberg adjusted false discovery rate P value.

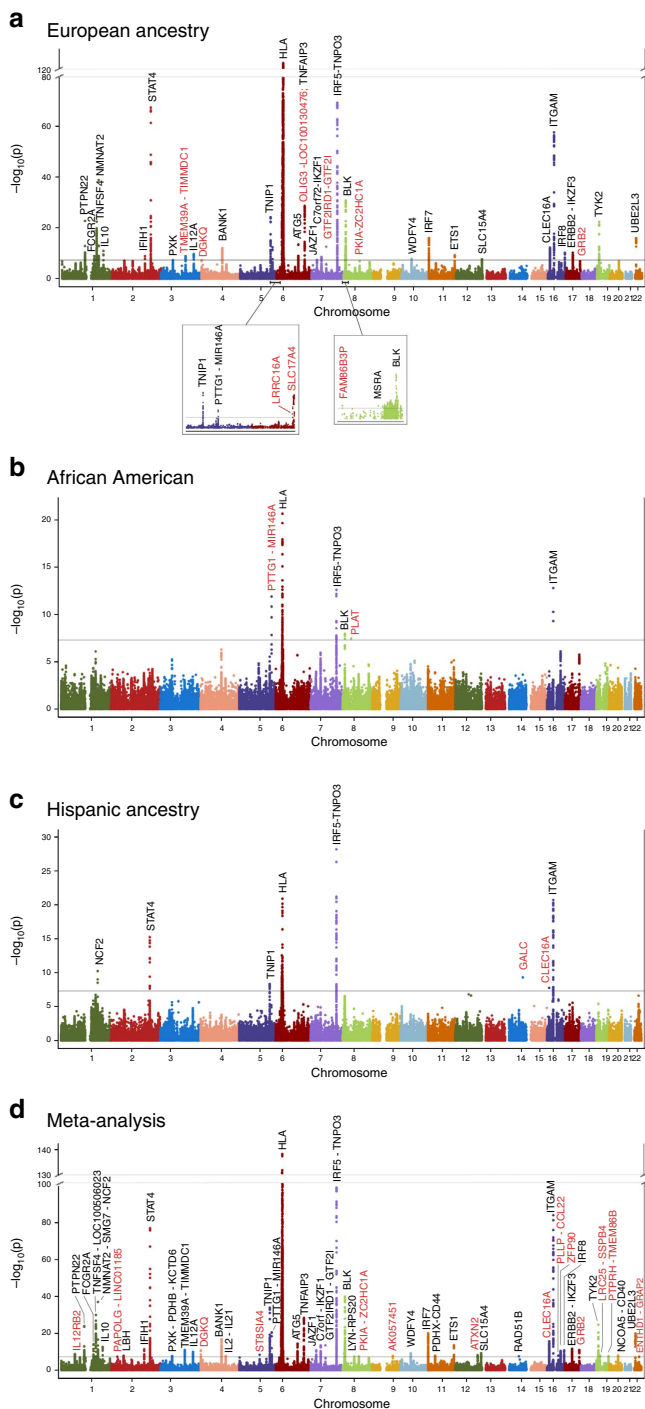


Figure 1 | Genome-wide associations in SLE. Manhattan plots for (a) European ancestry, (b) African American, (c) Hispanic ancestry, and the (d) meta-analysis. Tier 1 associations are labelled with novel regions highlighted in red. Genome-wide significance (5×10^{-8}) is indicated on each plot.

22q12 (*C1QTNF6*). Only the 16p13 locus is associated in AA and EA.

Chromosome X. None of the 442 chromosome X SNPs, predominantly in Xp22 and Xq28, met Tier 1 or Tier 2 thresholds of significance. The strongest evidence of association was in females at Xq28 within *GAB3* (Supplementary Fig. 4; rs2664170;

EA: OR = 0.89, $P = 0.0009$; AA: OR = 0.90, $P = 0.13$; HA: OR = 0.90, $P = 0.33$; Meta $P = 1.23 \times 10^{-4}$).

Two-way interactions among associated SNPs. No SNP-SNP interactions met the Bonferroni threshold ($P = 1 \times 10^{-9}$) (see Methods).

Human leukocyte antigen region. SNP analyses within the HLA region provided strong evidence of association with SLE across groups (Fig. 2). These associations are complicated by the region’s extended LD between SNPs and classical HLA alleles. Supplementary Data 3 and Supplementary Fig. 5 summarize the posterior probability distributions for the imputed four-digit HLA alleles in HLA-A, -B, -C, -DQA1, -DQB1, -DPB1 and -DRB1.

HLA allele associations. HLA allele associations for each ancestry and for multi-ancestral meta-analysis are shown in Supplementary Data 4. To disentangle regional LD effects, ancestry-specific stepwise logistic modelling was used to identify the set of alleles with unique HLA contributions to SLE risk (that is, risk or ‘protective’ alleles associated even after adjusting for other SLE-associated HLA alleles) (Supplementary Data 5). To account for HLA alleles contributing even nominal effects, the models’ entry and exit criteria were set to $P \leq 0.01$ (see Methods). The final models contained both risk and ‘protective’ alleles. In both the single-allele and multi-locus models, class II alleles exhibited the greatest association with SLE. The DR3 (DRB1*3:01-DQA1*05:01-DQB1*02:01) and DR15 (DRB1*15:01/03-DQA1*01:02-DQB1*06:01) haplotypes had the most significant class II risk alleles across populations.

SNP associations after adjusting for HLA alleles. Only two SNPs showed evidence of association with SLE (Supplementary Data 6) after adjusting for the HLA alleles identified in the stepwise modelling (Fig. 2). Specifically, for EA these SNPs are, rs1150755 (OR = 1.33, $P = 3.10 \times 10^{-8}$) within *TNXB* and rs9273448 (OR = 0.64, $P = 2.39 \times 10^{-8}$) within HLA-DQB1 (Supplementary Data 6 and Supplementary Fig. 6). These associations had comparable ORs in the AA and HA cohorts, except in HA for rs9273448. Transancestral meta-analysis showed stronger association at both loci (Supplementary Data 6 and Supplementary Fig. 6). Whether these residual associations reflect novel loci or imperfect imputation requires additional study.

Compound risk allele heterozygosity. In several autoimmune diseases, including lupus¹¹, having two different risk alleles (compound risk allele heterozygosity) generates greater disease risk than having two copies of the same risk allele^{12,13}. In SLE, there are two primary risk haplotypes (DRB1*3:01-DQA1*05:01-DQB1*02:01 and DRB1*15-DQA1*01:02-DQB1*06:01), which are comprised of alleles in strong linkage disequilibrium. Thus, we selected DRB1*03:01 and DR*15 (DRB1*15:01 in EA & HA; DRB1*15:03 in AA) as tagging alleles to evaluate risk allele heterozygosity. Supplementary Data 7 summarizes the genotypic associations and contrasts the effects of risk allele homozygosity, heterozygosity, and compound heterozygosity. In both EA and AA, compound risk allele heterozygosity (DRB1*03:01/*15 provided greater risk than homozygosity for either individual risk allele (that is, DRB1*03:01/03:01; 15/15); these effects are consistent in direction but not significant in HA. Transancestral meta-analysis strongly supports that the risk for compound heterozygotes is greater than homozygotes for any individual allele ($P_{03:01} = 1.79 \times 10^{-10}$; $P_{15:01} = 4.65 \times 10^{-28}$). While there was not conclusive evidence of a statistical interaction for people having these two risk alleles in EA ($P = 0.07$), AA ($P = 0.06$),

Table 2 (Continued).

Most significant SNP(s)	Chr.	Position (b37)	Gene region*	Region rank	Ref. allele	RAF case	RAF control	P value	OR (95% CI)	Regional stepwise P value	dbSNP function†
HA Tier 2											
rs73846279 ^j	3p11	89891345	EPHA3-PROS1	15	T	0.088	0.061	4.82×10^{-7}	1.57 (1.32-1.88)		
rs7653338 ^d	3p11	89938088	EPHA3-PROS1	15	A	0.087	0.060	1.64×10^{-6}	1.58 (1.31-1.91)	1.64×10^{-6}	
rs9394274	6p21	35114911	TCP11-SCUBE3 ¹	8	A	0.285	0.243	6.53×10^{-8}	1.34 (1.20-1.49)	6.53×10^{-8}	
rs12199481 ^d	6q25	159381492	RSPH3	16	C	0.430	0.396	5.72×10^{-4}	0.78 (0.67-0.90)	2.92×10^{-5}	
rs2092540 ^d	6q25	159416444	RSPH3	16	T	0.145	0.198	8.92×10^{-6}	0.73 (0.63-0.84)	6.23×10^{-7}	intron
rs2041862 ^d	12q15	68461697	DYRK2-IFNG	12	A	0.100	0.150	1.59×10^{-7}	0.66 (0.57-0.77)	1.59×10^{-7}	
rs17005500 ^d	12q21	79738884	SYT1	13	C	0.048	0.071	2.43×10^{-7}	0.58 (0.47-0.71)	2.43×10^{-7}	intron
rs2550333	16q21	58267472	CSNK2A2-CCDC113	10	G	0.374	0.292	9.52×10^{-7}	1.28 (1.16-1.41)	9.52×10^{-7}	
rs2731763 ¹	16q21	58280078	CSNK2A2-CCDC113	10	G	0.364	0.279	1.03×10^{-7}	1.31 (1.19-1.45)		
rs229533	22q12	37587111	CIQTNF6	14	C	0.488	0.437	4.61×10^{-6}	1.24 (1.13-1.35)	1.41×10^{-2}	
rs229541	22q12	37591318	CIQTNF6	14	T	0.486	0.427	2.46×10^{-7}	1.27 (1.16-1.39)	1.21×10^{-3}	

Novel regions have not previously been identified by SNP associations with P values $< 5 \times 10^{-8}$ and are highlighted in grey. Regions that are the first observed associations in a particular ancestry are indicated with a superscript¹ in the gene region.

¹: Imputed SNP.

^dor^r: Dominant, or recessive model; if not noted, additive model was used.

^P: Published association—this SNP has been identified as causal or as the most significant SNP in gene region.

*Named by the genes bounding the region of association, unless literature strongly implicated a specific gene.

[†]dbSNP's predicted functional effect.

[‡]The OLG3-LOC100130476 region reaches Tier 1 significance even after adjusting for the TNFAIP3 signal, an established SLE region.

[§]We validated that the MGAT5 region is distinct from the Lactase gene (LCT) on Chromosome 2, by adjusting for the top hit in the LCT gene (rs55634455¹, $P = 4.77 \times 10^{-4}$, OR = 0.87). After this adjustment, the top SNPs were still significant and minimally affected by the adjustment (rs2460382²: $P = 5.17 \times 10^{-3}$, OR = 0.91; rs10496726: $P = 6.65 \times 10^{-5}$, OR = 0.86).

or HA ($P = 0.50$), the lack-of-fit test supported the dominance model of risk (departure from additivity; see Methods) for an individual DR3 (EA $P = 7.90 \times 10^{-109}$; AA $P = 0.06$; HA $P = 5.14 \times 10^{-10}$) and DR15 (EA $P = 5.79 \times 10^{-26}$; AA $P = 3.99 \times 10^{-13}$; HA $P = 3.25 \times 10^{-11}$) SLE risk alleles.

HLA clustering by amino acid. HLA alleles with high sequence similarity, but contrasting ORs, suggest the potential presence of key amino acids influencing disease risk. As expected, clustering amino acid sequences resulted in most two-digit allele subtypes residing within the same clusters (Fig. 3 and Supplementary Fig. 7). When evaluating SLE associations of the three ancestries across these sequence clusters, several noteworthy patterns emerged.

The two primary DRB1 risk alleles, DR3 and DR15 clustered separately, suggesting comparative amino acid dissimilarity. Notably, the closest-clustered neighbours to each risk allele conferred non-risk in these three ancestries. Multi-sequence alignment distinguished the unique or less common amino acids among risk alleles (Supplementary Figs 8–10). Unique to risk alleles DRB1*15:01 and *15:03 were the amino acids Ser-1 (signal peptide), Phe47 and Ala71. Three-dimensional modelling of DRB1 (Supplementary Fig. 8b,c) reveals that these differences mostly reside within the peptide-binding pocket, creating a space of non-polar (hydrophobic) residues, unlike the polar-residue (hydrophilic) space of Tyr47 and Arg71 or Glu71 provided by non-risk alleles within this cluster (Supplementary Fig. 9). Residue 71, among the most variable residues in DRB1 (ref. 14), has been implicated in other diseases¹⁵. Among non-risk alleles with at least 95% identity to DRB1*03:01, the only amino acid unique to this risk allele was Tyr26 (Supplementary Fig. 10). DRB1*03:01 amino acids shared by less than half of the non-risk alleles in this cluster are highlighted in Supplementary Fig. 10 and are concentrated between positions 70–77, spanning the designated 'Shared Epitope' region^{16,17}.

One predominant DQA-DQB1 pair of SLE risk alleles exists per evolutionary DQ-sublineage (Fig. 3b,c)¹⁸. In the DQ2/3/4 sublineage, DQA1*05:01 confers risk across the three cohorts and its heterodimer counterpart, DQB1*02:01, confers risk in EA and HA, but not significantly in AA. Within the DQ5/6 sublineage, both DQA1*01:02 and DQB1*06:02 yield SLE risk across all three cohorts. Comparison of DQA1*01:02 to its closest-related alleles (Supplementary Fig. 11) reveals that DQA1*01:02 (DR15) uniquely encodes a Met207 versus Val207. DQA1*05:01

encodes a polar Thr13 compared to the non-polar Ala13 found in DQA1*05:05 (DR3) and DQA1*05:03 (Supplementary Fig. 12). Identification of specific risk residues was less distinct for the DQB1 risk alleles.

Gender-HLA and genome-wide SNP-HLA interaction. There was no evidence that the risk of SLE differed by gender at any HLA alleles or of a significant SNP-by-HLA allele interaction anywhere across the genome ($P_{FDR} > 0.05$).

Transancestral mapping and top meta-analysis regions. The three-ancestry meta-analysis identified additional SLE-associated regions and was particularly informative for 22 regions, including 11 novel regions, 3 published regions that now meet genome-significance, a complex multigenic region identified by adjusting for HLA alleles and 7 well-established regions more sharply localized by transancestral mapping or novel to these ancestries (Tables 3 and 4; Supplementary Figs 13–15). Supplementary Data 8 and Supplementary Fig. 16 show additional regions that only met genome-wide significance in the meta-analysis. Supplementary Data 9 lists any region with meta-analysis $P_{FDR} < 0.001$.

On 1p31, rs3828069 is within an intron of *IL12RB2* (OR = 0.85, $P = 1.77 \times 10^{-9}$) and has evidence of association in all three ancestries. Although *IL12RB2* is implicated in multiple autoimmune diseases^{19,20}, this specific SNP association with SLE is novel. The 2p16 region exhibited a novel SLE association at rs1432296 (OR = 1.18, $P = 1.34 \times 10^{-8}$) near *PAPOLG-LINC01185*, which includes *REL*. A linkage region at 4p16 (ref. 21) contained a strong novel association for rs3733345 (OR = 0.89, $P = 1.83 \times 10^{-11}$); EA dominated the association, but with significant support from HA and AA. On 8q21, rs4739134 is near *PKIA-ZC2HC1A* (OR = 1.12, $P = 3.47 \times 10^{-8}$) and the AA helped localize the association. The region about 16q13 (*PLLP-CCL22*) exhibited modest association in individual ancestries, but reached genome-wide significance for rs223889 (OR = 1.21, $P = 1.08 \times 10^{-8}$) in the meta-analysis. Similarly, rs137956 (OR = 0.88, $P = 5.0 \times 10^{-8}$) on 22q13 between *ENTHD1* and *GRAP2* was supported across all three ancestries. We bioinformatically explore three additional novel regions.

The meta-analysis about 16q22 (rs1749792; OR = 1.14, $P = 3.66 \times 10^{-11}$) near *ZFP90* had strong support from both EA and AA, with AA samples localizing the association

(Supplementary Fig. 13l). While previously identified in a Chinese cohort, this is the first significant association within EA and AA⁸. Within this region, 27 additional SNPs had a meta-analysis *P* value within one order of magnitude of the maximum association, rs1749792. These 28 SNPs span an interval of 44.6 kb, narrowed from the 100 kb associated region in EA. RegulomeDB²² and HaploReg4.1 (ref. 23) identified 4 of these SNPs with a RegulomeDB score of 1f and 1 with a RegulomeDB score of 2f, indicating they were eQTLs and transcription factor binding sites. HaploReg4.1 showed these five SNPs were enhancers and promoter histone marks in multiple tissues. Interestingly, one of these five, rs1170445, is in high LD with rs1749792 ($R^2_{EA} = 0.99$, $R^2_{AA} = 0.84$, $R^2_{HA} = 0.99$). Here, the G allele is the risk allele and creates a CpG site in the promoter region. In GTEx, the G allele corresponds to lowest gene expression. Hence, when methylated, this variant should result in decreased gene expression of *ZFP90*. The rs1170445-*ZFP90* expression association was reported in GTEx for whole blood ($P = 1 \times 10^{-47}$) and several other tissues (that is, spleen, skeletal muscle, brain cortex, lung, testis and EBV-transformed lymphocytes). Huang *et al.*²⁴ found expression of *ZFP90* in Jurkat T cells led to decreased expression of IL2 and interferon. Furthermore, they found that *ZFP90* protein binds to IL2 and interferon gamma promoters.

SLC15A4 was associated with SLE in the EA cohort and localized by the AA signal in the meta-analysis. The top EA signal was supported by a 43.7 kb region of SLE-associated SNPs exhibiting *P* values within one order of magnitude of the top signal. The meta-analysis narrowed the region of association to four SNPs, spanning 9.5 kb around rs1059312 (Supplementary Fig. 15j). rs1059312 is an eQTL for *SLC15A4* and three supporting SNPs (rs2291349, rs4760593 and rs11059916) altered CpG sites. The region has been previously reported in Asian populations^{25,26}; but this is the first instance of genome-wide significance in EA ($P < 5 \times 10^{-8}$)²⁶.

On 17q25 near *GRB2*, rs8072449 (OR = 0.84, $P = 1.19 \times 10^{-11}$) had modest support in each ancestry, but met genome-wide significance and better localization in the meta-analysis. rs8072449 is an eQTL for *GRB2* (Supplementary Fig. 13m). There were eight additional SNPs with a meta-analysis *P* value within one order of magnitude of the maximum association, and the transancestral analysis reduced the interval of association from 93 to 82 kb. The best RegulomeDB scores for these 9 SNPs was 1f for rs7219, reflecting rs7219 as a known *cis*-eQTL (*NUP85*, *MIF4GD*, *MRPS7*), a transcription binding site and within a DNase peak; in total 7 of the 9 SNPs were reported in transcription binding sites. Interestingly, the top associated SNP, rs8072449, breaks a CpG site and 6 others either end or begin a CpG site. Hence, 7 of the 9 top associated SNPs make or break a CpG site and several are transcription binding sites. Of the 147,111 ImmunoChIP SNPs that passed quality control analyses, only 30% begin or end a CpG site. Although this is a novel SLE association, *GRB2* reportedly regulates SHP2 activity^{27,28}, a potential contributor to SLE pathogenesis²⁹.

A few novel regions, sparsely mapped on the ImmunoChIP, reached genome-wide significance in the meta-analysis and merit further fine-mapping efforts. These include rs6886392 on 5q21 (OR = 1.13, $P = 4.08 \times 10^{-9}$), rs11788118 on 9q22 (OR = 0.88, $P = 1.53 \times 10^{-8}$) and rs13344313 on 19p13 (OR = 0.90, $P = 1.07 \times 10^{-8}$).

Additional loci not previously reported as having genome-wide significance for SLE in these ancestries now do so in the meta-analysis (Table 4). On 4q27, rs11724582 (OR = 0.88, $P = 1.71 \times 10^{-8}$) is near *IL21*, a known SLE risk locus^{30,31}. *IL21* is up-regulated by oestrogen and is produced by T follicular helper cells which stimulates B-cells to differentiate into

autoantibody-secreting cells; however, there was no evidence of a SNP-by-gender interaction in any ancestry ($P > 0.40$). The SNP rs2431098 (OR = 1.19, $P = 3.29 \times 10^{-21}$) at 5q33 between *PTTG1* and *MIR146A* has an $r^2 = 0.52$ with rs2431697, a SNP correlated with down-regulation of *MIR146A*³².

The 6p21 region is potentially confounded with nearby HLA associations. The advantages of using multiple ancestries in this study are exemplified by modelling of SNPs in the 6p21 region where three separate ancestry-specific signals were identified after adjusting for HLA alleles. The results show associations at previously reported *UHRF1BP1* and two novel loci within the *SCUBE3-DEF6* region (Fig. 2 and Supplementary Fig. 13e,f).

The transancestral meta-analyses of several previously established SLE associations provided important localization, and increased the number of independent signals or novel trans-ancestral effects. These included: 1q25 (*TNFSF4-LOC100506023*), 1q25 (*NMNAT2-SMG7-NCF2*), 7q32 (*IRF5-TNPO3*), 8q12 (*LYN-RPS20*), 11p13 (*PDHX-CD44*) and 20q13 (*NCOA5-CD40*) (Table 4, Supplementary Fig. 15).

Admixture and population frequencies of SLE-associated SNPs.

Clustering risk allele frequencies for Tier 1 and 2 SNPs in cases across EA, AA, and HA yielded three groups of SNPs: comparable allele frequencies in all three ancestries (75 SNPs), increased frequency in AA cases (40 SNPs), and reduced frequency in AA cases (66 SNPs) (Fig. 4); the latter two clusters show increased and decreased AA-ancestral contribution, respectively. Higher frequency risk alleles tend to exhibit comparable frequencies across ancestries; the rarest alleles were largely grouped in the reduced AA-ancestral cluster. When comparing admixture averages for risk alleles, AA exhibited the highest deviations from mean admixture estimates and EA, the lowest (Fig. 4; Supplementary Data 10). Deviations from average admixture in risk alleles were significantly weighted to higher proportions of CEU versus YRI in AA ($P = 8.36 \times 10^{-12}$) and HA ($P = 2.44 \times 10^{-4}$) (Supplementary Data 11), further suggesting increased European ancestry for risk alleles. When aligned to allele frequency information, highest CEU proportion deviations in AA and HA resided in the decreased-AA cluster, while the YRI proportion deviations resided in the increased-AA cluster. Thus, SLE risk alleles with a low frequency in AA are correlated with European admixture. Of the 181 Tier 1 and 2 SNPs, only in two regions were the top associated SNP (rs1804182 AA Tier 1 and rs11845506 HA Tier 2) nearly monomorphic (frequency < 0.003) in the other ancestral cohorts. This suggests that most of the ancestry-specific SNP associations were not driven by the presence of monomorphic alleles in the non-discovery cohorts. These allele patterns are further illustrated in Fig. 4.

Genetic load and SLE risk. To explore effects of the number of risk polymorphisms on SLE risk, we computed the genetic risk allele load (unweighted and β -weighted ($\beta = \log(\text{OR})$), see Methods). Here, a set of ORs that contrasted the lowest 10% of the risk-allele count distribution with a sliding window of 20 unweighted, or 4 weighted, counts was computed; these logistic models adjusted for admixture. The pattern of the sliding window ORs was different across ancestries (Fig. 5 and Table 5). Specifically, in 2,000 EA cases and 2,000 EA controls that were independent from the discovery set, a strong and nonlinear effect emerged, with $\text{OR}_{\text{unweighted}} > 30$ and $\text{OR}_{\text{weighted}} > 100$ for the highest load groups. In fact, there was a nonlinear trend in the $\log(\text{OR})$ (that is, β parameter denoting slope) with a greater than additive effect at the highest quarter of the genetic load range (Supplementary Fig. 17); this pattern suggests that the effect of at

Table 3 | Novel non-HLA associated regions identified by transancestral meta-analysis.

SNP	Chr.	Position (b37)	Gene region*	Ref. allele	Ancestry	RAF case	RAF control	P value	OR (95% CI)	dbSNP function†
<i>Tier 1 Meta-Analysis</i>										
rs3828069	1p31	67839573	IL12RB2	G	Meta	-	-	1.77×10^{-9}	0.85 (0.79-0.90)	intron
					EA	0.164	0.182	3.37×10^{-6}	0.87 (0.82-0.92)	
					AA	0.042	0.050	3.13×10^{-2}	0.82 (0.68-0.98)	
					HA	0.200	0.225	6.43×10^{-4}	0.82 (0.74-0.92)	
rs1432296	2p16	61068167	PAPOLG-LINC01185	A	Meta	-	-	1.34×10^{-8}	1.18 (1.10-1.26)	
					EA	0.165	0.147	5.31×10^{-7}	1.17 (1.10-1.24)	
					AA	0.049	0.042	6.41×10^{-2}	1.19 (0.99-1.42)	
					HA	0.083	0.083	3.80×10^{-2}	1.19 (1.01-1.41)	
rs3733345	4p16	954247	DGKQ	G	Meta	-	-	1.83×10^{-11}	0.89 (0.85-0.92)	untranslated-3
					EA	0.444	0.466	5.84×10^{-8}	0.89 (0.85-0.93)	
					AA	0.455	0.482	3.56×10^{-3}	0.89 (0.83-0.96)	
					HA	0.358	0.403	7.03×10^{-3}	0.88 (0.80-0.97)	
rs6886392	5q21	100135865	ST8SIA4	C	Meta	-	-	4.08×10^{-9}	1.13 (1.08-1.18)	
					EA	0.314	0.297	5.16×10^{-6}	1.12 (1.06-1.17)	
					AA	0.257	0.235	8.53×10^{-3}	1.12 (1.03-1.23)	
					HA	0.236	0.224	7.44×10^{-3}	1.16 (1.04-1.29)	
rs34840245	6p21	34812701	UHRF1BP1-DEF6 ¹	G	Meta	-	-	2.37×10^{-11}	1.20 (1.14-1.27)	intron
					EA	0.130	0.106	4.03×10^{-6}	1.18 (1.10-1.26)	
					AA	0.273	0.237	9.06×10^{-5}	1.20 (1.09-1.31)	
					HA	0.122	0.101	1.47×10^{-3}	1.27 (1.10-1.48)	
rs4739134	8q21	79556148	PKIA-ZC2HC1A	T	Meta	-	-	3.47×10^{-8}	1.12 (1.07-1.17)	
					EA	0.290	0.255	2.80×10^{-5}	1.11 (1.06-1.17)	
					AA	0.408	0.377	1.30×10^{-3}	1.14 (1.05-1.23)	
					HA	0.195	0.197	7.00×10^{-2}	1.11 (0.99-1.25)	
rs11788118	9q22	102337331	AKO57451	A	Meta	-	-	1.53×10^{-8}	0.88 (0.84-0.92)	
					EA	0.205	0.224	1.07×10^{-6}	0.88 (0.83-0.92)	
					AA	0.160	0.167	4.12×10^{-1}	0.96 (0.87-1.06)	
					HA	0.146	0.187	4.23×10^{-4}	0.80 (0.71-0.91)	
rs653178	12q24	112007756	ATXN2	C	Meta	-	-	7.39×10^{-9}	1.14 (1.08-1.20)	intron
					EA	0.522	0.491	2.02×10^{-5}	1.10 (1.05-1.15)	
					AA	0.083	0.075	9.97×10^{-2}	1.13 (0.98-1.30)	
					HA	0.288	0.282	2.51×10^{-5}	1.25 (1.13-1.38)	
rs2041670	16p13	11174652	CLEC16A ¹	T	Meta	-	-	2.14×10^{-16}	0.85 (0.82-0.89)	intron
					EA	0.287	0.316	2.34×10^{-12}	0.84 (0.80-0.88)	
					AA	0.536	0.564	2.78×10^{-3}	0.89 (0.83-0.96)	
					HA	0.201	0.251	1.67×10^{-3}	0.84 (0.75-0.94)	
rs223889 ^d	16q13	57392241	PLL-CLL22	T	Meta	-	-	1.08×10^{-8}	1.21 (1.13-1.29)	near-gene-5
					EA	0.307	0.285	1.16×10^{-4}	1.13 (1.06-1.20)	
					AA	0.697	0.685	5.56×10^{-3}	1.29 (1.08-1.55)	
					HA	0.429	0.381	3.22×10^{-4}	1.28 (1.12-1.47)	
rs1749792	16q22	68569440	ZFP90 ¹	T	Meta	-	-	3.66×10^{-11}	1.14 (1.10-1.19)	
					EA	0.253	0.227	4.32×10^{-6}	1.13 (1.07-1.18)	
					AA	0.320	0.274	1.64×10^{-6}	1.21 (1.12-1.30)	
					HA	0.298	0.273	4.50×10^{-2}	1.11 (1.00-1.23)	
rs8072449	17q25	73312184	GRB2	G	Meta	-	-	1.19×10^{-11}	0.84 (0.80-0.89)	
					EA	0.159	0.164	1.08×10^{-6}	0.86 (0.81-0.91)	
					AA	0.804	0.835	6.76×10^{-6}	0.79 (0.71-0.88)	
					HA	0.168	0.193	2.97×10^{-2}	0.88 (0.78-0.99)	
rs13344313	19p13	18517767	LRRC25-SSBP4	A	Meta	-	-	1.07×10^{-8}	0.90 (0.86-0.93)	
					EA	0.259	0.29	7.03×10^{-7}	0.88 (0.84-0.93)	
					AA	0.391	0.407	8.50×10^{-2}	0.93 (0.86-1.01)	
					HA	0.232	0.257	1.45×10^{-2}	0.88 (0.79-0.97)	
rs56154925	19q13	55737798	PTPRH-TMEM86B	T	Meta	-	-	2.27×10^{-8}	0.88 (0.84-0.92)	near-gene-3
					EA	0.164	0.178	1.17×10^{-5}	0.88 (0.83-0.93)	
					AA	0.187	0.207	9.51×10^{-3}	0.88 (0.80-0.97)	
					HA	0.200	0.218	2.02×10^{-2}	0.88 (0.78-0.98)	
rs137956	22q13	40293463	ENTHD1-GRAP2	G	Meta	-	-	5.00×10^{-8}	0.88 (0.84-0.92)	
					EA	0.424	0.446	3.36×10^{-4}	0.92 (0.88-0.96)	
					AA	0.099	0.114	8.50×10^{-3}	0.85 (0.75-0.96)	
					HA	0.475	0.502	2.75×10^{-4}	0.84 (0.77-0.92)	
<i>Tier 2 Meta-Analysis</i>										
rs6662618 ^d	1p22	92935411	GFI1-EVI5	T	Meta	-	-	1.02×10^{-7}	1.14 (1.08-1.21)	
					EA	0.182	0.157	1.54×10^{-6}	1.18 (1.10-1.26)	
					AA	0.450	0.442	1.18×10^{-1}	1.10 (0.98-1.23)	
					HA	0.250	0.229	5.55×10^{-2}	1.14 (1.00-1.29)	

Table 3 (Continued).

SNP	Chr.	Position (b37)	Gene region*	Ref. allele	Ancestry	RAF case	RAF control	P value	OR (95% CI)	dbSNP function†
rs835573	1p12	120464165	NOTCH2	A	Meta	-	-	2.23×10^{-7}	1.13 (1.08-1.18)	intron
					EA	0.138	0.122	2.74×10^{-5}	1.15 (1.08-1.22)	
					AA	0.438	0.415	1.57×10^{-2}	1.10 (1.02-1.19)	
					HA	0.180	0.159	6.44×10^{-2}	1.12 (0.99-1.27)	
rs12068671	1q24	172681031	FASLG	G	Meta	-	-	6.66×10^{-8}	0.88 (0.84-0.92)	
					EA	0.171	0.186	5.92×10^{-6}	0.88 (0.83-0.93)	
					AA	0.383	0.393	2.47×10^{-1}	0.96 (0.88-1.03)	
					HA	0.104	0.138	1.39×10^{-3}	0.79 (0.69-0.91)	
rs7579944	2p23	30445026	LBH ¹	A	Meta	-	-	5.11×10^{-8}	0.90 (0.86-0.93)	
					EA	0.340	0.364	2.03×10^{-4}	0.92 (0.88-0.96)	
					AA	0.698	0.727	5.49×10^{-4}	0.86 (0.79-0.94)	
					HA	0.340	0.367	1.76×10^{-2}	0.89 (0.81-0.98)	
rs461193	5p15	1368997	BC034612	G	Meta	-	-	1.20×10^{-7}	1.13 (1.08-1.18)	intron
					EA	0.214	0.187	1.18×10^{-5}	1.13 (1.07-1.19)	
					AA	0.255	0.241	9.94×10^{-2}	1.08 (0.99-1.18)	
					HA	0.169	0.162	7.95×10^{-3}	1.19 (1.05-1.35)	
rs909788	6p22	16636461	ATXN1 ¹	A	Meta	-	-	1.14×10^{-7}	1.10 (1.06-1.15)	intron
					EA	0.455	0.432	3.26×10^{-5}	1.10 (1.05-1.15)	
					AA	0.676	0.662	1.31×10^{-1}	1.06 (0.98-1.15)	
					HA	0.432	0.402	8.56×10^{-4}	1.17 (1.07-1.29)	
rs11154801	6q23	135739355	AHI1	T	Meta	-	-	3.60×10^{-7}	1.11 (1.06-1.16)	intron
					EA	0.384	0.363	1.12×10^{-5}	1.11 (1.06-1.16)	
					AA	0.115	0.104	5.36×10^{-2}	1.13 (1.00-1.28)	
					HA	0.307	0.298	7.98×10^{-2}	1.09 (0.99-1.21)	
rs7795074	7p15	26742154	SKAP2	A	Meta	-	-	1.93×10^{-7}	0.89 (0.86-0.93)	intron
					EA	0.327	0.339	1.49×10^{-3}	0.93 (0.89-0.97)	
					AA	0.301	0.328	3.43×10^{-3}	0.88 (0.81-0.96)	
					HA	0.281	0.324	4.26×10^{-4}	0.84 (0.76-0.92)	
rs6601327	8p23	9395532	TNKS	C	Meta	-	-	1.46×10^{-7}	1.10 (1.06-1.14)	
					EA	0.398	0.382	5.36×10^{-6}	1.11 (1.06-1.16)	
					AA	0.517	0.493	1.35×10^{-2}	1.10 (1.02-1.19)	
					HA	0.526	0.492	2.21×10^{-1}	1.06 (0.97-1.16)	
rs17630235	12q24	112591686	TRAFD1—C12orf51	T	Meta	-	-	2.50×10^{-7}	1.12 (1.06-1.18)	intron
					EA	0.457	0.422	2.82×10^{-5}	1.10 (1.05-1.15)	
					AA	0.069	0.063	1.91×10^{-1}	1.11 (0.95-1.30)	
					HA	0.256	0.257	1.83×10^{-3}	1.18 (1.06-1.31)	

Novel regions have not previously been identified by SNP associations with P values $< 5 \times 10^{-8}$ and are highlighted in gray

Regions that are the first observed associations in these ancestries are indicated with a superscript ¹ in the gene region.

*Named by the genes bounding the region of association, unless literature strongly implicated a specific gene;

† or ‡: Dominant, or recessive model; if not noted, additive model was used;

‡dbSNP's predicted functional effect.

least a subset of the alleles is greater when the overall genetic load is high. HA and AA showed markedly smaller ORs (between 3 and 10), reflecting the reduced predictive ability of EA-identified SLE risk loci in non-EA populations and the lack of capturing non-EA SLE risk loci on the Immunochip.

The total non-HLA weighted genetic load was correlated with an earlier age at SLE diagnosis in EA ($r_{\text{Spearman}} = -0.14$, $P = 0.0001$), and HA ($r_{\text{Spearman}} = -0.10$, $P = 0.0012$), but not AA ($r_{\text{Spearman}} = 0.04$, $P = 0.54$). Kaplan–Meier curves in the EA showed separation accelerates at ~ 35 years (Supplementary Fig. 18). The HLA-based genetic load was not correlated with age of onset ($P > 0.05$) in any ancestry.

Mapping SNP associations to eQTLs. Many SLE-associated SNPs are, or are in LD with, *cis* eQTLs (Supplementary Data 12 and Supplementary Figs 13–16) and potentially link associations with specific genes. In ancestry-specific eQTL analyses (Supplementary Data 12), EA yielded 96 unique SNPs or their proxies mapping to 193 unique genes, followed by HA (22 unique SNPs; 34 genes) and AA (10 unique SNPs; 17 genes). eQTL analyses based on the meta-analysis SNPs yielded 107 unique genes, identified by 40 SNPs (or their proxies), mostly from

whole blood, monocytes or B-cell derived LCL (Supplementary Data 12). Novel and previously implicated SLE genes were identified (for example, *BANK1*, *IRF5*). Interestingly, a number of SNPs were associated with expression levels for multiple genes. For example, four SNPs were associated with expression levels of at least three genes, and one SNP, newly associated in this study (rs8072449; 17q25), were associated with expression levels of eight genes. Thus, some associated SNPs, either directly or via LD with proxy SNPs, contribute to disease by modifying expression levels of multiple genes, potentially through transcription binding sites. Supplementary Data 13 and 14 provide predicted functional characterization of the 206 SNPs from Tiers 1 to 2 that are in RegulomeDB and HaploReg. These predictions are informative for generating hypotheses that can be experimentally tested.

Discussion

Applying the Immunochip to these multi-ancestral SLE case-control samples has identified 24 novel SLE-risk regions, replicated established SLE-risk loci and extended their impact into other ancestries, and refined association signals via

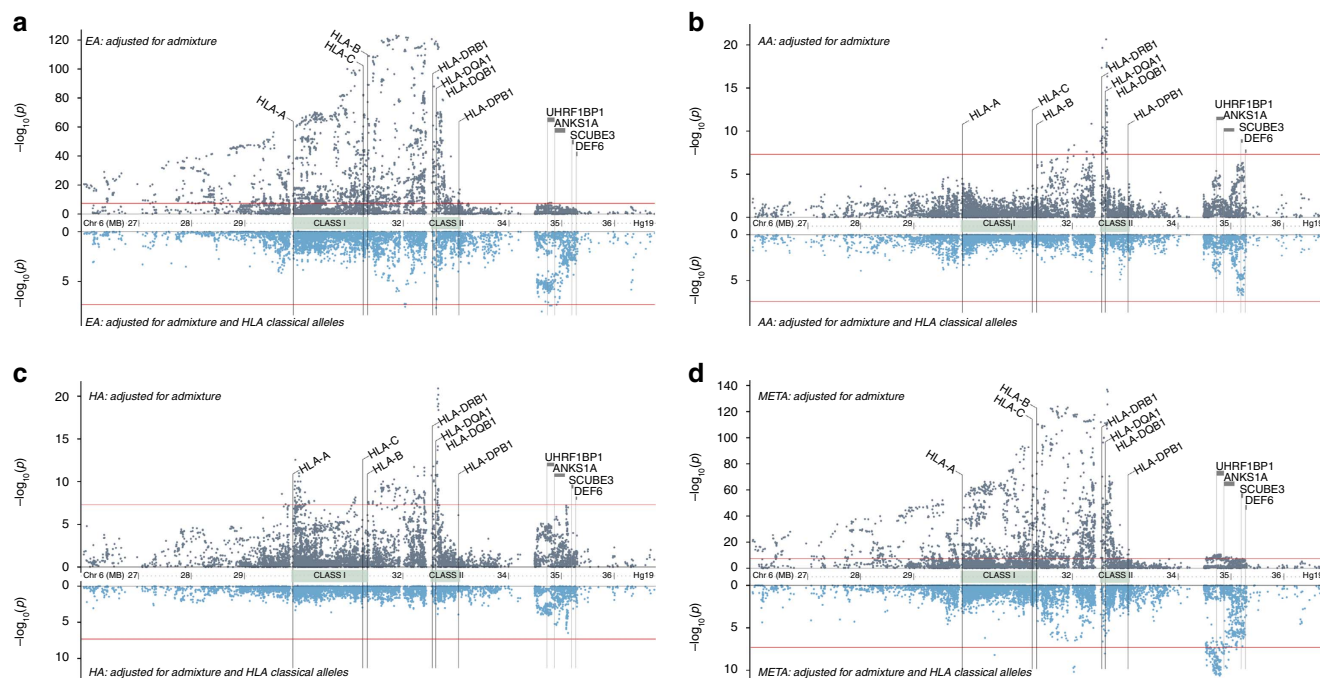


Figure 2 | HLA SNP associations with and without adjustment of classical HLA alleles. SNPs spanning the extended MHC region showed significant associations across (a) European ancestry, (b) African American, (c) Hispanic ancestry, and the (d) meta-analysis. The classical HLA alleles, from the ethnic-specific stepwise-models (Supplementary Data 5), accounted for a majority of the MHC SNP signals. For each plot, the Tier 1 threshold, $P \leq 5 \times 10^{-8}$, is indicated by the red line. Associations, downstream in 6p21 spanning UHRF1BP1-DEF6 were largely unaffected after adjusting for classical HLA alleles and appear independent of the MHC.

transancestral mapping. Over 50% of associated regions had multiple independent SNP associations. Many of these associations were linked via eQTL analysis to specific genes, a process that can accelerate discovery of critical pathways. The contrast of associations and genes across ancestries documents numerous ethnic-specific associations the ancestral diversity in SLE etiology; for example, HA regions not showing equivalent associations in EA include 3p11 (*EPHA3-PROS1*), 6q25 (*RSPH3*), 12q15 (*DYRK2-IFNG*), 12q21 (*SYT1*), 14q31 (*GALC*), 16q21 (*CSNK2A2-CCDC113*) and 22q12 (*CIQTNF6*). In total, these results underscore the shared and distinct genetic profiles of SLE relative to other autoimmune diseases.

To understand disease biology and prevalence across populations, distinguishing shared versus ancestry-specific associations is important because an allele identified in one population is likely relevant in others³³. Clustering by allele frequencies in cases and comparing risk allele admixture estimates, three clusters emerged: (1) alleles with comparable frequencies across populations without strong deviations in average admixture, (2) alleles with increased AA-ancestral contribution and (3) alleles with reduced AA-ancestral contribution and increased CEU admixture. The increased European ancestry observed in less common AA risk alleles likely reflects complex demographic histories and admixture patterns.

The nonlinear nature of how genetic load affects SLE risk leads us to posit the *cumulative hit hypothesis for autoimmune diseases*. That is, in our current environment the immune system can absorb, with a modest increase in risk, individual risk polymorphisms. But as the number of risk variants increases, the system becomes overwhelmed and immune dysregulation occurs. Currently, it is unclear whether it is the entire genetic load or only a subset of variants driving the nonlinear association. In addition, increasing genetic load correlates with an earlier age of disease onset. These hypotheses are testable within specific

and across autoimmune diseases given their shared genetic architecture.

Despite the large sample size, there was no robust evidence for SNP-gender, SNP-SNP or SNP-HLA allele interactions, suggesting that pairwise-interactions among these Immunochip loci are not a major source of missing heritability. While the lack of pairwise interactions across the immune-centric loci may be surprising given the statistical power of the study, the current analysis does not preclude higher-order interactions; albeit agnostic scans for such interactions are analytically challenging. Furthermore, given the nonlinear effect of genetic load on risk, explicit and strong pairwise interactions may not be the correct hypothesis—gene-based or pathway-based interactions may be more important. Because of limitations in the data, gene-environment interactions were not computed and this area needs study.

The individual roles of DR3 and DR15 haplotypes in SLE risk are well-established. However, in all three ancestries, having two different risk alleles yielded higher SLE risk than having two copies of the same risk allele. This is similar to type 1 diabetes, where heterozygotes for type 1 diabetes-associated haplotypes, DR3 and DR4, have shown higher risk of disease. It is hypothesized that this effect is driven via formation of DQA1 and DQB1 trans-heterodimers. In contrast, SLE risk alleles in DR3 and DR15 stem from divergent ancient haplotypes¹⁸; likewise, trans-pairing has not been shown between DQA and DQB in these two haplotypes^{34,35}.

Due to the highly polymorphic nature of HLA alleles and their protein products, it is important to consider high-order relationships among amino acids in three-dimensional space³⁶. Standard regression techniques using amino acids in isolation can be problematic and inappropriate for inference³⁷. To account for higher-order relationships among amino acids, we (1) clustered alleles by protein sequence similarity, (2) compared associations within and between clusters and (3) identified, when possible,

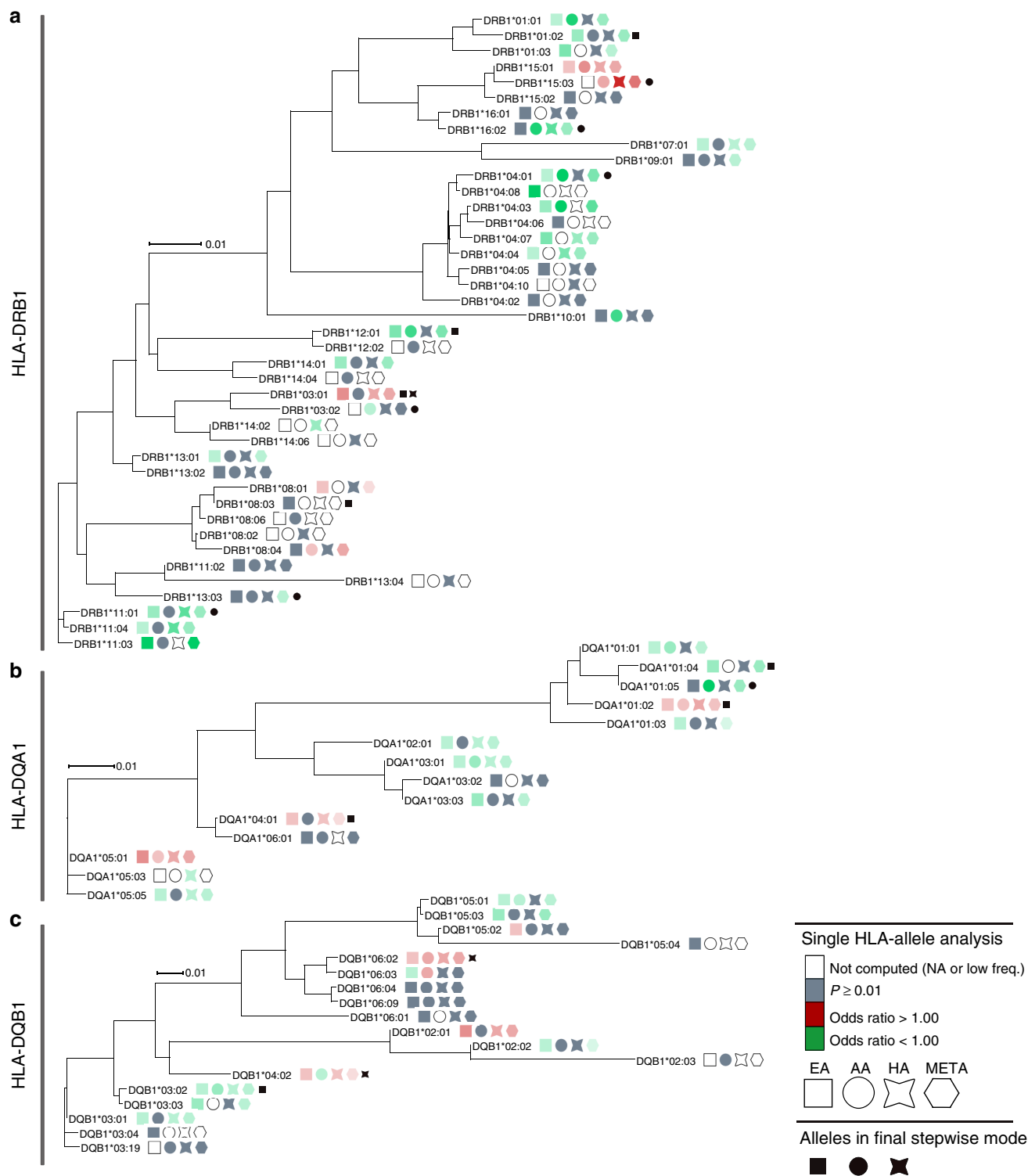


Figure 3 | Clustering of HLA Class II alleles by amino acid sequence similarity. For (a) DRB1, (b) DQA1, and (c) DQB1, the odds ratios for each cohort are superimposed on the cluster if the SLE association *P*-value was less than 0.01. Alleles that were present in the multi-locus model from the stepwise procedure are also denoted. This process aims to identify clusters with shared SLE risk or not-risk odds ratios across the three cohorts. Such clusters help identify potential amino acid sequences contributing to SLE risk. For example, DRB1*15:01 and 15:03 are clustered amongst protective alleles, suggesting presence of specific amino acids differentiating risk (Supplementary Figs 8 and 9).

amino acids that uniquely distinguished the risk alleles. This approach identified several examples of specific amino acids differentiating risk and protective HLA alleles. For example, the DRB15*01 amino acids – 1, 47 and 71 were unique to risk alleles. The combination of Ala71 and Phe47 create a hydrophobic space in the protein binding pocket compared to the alternatives

observed (Glu71 and Tyr47; or Arg71 and Tyr47). In addition to antigen binding, there is a vast array of HLA allele-specific properties, including surface expression stability³⁵, influence of DNA methylation³⁸ and DR-DQ heterodimers³⁹. Such findings may help prioritize functional experiments, as we work towards understanding the HLA mechanisms of SLE.

Table 4 | Tier 1 non-HLA meta-analysis regions noted for transracial mapping.

Gene region*	SNP	Chr.	Position (b37)	Ref. allele	Ancestry	RAF case	RAF control	P value	OR (95% CI)	dbSNP function†
TNFSF4-LOC100506023	rs2205960	1q25	173191475	A	Meta	-	-	1.16×10^{-30}	1.30 (1.23-1.38)	
					EA	0.267	0.225	3.84×10^{-23}	1.29 (1.23-1.36)	
					AA	0.066	0.050	7.46×10^{-4}	1.33 (1.13-1.56)	
					HA	0.390	0.314	2.01×10^{-7}	1.29 (1.17-1.43)	
TNFSF4-LOC100506023	rs1539255	1q25	173322660	T	Meta	-	-	1.60×10^{-19}	0.84 (0.81-0.87)	intron
					EA	0.311	0.350	2.37×10^{-12}	0.85 (0.81-0.89)	
					AA	0.428	0.460	7.24×10^{-4}	0.88 (0.81-0.95)	
					HA	0.273	0.326	1.06×10^{-6}	0.78 (0.71-0.86)	
NMNAT2-SMG7-NCF2‡	rs17484292	1q25	183300050	T	Meta	-	-	9.97×10^{-38}	1.59 (1.40-1.79)	intron
					EA	0.089	0.048	1.48×10^{-39}	1.77 (1.63-1.93)	
					AA	0.012	0.009	2.07×10^{-1}	1.27 (0.88-1.85)	
					HA	0.051	0.035	3.07×10^{-5}	1.61 (1.29-2.02)	
NMNAT2-SMG7-NCF2	rs10911363	1q25	183549757	A	Meta	-	-	2.52×10^{-17}	1.17 (1.13-1.22)	intron
					EA	0.317	0.275	1.18×10^{-13}	1.19 (1.14-1.25)	
					AA	0.329	0.319	2.40×10^{-1}	1.05 (0.97-1.14)	
					HA	0.399	0.333	3.73×10^{-7}	1.27 (1.16-1.39)	
IL2-IL21	rs11724582	4q27	123391464	C	Meta	-	-	1.71×10^{-8}	0.88 (0.84-0.93)	
					EA	0.262	0.280	2.49×10^{-6}	0.89 (0.84-0.93)	
					AA	0.129	0.152	9.86×10^{-4}	0.83 (0.75-0.93)	
					HA	0.142	0.166	3.58×10^{-1}	0.94 (0.83-1.07)	
PTTG1-MIR146A	rs2431098	5q33	159887336	C	Meta	-	-	3.29×10^{-21}	1.19 (1.14-1.23)	
					EA	0.532	0.497	5.09×10^{-13}	1.17 (1.12-1.23)	
					AA	0.404	0.351	1.49×10^{-8}	1.25 (1.16-1.36)	
					HA	0.470	0.441	4.62×10^{-3}	1.14 (1.04-1.25)	
IRF5-TNPO3	rs4728142	7q32	128573967	T	Meta	-	-	3.38×10^{-84}	1.44 (1.39-1.50)	
					EA	0.531	0.446	6.21×10^{-51}	1.40 (1.34-1.46)	
					AA	0.327	0.264	1.16×10^{-12}	1.35 (1.24-1.47)	
					HA	0.518	0.394	2.10×10^{-27}	1.65 (1.51-1.81)	
IRF5-TNPO3	rs35000415	7q32	128585616	T	Meta	-	-	1.17×10^{-99}	1.82 (1.69-1.96)	intron
					EA	0.184	0.118	5.67×10^{-70}	1.73 (1.63-1.84)	
					AA	0.040	0.022	1.80×10^{-7}	1.86 (1.47-2.34)	
					HA	0.298	0.158	7.11×10^{-33}	1.98 (1.77-2.22)	
LYN-RPS20	rs2953898	8q12	56980803	A	Meta	-	-	4.43×10^{-8}	0.84 (0.79-0.90)	untranslated-3
					EA	0.218	0.226	6.89×10^{-5}	0.90 (0.85-0.95)	
					AA	0.029	0.038	1.01×10^{-2}	0.77 (0.63-0.94)	
					HA	0.124	0.167	3.60×10^{-3}	0.82 (0.72-0.94)	
PDHX-CD44	rs353592	11p13	35119482	T	Meta	-	-	1.35×10^{-8}	0.89 (0.85-0.93)	
					EA	0.441	0.475	4.89×10^{-6}	0.90 (0.86-0.94)	
					AA	0.113	0.127	4.14×10^{-2}	0.88 (0.79-1.00)	
					HA	0.286	0.339	4.89×10^{-3}	0.87 (0.79-0.96)	
SLC15A4	rs1059312	12q24	129278864	G	Meta	-	-	6.53×10^{-10}	1.12 (1.07-1.16)	coding-synon
					EA	0.423	0.392	6.26×10^{-7}	1.12 (1.07-1.17)	
					AA	0.497	0.461	1.84×10^{-4}	1.16 (1.07-1.25)	
					HA	0.615	0.571	2.17×10^{-1}	1.06 (0.97-1.16)	
NCOA5-CD40	rs4810485 †	20q13	44747947	A	Meta	-	-	9.95×10^{-9}	1.43 (1.17-1.76)	intron
					EA	0.275	0.248	1.01×10^{-7}	1.38 (1.23-1.56)	
					AA	0.079	0.080	2.16×10^{-1}	1.53 (0.78-3.02)	
					HA	0.216	0.208	2.37×10^{-2}	1.42 (1.05-1.92)	

The corresponding plots can be found in Supplementary Fig. 15.

‡ or †: Dominant, or recessive model; if not noted, additive model was used.

*Named by the genes bounding the region of association, unless literature strongly implicated a specific gene.

†dbSNP's predicted functional effect.

‡SNP's association is not supported by LD SNPs. Cluster call plot was verified for quality control. Additional verification of association will be required.

Two major limitations of this study are the comparably fewer non-EA SLE cases and appropriate controls, and the strong EA bias in the Immunochip content. Power calculations using allele frequencies and ORs from EA, and the number of AA cases and controls, yielded 445.5 expected Tier 1 and 2 SNP associations; however, only 64 were observed. Although differences in LD contribute to this result, the highly reduced number of detected associations relative to expected, plus the genetic load analyses, strongly suggest that ancestry-specific and -independent loci contribute to SLE risk. It is imperative to recruit more non-EA populations for genetic studies.

In conclusion, SLE has a strong genetic contribution to risk with ancestry-dependent and ancestry-independent contributions. SLE risk has shared and independent genetic contributions relative to other autoimmune diseases. This genetic risk manifests itself as a nonlinear function of the cumulative risk allele load, a pattern potentially shared across autoimmune and non-autoimmune diseases.

Methods

Study cohort. Multiple studies provided de-identified DNA samples with approval from their respective institutional review boards or ethics committees. These ethics review committees included: Cedars-Sinai Medical Center

Table 5 | Genetic Load and SLE risk.

	non-HLA SNPs from EA Top Hits* (N = 545)			HLA classical alleles from EA top hits* (N = 10)			Combined (N = 555)		
	P value	OR (95% CI) [†]	c-statistic [‡]	P value	OR (95% CI) [†]	c-statistic [‡]	P value	OR (95% CI) [†]	c-statistic [‡]
Un-Weighted									
AA	5.31 × 10 ⁻³⁰	1.15 (1.13-1.18)	0.591 (0.590)	5.75 × 10 ⁻⁷	2.44 (1.72-3.45)	0.540 (0.538)	2.84 × 10 ⁻³¹	1.13 (1.11-1.16)	0.593 (0.592)
EA [§]	6.15 × 10 ⁻⁶²	1.23 (1.20-1.26)	0.694 (0.657)	7.20 × 10 ⁻⁴¹	5.15 (4.05-6.55)	0.670 (0.608)	5.24 × 10 ⁻⁶⁹	1.24 (1.21-1.28)	0.702 (0.665)
HA	1.76 × 10 ⁻²⁶	1.14 (1.12-1.17)	0.647 (0.624)	9.39 × 10 ⁻²⁰	3.41 (2.62-4.44)	0.638 (0.560)	7.02 × 10 ⁻³⁰	1.15 (1.12-1.18)	0.651 (0.630)
Weighted by Natural Log of the Odds Ratio (OR)									
AA	2.82 × 10 ⁻³⁶	2.59 (2.23-3.00)	0.602 (0.601)	9.34 × 10 ⁻¹¹	24.11 (9.20-63.17)	0.551 (0.550)	1.80 × 10 ⁻⁴⁰	2.71 (2.34-3.14)	0.608 (0.607)
EA [§]	1.37 × 10 ⁻¹⁰¹	7.83 (6.49-9.46)	0.738 (0.714)	7.60 × 10 ⁻⁴⁵	29.06 (18.16-46.52)	0.678 (0.618)	2.19 × 10 ⁻¹²¹	8.48 (7.09-10.15)	0.759 (0.734)
HA	2.09 × 10 ⁻⁴⁶	3.76 (3.14-4.51)	0.674 (0.660)	8.61 × 10 ⁻²⁴	34.33 (17.24-68.37)	0.645 (0.582)	1.98 × 10 ⁻⁵⁷	4.26 (3.57-5.09)	0.687 (0.675)

*Top hits from EA sample without validation set of 2,000 SLE cases, 2,000 controls.
[†]Per 5 alleles.
[‡]Whole model statistic, and in parentheses, the c-statistic for model without admixture factors.
[§]EA random sample with 2,000 SLE cases, 2,000 controls.

Institutional Review Board; Central Ethic Committee of Denmark; Centrala etikprövningsnämnden; Comité de Ética de la Investigación de Centro Hospital Universitario Virgen Macarena; Centro de Estudios Reumatológicos. Santiago de Chile; Centro Hospitalar Universitário do Porto, Unidade de Imunologia Clínica e Comissão de Ética; CEPI (Comite de Ética de Protocolos de Investigación) Institution: Hospital Italiano de Buenos Aires; Cincinnati Children's Hospital Medical Center Institutional Review Board; Clinical Research Unit, Padua University-Hospital, and Ethics Committee, Province of Padua; Comitato Etico Interaziendale AOU Maggiore della Carità Ethics Committee, Novara, Italy; Comité de Bioética del Consejo Superior de Investigaciones Científicas; Comité de Docencia e Investigación, Hospital Escuela Eva Perón, Gro Baigorria, Santa Fe, Argentina; Comité de Docencia e Investigación, Sanatorio Parque SA; Comité de ética de la investigación del HIGA San Martín de La Plata, Argentina; Comité de Ética en Investigación Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán; Comité de Ética en Investigación, Instituto Nacional de Medicina Genómica, Mexico; Comité de Ética en Investigación; Comité de Investigación de la Facultad de Medicina de la UANL y Hospital Universitario 'Dr José Eleuterio González'; Comité Docencia e Investigación H.I.G.A. Dr Oscar Alende Mar del Plata; Comité Étici d'Investigació Clínica de l'Hospital Clínic de Barcelona; Comités de Ética, Bio Ética y de Investigación. Hospital G. Almenara, Salud, Lima, Perú; Comites de Ética, Bioética y de Investigación Hospital Nacional Guillermo Almenara Irgoyen, Lima-Perú; Commission d'Éthique Hospitalo-Facultaire de l'Université catholique de Louvain; Duke University Health System Institutional Review Board; Ethics and Research Committee of Hospital General De Occidente; Fundacion Docencia e Investigación Hopsital Italiano de Cordoba; Institution of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden; Institutional Review Board of the University of Puerto Rico Medical Sciences Campus; Institutional Review Board Office Northwestern University; Johns Hopkins University School of Medicine Institutional Review Board; London Central Research Ethics Committee Study sponsor: King's College London; Medical Ethical Committee (METC) of the University Medical Center Groningen; Medical University of South Carolina Institutional Review Board for Human Research; Northwell Health Human Research Protection Program; Oklahoma Medical Research Foundation Institutional Review Board; omisión Nacional de Investigación Científica y Comisión de Ética en Investigación en Salud, Instituto Mexicano del Seguro Social, México; Regional Ethical Review Board at Karolinska Institutet, Stockholm, Sweden; Regional Ethics Review Board in Linköping; Regional Human Medical Research Ethics Committee of the University of Szeged; SickKids REB; The Institution Review Boards for human research at UCLA; The Local Ethics Committee of the Karolinska University Hospital/Karolinska Institutet, Stockholm Sweden; The University Health Network, Research Ethics Board; Institutional Review Board for Human Use University of Alabama at Birmingham; UC Davis Institutional Review Board; UCSF Human Research Protection Program Institutional Review Board; UHN REB; University Health Network Research Ethics Board and by the local ethics boards of the CaNIOS investigators at the following centres: Montreal General Hospital, St Joseph's Health Centre, Winnipeg Health Science Center, Queen Elizabeth II Health Sciences Centre, Ottawa Hospital, Hôpital Notre-Dame, Calgary Health Sciences Centre, Centre Hospitalier Universitaire de Sherbrooke, and Hôpital Maisonneuve-Rosemount; University Hospital of Gran Canaria Doctor Negrin Research Ethic Committee; University of Chicago Institutional Review Board; University of Southern California Health Sciences Institutional Review Board; University of Texas Southwestern Medical Center Institutional Review Board; Uppsala Ethical Review Board; Wake Forest University School of Medicine Institutional Review Board. All study participants provided written consent prior to study enrolment at the institution where the samples were collected. All SLE cases in this study were required to meet at least four of the eleven American College of Rheumatology classification criteria for SLE^{40,41}.

Genotyping and quality controls. Samples were genotyped on the custom-designed Immunochip Illumina Infinium Assay⁹ according to Illumina's protocols, using the Illumina iScan scanner at the following centres: Oklahoma Medical

Research Foundation, University of Texas Southwestern, HudsonAlpha Institute for Biotechnology, North Shore-LIJ Health System's Feinstein Institute for Medical Research. Intensity data were generated for all samples and sent to the Oklahoma Medical Research Foundation for genotype calling using OptiCall⁴². OptiCall default options were used with one exception: the 'nointcutoff' option was included to allow removal of intensity outliers. Subsequent genotype clusters were viewed against their intensity data using Evoker⁴³. Genotype calling was completed in four batches, keeping samples genotyped at the same center in the same batch. Batches were designed to include samples of multiple ancestries when possible to improve rare variant calling. The ancestry breakdown for the batches was: Batch I was 15% European ancestry (EA), 7% African American ancestry (AA), 55% Asian ancestry (ASA), 23% Hispanic ancestry (HA); Batch II was 44% EA, 18% AA, 1.4% ASA, 36% HA; Batch III was 48% EA, 38% AA, 1% ASA, 13% HA; and Batch IV was 92% EA, 8% AA. Some samples called with the SLE Immunochip study samples were used for other Immunochip studies.

Samples were excluded if their call rates were <98% across SNPs that passed quality control filters. Duplicates and first-degree relatives were removed, retaining the sample with the highest call rate. The Immunochip does not have sufficient markers in the non-pseudoautosomal regions of chromosome X to reliably complete gender checks. Admixture estimates were computed using the program ADMIXTURE⁴⁴. HapMap phase 2 individuals (CEU: Utah residents with ancestry from northern and western Europe; YRI: Yoruba in Ibadan, Nigeria; CHB: Han Chinese in Beijing, China) as anchoring populations. To facilitate testing for association between rare variants and SLE, and to improve multilocus modelling in regions of linkage disequilibrium (LD) among SNPs, a factor analysis was computed on the admixture estimates using principal component extraction and varimax rotation⁴⁵. The resulting factors are orthogonal (independent) and thereby remove collinearity among the admixture estimates when used as covariates in linear models. Reduced collinearity should facilitate more robust analysis of rare variants. In addition, principal component (PC) analysis was computed using Eigensoft v4.2 (refs 46,47) including HapMap phase 2 individuals (CEU, YRI and CHB) as reference populations. Both the admixture and PC analyses were completed using a subset of SNPs generated by removing SNPs in LD ($r^2 > 0.2$), with minor allele frequency (MAF) <0.01, or with low call rate (<95%).

The admixture estimates and PCs were used to identify and remove genetic outliers. A SNP was removed from the primary analysis if it had an overall call rate <95%, exhibited significant differential missingness between cases and controls ($P < 0.05$), had significant departure from Hardy-Weinberg equilibrium expectations ($P < 1 \times 10^{-6}$ in cases, $P < 0.01$ in controls) or a cluster separation score <0.40. SNPs violating the above Hardy-Weinberg equilibrium thresholds were retained if there was convincing evidence of association at SNPs in linkage disequilibrium (LD) and the cluster plots indicated that the pattern was not due to poor genotype calling. Primary inference was based on SNPs with MAF ≥ 0.01 . Finally, >10,000 SNP cluster plots were visually examined, including all SNPs reported, to remove results potentially based on poor genotyping.

To provide an estimate of the number of independent tests for multiple comparisons adjustment, the SNPs were LD pruned, $r^2 < 0.20$, within each ancestry. The union of these SNPs across ancestries was 46,744 uncorrelated SNPs, yielding a Bonferroni threshold of $P < 1.06 \times 10^{-6}$.

Statistical analysis. Regions in figures and tables are named by the genes bounding the regions of association or regions of significance for other statistical test, unless the literature strongly implicated a specific gene.

To test for an association between a SNP and case/control status within an ancestry, a logistic regression analysis was computed adjusting for admixture factors as covariates. Primary inference was based on the additive genetic model unless there was significant evidence of a lack-of-fit to the additive model ($P < 0.05$). If there was evidence of a departure from an additive model, then inference was based on the most significant of the dominant, additive, and recessive genetic models. The additive and recessive models were computed only if there were at least 10 and 30 individuals homozygous for the minor allele, respectively.

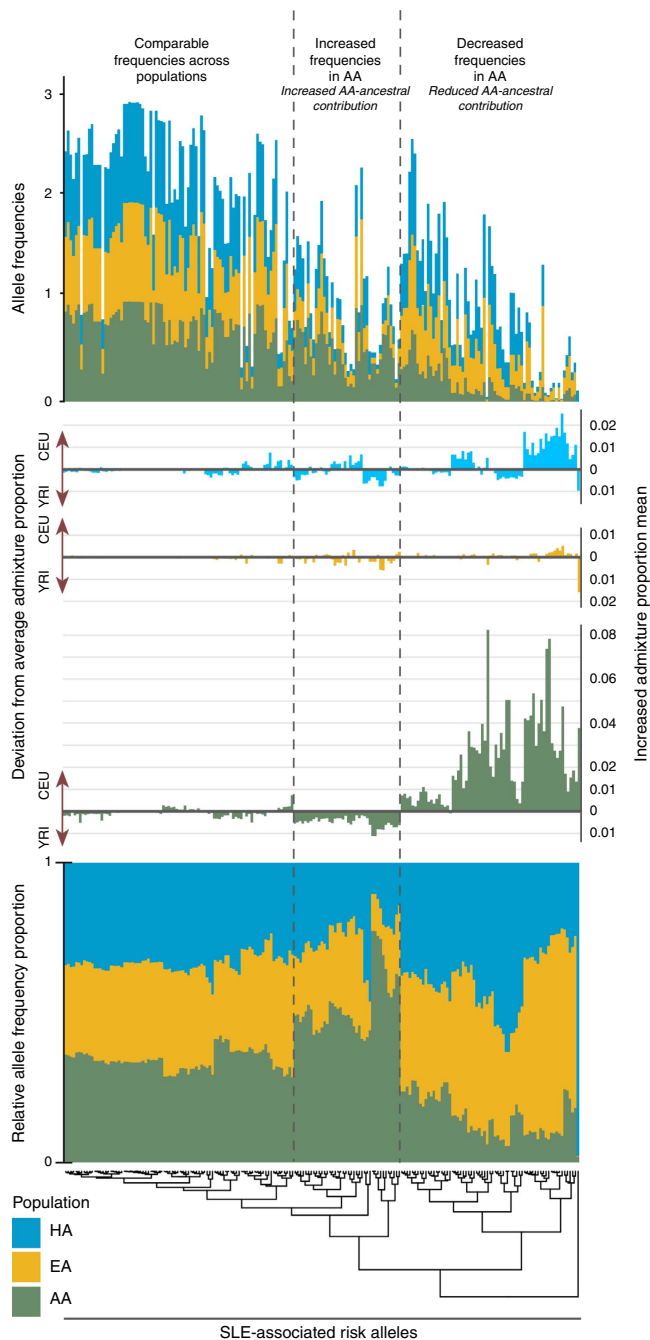


Figure 4 | Ancestral landscape of SLE risk alleles. Clustering by relative allele frequency yields three distinctive categories for SLE risk alleles: comparable frequencies across populations, increased frequencies in AA, and decreased frequencies in AA. The comparable frequency grouping contained the most risk alleles, of which, many were common alleles. This cluster had the smallest deviations from average admixture proportions, across the three cohorts. The increased frequencies in AA alleles exhibited moderate deviations towards greater AA-ancestral contribution. The largest deviations from average admixture were found within alleles exhibiting decreased frequencies in AA. These alleles were enriched for admixture deviations of increased CEU-ancestry. The patterns across relative allele frequencies reveal that ancestry-specific associations are largely not driven by monomorphic SNPs in other populations.

These tests of association were computed using the SNP-GWA version 4.0 module of SNPLASH (<https://www.phs.wakehealth.edu/public/bios/gene/downloads.cfm>). For ancestry-specific analysis of the X chromosome, the data were first stratified by gender and then meta-analysed using the weighted inverse normal method

(weighted by sample size). The genomic control inflation factor (λ_{GC}) was calculated using a set of SNPs included on the Immunochip for a study investigating the genetic basis for reading and writing ability. The resulting λ_{GC} was scaled to 1,000 cases and 1,000 controls to standardize comparisons across populations and studies.

Three tiers of statistical significance are reported. Tier 1 includes those SNPs that meet the literature-motivated genome-wide threshold of 5×10^{-8} . Tier 2 includes those SNPs that are not Tier 1 SNPs, but have a P value for association less than 1×10^{-6} . Tier 3 includes those SNPs that do not meet criteria for Tiers 1 or 2, but meet a genome-wide Benjamini–Hochberg false discovery rate⁴⁸ adjusted P value threshold of 0.05. The Tier 2 threshold meets the strict Bonferroni criteria for the number of uncorrelated SNPs ($r^2 < 0.20$).

Ancestry-specific logistic regression models were computed to test for evidence of interactions among all pairs of SNPs that had BH-FDR adjusted P value < 0.05 . Each logistic model contained the admixture factors, the two SNPs, and their centred cross-product term, with the latter term tested using the likelihood ratio test implemented in the Intertwolog module in SNPLASH. To adjust for the number of interactions tested, Bonferroni and BH-FDR adjusted P values were computed. To test for ancestry-specific gender-by-SNP interactions, a case-only autosomal scan was computed; here, gender was the outcome and admixture factors and SNP were the predictors. To adjust for the number of tests computed, the BH-FDR adjusted P values from the likelihood ratio test were computed for each SNP that passed quality control.

To determine how many distinct associations were within a genomic region, a manual stepwise procedure (that is, forward selection with backward elimination, entry and exit criteria of $P < 0.001$) was computed.

For the *transancestral meta-analyses*, three ancestries were examined for association and meta-analysed to better isolate shared SLE-risk loci by leveraging their LD pattern differences. For each SNP, a nonparametric meta-analysis, weighted inverse normal method (weighted by sample size), was computed as implemented in METAL⁴⁹. Regions of association were visually examined and tests of heterogeneity of the odds ratio were computed. Thus, for each region, ancestry-specific and meta-analytic tests of association and tests of heterogeneity are reported. The transancestral patterns of association and LD were visualized using LocusZoom⁵⁰. Results from the weighted inverse normal method were compared to random effects meta-analyses and results of the regions were comparable.

Classical HLA alleles at HLA-A, -B, -C, -DPB1, -DQA1, -DQB1 and -DRB1 were imputed using the program HIBAG⁵¹. HIBAG uses an ensemble classifier and bagging technique to arrive at an average posterior probability. Unlike alternative imputation software such as BEAGLE⁵², HLA*IMP⁵³ and SNP2HLA⁵⁴, HIBAG did not require training data for any of our three cohorts, as it provides multiple ancestry reference panels (European, African, Hispanic and Asian). This, combined with its accuracy rates being comparable to other approaches⁵¹, made HIBAG an ideal method for HLA imputation in our EA, AA, and HA cohorts. To account for imputation uncertainty, the allele dosage was utilized for all analyses. To filter out the lowest frequency alleles, a minimum best guess allele count of 10 was required in either the cases or controls for each allele, in each cohort.

For analysis of classical HLA alleles, single-allele associations were evaluated using logistic regression under the additive model and accounting for imputation uncertainty via allelic dose. To account for population substructure, cohort-specific factors were used as covariates (EA: factors 1–4; AA: factors 1–3; HA: factors 1–2) in each analysis. Meta-analysis was completed for any allele that had a single-allele analysis in at least two cohorts. Evidence of association from each cohort was combined using the weighted inverse normal method via METAL⁴⁹ and tests for heterogeneity of the odds ratio were computed.

To build multi-locus ancestry-specific models of classical HLA alleles for case/control status of SLE, stepwise regression models were computed. Stepwise logistic modelling (forward selection with backward elimination) was computed using all of the classical HLA alleles that met the QC criteria, including requiring at least a count of 10 alleles from the best guess allele count cross the individuals within an ancestry. The entry and exit criteria were set to $P < 0.01$ for each of the three cohorts. As in the single-allele analysis, the logistic models tested for an additive effect of the alleles and accounted for imputation uncertainty via allelic dose.

To evaluate and compare classical HLA allele associations across the three cohorts, the results from the single-allele and multilocus modelling were visualized in the context of classical HLA protein sequence similarity. Protein sequences for all observed HLA-imputed alleles were retrieved from the EMBL-EBI Immunogenetics HLA Database⁵⁵. Sequences within an HLA-gene were aligned using ClustalOmega⁵⁶. Unrooted phylogenetic trees for each of the HLA loci were then generated by Clustal-W2 via the aligned amino acid sequences. The neighbour-joining method, a distance matrix method, utilized a Markov chain of nucleotide or amino acid substitution⁵⁷. The neighbour-joining method uses this distance information to iteratively evaluate all pairings of neighbours in order to construct a tree that minimizes the branch length at each stage of clustering⁵⁸. The resulting trees were visualized using Dendroscope⁵⁹. All results from the single-allele and multilocus classical HLA associations from the three cohorts were graphically displayed on the unrooted trees.

A second set of ancestry-specific single-SNP analyses was computed across the HLA locus and surrounding region, while adjusting for the primary SLE-associated HLA risk alleles from the stepwise modelling. The logistic regression model was computed, as above, considering the fit to the three genetic models (dominant,

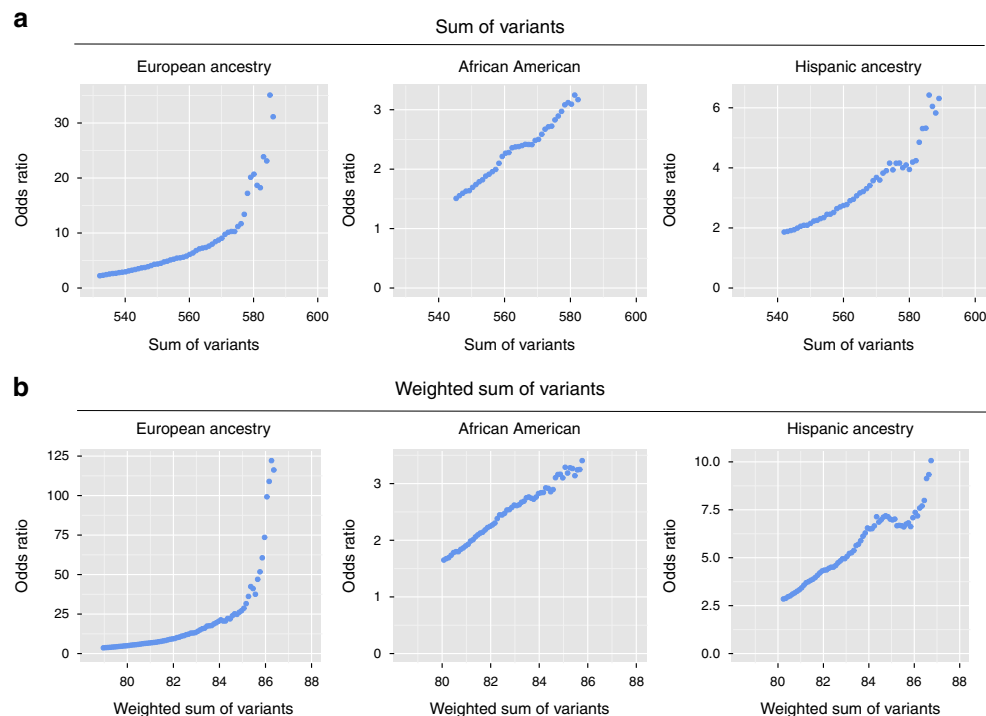


Figure 5 | The non-additive effect of EA risk-allele genetic load on SLE risk. The cumulative effect of EA SLE-risk alleles (cumulative hits) on an individual's risk of SLE is greater than if the individual SNPs were acting independently/additively. **(a)** The genetic load was computed as the sum of the number of EA risk variants from the Tier 1, 2 or 3 SNPs that met the region-specific stepwise modelling (see Online Methods). In the AA, HA and an independent set of 2,000 EA cases and 2,000 EA controls, the samples with the lowest 10% in risk-allele counts were identified and formed the baseline comparison group. Using a moving window of 10 in the allele count, the odds ratio for that window relative to the lowest 10% was computed and graphed. **(b)** The process was repeated for a weighted sum of the number of EA risk-allele variants. Here, the alleles are weighted by the natural logarithm of the odds ratio for that SNP's association with SLE. The corresponding moving window for the weighted genetic load used a window size of 3. Supplementary Fig. 17 plots the natural logarithm of the odds ratio (instead of the odds ratio) of genetic load versus SLE risk.

additive, recessive); the additive model required at least 10 homozygotes for the minor allele, while the recessive model required at least 30. The meta-analysis of these results was computed using METAL.

The Wald tests for HLA-by-SNP and HLA-by-gender interactions were computed using logistic regression models that adjusted for admixture factors and included both the main effects of the HLA allele and SNP (or gender) and their centred cross product as the multiplicative interaction term.

To test whether there was a difference in SLE risk between individuals homozygous for the same risk allele versus heterozygous for two different risk alleles, a Wald test from a logistic regression model was computed adjusting for admixture.

To examine ancestry of associated SLE risk alleles, genotyped SNPs from the population-specific (Tier1 and Tier 2) and the meta-analysis (primary and secondary) tables were compiled into a list of 205 unique SNPs. For evaluation, only SNPs of good quality across the three cohorts were retained. These criteria left 181 SNPs for comparison. In cases, admixture proportions of CEU and YRI were calculated using ADMIXTURE and then the average proportions were tallied for each cohort. Within each of the three populations and for each SNP, the risk allele's average admixture was computed. The resulting risk allele average admixture proportion was compared to the overall average sample admixture proportion in cases by computing the difference between risk allele and sample admixture proportion averages.

To evaluate the *SLE-risk allele genetic load*, the EA samples were partitioned into two groups: training (the entire EA sample minus 2,000 cases and 2,000 controls randomly chosen from the full EA cohort) and testing (the aforementioned 2,000 cases and 2,000 controls). In the training samples, the single SNP association and stepwise analyses were repeated to obtain a training set of SNPs that had BH-FDR adjusted P -value < 0.05 . From these results, the EA SLE-risk genetic load was calculated for each individual as the count of risk alleles from the training SNPs. Specifically, we define the EA SLE-risk allele genetic load as:

$$GRS_i = \sum_{k=1}^N \gamma_k RA_k,$$

where, GRS_i is the genetic risk score for individual i ; γ_k is the beta coefficient for the k th SNP association with SLE and serves as the weight for that risk allele; RA_k is the number of risk alleles for the k th SNP (0, 1, 2); and N is the number of SNPs. By definition of parameterizing relative to the risk allele, $\gamma_k > 0$ for all k . The EA SLE-risk genetic load was computed for AA, HA, and the EA testing samples. Individuals

whose genetic load (risk allele count) was in the lower 10% of the count distribution were used as the reference sample. A logistic regression model, including admixture factors as covariates, computed the odds ratio comparing the reference sample to samples within a moving window of 20 unweighted risk allele counts for the unweighted analysis and moving window of 4 for the weighted analysis). For example, a logistic model compared the risk of SLE for those in the lowest 10% to those whose risk allele counts ranged from 940 to 960 in the unweighted analysis. The next model and odds ratios were then computed, sliding the allele count up one (for example, 941–961). A plot of these odds ratios for moving windows of 20 counts was constructed to illustrate the pattern. The corresponding plot of the $\log(OR) = \beta$ from the genetic load association with SLE was generated to show that the nonlinearity was not due to the scale; that is, it documents a departure from linearity on the logit scale. A similar approach was completed for a weighted risk allele count, where each risk allele was weighted by the natural logarithm of the odds ratio from the EA SNP association analysis. Plots of the odds ratio effect of the EA genetic load (weighted and unweighted) were generated for AA, HA and the independent EA set.

Finally, for each ancestry an admixture-adjusted regression model was computed to test whether genetic load was associated with age of SLE onset. For ease of interpretation, the strength of the association was reported as the Spearman's rank correlation coefficient, but the P value is from the admixture-adjusted linear regression model.

Functional annotation analysis. To identify eQTLs for SLE-associated SNPs, all 1,000 Genomes SNPs in LD with the SLE-associated SNP were identified using SNAP⁶⁰. Specifically, LD was computed using the CEU (for EA and HA) or YRI (for AA) data with an $r^2 \geq 0.5$ for Tier 1 and 2 SNPs. SNPs and their proxies were then queried in a data set downloaded from the eQTL Browser (<http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/>; Pritchard lab, University of Chicago) and the GTEx Portal (<http://www.gtexportal.org>). The eQTL Browser contains eQTL data surveyed from 17 eQTL studies, and the Blood eQTL Browser⁶¹. The GTEx Portal is a comprehensive resource, with eQTL data from 44 different tissues. When multiple proxies existed for the same eQTL (that is, same SNP and same gene), only the proxy with the lowest P value was retained.

RegulomeDB is a database that annotates SNPs with known and predicted regulatory elements (eQTLs, DNAase hypersensitivity, binding sites of transcription factors) in the intergenic regions of the human genome²². It includes high-throughput, experimental data sets from GEO, the ENCODE project,

published literature, as well as computational predictions and manual annotations to identify putative regulatory potential and identify functional variants²². The variants associated with SLE (identified in Tier 1 and 2 in any ancestry cohort) were queried in RegulomeDB.

HaploReg v2 is a tool for exploring annotations of the noncoding genome at variants on haplotype blocks²³ and uses LD information from the 1,000 Genomes Project Phase 1 individuals. It analyzes sets of SNPs for an enrichment of cell type-specific enhancers, and includes all dbSNP build 137 SNPs, predicted chromatin state in nine cell types, conservation across mammals, motif instances from ENCODE experiments, enhancer annotations on 90 cell types from the Roadmap Epigenome Mapping Consortium and eQTLs from the GTEx eQTL browser²³. The query was performed using default settings, including LD calculations based on the 1,000 Genomes Phase 1 EUR individuals, and epigenome data from both the ENCODE and Roadmap Epigenome Mapping Consortium projects.

SNPs associated with SLE (Tiers 1 and 2) were annotated with the eQTL data and HaploReg v2 (ref. 23) to prioritize those with the highest biological potential. The top summary gene scores were summed across individual criteria (presence of an eQTL, presence of a nonsense or missense variant, promoter and enhancer status in a lymphoblastoid B-cell line (B-LCL), the presence of a DNase hypersensitivity site in any of five immune-related cell lines, presence of a conserved region, the presence of any bound protein, and transcription start site and enhancer status in any of 15 immune cell types), in the haplotype block of each SNP. In the calculation of the biological scores, each functional annotation was given a weight according to their regulatory potential. A score of '3' was given to SNPs in an LD block with any variant that mapped within an active or poised TSS in any of 15 immune cell types, was an eQTL, was non/missense, or mapped within an active promoter in a B-LCL. A score of '2' was given to SNPs in an LD block with any variant that mapped within an active upstream flanking TSS in any of 15 immune cell types or mapped within a conserved region. A score of '1' was given to SNPs in an LD block with any variant that mapped within a weak TSS or any enhancer in any of 15 immune cell types, mapped within a weak promoter or weak enhancer in a B-LCL, mapped within a DNase hypersensitivity site in any of 5 cell lines, or had any bound protein. The sum of these annotations resulted in a final biological score, ranging from zero to fifteen.

For each of the 146,111 (145,278 unique) SNPs that met quality control standards in at least one population, the flanking base pairs were identified using the UCSC reference genome (build 37). Once strand alignment was confirmed between the Immunochip and UCSC reference genome, it was evaluated whether either (or both) of a SNP's alleles created a CpG site in the 5'-3' direction.

Data availability. The summary data are available at www.immunobase.org. Individual genotype data, consistent with the respective Institutional Review Board approval and subject consent, are available from the corresponding authors.

References

- Danchenko, N., Satia, J. A. & Anthony, M. S. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* **15**, 308–318 (2006).
- International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN) *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.* **40**, 204–210 (2008).
- Chung, S. A. *et al.* Lupus nephritis susceptibility loci in women with systemic lupus erythematosus. *J. Am. Soc. Nephrol.* **25**, 2859–2870 (2014).
- Hom, G. *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.* **358**, 900–909 (2008).
- Rullo, O. J. & Tsao, B. P. Recent insights into the genetic basis of systemic lupus erythematosus. *Ann. Rheum. Dis.* **72**, ii56–ii61 (2013).
- Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat. Genet.* **47**, 1457–1464 (2015).
- Sun, C. *et al.* High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat. Genet.* **48**, 323–330 (2016).
- Morris, D. L. *et al.* Genome-wide association meta-analysis in Chinese and European individuals identifies ten new loci associated with systemic lupus erythematosus. *Nat. Genet.* **48**, 940–946 (2016).
- Cortes, A. & Brown, M. A. Promise and pitfalls of the Immunochip. *Arthritis Res. Ther.* **13**, 101 (2010).
- Chung, S. A. *et al.* Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet.* **7**, e1001323 (2011).
- Graham, R. R. *et al.* Specific combinations of HLA-DR2 and DR3 class II haplotypes contribute graded risk for disease susceptibility and autoantibodies in human SLE. *Eur. J. Hum. Genet.* **15**, 823–830 (2007).
- Lenz, T. L. *et al.* Widespread non-additive and interaction effects within HLA loci modulate the risk of autoimmune diseases. *Nat. Genet.* **47**, 1085–1090 (2015).
- Erlich, H. *et al.* HLA DR-DQ haplotypes and genotypes and type 1 diabetes risk analysis of the type 1 diabetes genetics consortium families. *Diabetes* **57**, 1084–1092 (2008).
- Reche, P. A. & Reinherz, E. L. Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. *J. Mol. Biol.* **331**, 623–641 (2003).
- Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296 (2012).
- Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum.* **30**, 1205–1213 (1987).
- du Montcel, S. T. *et al.* New classification of HLA-DRB1 alleles supports the shared epitope hypothesis of rheumatoid arthritis susceptibility. *Arthritis Rheum.* **52**, 1063–1068 (2005).
- Raymond, C. K. *et al.* Ancient haplotypes of the HLA Class II region. *Genome Res.* **15**, 1250–1257 (2005).
- Bossini-Castillo, L. *et al.* A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. *Hum. Mol. Genet.* **21**, 926–933 (2012).
- Hirschfield, G. M. *et al.* Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N. Engl. J. Med.* **360**, 2544–2555 (2009).
- Gray-McGuire, C. *et al.* Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. *Am. J. Hum. Genet.* **67**, 1460–1469 (2000).
- Boyle, A. P. *et al.* Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* **22**, 1790–1797 (2012).
- Ward, L. D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* **40**, D930–D934 (2012).
- 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**, 4470–4473 (2013).
- Han, J.-W. *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.* **41**, 1234–1237 (2009).
- Wang, C. *et al.* Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. *Eur. J. Hum. Genet.* **21**, 994–999 (2013).
- Ahmed, Z. *et al.* Direct binding of Grb2 SH3 domain to FGFR2 regulates SHP2 function. *Cell. Signal.* **22**, 23–33 (2010).
- Sun, J. *et al.* Antagonism between binding site affinity and conformational dynamics tunes alternative cis-interactions within Shp2. *Nat. Commun.* **4**, 2037 (2013).
- Wang, J. *et al.* Inhibition of SHP2 ameliorates the pathogenesis of systemic lupus erythematosus. *J. Clin. Invest.* **126**, 2077–2092 (2016).
- Sawalha, A. H. *et al.* Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. *Ann. Rheum. Dis.* **67**, 458–461 (2008).
- Choi, J.-Y. *et al.* Circulating follicular helper-like T cells in systemic lupus erythematosus: association with disease activity: circulating Tfh-like cells in SLE. *Arthritis Rheumatol.* **67**, 988–999 (2015).
- Löfgren, S. E. *et al.* Genetic association of miRNA-146a with systemic lupus erythematosus in Europeans through decreased expression of the gene. *Genes Immun.* **13**, 268–274 (2012).
- Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986 (2015).
- Kwok, W. W., Kovats, S., Thurtle, P. & Nepom, G. T. HLA-DQ allelic polymorphisms constrain patterns of class II heterodimer formation. *J. Immunol.* **150**, 2263–2272 (1993).
- Miyadera, H., Ohashi, J., Lernmark, Å., Kitamura, T. & Tokunaga, K. Cell-surface MHC density profiling reveals instability of autoimmunity-associated HLA. *J. Clin. Invest.* **125**, 275–291 (2015).
- van Heemst, J., Huizinga, T. J., van der Woude, D. & Toes, R. E. Fine-mapping the human leukocyte antigen locus in rheumatoid arthritis and other rheumatic diseases: identifying causal amino acid variants? *Curr. Opin. Rheumatol.* **27**, 256–261 (2015).
- Segal, M. R., Cummings, M. P. & Hubbard, A. E. Relating amino acid sequence to phenotype: analysis of peptide-binding data. *Biometrics* **57**, 632–643 (2001).
- Majumder, P. & Boss, J. M. DNA methylation dysregulates and silences the HLA-DQ locus by altering chromatin architecture. *Genes Immun.* **12**, 291–299 (2011).
- Kaushansky, N. *et al.* Role of a novel HLA-DQA1* 01: 02; DRB1* 15: 01 mixed-isotype heterodimer in the pathogenesis of 'humanized' multiple sclerosis-like disease. *J. Biol. Chem.* **290**, 15260–15278 (2015).
- Tan, E. M. *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **25**, 1271–1277 (1982).
- Hochberg, M. C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1725 (1997).
- Shah, T. S. *et al.* optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* **28**, 1598–1603 (2012).

43. Morris, J. A., Randall, J. C., Maller, J. B. & Barrett, J. C. Evoker: a visualization tool for genotype intensity data. *Bioinformatics* **26**, 1786–1787 (2010).
44. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).
45. Johnson, R. A. & Wichern, D. W. *Applied Multivariate Statistical Analysis* (Pearson Prentice Hall, 2007).
46. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
47. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
48. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57**, 289–300 (1995).
49. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
50. Pruim, R. J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010).
51. Zheng, X. *et al.* HIBAG—HLA genotype imputation with attribute bagging. *Pharmacogenomics J.* **14**, 192–200 (2014).
52. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
53. Dilthey, A. *et al.* Multi-population classical HLA type imputation. *PLoS Comput. Biol.* **9**, e1002877 (2013).
54. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
55. Robinson, J. *et al.* The IMGT/HLA database. *Nucleic Acids Res.* **41**, D1222–D1227 (2013).
56. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).
57. Yang, Z. & Rannala, B. Molecular phylogenetics: principles and practice. *Nat. Rev. Genet.* **13**, 303–314 (2012).
58. Saitou, N. & Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
59. Huson, D. H. & Scornavacca, C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067 (2012).
60. Johnson, A. D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
61. Westra, H.-J. *et al.* Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat. Genet.* **45**, 1238–1243 (2013).

Acknowledgements

We gratefully acknowledge the Alliance for Lupus Research for funding and support. The research was supported in part by awards from the Arthritis Research UK Special Strategic Award (ref. 19289) and from George Koukis (T.J.V.). In addition, 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 (T.J.V.). The work would not be possible without funding from the NIH grants AR049084 (RPK, EEB); the International Consortium on the Genetics of Systemic Lupus Erythematosus (SLEGEN) AI083194 (J.B.H.); CA141700, AR058621 Proyecteo de Excelencia, Consejería de Andalucía (M.E.A.R.); AR043814 and AR-065626 (B.P.T.); AR060366, MD007909, AI107176 (S.K.N.); AR-057172 (C.O.J.); RC2 AR058959, U19 AI082714, R01 AR063124, P30 GM110766, R01 AR056360 (P.M.G.); P60 AR053308 (L.A.C.), MUSC part is from UL1RR029882 (G.S.G., D.L.K.) and 5P60AR062755 (G.S.G., D.L.K., P.R.R.). Oklahoma Samples U19AI082714, U01AI101934, P30GM103510, U54GM104938 and P30AR053483 (J.A.J., J.M.G.); Northwestern P60 AR066464 and U54TR001018 (R.R.G.); This study was supported by the US National Institute of Health and Musculoskeletal and Skin Diseases of the National Institutes of Health (NIH) under Award Numbers K01 AR067280 and P60 AR062755 (PSR); N01AR22265 (funded collection of APPLE samples) (LES) and the APPLE Investigators; R01AR43727, NIH AR 043727 and 069572 (M.P.); NIAMS/NIH P50-AR055503 (D.R.K.).

We would like to also thank the RILITE foundation for financial support (C.D.L.). Additional funding for Immunochip genotyping was provided by Genentech.

Author contributions

H.C.A., D.S.C.G. and J.A.K. contributed equally. P.M.G., R.R.G., C.D.L. and T.J.V. jointly supervised research. P.M.G., R.R.G., C.D.L., T.J.V., D.S.C.G., J.A.K., M.E.A., T.W.B., L.A.C., J.B.H., T.D.H., C.O.J., R.P.K., P.S.R., E.D.S., K.L.S., B.P.T. and E.K.W. conceived and designed the experiments. P.M.G., R.R.G., J.A.K., C.D.L., E.D.S., T.J.V. and E.K.W. performed experiments. H.C.A., M.E.C., T.D.H., J.A.K., C.D.L., M.C.M., D.R.M. and E.K.W. performed statistical analysis. H.C.A., M.E.C., D.S.C.G., T.D.H., K.M.K., J.A.K., L.C.K., C.D.L., M.C.M., D.R.M., P.S.R. analysed the data. P.M.G., R.R.G., R.P.K., C.D.L., E.D.S., T.J.V. and E.K.W. contributed reagents, materials, and analysis tools. H.C.A., M.E.C., P.M.G., R.R.G., T.D.H., C.D.L., M.C.M. and T.J.V. wrote the manuscript. E.M.A.-V., G.S.A., M.E.A., A.M.B., V.B., T.W.B., A.A.B., G.A.B., T.B., M.B., E.E.B., H.I.B., M.H.C., J.C.A., L.C., R.C., L.A.C., J.M.C.-V., S.D., B.M.D.S., S.R.D., I.D., A.D., J.C.E., E.E., J.A.E.-V., P.R.F., B.I.F., J.F., M.A.G., I.G., G.G., D.D.G., P.K.G., I.G.d.I.T., J.M.G., J.L.H., C.O.J., J.A.J., C.G.M.K., D.L.K., D.R.K., R.P.K., L.K., H.L., B.R.L., Q.Z.L., M.A.M., J.M., J.M.M., J.T.M., P.M., J.F.M., S.K.N., T.B.N., J.R.O., L.O., N.O., M.P., C.A.P., B.A.P., J.P., P.R., R.R., J.D.R., L.R., J.M.S., C.A.S., J.K.S., L.E.S., H.R.S., R.S., M.F.S., E.D.S., C.S., E.S., A.C.S., S.D.T., S.M.A.T., L.T., B.P.T., T.T., C.V., L.M.V., D.J.W., M.H.W. and J.E.W. contributed samples. E.M.A.-V., G.S.A., M.E.A., A.M.B., V.B., T.W.B., A.A.B., G.A.B., T.B., M.B., E.E.B., H.I.B., M.H.C., J.C.A., L.C., R.C., L.A.C., J.A.C., J.M.C.-V., D.S.C.G., S.D., B.M.D.S., S.R.D., I.D., A.D., J.C.E., E.E., J.A.E.-V., P.R.F., B.I.F., J.F., P.M.G., M.A.G., I.G.d.I.T., G.G., D.D.G., R.R.G., P.K.G., I.G., J.M.G., J.L.H., J.B.H., C.O.J., J.A.J., C.G.M.K., D.L.K., D.R.K., K.M.K., J.A.K., R.P.K., L.C.K., L.K., H.L., B.R.L., Q.Z.L., M.A.M., J.M., J.M.M., D.R.M., J.T.M., P.M., J.F.M., D.L.M., S.K.N., T.B.N., J.R.O., L.O., N.O., M.P., C.A.P., B.A.P., J.P., P.R., P.S.R., R.R., J.D.R., J.D.R., L.R., L.P.R., J.M.S., C.A.S., J.K.S., L.E.S., H.R.S., R.S., M.F.S., E.D.S., K.L.S., C.S., E.S., A.C.S., S.D.T., S.M.A.T., L.T., B.P.T., T.T., C.V., L.M.V., T.J.V., E.K.W., J.E.W., M.H.W. and D.J.W. revised the manuscript.

Additional information

Supplementary Information accompanies this paper at <http://www.nature.com/naturecommunications>

Competing interests: R.R.G., T.B. and T.W.B. are employees of Genentech, Inc. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

How to cite this article: Langefeld, C. D. *et al.* Transancestral mapping and genetic load in systemic lupus erythematosus. *Nat. Commun.* **8**, 16021 doi: 10.1038/ncomms16021 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

© The Author(s) 2017

Carl D. Langefeld^{1,2}, Hannah C. Ainsworth^{1,2,*}, Deborah S. Cunninghame Graham^{3,*}, Jennifer A. Kelly^{4,*}, Mary E. Comeau^{1,2}, Miranda C. Marion^{1,2}, Timothy D. Howard^{1,5}, Paula S. Ramos^{6,7}, Jennifer A. Croker⁸, David L. Morris³, Johanna K. Sandling⁹, Jonas Carlsson Almlöf⁹, Eduardo M. Acevedo-Vásquez¹⁰, Graciela S. Alarcón⁸, Alejandra M. Babini¹¹, Vicente Baca¹², Anders A. Bengtsson¹³, Guillermo A. Berbotto¹⁴, Marc Bijl¹⁵, Elizabeth E. Brown⁸, Hermine I. Brunner¹⁶, Mario H. Cardiel¹⁷, Luis Catoggio¹⁸, Ricard Cervera¹⁹, Jorge M. Cucho-Venegas¹⁰, Solbritt Rantapää Dahlqvist²⁰, Sandra D'Alfonso²¹, Berta Martins Da Silva²²,

Iñigo de la Rúa Figueroa²³, Andrea Doria²⁴, Jeffrey C. Edberg⁸, Emőke Endreffy²⁵, Jorge A. Esquivel-Valerio²⁶, Paul R. Fortin²⁷, Barry I. Freedman^{1,28}, Johan Frostegård²⁹, Mercedes A. García³⁰, Ignacio García de la Torre³¹, Gary S. Gilkeson⁷, Dafna D. Gladman³², Iva Gunnarsson³³, Joel M. Guthridge⁴, Jennifer L. Huggins¹⁶, Judith A. James^{4,34}, Cees G.M. Kallenberg³⁵, Diane L. Kamen⁷, David R. Karp³⁶, Kenneth M. Kaufman³⁷, Leah C. Kottyan³⁷, László Kovács³⁸, Helle Laustrop³⁹, Bernard R. Lauwerys⁴⁰, Quan-Zhen Li³⁶, Marco A. Maradiaga-Ceceña⁴¹, Javier Martín⁴², Joseph M. McCune⁴³, David R. McWilliams^{1,2}, Joan T. Merrill⁴, Pedro Miranda⁴⁴, José F. Moctezuma⁴⁵, Swapan K. Nath⁴, Timothy B. Niewold⁴⁶, Lorena Orozco⁴⁷, Norberto Ortego-Centeno⁴⁸, Michelle Petri⁴⁹, Christian A. Pineau⁵⁰, Bernardo A. Pons-Estel⁵¹, Janet Pope⁵², Prithvi Raj³⁶, Rosalind Ramsey-Goldman⁵³, John D. Reveille⁵⁴, Laurie P. Russell^{1,2}, José M. Sabio⁵⁵, Carlos A. Aguilar-Salinas⁵⁶, Hugo R. Scherbarth⁵⁷, Raffaella Scorza⁵⁸, Michael F. Seldin⁵⁹, Christopher Sjöwall⁶⁰, Elisabet Svenungsson³³, Susan D. Thompson³⁷, Sergio M.A. Toloza⁶¹, Lennart Truedsson⁶², Teresa Tusié-Luna⁶³, Carlos Vasconcelos⁶⁴, Luis M. Vilá⁶⁵, Daniel J. Wallace⁶⁶, Michael H. Weisman⁶⁶, Joan E. Wither³², Tushar Bhangale⁶⁷, Jorge R. Oksenberg⁶⁸, John D. Rioux⁶⁹, Peter K. Gregersen⁷⁰, Ann-Christine Syvänen⁹, Lars Rönnblom⁷¹, Lindsey A. Criswell⁷², Chaim O. Jacob⁷³, Kathy L. Sivils⁴, Betty P. Tsao⁷, Laura E. Schanberg⁷⁴, Timothy W. Behrens⁶⁷, Earl D. Silverman⁷⁵, Marta E. Alarcón-Riquelme^{4,76,77}, Robert P. Kimberly⁸, John B. Harley³⁷, Edward K. Wakeland³⁶, Robert R. Graham⁶⁷, Patrick M. Gaffney⁴ & Timothy J. Vyse³

¹ Center for Public Health Genomics, Wake Forest School of Medicine, Winston-Salem, North Carolina 27101, USA. ² Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina 27101, USA. ³ Divisions of Genetics and Molecular Medicine and Immunology, Infection and Inflammatory Diseases, King's College London, Guy's Hospital, London SE1 9RT, UK. ⁴ Arthritis & Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104, USA. ⁵ Center for Human Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina 27101, USA. ⁶ Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ⁷ Department of Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ⁸ Division of Clinical Immunology and Rheumatology, UAB School of Medicine, Birmingham, Alabama 35294, USA. ⁹ Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala 752 36, Sweden. ¹⁰ Departamento de Reumatología, Hospital G. Almenara y Facultad de Medicina, Universidad Nacional Mayor de San Marcos, Lima 15081, Perú. ¹¹ Hospital Italiano de Córdoba, Córdoba X5004BAL, Argentina. ¹² Hospital de Pediatría, Centro Médico Nacional Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City 06720, Mexico. ¹³ Department of Clinical Sciences, Rheumatology, Lund University, Lund 22362, Sweden. ¹⁴ Hospital Eva Perón, Granadero Baigorria S2152EDD, Argentina. ¹⁵ Department of Internal Medicine and Rheumatology, Martini Hospital, Van Swietenplein 1, 9728, NT, Groningen, The Netherlands. ¹⁶ Division of Rheumatology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center and the University of Cincinnati, Cincinnati, Ohio 45229, USA. ¹⁷ Centro de Investigación Clínica de Morelia, Morelia, Michoacán 58070, Mexico. ¹⁸ Hospital Italiano de Buenos Aires, 1181, Buenos Aires C1181ACH, Argentina. ¹⁹ Department of Autoimmune Diseases, Hospital Clínic, University of Barcelona, Barcelona, Catalonia 08007, Spain. ²⁰ Department of Public Health and Clinical Medicine, Division of Rheumatology, Umeå University, Umeå 901 87, Sweden. ²¹ Department of Health Sciences and Institute of Research in Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara 28100, Italy. ²² Unidade Multidisciplinar em Investigação Biomédica/Instituto de Ciências Biomédicas de Abel Salazar—Universidade do Porto, Porto 4099-003, Portugal. ²³ Department of Rheumatology, Hospital Universitario de Gran Canaria Dr Negrín, Las Palmas de Gran Canaria 35010, Spain. ²⁴ Division of Rheumatology, Department of Medicine (DIMED), University of Padua, Padua 35122, Italy. ²⁵ Department of Pediatrics and Child Health Center, Albert Szent-Györgyi Medical Center, Faculty of Medicine, University of Szeged, Szeged H-6720, Hungary. ²⁶ Hospital Universitario 'Dr José Eleuterio González' Universidad Autónoma de Nuevo León, Monterrey 64020, México. ²⁷ CHU de Québec Université Laval, Québec, Canada G1R 2JG. ²⁸ Section on Nephrology, Wake Forest School of Medicine, Winston-Salem, North Carolina 27101, USA. ²⁹ Institute of Environmental Medicine, Unit of Immunology and Chronic diseases, Karolinska Institutet, Stockholm 171 77, Sweden. ³⁰ Division of Rheumatology, Hospital Interzonal General de Agudos General San Martín, La Plata 1900, Argentina. ³¹ University of Guadalajara, Departamento de Fisiología, Guadalajara, Jalisco 44100, Mexico. ³² Centre for Prognosis Studies in The Rheumatic Diseases, Krembil Research Institute, Toronto Western Hospital, Toronto, Ontario M5T 2S8, Canada. ³³ Unit of Rheumatology, Department of Medicine Solna, Karolinska Institutet, Karolinska University Hospital, Stockholm SE-171 76, Sweden. ³⁴ Departments of Medicine and Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA. ³⁵ Department of Rheumatology and Clinical Immunology, University Medical Center Groningen, University of Groningen, Groningen 9713 GZ, The Netherlands. ³⁶ Department of Immunology, University of Texas SouthWestern Medical Center, Dallas, Texas 75235, USA. ³⁷ Department of Pediatrics, Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ³⁸ Department of Rheumatology, Albert Szent-Györgyi Medical Centre, University of Szeged, Szeged H-6720, Hungary. ³⁹ Department of Rheumatology, Odense University Hospital, Odense 5000, Denmark. ⁴⁰ Rheumatology, Cliniques Universitaires Saint-Luc & Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Louvain-la-Neuve 1348, Belgium. ⁴¹ Hospital General de Culiacán, Sinaloa 80220, Mexico. ⁴² Instituto de Parasitología y Biomedicina López Neyra, CSIC, Granada 18100, Spain. ⁴³ University of Michigan Medical Center, Ann Arbor, Michigan 48103, USA. ⁴⁴ Centro de Estudios Reumatológicos, Santiago de Chile, Santiago 7500000, Chile. ⁴⁵ Departamento de Reumatología, Hospital General de México, Mexico D.F., 06726 Mexico. ⁴⁶ Department of Rheumatology, Mayo Clinic, Rochester, Minnesota 55905, USA. ⁴⁷ Instituto Nacional de Medicina Genómica (INMEGEN), México City 14610, México. ⁴⁸ Unidad de Enfermedades Autoinmunes Sistémicas, UGC Medicina Interna, Hospital Universitario San Cecilio, Granada 18007, Spain. ⁴⁹ Division of Rheumatology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21218, USA. ⁵⁰ Rheumatology Division, McGill University, Montreal, Quebec H3A 0G4, Canada. ⁵¹ Department of Rheumatology, Sanatorio Parque, Rosario S2000, Argentina. ⁵² University of Western Ontario, London, Ontario, Canada M5T 2S8. ⁵³ Division of Rheumatology, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. ⁵⁴ The University of Texas Health Science Center at Houston (UTHealth) Medical School, Houston, Texas 77030, USA. ⁵⁵ Hospital Universitario Virgen de las Nieves, Granada 18014, Spain. ⁵⁶ Instituto Nacional de Ciencias Médicas y Nutrición, Department of Endocrinology and Metabolism, Vasco de Quiroga 15, Mexico City 14080, Mexico. ⁵⁷ Unidad Reumatología y Enfermedades

Autoinmunes H.I.G.A. Dr Alende Mar del Plata, Buenos Aires B7600, Argentina. ⁵⁸Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca'Granda Ospedale Ma Repiore Policlinico and University of Milan, Milan 20122, Italy. ⁵⁹Department of Biochemistry and Molecular Medicine, UC Davis School of Medicine, Sacramento, California 95616, USA. ⁶⁰Rheumatology Division of Neuro and Inflammation Sciences, Department of Clinical and Experimental Medicine, Linköping University, Linköping 581 83, Sweden. ⁶¹Ministry of Health, San Fernando del Valle de Catamarca, Catamarca K4700, Argentina. ⁶²Department of Laboratory Medicine, Section of Microbiology, Immunology and Glycobiology, Lund University, Lund 221 00, Sweden. ⁶³Unidad de Biología Molecular y Medicina Genómica Instituto de Investigaciones Biomédicas/UNAM Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, Mexico City 14080, Mexico. ⁶⁴Hospital Santo Antonio, Universidade do Porto, Porto 4099-003, Portugal. ⁶⁵University of Puerto Rico School of Medicine, San Juan 00936, Puerto Rico. ⁶⁶Department of Medicine, Cedars Sinai Medical Center, Los Angeles, California 90048, USA. ⁶⁷Human Genetics, Genentech Inc, South San Francisco, California 94080, USA. ⁶⁸Department of Neurology and Institute of Human Genetics, University of California at San Francisco, San Francisco, California 94158, USA. ⁶⁹Université de Montréal and the Montreal Heart Institute, Montreal, Quebec, Canada H1T 1C8. ⁷⁰Center for Genomics & Human Genetics, The Feinstein Institute for Medical Research, Manhasset, New York 11030, USA. ⁷¹Department of Medical Sciences, Rheumatology, Uppsala University, 752 36, Sweden. ⁷²Rosalind Russell/Ephraim P Engleman Rheumatology Research Center, Division of Rheumatology, UCSF School of Medicine, San Francisco, California 94158, USA. ⁷³Keck School of Medicine of USC, Los Angeles, California 90033, USA. ⁷⁴Department of Pediatrics, Duke University, Durham, North Carolina 27708, USA. ⁷⁵Department of Pediatrics and the Institute of Medical Sciences, The Hospital for Sick Children, Hospital for Sick Children Research Institute and University of Toronto, Ontario, Canada M5G 1X8. ⁷⁶Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO), Granada 18007, Spain. ⁷⁷Unit of Institute of Environmental Medicine, Karolinska Institute, Solnavägen 171 77, Sweden. * These authors contributed equally to this work.