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.
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 ancestry. Candidate gene and genome-wide association studies have successfully identified ~90 SLE risk loci that explain a significant proportion of SLE's heritability. 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 loci.

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 effect just outside of the HLA region. The association at rs6681482 ($P = 1.35 \times 10^{-8}$) is significant in AA (OR = 1.27) 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 ethnic group. 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 LRRC16A), 6q23 (OLIG3-LOC100130476), 8p23 (FAM86B3P), 8q21 (PKIA-ZC2H1A) 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 $\alpha = 1 \times 10^{-8}$. In addition to known regions in AA, novel AA regions identified include 5q33 (PTTGI-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-rs205960 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, $P = 7.39 \times 10^{-7}$). The association with rs2431697 (OR = 0.76, $P = 1.27 \times 10^{-10}$) 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-PROSI), 6p21 (TCP11-SCUBE3), 6q25 (RSPH3), 12q15 (DYSK-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 | 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 $P$ value $< 0.001$ | Not cumulative | 55 | 312 | 154 |

*For Tier 1 and Tier 2 regions only, novel to SLE or first observed in specific ancestral cohort. 
1Not cumulative. 
2FDR $P$ value represents the Benjamini–Hochberg adjusted false discovery rate $P$ value.
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 disenable 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 lupus, 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*15:01/03-DQA1*01:02-DQB1*06:01), which are comprised of alleles in strong linkage disequilibrium. Thus, we selected DRB1*03:01 and 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 \((P_{\text{AA}} = 1.79 \times 10^{-10}, P_{\text{HA}} = 4.65 \times 10^{-28})\). 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 | Novel ancestry-specific non-HLA associated regions.

| SNP(s) | Region rank | Ref. allelle | RAF case | RAF control | P value | OR (95% CI) | Regional stepwise P value | dbSNP function |
|--------|-------------|-------------|----------|-------------|---------|-------------|--------------------------|-----------------|
| rs1550283 | 3q13 | 1245368 | CPFS3L | G | 0.188 | 0.211 | 1.27 | 0.226–2.18 | intron |
| rs11681718 | 7q11 | 10301544 | IL1RAP | C | 0.252 | 0.287 | 1.80 | 0.88–3.93 | coding-synon |
| rs7135569 | 19p13 | 18517331 | LRRC25-SSBP4 | A | 0.139 | 0.165 | 1.51 | 0.50–4.64 | missense |
| rs5444697 | 1q22 | 155034632 | ADAM15-ENF1 | G | 0.225 | 0.210 | 1.39 | 0.08–2.17 | intron |
| rs9701606 | 10p15 | 6070276 | IL2RA | A | 0.256 | 0.307 | 1.31 | 0.13–17.43 | missense |
| rs11587156 | 2q11 | 9911910 | GTF2I-ENF1 | A | 0.273 | 0.291 | 1.57 | 0.23–11.69 | missense |
| rs13300438 | 1q22 | 5654227 | IL2RA | A | 0.208 | 0.182 | 1.15 | 0.69–1.88 | missense |
| rs3519297 | 10p15 | 6100304 | PAPOLG | C | 0.077 | 0.070 | 1.49 | 0.14–15.63 | missense |
| rs2600669 | 10p15 | 6140129 | PAPOLG | T | 0.366 | 0.387 | 1.01 | 0.13–11.32 | missense |
| rs11861718 | 7q11 | 9911910 | GTF2I-ENF1 | T | 0.275 | 0.270 | 1.43 | 0.09–2.35 | coding-synon |
| rs4264825 | 1q22 | 15503091 | ADAM15-ENF1 | T | 0.226 | 0.211 | 1.38 | 0.12–15.43 | missense |
| rs12971295 | 19p13 | 18191621 | LRRC25-SSBP4 | T | 0.055 | 0.030 | 3.03 | 0.07–1.28 | intron |
| rs11673485 | 19p13 | 18337058 | LRRC25-SSBP4 | G | 0.177 | 0.191 | 1.17 | 0.16–9.49 | missense |
| rs2318463 | 19p13 | 18517331 | LRRC25-SSBP4 | G | 0.177 | 0.165 | 1.20 | 0.06–1.17 | missense |
| rs4923177 | 5q11 | 17370148 | LRRC25-SSBP4 | C | 0.480 | 0.470 | 1.48 | 0.17–20.90 | missense |
| rs6896898 | 5q13 | 15888037 | IL1RAP | C | 0.252 | 0.287 | 1.47 | 0.88–2.46 | missense |
| rs7170664 | 5q13 | 15888037 | IL1RAP | T | 0.465 | 0.500 | 1.35 | 0.68–2.62 | missense |
| rs2052993 | 17q25 | 73404537 | GRB2 | T | 0.061 | 0.106 | 3.24 | 0.27–36.40 | intron |
| rs1275926 | 17q25 | 73404537 | GRB2 | C | 0.233 | 0.197 | 1.22 | 0.10–22.14 | intron |
| rs33304138 | 1q22 | 5654227 | IL2RA | A | 0.208 | 0.182 | 1.15 | 0.69–1.88 | missense |
| rs11673485 | 19p13 | 18337058 | LRRC25-SSBP4 | G | 0.177 | 0.191 | 1.17 | 0.16–9.49 | missense |
| rs33304138 | 1q22 | 5654227 | IL2RA | A | 0.208 | 0.182 | 1.15 | 0.69–1.88 | missense |
| rs4264825 | 1q22 | 15503091 | ADAM15-ENF1 | T | 0.226 | 0.211 | 1.38 | 0.12–15.43 | missense |
| rs4544697 | 1q22 | 155034632 | ADAM15-ENF1 | G | 0.225 | 0.210 | 1.39 | 0.08–2.17 | intron |
| rs9701606 | 10p15 | 6070276 | IL2RA | A | 0.256 | 0.307 | 1.31 | 0.13–17.43 | missense |
| rs4264825 | 1q22 | 15503091 | ADAM15-ENF1 | T | 0.226 | 0.211 | 1.38 | 0.12–15.43 | missense |

*Note: The table includes various SNPs associated with different non-HLA related regions, showing their chromosome location, position, region, rank, reference allele, RAF case, RAF control, p-value, OR (95% CI), and regional stepwise P value. Some SNPs are marked with functional annotations such as intron or missense.*
or HA (P = 0.50), the lack of fit-supported the dominance model of risk (departure from additivity; see Methods) for an individual DR3 (EA P = 7.90 × 10−99; AA P = 0.06; HA P = 5.14 × 10−10) and DR15 (EA P = 5.79 × 10−23; AA P = 3.99 × 10−11; HA P = 3.25 × 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 SSE clustering 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 diseases.15 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’ region16,17.

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 DQ6/5 sublineage, both DQA1*01:02 and DQB1*06:02 yield 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).

**Gender-HLA and genome-wide SNP-HLA interaction.** There was evidence that the risk of SLE differs 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 × 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 × 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 × 10−1); 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 × 10−8) and the AA helped localize the association. The region about 16q13 (PLLP-CLC22) exhibited modest association in individual ancestries, but reached genome-wide significance for rs223889 (OR = 1.21, P = 1.08 × 10−8) in the meta-analysis. Similarly, rs137956 (OR = 0.88, P = 5.00 × 10−8) on 22q13 between ENTHDI 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 × 10−11) near ZFP90 had strong support from both EA and AA, with AA samples localizing the association...
expression association was reported in GTEx for whole blood CpG sites. The region has been previously reported in Asian lymphocytes). Huang muscle, brain cortex, lung, testis and EBV-transformed ZFP90 region. In GTEx, the G allele corresponds to lowest gene G allele is the risk allele and creates a CpG site in the promoter. ZFP90 helper cells which stimulates B-cells to differentiate into IL21. The top RegulomeDB scores for these 9 SNPs were 1f for rs7219, reflecting rs7219 as a known cis from 93 to 82 kb. The best RegulomeDB scores for these 9 SNPs is an eQTL for SLE-associated SNPs localized by the AA signal in the meta-analysis. The top EA signal was associated with SLE in the EA cohort and further fine-mapping efforts. These include rs6886392 on 5q21 (frequency 0.003) in the other ancestral cohorts. This suggests 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 \( \beta \)-weighted (\( \beta = \log(OR) \)), 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 ORunweighted > 30 and ORweighted > 100 for the highest load groups. In fact, there was a nonlinear trend in the log(OR) (that is, \( \beta \) 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 additional loci not previously reported as having genome-wide significance for SLE. Of the 147,111 Immunochip SNPs that passed quality control analysis, only 30% begin or end a CpG site. Although this is a novel SLE association, GRB2 reportedly regulates SHP2 activity, a potential contributor to SLE pathogenesis. 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, \( P = 4.08 \times 10^{-5} \)), 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 ancestry 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. 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 PTG1 and MIR146A has an \( r^2 = 0.52 \) with rs2431697, a SNP correlated with down-regulation of MIR146A.

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 UHFRFI 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 transancestral effects. These included: 1q25 (TNFSF4-LOC100506023), 1q25 (NMNAT2-SMG7-NC2), 7q32 (IRF5-TNP03), 8q12 (LYN-RPS20), 11p13 (PDHX-CD44) and 20q13 (NCOA5-CD40) (Table 4, Supplementary Fig. 15).
**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 |
|------------|------|----------------|-------------|-------------|-----------|----------|-------------|---------|-------------|----------------|
| rs3828069  | 1p31 | 67839573       | IL12RB2     | G           | Meta      | -        | -           | 1.77 x 10^-9 | 0.85 (0.79-0.90) | intron |
| rs1432296  | 2p16 | 61068167       | PAPOLG-LINC01185 | A           | Meta      | -        | -           | 1.34 x 10^-8 | 1.18 (1.10-1.26) | |
| rs3733345  | 4p16 | 954247         | DGKQ        | G           | Meta      | -        | -           | 1.83 x 10^-11 | 0.89 (0.85-0.92) | untranslated |
| rs6886392  | 5q21 | 100135865      | ST8SIA4     | C           | Meta      | -        | -           | 5.84 x 10^-8 | 0.89 (0.85-0.93) | |
| rs34840245 | 6p21 | 34812701       | UHRF1B1-DEF6 | G           | Meta      | -        | -           | 1.19 x 10^-3 | 1.16 (1.04-1.29) | |
| rs4739134  | 8q21 | 79556148       | PKIA-ZC2HC1A | T           | Meta      | -        | -           | 3.47 x 10^-6 | 1.12 (1.07-1.17) | |
| rs11788118 | 9q22 | 102337331      | AK057451    | A           | Meta      | -        | -           | 7.00 x 10^-2 | 1.11 (0.99-1.25) | |
| rs653178   | 12q24| 112007756      | ATXN2       | C           | Meta      | -        | -           | 2.14 x 10^-16 | 0.88 (0.84-0.92) | intron |
| rs2041670  | 16p13| 11174652       | CLEC16A     | T           | Meta      | -        | -           | 1.19 x 10^-11 | 0.84 (0.75-0.94) | |
| rs223889   | 16q13| 57392241       | PLLP-CCL22  | T           | Meta      | -        | -           | 1.08 x 10^-8 | 1.21 (1.13-1.29) | near-gene-5 |
| rs1749792  | 16q22| 68569440       | ZFP90       | T           | Meta      | -        | -           | 1.67 x 10^-3 | 1.11 (1.00-1.23) | |
| rs8072449  | 17q25| 73312184       | GRB2        | G           | Meta      | -        | -           | 1.19 x 10^-11 | 0.84 (0.80-0.89) | |
| rs13344313 | 19p13| 18517767       | LRR25-SSBP4 | A           | Meta      | -        | -           | 1.19 x 10^-11 | 0.84 (0.80-0.89) | |
| rs56154925 | 19q13| 55737798       | PTPRH-TMEM86B | T           | Meta      | -        | -           | 1.77 x 10^-9 | 1.14 (1.08-1.21) | near-gene-3 |
| rs137956   | 22q13| 40293463       | ENTHD1-GRAP2 | G           | Meta      | -        | -           | 1.02 x 10^-7 | 1.14 (1.08-1.21) | |

**Tier 2 Meta-Analysis**

| SNP        | Chr. | Position (b37) | Gene region | Ref. allele | Ancestry | RAF case | RAF control | P value | OR (95% CI) |
|------------|------|----------------|-------------|-------------|-----------|----------|-------------|---------|-------------|
| rs6662618  | 1p22 | 92935411       | GF11-EV15   | T           | Meta      | -        | -           | 1.02 x 10^-7 | 1.14 (1.08-1.21) | |

**Notes:**
- HA: Haplogroup A
- AA: Haplogroup A
- EA: Haplogroup E

**P values and ORs:**
- HA: Haplogroup A
- AA: Haplogroup A
- EA: Haplogroup E

**Control:**
- P: significant at 5% level.
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 (rSpearman = -0.14, P = 0.0001), and HA (rSpearman = -0.10, P = 0.0012), but not AA (rSpearman = 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 at least 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
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-PROSI), 6q25 (RSPH3), 12q15 (DYNK2-IFNG), 12q21 (SYT1), 14q31 (GALC), 16q21 (CSNK2A2-CCDC113) and 22q12 (C1QTNF6). 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 haplotypes34,35.

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,
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 DRB1*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 stability, influence of DNA methylation and DR-DQ heterodimers. Such findings may help prioritize functional experiments, as we work towards understanding the HLA mechanisms of SLE.

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).
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.

Two major limitations of this study are the comparatively 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.

**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 x 10^-30 | 1.30 (1.23-1.38) |
| TNFSF4-LOC100506023 | rs1539255 | 1q25 | 173322660 | T | Meta | - | - | 1.60 x 10^-19 | 0.84 (0.81-0.87) |
| MNAT2-SMG7-NCF2 | rs17484292 | 1q25 | 183300050 | T | Meta | - | - | 9.97 x 10^-38 | 1.59 (1.40-1.79) |
| IL2-IL21 | rs11724582 | 4q27 | 123391464 | C | Meta | - | - | 2.52 x 10^-17 | 1.17 (1.13-1.22) |
| PTTG1-MIR146A | rs2431098 | 5q33 | 159887336 | C | Meta | - | - | 2.39 x 10^-6 | 0.89 (0.84-0.93) |
| IRF5-TNPO3 | rs4728142 | 7q32 | 128573967 | T | Meta | - | - | 1.71 x 10^-8 | 1.03 (0.97-1.09) |
| IRF5-TNPO3 | rs35000415 | 7q32 | 128585616 | T | Meta | - | - | 1.17 x 10^-9 | 1.82 (1.69-1.96) |
| LYN-RPS20 | rs2953898 | 8q12 | 56980803 | A | Meta | - | - | 4.43 x 10^-8 | 0.84 (0.79-0.90) |
| PDHX-CD44 | rs353592 | 11p13 | 35119482 | T | Meta | - | - | 1.35 x 10^-8 | 0.89 (0.85-0.93) |
| SLC15A4 | rs1059312 | 12q24 | 129278864 | G | Meta | - | - | 6.53 x 10^-10 | 1.12 (1.07-1.16) |
| NCOA5-CD40 | rs4810485 | 20q13 | 44747947 | A | Meta | - | - | 9.95 x 10^-9 | 1.43 (1.17-1.76) |

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

The ethics review committees included: Cedars-Sinai Medical Center.
Institutional Review Board; Central Ethic Committee of Denmark; Centrala etik-tigacie en Salud, Instituto Mexicano del Seguro Social, Mexico; Regional Ethical Committee Office Northwestern University; Johns Hopkins University School of Medicine, Rheumatology, Umeå University, Umeå, Sweden; Institutional Review Board Office Northwestern University; Johns Hopkins University School of Medicine, Institutional Review Board; University Medical Center Groningen; Medical University of South Carolina Institutional Review Board; Ethics and Research Protection Program Institutional Review Board; UHN REB; University of Alabama at Birmingham; UC Davis Institutional Review Board; UCSF Human Research Ethics Board; Institutional Review Board for Human Use 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; Regional Ethic Committee 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 ‘noinput’ 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 for the batches was: Batch I was 15% European ancestry (EA), 7% African American ancestry (AA), 55% Asian ancestry (ASA), 23% Hispanic ancestry (HHA); Batch II was 44% EA, 14% AA, 1.4% ASA, 36% HHA; Batch III was 48% EA, 38% AA, 1% ASA, 13% HHA; 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-pseudautosomal regions of chromosome X to reliably complete gender checks. Admixture estimates were computed using the program ADMIXTURE44. HapMap phase 2 individuals (CEU: Utah residents from 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 v4.4 (refs 46,47) including HapMap phase 2 individuals (CEU, YRI and CHB). Reference population ancestry estimates 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, or with low call rate (<95%).

The admixture estimates and PCs were used to identify and remove genetic confounders. Samples were retained from the primary analysis if it had an overall call rate >95%, exhibited significant differential miscallings between controls and cases (P <0.05), had significant departure from Hardy-Weinberg equilibrium expectations (P <1×10−8 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, >1,000 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, P <0.20, within each ancestry. The union of these SNPs across ancestries was 46,744 uncorrelated SNPs, yielding a Bonferroni threshold of P <1.06×10−6.

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. A SNP test for an association with a specific patient 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 of 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.

Table 5 | Genetic Load and SLE risk.

|                  | P value | OR (95% CI) | c-statistic | P value | OR (95% CI) | c-statistic |
|------------------|---------|-------------|-------------|---------|-------------|-------------|
| Non-HLA SNPs     |         |             |             | Combined |             |             |
| Un-Weighted      |         |             |             | Combined |             |             |
| AA               | 5.31×10−30 | 1.15 (1.13-13) | 0.599 (0.590) | 3.67×10−3 | 0.724 (0.718) | 1.21×10−3 |
| EA               | 5.31×10−30 | 1.15 (1.13-13) | 0.604 (0.605) | 3.67×10−3 | 0.724 (0.718) | 1.21×10−3 |
| HA               | 1.16×10−26 | 1.14 (1.12-15) | 0.647 (0.624) | 3.41×10−3 | 0.638 (0.560) | 7.02×10−3 |

Weighted by Natural Log of the Odds Ratio (OR)

|                  | P value | OR (95% CI) | c-statistic | P value | OR (95% CI) | c-statistic |
|------------------|---------|-------------|-------------|---------|-------------|-------------|
| Non-HLA SNPs     |         |             |             | Combined |             |             |
| Un-Weighted      |         |             |             | Combined |             |             |
| AA               | 2.28×10−26 | 2.59 (2.33-3.00) | 0.602 (0.601) | 2.24×10−3 | 0.551 (0.550) | 1.80×10−3 |
| EA               | 1.37×10−01 | 0.64 (0.49-0.96) | 0.738 (0.714) | 2.96×10−3 | 0.678 (0.618) | 2.19×10−2 |
| HA               | 2.09×10−16 | 3.76 (3.14-4.51) | 0.674 (0.660) | 3.43×10−3 | 0.645 (0.582) | 1.98×10−3 |

*Top hits from EA sample without validation set of 2,000 SLE cases, 2,000 controls.

**Her 5 alleles.****

†Whole model statistic; in parentheses, the c-statistic for model without admixture factors.

§EA random sample with 2,000 SLE cases, 2,000 controls.

**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. A SNP test for an association with a specific patient 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 of 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.
For ancestry-specific analysis of the X chromosome, the data were first stratified by SNPLASH (https://www.phs.wakehealth.edu/public/bios/gene/downloads.cfm). These tests of association were computed using the SNPGWA version 4.0 module. Increased frequencies in AA yielded three distinctive categories for SLE risk alleles: increased AA-ancestral contribution, decreased frequencies in AA, and reduced AA-ancestral contribution. These alleles were enriched for admixture deviations from average admixture proportions. The largest cluster had the smallest deviations from average admixture proportions, contained the most risk alleles, of which many were common alleles. This allele frequency yields three distinctive categories for SLE risk alleles: increased AA-ancestral contribution, decreased frequencies in AA, and reduced AA-ancestral contribution.

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 driven by monomorphic SNPs in other populations.

These tests of association were computed using the SNPGWA 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 ($\lambda_{GC}$) 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 $\lambda_{GC}$ 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 $5 \times 10^{-8}$. Tier 2 includes those SNPs that are not Tier 1 SNPs, but have a $P$ value for association less than $1 \times 10^{-6}$. Tier 3 includes those SNPs that do not meet criteria for Tiers 1 or 2, but meet a genome-wide Benjamini–Hochberg false discovery rate 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 METAL. Regions of association were visually examined and tests of heterogeneity of the odds ratio were computed. Thus, for each region, ancestry-specific meta-analysis and meta-analytic tests of association and tests of heterogeneity were reported. The transancestral patterns of association and LD were visualised using LocusZoom. 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 HIBAG. HIBAG uses an ensemble classifier and bagging technique to arrive at an average posterior probability. Unlike an average posterior probability, our alternative imputation software such as BEAGLE, HLA*IMP and SNP2HLA, 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 approaches, 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 uncertainty via allelic dose. 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 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-analogue 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 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 Database. Sequences within an HLA-gene were aligned using ClustalOmega. 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 substitution. 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 clustering. The resulting trees were visualized using Dendroscope. 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,
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 SNPs of good quality across the three populations and for each SNP, the risk allele's proportion was compared to the overall average sample admixture proportion in each cohort. Within each of the three populations and for each SNP, the risk allele's average admixture was computed. The resulting 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:

$$G_{RS_i} = \sum_{k=1}^{N} \beta_k R_{Ak}$$

where, $G_{RS_i}$ is the genetic risk score for individual i; $\beta_k$ is the beta coefficient for the kth SNP association with SLE and serves as the weight for that risk allele; $R_{Ak}$ is the number of risk alleles for the kth SNP (0, 1, 2); and N is the number of SNPs. By definition of parameterizing relative to the risk allele, $\beta_k > 0$ 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) = β 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 SNAP”. Specifically, LD was computed using the CEU (for EA and HA) or YRI (for AA) data with an $r^2 > 0.5$ 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 Browser . 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, DNA-Ase hypersensitivity, binding sites of transcription factors) in the intergenic regions of the human genome. 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 variants. 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 blocks 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 browser.

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, or non-missense, or mapped within an active enhancer 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**

1. Danchenko, N., Satia, J. A. & Anthony, M. S. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. *Lupus* 15, 308–318 (2006).

2. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGENS) et al. Genome-wide association study of systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat. Genet.* 40, 204–210 (2008).

3. Chung, S. A. et al. Lupus nephritis susceptibility loci in women with systemic lupus erythematosus. *J. Am. Soc. Nephrol.* 25, 2859–2870 (2014).

4. Hom, G. et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX, N. Eng. J. Med. 358, 900–909 (2008).

5. Rullo, O. J. & Tsao, B. P. Recent insights into the genetic basis of systemic lupus erythematosus. *N. Engl. J. Med.* 360, 2544–2555 (2009).

6. Boyle, A. P. et al. Annotation of functional variation in personal genomes using chromatin state, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.* 40, D930–D934 (2012).

7. 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).

8. Han, I.-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).

9. Wang, C. et al. Genes identified in Asian SLE GWAS are also associated with SLE in Caucasian populations. *Eur. J. Hum. Genet.* 21, 994–999 (2013).

10. Ahmed, Z. et al. Direct binding of Grb2 SH3 domain to FGFR2 regulates SHP2 function. *Cell. Signal.* 22, 23–33 (2010).

11. Sun, J. et al. Antagonsism between binding site affinity and conformational dynamics tunes alternative cis-interactions within Shp2. *Nat. Commun.* 4, 2037 (2013).

12. Wang, J. et al. Inhibition of SHP2 ameliorates the pathogenesis of systemic lupus erythematosus. *J. Clin. Invest.* 126, 2077–2092 (2016).

13. Sawalha, A. H. et al. Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. *Ann. Rheum. Dis.* 67, 458–461 (2008).

14. Choi, J.-Y. et al. Circulating follicular helper-like T cells in systemic lupus erythematosus: association with disease activity: circulating Th1-like cells in SLE. *Arthritis Rheumatol.* 67, 988–999 (2015).

15. Lofgren, S. E. et al. Genetic association of miRNA-146a with systemic lupus erythematosus in European women with decreased expression of the gene. *Genes Immun.* 13, 268–274 (2012).

16. 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).

17. Kwkow, W. W., Kovats, S., Thorle, P. & Nepom, G. T. HLA-DQA1 allelic polymorphisms constrain patterns of class II heterodimer formation. *Hum. Immunol.* 150, 2263–2272 (1993).

18. 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).

19. van Heerst, J., Huttinga, T. J., van der Wooude, 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).

20. 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).

21. Majumder, P. & Boss, J. M. DNA methylation disregulates and silences the HLA-DQ locus by altering chromatin architecture. *Genes Immun.* 12, 291–299 (2011).

22. Rutkovsky, N. et al. 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).

23. Tan, E. M. et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 25, 1271–1277 (1982).

24. Hochberg, M. C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* 40, 1725 (1997).

25. Shah, T. S. et al. OptiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants. *Bioinformatics* 28, 1598–1603 (2012).
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 NIHR grants AR049084 (RPK, EEB) and the International Consortium on the Genetics of Systemic Lupus Erythematosus (SLEGEN) ABO38194 (J.H.B.); CA141700, AR058621 Proyecto de Excelencia, Consejería de Andalucía (M.E.R.A.R); AR043814 and AR-065626 (R.P.T.); AR060366, MD007909, AI107176 (S.K.N.); AR-057172 (C.O.J.); AR053038 (L.A.C), MUSC part is from UL1RR029882 (G.S.G., D.L.K.) and 5P60AR062755 (G.S.G., D.L.K.), MUSC part is from UL1RR029882 (G.S.G., D.L.K.) and 5P60AR062755 (G.S.G., D.L.K.), MUSC part is from UL1RR029882 (G.S.G., D.L.K.) and 5P60AR062755 (G.S.G., D.L.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.K.W., D.K.S. and F.P.T. and E.K.W. conceived and designed the experiments. P.M.G., R.R.G., J.A.K., C.D.L., E.S.K., T.J.V. and E.K.W. performed experiments. H.C.A., M.E.C., T.D.H., 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., M.C.M., D.R.M., P.S.R. performed the data analysis. P.M.G., R.R.G., R.P.K., C.D.L., E.S.K. and T.J.V. performed 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., M.A.B., V.B., T.W.B., A.A.B., G.A.B., T.B., M.B., E.B., E.R., H.B., L.B., J.K., M.H.C., J.C.A., L.C., R.C., L.A.C., I.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.L.T., J.M.G., J.L.H., C.O.I., J.A.I., C.G.M.K., D.L.K., D.R.K., P.K., L.H., 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., R.R., J.D.R., L.R., J.M.S., C.A.S., I.S., H.S., R.S., M.F.S., E.D.S., K.L.S., C.S., E.A., S.C.T., S.D.T., S.M.A.T., L.T., B.P.T., T.T., C.V., L.M.V., D.J.W., H.M.W. and J.E.W. contributed samples. E.M.A.-V., G.S.A., M.E.A., M.A.B., V.B., T.W.B., A.A.B., G.A.B., T.B., M.B., E.B., E.R., H.B., L.B., J.K., M.H.C., J.C.A., L.C., R.C., L.A.C., I.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.L.T., J.M.G., J.L.H., C.O.I., J.A.I., C.G.M.K., D.L.K., D.R.K., K.M.K., J.A.K., R.P.K., L.C.K., L.H., 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., S.K.N., T.B.N., J.R.O., L.O., N.O., M.P., C.A.P., B.A.P., J.P., P.S.R., B.R., J.D.R., L.R., L.P.R., J.M.S., C.A.S., I.S., H.S., R.S., M.F.S., E.D.S., K.L.S., C.S., E.A., S.C.T., S.D.T., S.M.A.T., L.T., B.P.T., T.T., C.V., L.M.V., D.J.W., E.K.W., J.E.W., M.H.W. and D.J.W. revised the manuscript.

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