Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus

James Bentham¹,¹²,¹³, David L Morris¹,¹³, Deborah S Cunninghame Graham¹, Christopher L Pinder¹, Philip Tombleson¹, Timothy W Behrens², Javier Martín³, Benjamin P Fairfax⁴, Julian C Knight⁴, Lingyan Chen¹, Joseph Replogle⁵, Ann-Christine Svyänen⁶, Lars Rönnblom⁶, Robert R Graham², Joan E Wither⁷, John D Rioux⁸,⁹, Marta E Alarcón-Riquelme¹⁰ & Timothy J Vyse¹,¹¹

Systemic lupus erythematosus (SLE) is a genetically complex autoimmune disease characterized by loss of immune tolerance to nuclear and cell surface antigens. Previous genome-wide association studies (GWAS) had modest sample sizes, reducing their scope and reliability. Our study comprised 7,219 cases and 15,991 controls of European ancestry, constituting a new GWAS, a meta-analysis with a published GWAS and a replication study. We have mapped 43 susceptibility loci, including ten new associations. Assisted by dense genome coverage, imputation provided evidence for missense variants underpinning associations in eight genes. Other likely causal genes were established by examining associated alleles for cis-acting eQTL effects in a range of ex vivo immune cells. We found an over-representation (n = 16) of transcription factors among SLE susceptibility genes. This finding supports the view that aberrantly regulated gene expression networks in multiple cell types in both the innate and adaptive immune response contribute to the risk of developing SLE.

SLE (MIM 152700) is a clinically heterogeneous disease with a strong genetic component, as demonstrated by the tenfold increase in concordance rates between monozygotic and dizygotic twins¹ and familial aggregation (sibling risk ratio, λs = 29)². Since 2008, the field of SLE genetics has been transformed by genome-wide association studies³–⁸ and independent replication studies⁹,¹⁰. However, although the pace of discovery has been unprecedented, providing a richer understanding of lupus genetic etiology, the findings were driven by modestly sized GWAS, using 1,800 European cases¹¹ and slightly fewer Asian cases⁵,⁶; these studies therefore had limited power to detect loci with relatively low odds ratios and/or minor allele frequencies (MAFs)¹¹. The size of our study, coupled with a meta-analysis and replication study, has greatly increased power to detect susceptibility loci.

We genotyped 4,946 individuals with SLE (cases) and 1,286 healthy controls using the Illumina HumanOmni1-Quad BeadChip. These data were combined with the genotypes of 5,727 healthy controls taken from the University of Michigan Health and Retirement Study (HRS), which were generated using the Illumina HumanOmni2.5 BeadChip. After quality control analyses, our data comprised 4,036 SLE cases and 6,959 controls (1,260 controls mainly from southern Europe genotyped using the HumanOmni1-Quad chip and 5,699 controls from the HRS cohort). The final SNP set comprised 644,674 markers that were present on both the HumanOmni1-Quad and HumanOmni2.5 chips (Online Methods). Four principal components were used as covariates to correct for population structure¹²,¹³. The genomic inflation factor¹⁴,¹⁵ for our data, λ₁,000⁰ was 1.02, with λGCC = 1.16.

Our analysis strategy is described in detail in the Online Methods and is shown schematically in Supplementary Figure 1. The GWAS identified 25 loci (Table 1 and Supplementary Fig. 2a) of genome-wide significance (P < 5 × 10⁻⁸). Three of these associations are new in SLE: rs6740462 and rs3768792 on chromosome 2p14 and 2q34, respectively, and rs7726414 on chromosome 5q31.1.

To validate these findings and to search for additional susceptibility loci, we carried out a meta-analysis of our GWAS results and those from an independent European SLE GWAS comprising 1,165 cases and 2,107 controls (the study by Hom et al.⁴). Each of the 25 loci mapped in the original GWAS had a genome-wide significant P value in this meta-analysis (Supplementary Table 1), and these loci are therefore considered to be associated with SLE. We then designed a replication study, with locus inclusion based on the meta-analysis of the two GWAS. At loci with no published association in SLE, we adopted a threshold of .000001.
for inclusion of \( P < 2.5 \times 10^{-5} \), whereas, for loci with previously reported associations, the threshold was set at \( P < 1 \times 10^{-8} \) (see the Online Methods for rationale). The 33 SNPs with \( P \) values meeting these criteria were genotyped in our replication study (Supplementary Table 2), using a custom panel that also included 53 ancestry-informative markers (Online Methods). After applying quality control measures, the replication data comprised 2,018 cases and 6,925 controls, none of which had been included in either GWAS (Online Methods).

Finally, we carried out a post-replication meta-analysis of the results of our GWAS, the study by Hom et al. and the replication study for the 33 SNPs, again applying the standard measure of genome-wide significance. The 18 SNPs (in addition to the 25 already mapped) with \( P \) values <5 \( \times 10^{-8} \) in this meta-analysis were also considered to be associated with SLE (Table 1 and Supplementary Fig. 2b).

In addition to the three new loci mapped in the GWAS, seven to be associated with SLE (Hom et al., 2014), the study by Hom et al. and the replication study for the 33 SNPs, again applying the standard measure of genome-wide significance. The 18 SNPs (in addition to the 25 already mapped) with \( P \) values <5 \( \times 10^{-8} \) in this meta-analysis were also considered to be associated with SLE (Table 1 and Supplementary Fig. 2b).
association in SLE, were mapped in the overall meta-analysis: rs564799 (3q25.33), rs3794060 (11q13.4), rs10774625 (12q14.1), rs4902562 (14q24.1), rs9652601 (7q32.1), rs2286672 (17p13.2) and rs887369 (Xp21.2). The heritability explained by the risk alleles mapped at the 43 validated susceptibility loci is 15.3% (95% confidence interval (CI) = 10.6–21.2%), where the total heritability of lupus is estimated to be 66% (ref. 16). This is a large increase over the 8.7% (5.33–12.96%) reported by So et al.17 in 2011 using the same measure.

We imputed data from both the main GWAS and the study by Hom et al. to the density of the 1000 Genomes Project study18 and reanalyzed the data (Online Methods). Although no additional loci were identified, we did obtain stronger evidence in support of some loci, for example, the signal at the SPRED2 locus, at which the most strongly associated 1000 Genomes Project variant, rs268134, was strongly replicated. In addition, imputation enabled us to fine map associated loci and to determine whether multiple signals were present at each locus (Supplementary Table 3a,b). We identified multiple independent association signals at the TNFSF4, STAT4 and IRF5 loci, as well as five independently associated SNPs at the major histocompatibility complex (MHC) locus.

Given that the SNP with the smallest $P$ value at a locus is not necessarily the true causal variant, we considered SNPs from the most associated to a defined cutoff of association as potentially causal in our subsequent analyses. Specifically, guided by previous work on functional annotation19 (Online Methods), the cutoff was defined as a Bayes factor with respect to the most significantly associated SNP equal to 0.34. Any SNPs in the resulting set for each locus that were missense variants were considered to be more likely candidates than the most significantly associated SNP. The results are summarized in Supplementary Tables 3c and 4, listing candidate causal missense variants in PTPN22, FCGR2A, NCF2, TNFAIP3, WDFY4, IRF7, ITGAM and TYK2.

MHC polymorphisms, including SNPs and classical human leukocyte antigen (HLA) alleles, have consistently been observed to be associated with SLE20. We imputed HLA alleles21 in data from both the main GWAS and the study by Hom et al. and incorporated them into our analysis of 1000 Genomes Project–imputed data across the MHC locus (Online Methods). Of the five MHC SNPs we found to be independently associated with SLE (Supplementary Table 3a,b), the class III SNP in SLC44A4 (rs74290525) was the only association signal that was clearly independent of any HLA alleles. We found that rs74290525 was significantly associated not only when conditioning on each of the HLA genes separately but even when conditioning on all 199 HLA alleles (Supplementary Table 5a–e), and this SNP is not in linkage disequilibrium (LD) with any HLA alleles ($r^2 < 0.1$ with each HLA allele). We found that the best model for association included the HLA class I alleles HLA-B*08:01 and HLA-B*18:01, the class II alleles HLA-DRB1*04:05 and HLA-DQB1*06:02, and the class III SNP rs74290525, consistent with previous findings suggesting multiple SLE associations in the MHC region20 (Supplementary Table 6a,b). The LD between the five HLA SNPs and HLA alleles on known SLE risk-associated haplotypes is shown in Supplementary Table 6c.

To highlight potential causal genes at the susceptibility loci, we tested the associated SNPs at each of the loci for correlation with cis-acting gene expression in ex vivo cells. The heat map includes all genes with evidence of cis-regulatory action by SLE-associated SNPs (within 1 Mb of the target gene) in at least one cell type. Color represents a signed RTC score: a positive score indicates that the associated allele in the GWAS is positively correlated with gene expression, and a negative score indicates that the associated allele in the GWAS is negatively correlated with gene expression. We set the RTC score to zero if the $P$ value for association was >0.001. Colors represent the RTC score as follows: blue, $RTC < -0.9$ (the GWAS risk allele reduces expression); green, $RTC < -0.5$ (the GWAS risk allele reduces expression); yellow, $-0.5 < RTC < 0.5$; orange, $RTC > 0.5$ (the GWAS risk allele increases expression); red, $RTC > 0.9$ (the GWAS risk allele increases expression). A white block indicates that data were not available for this cell type (see Supplementary Fig. 4) for results in lymphoblastoid cell lines (LCLs), either because the probe data failed quality control or the probe was not present on the experimental platform. Clustering was performed on the cell types, including only genes with data observed for all cell types (missing data did not inform cell type clustering). Genes were clustered using all available data across cell types (missing data were not included when determining the distance between pairs of genes if eQTL results were not observed for one of the pair). LPS, lipopolysaccharide.

Supplementary Table 3a

Supplementary Table 3b

Supplementary Table 3c

Supplementary Table 3d

Supplementary Table 3e

Supplementary Table 4

Supplementary Table 5a

Supplementary Table 5b

Supplementary Table 5c

Supplementary Table 5d

Supplementary Table 5e

Supplementary Table 6a

Supplementary Table 6b

Supplementary Table 6c

Supplementary Table 6d

Supplementary Table 6e

Supplementary Fig. 3a

Supplementary Fig. 3b

Supplementary Fig. 4

Supplementary Fig. 5a

Supplementary Fig. 5b

Supplementary Fig. 5c

Supplementary Fig. 5d

Supplementary Fig. 5e

Supplementary Fig. 6a

Supplementary Fig. 6b

Supplementary Fig. 6c

Supplementary Fig. 6d

Supplementary Fig. 6e

Supplementary Fig. 7a

Supplementary Fig. 7b

Supplementary Fig. 7c

Supplementary Fig. 7d

Supplementary Fig. 7e

Supplementary Fig. 8a

Supplementary Fig. 8b

Supplementary Fig. 8c

Supplementary Fig. 8d

Supplementary Fig. 8e

Supplementary Fig. 9a

Supplementary Fig. 9b

Supplementary Fig. 9c

Supplementary Fig. 9d

Supplementary Fig. 9e

Supplementary Fig. 10a

Supplementary Fig. 10b

Supplementary Fig. 10c

Supplementary Fig. 10d

Supplementary Fig. 10e

Supplementary Fig. 11a

Supplementary Fig. 11b

Supplementary Fig. 11c

Supplementary Fig. 11d

Supplementary Fig. 11e

Supplementary Fig. 12a

Supplementary Fig. 12b

Supplementary Fig. 12c

Supplementary Fig. 12d

Supplementary Fig. 12e

Supplementary Fig. 13a

Supplementary Fig. 13b

Supplementary Fig. 13c

Supplementary Fig. 13d

Supplementary Fig. 13e

Supplementary Fig. 14a

Supplementary Fig. 14b

Supplementary Fig. 14c

Supplementary Fig. 14d

Supplementary Fig. 14e

Supplementary Fig. 15a

Supplementary Fig. 15b

Supplementary Fig. 15c

Supplementary Fig. 15d

Supplementary Fig. 15e

Supplementary Fig. 16a

Supplementary Fig. 16b

Supplementary Fig. 16c

Supplementary Fig. 16d

Supplementary Fig. 16e

Supplementary Fig. 17a

Supplementary Fig. 17b

Supplementary Fig. 17c

Supplementary Fig. 17d

Supplementary Fig. 17e

Supplementary Fig. 18a

Supplementary Fig. 18b
| Associated SNP | Chr. | Gene within 200 kb of SNP | Gene within same LD block as SNP | Immune phenotype in mouse model | Coding variant with SNP | cis eQTL with SNP | Functional and/or fine-mapping studies | Likely causal gene(s) |
|---------------|------|--------------------------|-------------------------------|-------------------------------|------------------------|-------------------|----------------------------------------|----------------------|
| rs2476601     | 1    | MAGI3, PHTF1, RSBN1, PTN22, BCL2L15, AP4A1, DCLRE1B, HIPK1, OLFM3 | RSBN1, PTN22 | PTPN22 | PTPN22 | PTPN22 | 32 | PTPN22 |
| rs1801274     | 1    | MP2, SDHC, CCn192, FCRG2A, HSPA6, FCRG3A, FCRG2B, FCRG2C, FCRG3B, FCRG3A | FCRG2A | FCRG2A, FCRG2B, FCRG3B | FCRG2A, FCRG2B, FCRG3B | FCRG2A, FCRG2B, FCRG3B | 33 | FCRG2A, FCRG2B, FCRG3B |
| rs704840      | 1    | TNFSF4 | TNFSF4 | TNFSF4 | TNFSF4 | TNFSF4 | 36 | TNFSF4 |
| rs17849501    | 1    | NNNAT2, SMG7, NCF2, ARPC5, RGL1, APOBEC4 | SMG7, NCF2 | NCF2 | SMG7 | NCF2 | 37 | SMG7, NCF2 |
| rs3024505     | 1    | RASSF5, E2F3D, DURK3, MAPKAPK2, IL10, IL19, IL20, IL24, FAIM3, PGR, FCAMR | IL10 | RASSF5, MAPKAPK2, IL10, FAIM3, FCAMR | IL10 | IL10 | 38 | IL10 |
| rs9782955     | 1    | LYST, NID1 | LYST | LYST | LYST | LYST | 39 | LYST |
| rs6740642     | 2    | ACTR2, SPRED2 | SPRED2 | SPRED2 | SPRED2 | SPRED2 | 40 | SPRED2 |
| rs2111485     | 2    | DPP4, CCG, FAP, IFI1H, GCA, KCNH7 | IFI1H | IFI1H | IFI1H | IFI1H | 41 | IFI1H |
| rs11889342    | 2    | GLS, STAT1, STAT4, MYO1B | STAT4 | STAT1, STAT4 | STAT4 | STAT4 | 42 | STAT4 |
| rs3768792     | 2    | IKZF2 | IKZF2 | IKZF2 | IKZF2 | IKZF2 | 43 | IKZF2 |
| rs9311676     | 3    | ABHD6, RPP14, PKX, PDHB, KCTD6, ACOX2, FAM107A, FAM3D | ABHD6, PKX | ABHD6, PKX | ABHD6, PKX | ABHD6, PKX | 44 | ABHD6, PKX |
| rs564799      | 3    | SCH1P1, IL12A | IL12A | IL12A | IL12A | IL12A | 45 | IL12A |
| rs10028805    | 3    | BANK1 | BANK1 | BANK1 | BANK1 | BANK1 | 46 | BANK1 |
| rs7726415     | 3    | C5orf15, VDAC1, TC7F, SKP1 | TC7F | TC7F | TC7F | TC7F | 47 | TC7F, SKP1 |
| rs10036748    | 5    | IRGM, ZNF300, GPX3, TNIP1, ANX6, C6DC69, GM2A, SLC36A3 | TNIP1 | TNIP1, ANX6 | TNIP1 | TNIP1 | 48 | TNIP1 |
| rs2431697     | 5    | C1QTNF2, C5orf54, SL7, PTG1, MIR146A, 3142 | Intergenic | PTTG1 | PTTG1 | PTTG1 | 49 | MIR146A |
| rs1270942     | 6    | MHCd | Intergenic | UHRF1BP1, ANK51A, C6orf106 | UHRF1BP1 | UHRF1BP1 | 50 | UHRF1BP1 |
| rs9462027     | 6    | C6orf106, SNRPQ, UHRF1BP1, TAF11, ANK31A | Intergenic | UHRF1BP1, ANK51A, C6orf106 | UHRF1BP1 | UHRF1BP1 | 51 | UHRF1BP1 |
| rs6568431     | 6    | PRDM1, ATG5 | Intergenic | TNFAIP3 | TNFAIP3 | TNFAIP3 | 52 | TNFAIP3 |
| rs6932056     | 6    | TNFAIP3, PERP | Intergenic | TNFAIP3 | TNFAIP3 | TNFAIP3 | 53 | TNFAIP3 |
| rs849142      | 7    | JAZF1, CREB5 | JAZF1 | JAZF1 | JAZF1 | JAZF1 | 54 | JAZF1 |
| rs4917014     | 7    | ZPB2, C7orf72, IKZF1 | IKZF1 | IKZF1 | IKZF1 | IKZF1 | 55 | IKZF1 |
| rs10488631    | 7    | CALU, OPN1SW, CCDC136, FLNC, ATP6V1F, IRF5, TNP3, TSPAN33 | IRF5, TNP3 | IRF5 | IRF5 | IRF5 | 56 | IRF5 |
| rs2736340     | 8    | MTMR9, SLC355S, C8orf12, FAM167A, BLK, GATA4 | BLK | BLK | BLK | BLK | 57 | BLK |
| rs2663052     | 10   | WDFY4, LRR18C, VSTM4 | WDFY4 | WDFY4 | WDFY4 | WDFY4 | 58 | WDFY4 |
| rs4984969     | 10   | ARID5B, RTKN2 | ARID5B | ARID5B | ARID5B | ARID5B | 59 | ARID5B |
| rs12802200    | 11   | B4AGNT4, PKP3, SIGIRR, AN09, PTDSS2, RNH1, HRAS, LRRCC6, C11orf35, RASSF7, PRR17, IRF7, CDHR3, SCOT, DRR4, DEAF1, EPSL1, TMEM80, TALDO1 | LRRCC6, LMNTO2, RASSF7, MIR210HG, PRR17, IRF7, CDHR3 | LRRCC6, LMNTO2, RASSF7, MIR210HG, PRR17, IRF7, CDHR3 | LRRCC6, LMNTO2, RASSF7, MIR210HG, PRR17, IRF7, CDHR3 | LRRCC6, LMNTO2, RASSF7, MIR210HG, PRR17, IRF7, CDHR3 | 60 | LRRCC6, LMNTO2, RASSF7, MIR210HG, PRR17, IRF7, CDHR3 |
| rs2732549     | 11   | APIP, PDHX, CD44, SLC1A2 | Upstream | CD44 | CD44 | CD44 | 61 | CD44 |
| rs3790600     | 11   | DHCRR7, NADSYN1, KRTAP5 | DHCRR7, NADSYN1 | DHCRR7, NADSYN1 | DHCRR7, NADSYN1 | DHCRR7, NADSYN1 | 62 | DHCRR7, NADSYN1 |
| rs7941765     | 11   | ETS1, FL1, CUX2 | Intergenic | ETS1, FL1 | ETS1, FL1 | ETS1, FL1 | 63 | ETS1, FL1 |
| rs10774625    | 12   | FAM109A, SH2B3, ATXN2, BRAP | SH2B3, ATXN2 | SH2B3, ATXN2 | SH2B3, ATXN2 | SH2B3, ATXN2 | 64 | SH2B3 |
| rs1059312     | 12   | TME132C, SLC15A4, GLT1D1 | SLC15A4 | SLC15A4 | SLC15A4 | SLC15A4 | 65 | SLC15A4 |
| rs4902562     | 14   | RAD51B | SCAMP5, PPCDC | SCAMP5, PPCDC | SCAMP5, PPCDC | SCAMP5, PPCDC | 66 | SCAMP5, PPCDC |
| rs2289585     | 15   | LMAN1, CPLX3, ULK3, SCAMP2, MPI, FAM219B, COX5A, RPP25, SCAMP5, PPCDC, C15orf39 | CSDK, ULK3, MPI, FAM219B, C15orf39 | CSDK, ULK3, MPI, FAM219B, C15orf39 | CSDK, ULK3, MPI, FAM219B, C15orf39 | CSDK, ULK3, MPI, FAM219B, C15orf39 | 67 | CSDK, ULK3, MPI, FAM219B, C15orf39 |

(continued)
Some were common to all cell types, such as loci for $UBE2L3$ and $UHRF1BP1$, whereas others were more cell type specific, including loci for $BLK$ in B cells and $JAZF1$ in T cells. In general, directionality was consistent, although not in all cases: for example, $ABHD6$ showed reduced expression in monocytes and elevated expression in lymphocytes.

We note that some caution must be used when inferring causality, as RTC scores have a uniform distribution and setting an RTC score threshold of 0.9, for example, sets the type I error rate to be 0.1. Furthermore, some low RTC scores were found in genes (for example, $UBE2L3$) where the associated allele resides in a region with strong LD and the haplotype bearing the associated allele shows robust evidence of functional effects on gene expression. We suggest that the gene expression analyses provide some support for likely causal genes, but we note that proof of true causality through altered gene expression will only be achieved by additional experimentation.

We then integrated the results of these eQTL analyses and the coding variant analysis with an in silico survey of mouse phenotype data from knockouts targeting genes within the associated SLE loci (Table 2). At some loci, these lines of evidence point to one likely causal gene: examples include $IFIH1$, $LYST$, $WDFY4$ and $BANK1$. In other instances, we found evidence that supports the role of multiple genes as candidates at a given locus; for example, $ABHD6$ (encoding an enzyme involved in the endocannabinoid pathway) and $PAK$ (encoding a lymphocyte protein kinase) both exhibit correlation of their expression with the associated SNP. Similarly, $TCF7$ (encoding a T cell transcription factor) implicated by the rs7726414 association, has been associated with type 1 diabetes; however, we show that $SKP1$ (which encodes a protein involved in the regulation of ubiquitination) within the same LD block exhibits a strong cis-eQTL effect in monocytes and NK cells. rs9652601 resides within $CLEC16A$, a gene previously reported in association studies for other autoimmune diseases; we present evidence suggesting that $SOCS1$ (encoding suppressor of cytokine signaling 1) is a causal gene at this locus in SLE rather than $CLEC16A$. Our analyses have the advantage of including cis-eQTLs identified on the basis of data from ex vivo cells rather than cell lines alone. Nevertheless, we acknowledge the restricted range and activation states of the immune cell types available for

---

**Table 2 Candidate genes at SLE-associated loci (continued)**

| Associated SNP | Chr. | Gene within 200 kb of SNP | Gene within same LD block as SNP | Immune phenotype in mouse model | Coding variant with SNP | Functional and/or fine-mapping studies | Likely causal gene |
|----------------|------|--------------------------|---------------------------------|---------------------------------|------------------------|----------------------------------------|-------------------|
| rs9652601      | 16   | $CITTA$, DEX1, CLEC16A, RMI2, SOCS1, TNP2, PRM3, PRM2 | $CLEC16A$ | $CITTA$, SOCS1 | SOCS1, RMI2 | $CITTA$, SOCS1 | $CITTA$, SOCS1 |
| rs34572943     | 16   | ZNF668, ZNF646, PRSS53, VKORC1, BCKD KATB PRSS8, PRSS36, FUS, PYSRD, C16orf98, TRIM72, PYSRD1, ITGAM, ITGAX, ITGAD, COX6A2, ZNF943, ARM95 | $ITGAM$ | $ITGAM$, ITGAX, ITGAD, PYCARD | $ITGAM$, ITGAD, PYCARD | $ITGAM$, ITGAD, PYCARD | $ITGAM$, ITGAD, PYCARD |
| rs11644034     | 16   | C16orf74, EMCR, COX41, IRF8, ALOX15, PELP1, ARRB2, MED11, CLX16 ZMYND15, TM4SF5, VM01, GLTP02, PSMB6, PLD2, MINK1, CHRNA7, C17orf107, GP1BA, SLC25A11, RNFL67, PPFN1, EN03, SPAG7, CAMT2A, INCA1, KIF1C | Intergenic | PLD2 | $RNF167$ | $RNF167$ | $RNF167$ |
| rs2941509      | 17   | NEUROD2, PPP1R18B, STAR03, TCA5, PNMT, PAG3, ERBB2, MIEN1, GRB7, IKF3, ZPB2, GSDMB, ORMEL3, LRRCC3, GSDMA | $ERBB2$, HER-2, C17orf37, GRB7, IKF3, ZNF12, ZPB2, GSDMB | IKZF3 | IKZF3 | IKZF3 | IKZF3 |
| rs2304256      | 19   | DNMT1, S1PR2, MRPL4, ICAM1, ICAM4, ICAM5, ZGLP1, FD1X1, RAVER1, ICAM3, TYK2, CDC37, PD04A, KEAP1, S1PR5, ATG4D, KRI1 | $TYK2$ | $TYK2$, ICAM3 | $TYK2$, ICAM3 | $TYK2$, ICAM3 | $TYK2$, ICAM3 |
| rs7444         | 22   | HIC2, RIMBP3C, UBE2L3, YD1C, CDCC116, SDF2L1, SN1L2, YP1L1, CFPK1 | $UBE2L3$, YD1C | MAPK1 | UBE2L3 | UBE2L3 | UBE2L3 |
| rs887369       | X    | CXorf21, G6 | $CXorf21$ | $ARHGA4$, NAA10, RENBP, HCFIC1, TMEM187, IRAK1, MECP2, OPN1LW, TEX28P2, OPN1MW, TEX28P2, TEX28, KTL1 |
| rs1734787      | X    | LICAM, LCA10, AVPR2, ARHGAP4, NAA10, RENBP, HCFIC1, TMEM187, IRAK1, MECP2, OPN1LW, TEX28P2, OPN1MW, TEX28, KTL1 |

*An LD block is defined by SNPs showing a correlation ($r^2$) of 0.75 with the associated SNP. The immune phenotype designation is taken from http://www.informatics.jax.org/phenotypes.shtml for genes within 200 kb up- or downstream of the associated SNP. The gene(s) implicated at each locus as potentially causal. The MHC region is not included because of extended LD and gene density at the locus.*
The ten previously unmapped SLE-associated loci (shown in bold type in Table 1 and Supplementary Table 3a) encompass genes of diverse function. Those of note include IKZF2 (Helios), which represents the third member of the Ikaros transcription factor family to be associated with SLE (in addition to IKZF1 and IKZF3). The association signal in the PLD2 gene (encoding phospholipase D2) corresponds to a missense variant (p.Arg172Cys), which may alter the function of this enzyme that has a role in leukocyte migration and apoptosis. The importance of interleukin (IL)-12, a cytokine that has a critical role in the generation of \( \gamma \)-interferon from T helper type 1 (TH1) and NK cells, is highlighted by the association with IL12A (Table 1) and by the suggestive associations at IL12B and the locus encoding the IL-12 receptor, IL12RB2 (Supplementary Table 2).

In view of the sexual dimorphism of SLE, the newly identified X-chromosome association at rs887369 is of note. We suggest that the gene CXorf21 is likely to be etiological. Although the function of this gene is unknown, it is among a limited set of genes that largely escape X inactivation\(^{29}\). Sex-chromosome dosage has been implicated in the genetic risk of SLE\(^{30}\). We observed an elevated prevalence of Klüver-Barrera’s syndrome\(^{31}\) in male cases in our GWAS in comparison with the general population (Online Methods), strengthening the sex-chromosome dosage hypothesis. The only other gene close to rs887369 (Table 2) is GK (encoding glycerol kinase), which does not escape X inactivation, supporting CXorf21 as a candidate gene.

Five other genes (TNIP1, IKZF1, ETS1, WDFY4 and ARID5B) that we mapped are new in European SLE but had been previously shown to be associated with SLE in Chinese subjects\(^{5,6}\). SLE is more prevalent in non-European populations; our data suggest that locus heterogeneity among common genetic variants is unlikely to explain this differential prevalence.

We present all of our principal findings in Figure 2. This figure includes ten missense coding variants that likely contribute to SLE risk; these occur largely in genes encoding kinases and other enzymes. We also note that 16 of the genes shown are transcription factors, an enrichment above the nine expected (\( P = 2.3 \times 10^{-3}, \chi^2 \) test). We studied the distribution of the expression for these transcription factors in the ex vivo immune cell types examined for eQTLs; we found no evidence of skewed expression in any cell type. Our results suggest that an important facet in the future exploration of SLE pathogenesis will be detailed scrutiny of trans eQTLs and regulatory expression networks in multiple immune cells.

The research was funded and supported by the National Institute for Health Research UK Special Strategic Award, the Arthritis Research UK Special Strategic Award, the Lupus Association, Lupus Research, the Robert Welch Foundation, the George Koukis Foundation and the Arthritis Research UK Special Strategic Award. J.E.W. was funded by the George Koukis Foundation and an Arthritis Research UK Special Strategic Award. T.J.V., J.D.R. and M.E.A.-R. were awarded funding to carry out genotyping and analysis from the George Koukis Foundation and an Arthritis Research UK Special Strategic Award (2013). M.E.A.-R. received grants from the Instituto de Salud Carlos III (PS09/00129), co-financed by the FEDER funds of the European Union, the Consejería de Salud de Andalucía (P10012) and the Swedish Research Council of Medicine, and from the European Science Foundation to the BIoLUPUS network. J.B. was funded by the George Koukis Foundation and the Arthritis Research UK Special Strategic Award. J.E.W. was funded by the Canadian Institutes of Health Research (94825). C.L.P. was funded by a Wellcome Trust grant (085492). P.T. is employed by the Biomedical Research Centre. L.C. was funded by the China Scholarship Council (201406830127). The research was funded and supported by the National Institute for Health
Research (NIHR) Biomedical Research Centre based at Guy’s and St. Thomas’ National Health Service (NHS) Foundation Trust and King’s College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the UK Department of Health.

HRS genetic data were obtained from the database of Genotypes and Phenotypes (dbGaP) under accession phs000187.v1.p1; the study is sponsored by the National Institute on Aging (grants U01AG097940, R2CA036495 and RC4AG093029) and was conducted by the University of California, San Francisco. The melanoma study data were obtained from dbGaP under accession phs000107.v1.p1. Research support to collect data and develop an application to support this project was provided by US National Institutes of Health (NIH) grants 3P5CA093459, 5P5CA097007, SR01ES017140 and SR01CA133996. Funding support for the Genes and Blood Clotting Study was provided through the US NIH/National Heart, Lung, and Blood Institute (NHLBI) (R37HL039693). The Genes and Blood Clotting Study is one of the Phase 3 studies that are part of the Gene-Environment Association Studies (GENEVA) under the Genes, Environment and Health Initiative (GEI). Assistance with genotype cleaning was provided by the GENEVA Coordinating Center (U01HG004446, US NIH). Funding support for DNA extraction and genotyping, which were performed at the Broad Institute, was provided by the US NIH/NHLBI (R37HL039693). Additional support was provided by the Howard Hughes Medical Institute. The data sets used for the analyses described in this manuscript were obtained from dbGaP under accession phs000540.v2.p1.

CGEMS prostate cancer study data were obtained from dbGaP under accession phs000207.v1.p1. We thank Genentech, Inc., for providing the genotype data from their GWAS. We thank V. Anand and S. Ragan for their help in coordinating data collection. We thank T. Axelsson, B. Fürnrohr, S. Ragan and J. Kelly for their help with the replication study. A large number of people contributed samples or clinical data to the GWAS. The following samples were obtained via the BIOUPUS network coordinated by M.E.A.-R., Belgium: B. Lawers and F. Houssiau (Université Catholique de Louvain). Denmark: S. Jacobsen (University of Copenhagen), P. Junker and H. Lastrup (Odense University Hospital). Germany: T. Witte (Medizinische Hochschule Hannover). Greece: H. Moutsopoulos and E.K. Kapsogeorgou (National University of Athens). Hungary: E. Endrely and L. Kovacs (Albert Szent-Györgyi Medical University). Iceland: K. Steinsson (Landspitali National University Hospital). Italy: A. Doria (University of Padova), P.L. Meroni (IRCCS Istituto Auxologico Italiano, Milan), R. Scorza (University of Milan), M. Zummer (University of Siena). Japan: N. Costedoat-Chalumeau (AP-HP, Hôpital Cochin, Centre de Référence Maladies Auto-immunes et Systémiques Rares, France; Université Paris Descartes–Sorbonne Université Paris Cité, France). The Netherlands: M. Bijl and C. Kallenberg (University of Groningen). Portugal: C. Vasconcelos (Hospital Santo António, Porto), B. Martins Silva (University of Porto). Spain: J. Martin and E. Martin Rodriguez (Instituto de Parastasología y Biomedicina Lopez Neyra), A. Suarez (Hospital Universitario Central de Asturias), I. Rua Figueroa (Hospital Dr. Negrín, Gran Canaria), G. Pons-Estel (Hospital Clinic, Barcelona). From the GENEVS collaboration: Argentina: B. Pons-Estel (Hospital Provincial de Rosario). Other contributors: Canada: P. Fortin, J. Wither, D. Gladman and M. Urowitz (Toronto Western Hospital, University Health Network), A. Clarke, S. Bernatsky, C. Pineau and J. Rauch (McGill University), T. Hudson (Ontario Institute for Cancer Research), J. Pope (University of Western Ontario), C. Peschken and C. Hitchon (University of Manitoba). Denmark: M.-A. Lee-Kirsch (Technische Universität München), E. Rich and J.-L. Senécal (Centre Hospitalier de l’Université de Montréal), M. Zummer (Universität Freiburg). The following samples were obtained via the BIOLUPUS network coordinated by M.-A. Lee-Kirsch (Technische Universität München), E. Rich and J.-L. Senécal (Centre Hospitalier de l’Université de Montréal), M. Zummer (Universität Freiburg). The following samples were obtained via the BIOLUPUS network coordinated by M.-A. Lee-Kirsch (Technische Universität München), E. Rich and J.-L. Senécal (Centre Hospitalier de l’Université de Montréal), M. Zummer (Universität Freiburg).

The authors declare no competing financial interests.

D.L.M., J.B., D.S.C.G., J.D.R. and T.J.V. wrote the manuscript. All authors have read and contributed to the manuscript.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Deanten, D. et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis Rheum. 35, 311–318 (1992).
2. Alarcón-Segovia, D. et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis Rheum. 52, 1138–1147 (2005).
3. Harty, J.B. et al. Genome-wide association study in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat. Genet. 40, 204–210 (2008).
4. Horn, G. et al. Association of systemic lupus erythematosus with C8orf13-3BLK and ITGAM-1TGNX. N. Engl. J. Med. 358, 909–908 (2009).
5. Yang, W. et al. Genome-wide association study in Asian populations identifies variants in ETSL and WDF4 associated with systemic lupus erythematosus. PLoS Genet. 6, e1000841 (2010).
6. 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).
7. Graham, R.R. et al. Genetic variants near TNFAP3 on 6q23 are associated with systemic lupus erythematosus. Nat. Genet. 40, 1059–1061 (2008).
8. Okada, Y. et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus. PLoS Genet. 8, e1002545 (2012).
9. Gaveta, V. et al. A large-scale replication study identifies TNIP1, PARD1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat. Genet. 41, 1228–1233 (2009).
10. Cunningham Graham, D.S. et al. Association of NCF2, IKZF1, IRFB, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet. 7, e1002341 (2011).
11. Hofhorn, T.N. & Daly, M.J. Genome-wide association studies for common diseases and complex traits. Nat. Rev. Genet. 6, 95–108 (2005).
12. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. 38, 904–909 (2006).
13. Deo, R. et al. Roeder, K. Genetic control for association studies. Biometrics 55, 997–1004 (1999).
14. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. Recent approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11, 459–463 (2010).
15. de Bakker, P.I.W. et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 17, R122–R128 (2008).
16. Lawrence, J.S., Martins, C.L. & Drake, G.L. A family survey of lupus erythematosus. Am. J. Hum. Genet. 91, 913–921 (1987).
17. So, H.-C., Gui, A.H.S., Cherry, S.S. & Sham, P.C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. Genet. Epidemiol. 35, 310–317 (2011).
18. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature 497, 65–66 (2012).
19. Knight, J., Barnes, M.R., Breen, G. & Weale, M.E. Using functional annotation for the empirical determination of Bayes Factors for genome-wide association study analysis. PLoS ONE 6, e14808 (2011).
20. Morris, D.L. et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. Am. J. Hum. Genet. 91, 778–793 (2012).
21. Dillthey, A.T., Moutsianas, L., Leslie, S. & McLeave, G. HLA*IMP—an integrated framework for imputing classical HLA alleles from SNP genotype data. Bioinformatics 27, 968–972 (2011).
22. Fairfax, B.P. et al. Innate immune activity conditions the effect of regulatory variants on monocyte gene expression. Science 343, 124694 (2014).
23. Fairfax, B.P. et al. Genetics of gene expression in primary immune cells identifies cell-type-specific master regulators and roles of HLA alleles. Nat. Genet. 44, 502–510 (2012).
42. Alexander, T. et al. Polymorphism in the NCF2 gene, C883A, is associated with type 1 diabetes. Science 344, 519–523 (2014).
43. Manku, H. et al. The Chediak-Higashi protein interacts with SNARE complex proteins to bring unique insights to the structure and function of NADPH oxidase. PLoS Genet. 9, e1003554 (2013).
44. Wright, W. et al. Systematic identification of trans-eQTLs as putative drivers of known disease associations. Nat. Genet. 45, 1238–1243 (2013).
45. Kotayan, L.C. et al. The IRS5-TNP3 association with systemic lupus erythematosus (SLE) has two components that other autoimmune disorders variably share. Hum. Mol. Genet. 24, 582–596 (2015).
46. Guthridge, J.M. et al. Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. Am. J. Hum. Genet. 94, 586–598 (2014).
47. Zhao, H. et al. An intronic variant associated with systemic lupus erythematosus changes the binding affinity of Yin yang 1 to downregulate WDFY4. Genes Immun. 13, 536–542 (2012).
48. Hering, M. et al. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature 467, 460–464 (2010).
49. Grispin, J.C. et al. Expression of CD44 variant isoforms CD44v3 and CD44v6 is increased on T cells from patients with systemic lupus erythematosus and is correlated with disease activity. Arthritis Rheum. 62, 1431–1437 (2010).
50. Zhang, J. et al. Epistatic interaction between genetic variants in susceptibility gene ET31 correlates with IL-17 levels in SLE patients. Ann. Hum. Genet. 77, 344–350 (2013).
51. Morris, E.E. et al. A GA microsatellite in the FLII promoter modulates expression and is associated with systemic lupus erythematosus patients without nephritis. Arthritis Res. Ther. 12, R212 (2010).
52. Mori, T. et al. LinkSH2bp3 controls the production and function of dendritic cells and regulates the induction of IFN-γ-producing T cells. J. Immunol. 193, 1728–1736 (2014).
53. Manjarrez-Orduño, N. et al. CSA regulatory polymorphism is associated with systemic lupus erythematosus and influences B-cell signaling and activation. Nat. Genet. 44, 1227–1230 (2012).
54. Bronson, P.G. et al. The rs4774 C/T allele is associated with risk of systemic lupus erythematosus. Genes Immun. 12, 667–671 (2011).
55. Fujimoto, M. et al. Inadequate induction of suppressor of cytokine signaling-1 causes systemic autoimmune diseases. Int. Immunol. 16, 303–314 (2004).
56. Rhodes, B. et al. The rs1143679 (R77H) lupus associated variant of IRAK1 complements receptor 3 mediated functions in human monocytes. Ann. Rheum. Dis. 71, 2028–2034 (2012).
57. Chrabot, B.S. et al. Genetic variation near IL23RA is associated with serologic and cytokine profiles in systemic lupus erythematosus and multiple sclerosis. Genes Immun. 14, 471–478 (2013).
58. Sun, J. et al. A natural mutation in the Tyk2 pseudokinase domain underlies altered susceptibility of B10.QJ mice to infection and autoimmunity. Proc. Natl. Acad. Sci. USA 100, 11594–11599 (2003).
59. Kaufman, K.M. Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. Ann. Rheum. Dis. 72, 437–444 (2013).
ONLINE METHODS

**Genome-wide association study.** We genotyped 4,946 SLE cases and 1,286 healthy controls using the Illumina HumanOmni1-Quad BeadChip (1,140,419 markers) (for details on genotyping quality control checks and the origins of the genotyped subjects, see the Supplementary Note and Supplementary Tables 7 and 8). The genotyped controls were mostly from southern Europe and Canada, matching our Spanish, Italian and Turkish cases with controls from the same countries. We also used data for 5,727 previously genotyped controls taken from the University of Michigan HRS. These subjects were genotyped using the Illumina HumanOmni2.5 BeadChip (2,443,179 markers).

The clinical features of our GWAS cohort were documented on the basis of standard American College of Rheumatology (ACR) classification criteria. The experiment was designed to avoid batch effects to the greatest extent possible. All DNA samples were sent to the laboratory at King’s College London, where the integrity of the DNA was checked. The GWAS samples were then genotyped in a single laboratory. All data analysis was carried out in the laboratory at King’s College London.

Genotyping for the GWAS was carried out using 82 plates, processed in 13 batches. Duplicate samples taken from HapMap Phase 3 were added to each plate to check genotyping quality. Case-control status and country of recruitment were randomized across plates as far as possible, to avoid artifactual differences to check genotyping quality. Case-control status and country of recruitment were randomized across plates as far as possible, to avoid artifactual differences in genotyping between plates having an effect on the association statistics. Our final data set comprised genotyping of 644,674 SNPs for 4,036 SLE cases and 6,895 controls (1,260 controls of mainly southern European ancestry and 5,699 controls from the HRS cohort). Power calculations are shown in Supplementary Figure 5.

**Study by Hom et al.** We analyzed data from a previous GWAS of SLE (Hom et al.), which comprised 1,165 cases after our quality control analysis (Supplementary Note). We used a further 2,107 previously genotyped controls from the US National Institutes of Health CGEMS study, which were genotyped using the Illumina HumanHap550 chip. Owing to the lower density of genotyping, in some cases, data imputed to the density of the 1000 Genomes Project study was used in the analysis of the study by Hom et al. and the subsequent meta-analysis. Imputed data are indicated in the tables.

**Replication study.** A cohort of 2,310 cases not included in any previous genetic study of SLE was genotyped using a custom array. The largest group of samples was from the UK, followed by cohorts from France, the United States, Germany and Canada.

The control data for the replication study comprised 3,672 subjects from the HRS cohort (independent of those used in the GWAS), 3,102 subjects from a study of melanoma and 1,202 subjects from a study of blood clotting. These control subjects were genotyped using the Illumina HumanOmni2.5 BeadChip. After quality control procedures (Supplementary Note), the final control data set comprised 6,925 individuals: 3,668 from the HRS cohort, 2,889 from the melanoma study and 368 from the blood clotting study. The final case data set consisted of 2,018 samples.

In some cases, SNPs identified by our GWAS as genome-wide significant were not present in the replication control data (owing to the absence of genotyping in one of the three control sets after quality control), and genotypes for these SNPs were therefore imputed. We indicate these SNPs in our results tables.

**Ethical approval.** The UK subjects with SLE in the study were recruited with the study having obtained ethical approval from the London Ethics Committee (MREC/98/2/06 and MREC/98/2/0). Individuals were invited into the study and given information sheets as well as verbal explanations of what the research entailed. For individuals willing to participate, informed written consent was obtained. The recruitment protocols in continental Europe and Canada were subject to local review and ethical approval. Copies of the relevant supporting documentation were sent to the investigators at King's College London at the commencement of the study.

**Quality control.** Initial quality control analysis of the genotype data was carried out in accordance with Illumina's Technical Note on Infinium Genotyping Data Analysis (Supplementary Table 7). In silico quality control checks were carried out for the following criteria: (i) individual missingness (3% threshold), (ii) SNP missingness (3% threshold), (iii) identity by descent (IBD; 0.125 threshold), (iv) population structure, (v) MAF (0.002 threshold), (vi) autosomal heterozygosity, (vii) X-chromosome heterozygosity, (viii) Y-chromosome calling and homozygosity, and (ix) Hardy-Weinberg equilibrium (control data only). IBD analysis included checks both within and across cohorts; no subject in the main GWAS or the study by Hom et al. is related to any other subject in either cohort. We calculated principal components for the GWAS data using the EIGENSTRAT algorithm (Supplementary Fig. 6) and derived the empirical genomic inflation factor (for these data. As noted by Price et al., the definition of genomic control means that \( \hat{\lambda}_{GC} \) is proportional to sample size. We therefore report \( \hat{\lambda}_{GC} \), the inflation factor for an equivalent study of 1,000 cases and 1,000 controls, in the main text as well as \( \hat{\lambda}_{GC} \). For the replication cohort, population structure was estimated using 46 ancestry-informative markers (after quality control measures on these SNPs).

As described in the Supplementary Note, we merged these data with HapMap data to help identify non-European samples. Principal components were calculated using the EIGENSTRAT algorithm (Supplementary Figs. 7–12). We removed from the analysis 120 subjects who clustered with the non-European HapMap populations.

**Klinefelter’s syndrome.** During quality control analysis, we identified subjects in our GWAS cohort with abnormal karyotypes consistent with Klinefelter’s syndrome (47,XXY). Three of the 365 male cases in our main GWAS have clinical and genetic data that confirm their status as individuals with Klinefelter’s syndrome (Supplementary Figs. 13–16 and Supplementary Note). Given that the prevalence of Klinefelter’s syndrome in the general population is estimated to be 0.1–0.2% (ref. 31), this estimate suggests an approximately four- to eightfold increase in prevalence in comparison with 46,XY males, consistent with Klinefelter’s males and 46,XX females having a similar risk of developing SLE.

**Association analysis.** All case-control analysis was carried out using the SNPTTEST algorithm; we used a standard threshold of \( P = 5 \times 10^{-8} \) for reporting genome-wide significance throughout. The inverse variance method was used for meta-analysis. All markers were fully genotyped in the main GWAS (no imputation was carried out). The imputation performed for the Hom et al. and replication studies and the fine-mapping imputation are described below.

For all SNPs at which we report a new association with SLE, we compared the allele frequencies in the main GWAS controls with those in publically available control cohorts (1000 Genomes Project European samples, Wellcome Trust Case Control Consortium (WTCCC) genomes, TwinsUK samples, HapMap CEU (European-ancestry) population data and sample genotypes from Knight laboratory expression data). We tested for a statistically significant (\( \alpha = 0.01 \)) difference in allele frequency between our GWAS controls and the public controls, using a 1-degree-of-freedom \( \chi^2 \) test of allele frequency. One SNP failed this test (rs1439112, MGATS5) and was removed from further analysis. In three further cases, the difference in allele frequency strengthened our observed association. These data are presented in Supplementary Table 9.

**Annotation of results.** The gene names listed in results tables were obtained by overlaying the GWAS results onto the UCSC Genome Browser. We adopted a threshold for inclusion based on LD: for each SNP, we noted the set of markers with \( r^2 \geq 0.75 \) with respect to the SNP of interest (Table 2).

**Post-hoc quality control.** Checks carried out following case-control analysis included examination of plots of raw genotype intensity data; this was of particular relevance given the increase in the numbers of relatively rare variants due to the higher density of genotyping (as with imputation, genotype calling is by definition more difficult for rarer variants). We checked that the intensity plots showed clusters of genotypes (for homozygotes or heterozygotes) that were compact and well discriminated. This check was also carried out with stratification by quality control group. Plots of intensity were examined for each associated SNP and for all of the SNPs in the replication study.
Replication study chip design. We selected SNPs for the replication study on the basis of the results of the meta-analysis of the two GWAS. At loci with no known association in SLE, we adopted an inclusion threshold of \( P = 2.5 \times 10^{-5} \), whereas, for loci with previously reported associations, the threshold was set at \( P = 1 \times 10^{-4} \). This approach follows the methodology used in Box 1 of the WTCCC study of seven common diseases, in which SNPs were declared to be associated if the posterior odds of association were greater than 10. In that study, the assumption was made that ten detectable genes were present, so the prior odds of a true association would be on the order of 100,000:1, assuming 1,000,000 independent regions in the genome. On the basis of the autoimmune genetics literature, we have made the assumption that there are likely to be as many as 500 genes associated with SLE. We have required the posterior odds in favor of a SNP being associated to be >1 (as opposed to >10, which would be advisable if declaring an association rather than choosing SNPs for replication). This corresponds to a \( P \)-value threshold of \( 2.5 \times 10^{-5} \). For SNPs at loci with previously published SLE associations, we reduced our threshold for inclusion in the replication study to \( P = 1 \times 10^{-4} \). This is because we believe a priori that these SNPs are more likely to be at susceptibility loci than with those with no evidence of association, increasing the prior odds by at least a factor of 4.

1000 Genomes Project imputation. For imputation, data from both the main GWAS and the Hom et al. study were prephased using the SHAPEIT algorithm20 and then imputed to the density of the 1000 Genomes Project data using IMPUTE21,22,23. Only markers with an IMPUTE info score >0.7 were used in analysis. For SNPs identified in our GWAS as genome-wide significant at which data were absent in the replication study controls, we imputed over a 2-Mb genomic region centered on the SNP of interest.

We used 1000 Genomes Project data both to fine map loci and to determine whether multiple signals were present at each locus. For this analysis, we carried out a meta-analysis of 1000 Genomes Project–imputed data for the main GWAS and the study by Hom et al. Association testing was performed on the 1000 Genomes Project data within a 2-Mb window centered on the reported SNP. For the MHC region, we included the complete 8-Mb region (26–34 Mb on chromosome 6) in our analysis. To scan for further independent signals, association tests were performed including the genotype data for the most highly associated SNP as a covariate. If secondary signals were found to be associated by this analysis (with a \( P \)-value threshold of \( 5 \times 10^{-5} \)) and the odds ratios were consistent across the single-marker and conditional analyses, the secondary signals were reported as independent associations.

To address the problem that the most strongly associated variant (lead SNP: assuming the marker with the lowest \( P \) value) is not necessarily the best candidate for the true causal variant, we considered markers from the strongest association down to a defined cutoff of association. The cutoff was defined as a Bayes factor against the most associated SNP equal to 0.34. This was derived from assuming a prior odds of causality for a nonsynonymous SNP equal to 3, taken from an empirical analysis of GWAS annotation19,24. Any SNPs above this Bayes factor cutoff that were missense variants were declared to be more likely candidates than the most significantly associated SNP: assuming the prior odds of a missense SNP (being causal) against a non-missense SNP to be equal to 3, any missense SNP with a Bayes factor >0.34 will have posterior odds >1 and will therefore have a higher posterior probability than the most significantly associated marker (if this marker is not a missense variant). Therefore, we searched for functional variants within a set of markers where inclusion in the set required a maximum Bayes factor >0.34 between the marker and the most significantly associated SNP in the 1000 Genomes Project–imputed data. We considered any marker that had a Bayes factor >0.34 with respect to the most significantly associated marker and noted whether any had functional effects. We calculated an approximate Bayes factor following the method of Wakefield25, using a prior distribution on effect sizes (odds ratios) that was proportional to MAF (as rare variants are believed to have large effects, whereas common variants are believed to exert small effects). The Bayes factor threshold implies that we believe that associations with functional variants, such as missense variants, are three times greater (for effects). The Bayes factor threshold implies that we believe that associations with functional variants, such as missense variants, are three times greater (for effects). The Bayes factor threshold implies that we believe that associations with functional variants, such as missense variants, are three times greater (for effects). We then calculated posterior model probabilities following the method of Maller et al.77, but with prior odds of 3 between missense SNPs and non-missense SNPs: Maller et al. use a uniform prior on all model probabilities (all SNPs are considered to have equal weights a priori, and the prior odds are therefore 1). We present these results in Supplementary Table 4 where we also, separately, display SNPs with a Bayes factor >0.1 (as a strict threshold of 0.34 does not reflect the uncertainty in the prior odds of causality and Bayes factor estimates). We also calculated the Bayes factors between the SNPs presented in Table 1 and the SNPs listed in Supplementary Table 3a and declared that the marker for association had changed if the Bayes factor was greater than 10 (equal to ‘strong’ evidence on the Jeffreys’ scale78). These SNPs are annotated in Supplementary Table 3a.

Analysis of MHC and HLA alleles. We included imputed HLA alleles in analysis of the MHC region, allowing us to determine the most likely model of association within this region. HLA imputation was performed using HLA*IMP (v2)21 with genotyped SNP data. To determine the best model for association when considering the HLA alleles alone, we ran forward stepwise regression. We then tested the five SNPs listed in Supplementary Table 6a-c for association, conditional on the HLA alleles. To test whether each of the five SNPs was independent of the HLA alleles (rather than just the alleles in the best HLA model), we carried out a test conditional on all alleles (where the HLA alleles were used as covariates) in each HLA gene and for all HLA alleles over all genes. We used a significance threshold at each stage of the stepwise regression of \( P = 5 \times 10^{-5} \), which takes into account a Bonferroni adjustment for 204 tests (for 199 HLA alleles and five SNPs), with a family-wise type I error rate of 0.01.

Analysis of gene expression data. Gene expression data were obtained from three sources. First, we obtained data from Fairfax et al.22,23 and unpublished data for NK cells, naive monocytes, monocytes stimulated by LPS (collected after 2 h and 24 h) or interferon, and B cells (B.P.F. and J.C.K., unpublished data). Second, we interrogated the Genevar database for LCL eQTL results, taking results from the MuTHER resource79. The CD4 (CD4+ T cells) and CD14 (CD14+ or CD16+ monocytes) data were obtained from a previous study of gene expression in immune-related cells80. An adjustment was made for multiple testing using a Bonferroni correction by counting the number of tests across all loci for genes within a 2-Mb window centered on the SLE-associated SNP. With a family-wise test size of 0.01, the \( P \)-value threshold was \( 1.41 \times 10^{-5} \).

To test whether observed associations between SNPs and the expression levels of cis-acting genes were purely due to chance, we calculated the RTC score24 for all SNP–gene eQTL results displayed in the heat map (Fig. 1). This score is generated by testing the null hypothesis that the GWAS-associated SNP and the best eQTL (within a recombination hotspot) are tagging two separate effects and the observed eQTL is purely due to the LD between the GWAS-associated SNP and the ‘true’ eQTL SNP. For our data, we were interested in the distribution of RTC scores, given that the eQTL results were generated in multiple cell types. Not all eQTLs were consistently present across all the cell types. We therefore plotted the RTC scores against the \(-\log_{10} (P \text{ values})\) supporting each cis eQTL in all cell types (Supplementary Fig. 3a,b). Three genes were outliers: ITGAM in two cell types and UBE2L3 and PLD2 in CD4+ T cells. However, we have strong a priori evidence of a true causal effect on expression by polymorphisms around UBE2L3 (ref. 26). For ITGAM, we note the low RTC scores in Figure 1, which includes all eQTL data for ITGAM given that the results are convincing for the eQTL in LPS-stimulated monocytes (\( P = 2.67 \times 10^{-19} \) and RTC = 0.85). We have removed the declaration of an eQTL for PLD2. Supplementary Figure 4 displays a heat map for the data using a \( t \) statistic.

69. Freedman, M.L. et al. Assessing the impact of population stratification on genetic association studies. Nat. Genet. 36, 388–393 (2004).
70. Reich, D.E. & Goldstein, D.B. Detecting association in a case-control study while correcting for population stratification. Genet. Epidemiol. 20, 4–16 (2001).
71. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. Nat. Genet. 39, 906–913 (2007).
72. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11, 499–511 (2010).
73. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447, 661–678 (2007).
74. Delaneau, O., Howie, B., Cox, A.J., Zagury, J.-F. & Marchini, J. Haplotype estimation using sequencing reads. Am. J. Hum. Genet. 93, 687–696 (2013).
75. Gagliano, S.A., Barnes, M.R., Weale, M. & Knight, J. A Bayesian method to incorporate hundreds of functional characteristics with association evidence to improve variant prioritization. PLoS ONE 9, e98122 (2014).
76. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. Genet. Epidemiol. 33, 79–86 (2009).
77. Maller, J.B. et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. 44, 1294–1301 (2012).
78. Jeffreys, H. Theory of Probability 3rd edn. (Oxford Univ. Press, 1961).
79. Grundberg, E. et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nat. Genet. 44, 1084–1089 (2012).