High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry

Celi Sun1,2, Julio E Molineros1,2, Loren L Looger2,2, Xu-jie Zhou1,2, Kwangwoo Kim1,4,2, Yukinori Okada5,6, Jinyang Ma7, Yuan-yuan Qi3, Xana Kim-Howard1, Prasenjeet Motghare1, Krishna Bhattachar1,1, Adam Adler1, So-Young Bang4, Hye-Soon Lee4, Tae-Hwan Kim4, Young Mo Kang8, Chang-Hee Suh9, Won Tae Chung10, Yong-Beom Park11, Jung-Yoon Choe12, Seung Cheol Shim13, Yuta Kochi14, Akari Suzuki14, Michiaki Kubo15, Takayuki Sumida16, Kazuhiko Yamamoto14,17, Shin-Seok Lee18, Young Jin Kim19, Bok-Ghee Han19, Mikhail Dzomorov20, Kenneth M Kaufman21, Jonathan D Wren1, John B Harley21, Nan Shen7,22, Ket Heng Chua23, Hong Zhang3, Sang-Cheol Bae4 & Swapan K Nath1

Systemic lupus erythematosus (SLE) has a strong but incompletely understood genetic architecture. We conducted an association study with replication in 4,478 SLE cases and 12,656 controls from six East Asian cohorts to identify new SLE susceptibility loci and better localize known loci. We identified ten new loci and confirmed 20 known loci with genome-wide significance. Among the new loci, the most significant locus was GTF2IRD1-GTF2I at 7q11.23 (rs73366469, \( P_{\text{meta}} = 3.75 \times 10^{-11} \), odds ratio (OR) = 2.38), followed by DEG6, IL12B, TERT, CD226, PCNLX3, RASGRP1, SYNGR1 and SIGLEC6. We identified the most likely functional variants at each locus by analyzing epigenetic marks and gene expression data. Ten candidate variants are known to alter gene expression in cis or in trans. Enrichment analysis highlights the importance of these loci in B cell and T cell biology. The new loci, together with previously known loci, increase the explained heritability of SLE to 24%. The new loci share functional and ontological characteristics with previously reported loci and are possible drug targets for SLE therapeutics.

SLE is a debilitating autoimmune disease characterized by pathogenic autoantibody production that can affect virtually any organ. Asians have higher SLE incidence, more severe disease manifestations and greater risk of organ damage (for example, lupus nephritis)1,2 than European-ancestry populations. SLE has a strong genetic component, constituting a sibling risk ratio3 (\( \lambda_s \)) of ~30, with ~40 susceptibility loci reported through candidate gene studies and genome-wide association studies (GWAS)4-6. However, only 8–15% of disease heritability7,8 is accounted for, leaving many contributing loci unidentified. Because multiple susceptibility loci are shared across autoimmune diseases and studying high-risk populations can facilitate the identification of new risk loci, we performed high-density association analysis in East Asians.

Our study was conducted in three stages (Fig. 1 and Online Methods). First, after quality control, we performed association analysis based on the Immunochip7 in 2,485 cases and 3,947 controls from Korean (KR), Han Chinese (HC) and Malaysian Chinese (MC) populations and identified 578 associated regions (\( P < 5 \times 10^{-5} \)) (Supplementary Tables 1–3). To increase statistical power, we included 3,669 out-of-study Korean controls (Supplementary Fig. 1 and Supplementary Table 1) and conducted imputation-based association analysis (Online Methods). Second, we followed up 16 newly associated loci with \( P_{\text{discovery-meta}} < 5 \times 10^{-5} \) in three replication cohorts: one Japanese cohort (JAP) and two independent cohorts of Han Chinese ancestry from Beijing (BHC).

Received 20 July 2015; accepted 23 December 2015; published online 25 January 2016; doi:10.1038/ng.3496
and Shanghai (SHC). We identified ten new loci associated at genome-wide significance ($P_{\text{meta}} < 5 \times 10^{-8}$; Figs. 2 and 3, Table 1, Box 1 and Supplementary Fig. 3) and six new suggestive loci (Supplementary Table 4). Third, we used a series of bioinformatic analyses, including two recently developed Bayesian-based tests\textsuperscript{10,11} (Online Methods), to identify the most likely functional variants at each locus. Because the lead SNPs might not be functional, we examined SNPs in high linkage disequilibrium (LD; $r^2 > 0.8$) with them. Variants were annotated using Encyclopedia of DNA Elements (ENCODE)\textsuperscript{12} and blood expression quantitative trait locus (eQTL) data\textsuperscript{13}. We estimated the proportion of the heritability and sibling risk ($\lambda_s$) explained by new and known SLE-associated loci.

The strongest new signal ($P_{\text{meta}} = 3.75 \times 10^{-117}$, $\text{OR}_{\text{meta}} = 95\%$ confidence interval (CI) = 2.38 (2.22–2.56)) was at rs73366469 between two ‘general transcription factor’ genes\textsuperscript{14}, GTF2I and GTF2IRD1 (Supplementary Table 5). Surprisingly, this signal was much stronger than variants in the human leukocyte antigen (HLA) region. Notably, rs117026326 within GTF2I (92 kb from rs73366469) was recently identified as a major risk locus for primary Sjögren’s syndrome, another autoimmune disease, in Han Chinese\textsuperscript{15} and southern Chinese\textsuperscript{16}. Two recent Sjögren’s syndrome GWAS\textsuperscript{15,17} showed substantial overlap with SLE\textsuperscript{18}, emphasizing the validity and immune relevance of this region.

To confirm the veracity of this association signal, we genotyped 2–6 SNPs (including rs73366469) in ~40% of our discovery samples and in two replication cohorts (Supplementary Table 6). Associations were consistently replicated; rs117026326 showed the strongest association but is in LD with rs73366469 ($r^2_{\text{KR}} = 0.76$, $r^2_{\text{BHC}} = 0.65$; and $r^2_{\text{SHC}} = 0.64$ in controls), making it difficult to separate the effects of these SNPs (Supplementary Table 6). Interestingly, conditional analysis on four SNPs showed that rs80346167 (GTF2IRD1) was independent in the KR cohort, supporting the involvement of both genes in the locus. However, because of the strong correlation structure among variants, genotyping and fine-mapping on a larger scale are required to further delineate this signal. ENCODE data indicate that the high-LD SNP rs7800325 ($r^2 = 0.99$) and indel SNP rs587608058 ($r^2 = 0.81$), ~1.000 bp from rs73366469, lie within conserved enhancers, active chromatin and transcription factor binding sites in CD4+ T cells and GM12878 lymphoblastoid cells (Supplementary Fig. 4a). Chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) and chromosome conformation capture (Hi-C) showed that this region overlaps transcription start sites for GTF2I and VGF (Supplementary Fig. 5 and Supplementary Tables 7 and 8).

The second strongest signal was at the intronic SNP rs10807150 ($\text{DEF6}$, $P_{\text{meta}} = 6.06 \times 10^{-10}$) and is correlated with rs8205 (ZNF76 promoter, $r^2 = 1$), a cis-eQTL altering expression of ZNF76 and $\text{DEF6}$ (Supplementary Tables 5 and 9). The nearby SNP rs4711414 ($r^2 = 0.91$) alters a highly conserved promoter and transcription factor binding site cluster (Supplementary Fig. 4b). The third strongest signal was near IL12B (encoding interleukin (IL)-12); rs2421184, $P_{\text{meta}} = 4.67 \times 10^{-12}$, in a highly conserved enhancer (Supplementary Fig. 4c and Supplementary Tables 5 and 7).

Among the other new signals, rs7726414 ($P_{\text{meta}} = 1.13 \times 10^{-11}$) in the distal promotor of $\text{TCF7}$ is highly linked to rs6874758 ($r^2 = 0.99$), located in a conserved enhancer (Supplementary Fig. 4d and Supplementary Tables 5 and 7). Nearby rs201806887 ($r^2 = 0.79$) alters a strong enhancer and transcription factor binding site cluster. The 5p15.33 signal was in an oncogene\textsuperscript{19} ($\text{TERT}$, intronic; rs7726159, $P_{\text{meta}} = 2.11 \times 10^{-11}$); this locus is tightly bound by the RNA-binding proteins PABPC1 and SLBP (Supplementary Fig. 4e). High-LD SNP rs7705526 ($r^2 = 0.94$) has been linked to chronic lymphocytic leukemia\textsuperscript{20}. The CD226 signal is explained by the intronic SNP rs1610555 ($P_{\text{meta}} = 4.50 \times 10^{-10}$), is linked ($r^2 = 0.74$) to the nonsynonymous SNP rs763361 (Supplementary Fig. 4f) and has been associated with multiple autoimmune diseases. rs763361 is a cis-eQTL for CD226 and also a trans-eQTL for ACBP and MAP3K7CL (Supplementary Table 9). The signal at PCNXL3 (rs2009453, $P_{\text{meta}} = 9.61 \times 10^{-11}$) is in strong LD ($r^2 = 0.95$) with rs931127 (Supplementary Fig. 4g), a cis-eQTL for PCNXL3, SIPA1 and RELA (Supplementary Table 9). The signal at $\text{RASGPR1}$ (rs12900339, $P_{\text{meta}} = 4.73 \times 10^{-10}$) is connected with multiple chromatin interactions (Supplementary Table 8) as well as correlated ($r^2 = 0.77$) with rs12324579, a cis-eQTL for C15orf53 (Supplementary Table 9). Intronic rs61616683 ($\text{SYNGR1}$, $P_{\text{meta}} = 5.73 \times 10^{-10}$) is in active chromatin (Supplementary Fig. 4i) and is a cis-eQTL for $\text{SYNGR1}$ (Supplementary Table 9). Correlated SNP ($r^2 = 0.86$) rs909685 is associated with rheumatoid arthritis in Koreans\textsuperscript{21}. Intronic rs2305772 ($\text{SIGLEC6}$, $P_{\text{meta}} = 1.34 \times 10^{-7}$) is a cis-eQTL for SIGLEC6–SIGLEC12 (Supplementary Table 9) and disrupts a conserved SIGLEC6 splice junction (Supplementary Fig. 4j).

We also confirmed association ($P < 0.005$) with 36 previously reported SLE susceptibility loci (Supplementary Fig. 6 and...
Figure 2  Manhattan plot of the meta-analysis results using the discovery sets. New significant loci are highlighted in red, suggestive loci are highlighted in blue and previously known SLE-associated loci are highlighted in black. The red line represents the threshold for genome-wide significance ($P = 5 \times 10^{-8}$), and the blue line represents the threshold for suggestive evidence of association ($P = 5 \times 10^{-5}$).

Supplementary Table 10. Conditional analysis (Online Methods) at each locus identified secondary associations in three new and ten reported loci (Supplementary Table 11).

As expected, HLA association was replicated in all cohorts (Supplementary Fig. 7a and Supplementary Table 10). The strongest signal was at the HLA class II locus (rs113164910, $P_{\text{discovery-meta}} = 2.48 \times 10^{-37}$, OR = 1.65), 14 kb 3′ of HLA-DRA. To further delineate the HLA signal, we imputed SNPs, classical HLA alleles and HLA amino acid residues in all three cohorts (KR, HC and MC; Online Methods). The most significant association was identified at HLA-DRB1 amino acid position 13 ($P = 9.5 \times 10^{-45}$) and the linked amino acid position 11 ($P = 7.37 \times 10^{-39}$), as shown in a recent HLA fine-mapping study using a subset (~60%) of our Korean samples22 (Supplementary Table 12). Our results also confirmed the reported associations of the two linked classical alleles HLA-DRB1*15:01 ($P = 4.19 \times 10^{-29}$) and HLA-DQB1*06:02 ($P = 6.46 \times 10^{-26}$) (Supplementary Fig. 7b and Supplementary Table 12). To investigate the secondary effect within and outside of HLA-DRB1, we performed a conditional analysis. Consistent with the recent study22, the associations at HLA-DRB1 were almost entirely explained by the primary effect of amino acid positions 11 and 13 and the secondary effect of amino acid position 26 ($P = 4.09 \times 10^{-17}$). After accounting for the effect of the HLA-DRB1 locus (Online Methods), no additional association signals were detected. Thus, the HLA-DRB1 locus explained most of the major histocompatibility complex (MHC) associations (Supplementary Fig. 7c and Supplementary Table 12). Comparing SNP and classical HLA allele associations, we found that both association results colocalized the strongest effects in the HLA-DRB1 region (Supplementary Fig. 7b), as evidenced by the signals for HLA-DRB1*15:01 and nearby rs113164910.

Figure 3  Meta-analysis of the lead SNPs from the ten newly identified loci. We identified ten new loci in the KR, HC and MC cohorts that were replicated in at least two independent cohorts. A partial discovery meta-analysis is presented in the middle of the plot, and the overall meta-analysis is presented below the replication cohorts. Diamonds are used to represent the meta-analysis odds ratios; 95% confidence intervals are represented for each cohort as tickmarks. KR, Korean; HC, Han Chinese; MC, Malaysian Chinese; JAP, Japanese; BHC, Beijing Han Chinese; SHC, Shanghai Han Chinese.
## Table 1 Meta-analysis results for newly identified and suggestive loci associated with SLE in Asian cohorts

| SNP (alleles) | Nearest gene | Discovery cohorts | Replication cohorts | Overall meta-analysis |
|--------------|--------------|-------------------|--------------------|-----------------------|
|              | KR           | HC                | MC                 | JAP                   | SHC           | BHC | P_meta | OR CI |
|              | MAF P        | MAF P             | MAF P             | MAF P                | MAF P        | MAF P |        |       |
| **Newly identified loci** | | | | | | | | |
| rs7366469 (CT) | GTF2IRD1-GTF2I | 0.290.13×10^-69 | 0.360.23 ×10^-10 | 0.250.12 | 1.53×10^-7 | NA | NA | 0.300.16 | 1.17×10^-16 | 0.330.18 | 3.57×10^-21 | 3.75×10^-117 | 2.38 |
| rs10807150 (C/T) | DEF6 | 0.360.32 | 9.21×10^-6 | 0.410.32 | 2.49×10^-4 | 0.370.36 | 5.81×10^-1 | 0.390.35 | 3.34×10^-4 | 0.410.32 | 2.90×10^-6 | 0.380.33 | 3.16×10^-3 | 6.06×10^-16 | 2.22-2.56 |
| rs726159 (A/C) | TERT | 0.070.05 | 1.20×10^-4 | 0.090.05 | 4.77×10^-3 | 0.130.10 | 8.17×10^-2 | 0.090.06 | 3.66×10^-4 | 0.100.08 | 4.21×10^-2 | 0.090.06 | 1.19×10^-2 | 1.13×10^-11 | 1.27-1.54 |
| rs1610555 (T/G) | CD226 | 0.430.39 | 2.10×10^-4 | 0.440.42 | 1.61×10^-3 | 0.480.42 | 7.30×10^-2 | 0.410.38 | 7.70×10^-3 | 0.380.34 | 4.77×10^-2 | 0.340.31 | 3.86×10^-2 | 4.50×10^-11 | 1.14-1.28 |
| rs2009453 (T/C) | PCNL3 | 0.400.45 | 6.13×10^-7 | 0.420.45 | 2.08×10^-3 | 0.430.47 | 1.52×10^-1 | 0.370.40 | 2.56×10^-3 | 0.430.48 | 2.43×10^-2 | 0.410.43 | 3.25×10^-1 | 9.61×10^-11 | 1.13-1.26 |
| rs11235604 (T/C) | ATG16L2 | 0.070.10 | 2.69×10^-9 | 0.090.10 | 5.19×10^-3 | 0.090.12 | 1.71×10^-1 | 0.070.09 | 2.45×10^-2 | 0.080.09 | 6.73×10^-1 | 0.090.10 | 5.87×10^-1 | 1.90×10^-9 | 0.76 |
| rs10936599 (G/A) | MYNN-LRRC34 | 0.440.39 | 5.39×10^-7 | 0.490.46 | 1.15×10^-3 | 0.460.48 | 3.16×10^-1 | 0.390.36 | 1.49×10^-2 | 0.490.45 | 9.13×10^-2 | 0.480.46 | 1.27×10^-1 | 4.93×10^-9 | 1.16 |
| rs2305772 (A/G) | SIGLEC6 | 0.420.46 | 2.33×10^-6 | 0.410.44 | 2.22×10^-3 | 0.390.41 | 5.17×10^-1 | 0.400.45 | 3.90×10^-4 | 0.420.41 | 6.97×10^-1 | 0.420.46 | 3.40×10^-2 | 1.34×10^-9 | 0.86 |
| **Suggestive loci** | | | | | | | | |
| rs11235604 (T/C) | ATG16L2-FCHSD2 | 0.070.10 | 2.69×10^-9 | 0.090.10 | 5.19×10^-3 | 0.090.12 | 1.71×10^-1 | 0.070.09 | 2.45×10^-2 | 0.080.09 | 6.73×10^-1 | 0.090.10 | 5.87×10^-1 | 1.90×10^-9 | 0.76 |
| rs10936599 (G/A) | MYNN-LRRC34 | 0.440.39 | 5.39×10^-7 | 0.490.46 | 1.15×10^-3 | 0.460.48 | 3.16×10^-1 | 0.390.36 | 1.49×10^-2 | 0.490.45 | 9.13×10^-2 | 0.480.46 | 1.27×10^-1 | 4.93×10^-9 | 1.16 |
| rs223881 (C/T) | CCL22 | 0.460.49 | 9.31×10^-5 | 0.460.53 | 1.29×10^-3 | 0.510.59 | 7.23×10^-3 | 0.360.40 | 2.60×10^-3 | 0.540.53 | 7.36×10^-1 | 0.500.51 | 5.94×10^-1 | 2.06×10^-8 | 0.87 |
| rs1885889 (A/G) | UBAC2 | 0.410.36 | 6.34×10^-7 | 0.400.37 | 1.38×10^-3 | 0.430.36 | 1.43×10^-2 | 0.430.41 | 2.25×10^-1 | 0.370.38 | 8.43×10^-1 | 0.380.37 | 4.47×10^-1 | 3.17×10^-7 | 1.14 |
| rs7556469 (G/A) | PTPrC | 0.120.15 | 2.33×10^-6 | 0.110.14 | 1.45×10^-3 | 0.110.10 | 4.94×10^-1 | 0.170.17 | 3.86×10^-1 | 0.110.13 | 1.61×10^-1 | 0.130.15 | 1.85×10^-1 | 1.95×10^-6 | 0.84 |
| rs1202418 (G/A) | RGS1 | 0.320.36 | 1.53×10^-4 | 0.290.32 | 1.90×10^-3 | 0.310.38 | 1.26×10^-2 | 0.350.37 | 1.95×10^-1 | 0.320.35 | 9.68×10^-2 | 0.330.32 | 6.82×10^-1 | 1.12×10^-5 | 0.88 |

Alleles are shown as minor/major alleles. KR, Korean; HC, Han Chinese; MC, Malaysian Chinese; JAP, Japanese; BHC, Beijing Han Chinese; SHC, Shanghai Han Chinese; MAF, minor allele frequency for cases/controls; OR, odds ratio; CI, confidence interval.

*Only from Immunochip controls.
Box 1 Ten loci newly associated with SLE

**GTF2IRD1-GTF2I** (7q11.23). Our top signal was at rs73366469 in the intergenic region between the critical general transcription factor (GTF) genes**GTF2IRD1** and **GTF2I**. Both proteins encode multifunctional phosphoproteins with roles in transcription and signal transduction. Both have been reported to be major genes responsible for neurocognitive defects in Williams-Beuren syndrome**34,35,** as well as associated with Sjögren’s syndrome. Deletion of this cytogenetic band reportedly alters craniofacial and neurocognitive characteristics**36,**. Several studies also reported connections (especially for **GTF2I**) to transcriptional regulation induced in response to various signaling pathways, including immune response in both B and T cells**37,38,**.

**DEF6** (6p21.31). **DEF6**, differentially expressed in FDCP (factor-dependent cell progenitors) 6 homolog, is a guanine nucleotide exchange factor for RAC and CDC42; it is highly expressed in B and T cells**39,**. **DEF6** is implicated in autoimmunity through regulation of **IRF4** (encoding interferon regulatory factor 4) in IL-12 responsiveness**40,**.

**IL12B** (5q33.3). **IL-12**β is a component of both IL-12 (made in B cells, macrophages, dendritic cells and neutrophils) and IL-23 (made in macrophages and dendritic cells). Both interleukins are critical secreted signals in T cell activation. Ustekinumab, a monoclonal antibody against IL-12β, recognizes both IL-12 and IL-23; it is used in the treatment of psoriasis and is in testing for other autoimmune diseases.

**TCF7** (5q31.1). **TCF7** is a T cell–specific transcription factor that regulates expression of CD3, the T cell co-receptor. **TCF7** is associated with type 1 diabetes risk**41,**. A mouse *Tcf7* knockout showed reduced immunocompetence of T cells in the periphery.

**TERT** (5p15.33). **TERT** (telomerase reverse transcriptase) has a critical role in DNA replication and chromosomal stability and is strongly associated with cancer**15,**. Telomerase activity was dramatically upregulated in leukocytes from patients with SLE, particularly in CD19+ B cells**42,** and in untreated patients and patients with nephritis. Immunoglobulin (Ig) recombination and telomere maintenance both function through non-homologous end joining, a core component of which is Ku70-Ku80, first discovered as a lupus autoantigen. The mechanisms by which SLE and telomerase activity interact remain unknown.

**CD226** (18q22.3). **CD226** is a glycoprotein expressed on the surface of natural killer cells, platelets, monocytes and a subset of T cells. The protein mediates cellular adhesion of platelets and megakaryocytic cells to vascular endothelial cells. **CD226** also mediates T cell and natural killer cell recognition and lysis of tumor cells**43,**. It is a member of the Ig superfamily, containing two Ig-like domains of the V set, and is strongly associated with multiple autoimmune diseases**44-49,**. Previous associations with SLE**50,** fell short of genome-wide significance.

**PCNL3** (11q13.1). **PCNL3** (pecanex-like 3) is a highly conserved, ubiquitously expressed membrane protein of unknown function that is known to affect Notch signaling. A previous study found that **PCNL3** was one of the four most diagnostic genes for psoriatic arthritis (where it is downregulated in symptomatic patients)**51,**.

**RASGRP1** (15q14). **RAS** guanyl-releasing protein 1 (calcium and DAG regulated) activates the ERK–MAP kinase cascade, which couples RAS to development, homeostasis and the differentiation of T and B cells. **RASGRP1** was found to be downregulated in symptomatic patients with SLE**52,**. The related gene **RASGRP3** has been associated with SLE and clinical features (discoid rash, malar rash and antinuclear antibodies) in a Han Chinese population**53,**.

**SYNGR1** (22q13.1). **SYNGR1** (synaptogyrin1) has primary roles in neuronal synaptic transmission and is implicated in schizophrenia**54,** and rheumatoid arthritis**55,** and possibly in primary biliary cirrhosis**56,**. The related protein **SYNGR2** is highly expressed in dendritic cells**56,**.

**SIGLEC6** (19q13.3). **SIGLEC6** (sialic acid-binding Ig-like lectin 6) encodes a transmembrane receptor that binds lectin. **SIGLEC6** mediates cell-cell adhesion by binding to glycans and is expressed almost exclusively in the placenta and in B cells**57,**.

Additionally, we identified six new suggestive loci (1.9 × 10−9 < *p*meta < 1.12 × 10−5) with three missense variants (Supplementary Tables 4 and 13). Although three of these loci (ATG16L2-FCHSD2, MYNN-LRRC34 and CCL22) passed genome-wide significance, further replications are needed to confirm their association.

We replicated most of the previously reported genes with the same published SNPs or with highly correlated SNPs (Supplementary Table 14). We also found four genes with new uncorrelated SNPs shifting the association peaks in Asians (Supplementary Table 15). Of these, **ARHGAP31-TMEM39A-CD80** was of special interest: previously reported association signals for **TMEM39A** (rs11322000) and **CD80** (rs6804441) were now explained by a synonymous SNP in **ARHGAP31** (rs2305249, *p*meta = 1.64 × 10−9), a cis-eQTL of **B4GALT4** and **POGLUTI** (a Notch signaling regulator)**24,**.

To identify the most likely functional variants within a locus, we used Bayesian-based analyses**10,11,**, eQTL searches and epigenetic analyses (Online Methods and Supplementary Tables 7–9 and 16). We found that the lead SNPs in the **GTF2I**, **IL12B**, **PCNL3**, **SYNGR1**, **RASGRP1** and **SIGLEC6** loci had a high probability of being functional (Supplementary Table 17).

To explore biological functions and pathways related to the SLE susceptibility loci (new and previously published), we performed gene set enrichment analysis (GSEA; Online Methods). We identified pathways and gene ontology categories (including immunity, inflammation and cytotoxicity) (Supplementary Fig. 8) in common between the new and published loci. Moreover, GSEA with a drug target database**25** identified a set of 56 significantly enriched drugs (adjusted *P* < 0.05; Supplementary Table 18), including SLE therapeutics**26** (cyclosporine, zinc acetate, hydrocortisone and methotrexate), that affected expression of the target loci. Of note was **GTF2I**, significantly enriched in drugs used for the treatment of leukemia (imatinib; adjusted *P* = 1.82 × 10−10) and lymphoma (cisplatin; adjusted *P* = 2.68 × 10−4). Immune system involvement was confirmed by enrichment analysis of SLE-associated loci in mouse immune phenotypes, with significant enrichment in abnormal lymphocyte, leukocyte and/or immune cell physiology and abnormal cell-mediated and/or adaptive immunity (Supplementary Table 19).

To understand the relationship between our newly identified loci and known SLE loci and to identify possible molecular mechanisms involved in SLE pathogenesis, we performed network interaction analysis**27,28** (Online Methods). We found that the new and replicated SLE loci are connected directly and indirectly to each other through gene regulation as well as protein and biochemical interactions (Supplementary Figs. 9 and 10). Text mining methods**29** confirmed that many of these loci have strong associations with one another in the literature and show how the new loci are related to the replicated loci (Supplementary Fig. 11). Within these relationships, we further identified subnetworks of molecules interacting with our new
loci in the context of known SLE-relevant genes (TERT, IL12B, GTF2I, RELA, SRC and NFKB2; Supplementary Fig. 12).

We identified only one nonsynonymous variant (rs2305772, p.Pro246Ser/splice junction, SIGLEC6) in LD ($r^2 \geq 0.8$) with the new SNPs (Supplementary Table 20), suggesting that other variants likely contribute to SLE pathogenesis through epigenetic regulation rather than alterations to protein structure or function. Joint analysis of lead and correlated ($r^2 > 0.8$) SNPs indicated a 13-fold enrichment in strong enhancers in K362 erythroblastemia cells and an up to 22-fold enrichment in DNase I hypersensitivity sites in MCF-7 breast cancer cells (Supplementary Table 21).

In six of the ten new loci (GTF2I, DEF6, CD226, PCNX1, RASGRP1 and SIGLEC6), the highly conserved, ancestral alleles were the risk allele. Except in the SIGLEC6 locus, all derived, protective alleles were the major allele in Asians and European-ancestry populations (Utah residents of Northern and Western European ancestry (CEU)); in SIGLEC6, the derived, protective allele was the major allele only in Europeans. Notably, derived risk alleles for SYNGR1 occur at a frequency of >80% in Asians (Han Chinese in Beijing, China (CHB) and Japanese in Tokyo, Japan (JPT)), as compared to a frequency of ~20% in the CEU population, suggesting that SYNGR1 is undergoing selection in Asian populations, as indicated by $F_{ST}$, IHS and XP-EHH analyses (Supplementary Table 22).

We assessed whether regions associated with these SNPs (new and replicated) harbored genes expressed in distinct immune cell types (Online Methods). We identified significant (1 × 10$^{-9}$ < $P < 4 × 10^{-6}$) cell type–specific expression of genes within the new loci in human B cells, T cells, natural killer cells and dendritic cells (Fig. 4 and Supplementary Fig. 13a). This result was further strengthened by replication with homologous mouse genes in mouse cell lines, with significant enrichment in CD19$^+$ B cells ($P = 1.0 \times 10^{-5}$) and transitional B cells ($P = 1.0 \times 10^{-5}$) (Supplementary Fig. 13b). Thus, our results point to a strong (and conserved) effect of gene expression in B and T cells during SLE pathogenesis.

Six of the ten newly identified loci are also associated with other autoimmune diseases, including celiac disease, rheumatoid arthritis, type 1 diabetes and multiple sclerosis (Supplementary Table 23), suggesting pleiotropic effects. This pattern extended to suggestive signals at ATG16L2, PTPRC, UBAC2 and RGS1, which are reportedly associated with other autoimmune diseases.

Collectively, these new and previously reported SLE susceptibility variants (47 SNPs) explain 24% of the total heritability of SLE in Asians (Supplementary Table 24). Among these loci, the HLA region explains 2% of the heritability and the ten new loci account for 6%. All loci together explain 24% of $\Lambda_p$ (Supplementary Table 25); new loci explain 7% and HLA loci explain 2%. To quantify the predictive ability of these variants, we estimated genetic risk through the weighted genetic risk score (wGRS). Newly identified risk alleles significantly ($P = 6.58 \times 10^{-39}$) increased the wGRS area under the curve (AUC) (95% CI) from 0.82 to 0.85 (0.85–0.86) (Supplementary Fig. 14a,b).

In summary, our results further define the genetic architecture and heritability of SLE risk (especially in Asians) and provide insights into disease pathogenesis. Through comprehensive analysis of multiple Asian populations, we identified ten new SLE-predisposing loci and validated association in 36 reported loci (often refining the associated intervals). We pinpointed and annotated independently associated variants at each locus. Further analysis in additional populations and experimental validation in cultured and patient-derived cells (as previously performed$^{31-33}$) will demonstrate which SNPs are causal and elucidate biochemical pathways through which genetic changes contribute to SLE. This study highlights the success of targeting high-risk populations for genetic analysis, followed by systematic bioinformatics analysis to set up future experimental validation.

**METHODS**

Methods and any associated references are available in the online version of the paper.
Accession codes. Summary-level association data for the discovery sets are provided as a Supplementary Data Set.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We are grateful to the affected and unaffected individuals who participated in this study. We thank the research assistants, coordinators and physicians who helped in the recruitment of subjects, including the individuals in the coordinating projects. A part of the Korean control data was provided from the Korean Biobank Project supported by the Korea Center for Disease Control and Prevention at the Korea National Institute of Health. Genomic DNA from ~100 Korean patients with SLE was obtained from the Korean National Biobank at Wonkwang University Hospital, which is supported by the Ministry of Health and Welfare, Republic of Korea.

This work was supported by grants from the US National Institutes of Health (AR060366, MD007909, AI013399, AI024717, AI083194, AI107176, TR001425, HG008666 and HG006828), the US Department of Defense (PR094002), the US Department of Veterans Affairs, the National Basic Research Program of China (973 program) (2014CB541902), the Research Fund of Beijing Municipal Science and Technology for the Outstanding PhD Program (20121000110), the National Natural Science Foundation of China (#81200524, #81230072) and High-Impact Research Ministry of Education Grant UMC/625/1/HIR/MoE/000044-20001, Malaysia. This study was also supported by a grant from the Korea Healthcare Technology R&D Project (HI13C2124), Ministry for Health and Welfare, Republic of Korea.

AUTHOR CONTRIBUTIONS

S.K.N., J.B.H. and S.-C.B. conceived and initiated the study. S.K.N. and J.B.H. conceived and initiated the study. S.K.N. designed, coordinated projects. A part of the Korean control data was provided from the Korean Biobank Project supported by the Korea Center for Disease Control and Prevention at the Korea National Institute of Health. Genomic DNA from ~100 Korean patients with SLE was obtained from the Korean National Biobank at Wonkwang University Hospital, which is supported by the Ministry of Health and Welfare, Republic of Korea.

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. Jakes, R.W. et al. Systematic review of the epidemiology of systemic lupus erythematosus in the Asia-Pacific region: prevalence, incidence, clinical features, and mortality. Arthritis Care Res. (Hoboken) 64, 159–168 (2012).
2. Danchenko, N., Satia, J.A. & Anthony, M.S. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus 15, 308–318 (2006).
3. Wandstrat, A. & Wakeland, E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat. Immunol. 2, 802–809 (2001).
4. Harley, I.T., Kaufman, K.M., Langefeld, C.D., Harley, J.B. & Kelly, J.A. Genetic epidemiology of systemic lupus erythematosus at 7q11.23. Nat. Genet. 45, 1361–1365 (2013).
5. Zheng, J. et al. The GTF2I rs117026326 polymorphism is associated with anti-SSA-positive primary Sjögren’s syndrome. Rheumatology (Oxford) 54, 562–564 (2015).
6. Lessard, C.J. et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjögren’s syndrome. Nat. Genet. 45, 1284–1292 (2013).
7. Chen, E.Y. Emerging new pathways of pathogenesis and targets for treatment in systemic lupus erythematosus and Sjögren’s syndrome. Curr. Opin. Rheumatol. 21, 443–447 (2009).
8. Johaafy, S.E. et al. Evaluation of candidate stromal epithelial cross-talk genes analysis tool. BMC Bioinformatics 14, 128 (2013).
9. Xiaoping, W. & Lahita, R.G. Pragmatic approaches to therapy for systemic lupus erythematosus. Nat. Rev. Rheumatol. 10, 97–107 (2014).
10. Tost, B., Arsenault, R., Griebl, P., Napper, S. & Kusakil, A. DAPPEL: a pipeline for the homology-based prediction of phosphorylation sites. Bioinformatics 29, 254–262 (2013).
11. Chen, E.Y. et al. Enrich: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128 (2013).
12. Xiong, W. & Lahita, R.G. Pragmatic approaches to therapy for systemic lupus erythematosus. Nat. Rev. Rheumatol. 10, 97–107 (2014).
13. Wren, J.D., Bekerejian, R., Stewart, J.A., Shohet, R.V. & Garner, H.R. Knowledge discovery by association-based identification and ranking of implicit relationships. Bioinformatics 20, 389–398 (2004).
14. Chen, E.Y. et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am. J. Hum. Genet. 89, 496–506 (2011).
15. Molineros, J.E. et al. Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production. PLoS Genet. 9, e1003222 (2013).
16. Math, A.K. et al. Combined protein- and nucleic acid–level effects of rs1143679 (R77T), a lupus-predisposing variant within ITGAM. Hum. Mol. Genet. 23, 4161–4176 (2014).
17. 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).
18. Vandeweyer, G., Van der Aa, N., Reyniers, E. & Kooy, R.F. The contribution of CLIP2 haplinsufficiency to the clinical manifestations of the Williams-Beuren syndrome. Am. J. Hum. Genet. 90, 1071–1078 (2012).
19. Howard, M.L. et al. Mutation of GIT2r from the Williams-Beuren syndrome critical region results in facial dysplasia, motor dysfunction, and altered vocalisations. Neurobiol. Dis. 45, 913–922 (2012).
20. Mital, P. et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IR1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. J. Med. Genet. 47, 312–320 (2010).
21. Roy, A.L. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I. 10 years later. Gene 492, 32–41 (2012).
22. McDonald, T., Kam, T., Pour, P.S. & Sadowski, I. Specific interaction of TFII-I with factor TFII-I: 10 years later. J. Biol. Chem. 287, 43541–43549 (2002).
23. Biswas, P.S. et al. Dual regulation of IRF4 function in T and B cells is required for the coordination of T-B cell interactions and the prevention of autoimmunity. J. Exp. Med. 209, 581–590 (2012).
24. Noble, J.A. et al. A polymorphism in the TCF7 gene, C883A, is associated with type 1 diabetes. Diabetes 52, 1579–1582 (2003).
42. Klapper, W. et al. Telomerase activity in B and T lymphocytes of patients with systemic lupus erythematosus. Ann. Rheum. Dis. 63, 1681–1683 (2004).
43. Iguchi-Manaka, A. et al. Accelerated tumor growth in mice deficient in DNAM-1 receptor. J. Exp. Med. 205, 2959–2964 (2008).
44. Alcina, A. et al. The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes Immun. 11, 439–445 (2010).
45. Deshmukh, H.A. et al. Evaluation of 19 autoimmune disease-associated loci with rheumatoid arthritis in a Colombian population: evidence for replication and gene-gene interaction. J. Rheumatol. 38, 1866–1870 (2011).
46. Haffner, J.P. et al. CD226 Gly307Ser association with multiple autoimmune diseases. Genes Immun. 10, 5–10 (2009).
47. Maiti, A.K. et al. Non-synonymous variant (Gly307Ser) in CD226 is associated with susceptibility to multiple autoimmune diseases. Rheumatology (Oxford) 49, 1239–1244 (2010).
48. Qiu, Z.X., Zhang, K., Qiu, X.S., Zhou, M. & Li, W.M. CD226 Gly307Ser association with systemic lupus erythematosus in the Chinese Han population. Tissue Antigens 77, 65–67 (2011).
50. Du, Y. et al. Association of the CD226 single nucleotide polymorphism with systemic lupus erythematosus in the Chinese Han population. Tissue Antigens 77, 583–591 (2006).
51. Yasuda, S. et al. Defective expression of Ras guanyl nucleotide–releasing protein 1 in a subset of patients with systemic lupus erythematosus. J. Immunol. 179, 4890–4900 (2007).
52. He, C.F. et al. TNIPI, SLC15A4, ETS1, RasGRP3 and IKZF1 are associated with clinical features of systemic lupus erythematosus in a Chinese Han population. Lupus 19, 1181–1186 (2010).
53. Patel, N. et al. OB-BP1/Siglec-6. A leptin- and sialic acid–binding protein of the immunoglobulin superfamily. J. Biol. Chem. 274, 22729–22738 (1999).
ONLINE METHODS

Study overview. This study was conducted in three stages (Fig. 1). In the first stage, we genotyped three Asian cohorts: Koreans (KR), Han Chinese (HC) and Malaysian Chinese (MC). This step was followed by quality control and preliminary association analysis to identify 578 regions with association \( P < 5 \times 10^{-3} \). We then increased the Korean sample size with out-of-study controls and performed imputation-based meta-analysis to discover 16 new regions with \( P_{\text{discovery-meta}} < 5 \times 10^{-5} \). In the second stage, we followed up these 16 new regions, performing in silico replication on a Japanese (JAP) GWAS data set and two independent replications on separate Beijing Han Chinese (BHC) and Shanghai Han Chinese (SHC) data sets to identify ten new loci with \( P_{\text{meta}} < 5 \times 10^{-8} \). In the third stage, we used bioinformatic databases to annotate the identified variants and carried out comprehensive analyses to uncover potential disease-predisposing variants involved in SLE pathogenesis (Supplementary Table 1 and Supplementary Note). Information on ethical approval is provided in the Supplementary Note.

Imputation-based association analysis, meta-analysis and conditional analysis. For the first stage of our study (Fig. 1), we performed single-SNP case-control association analysis based on Immunochip genotype data from each population that were subjected to quality control. We calculated association \( P \) values, standard errors, and odds ratios and 95% confidence intervals using PLINK. This identified 578 regions with \( P < 5 \times 10^{-3} \) in at least one Asian cohort for imputation (Supplementary Fig. 1 and Supplementary Table 3). To perform the imputation more intensively and accurately, we wrote a script based on a recursive algorithm to define imputation regions. Imputation regions were defined if they contained a peak SNP with \( P < 5 \times 10^{-3} \). Region size was defined by the length of the LD region \( (r^2 > 0.2) \) with respect to the peak SNP. To avoid edge effects, we extended a further 100 kb on each side for each region. The recursive algorithm to define imputed regions used the following steps:

1. Find the peak SNP with minimal \( P \leq 5 \times 10^{-3} \) in a region \( x(y) \) (the region starting with the whole chromosome \( x \) is the start position, \( y \) is the end position)). If such a peak SNP exists, continue; otherwise, stop.
2. Define the imputation region as \( d(u,v) = LD \text{ region} \ (r^2 > 0.2) \) with peak SNP \( \pm 100 \) kb.
3. If \( (x,u) \) exists, go to \( a(x,u) \) recursively (step 1); if \( (v,y) \) exists, go to \( a(v,y) \) recursively (step 1). Otherwise, stop.
4. Collect all regions \( d(u,v) \) for final imputation.

For the second stage of our study (Fig. 1), we integrated additional GWAS data from Korean out-of-study controls to increase both SNP density and statistical power. Because the Korean Immunochip and GWAS data sets were genotyped on two different platforms and the number of overlapping SNPs was less than the original number of SNPs for either the Korean Immunochip or Korean GWAS data set, we imputed each set separately on its original number of real genotyping SNPs using MACH-Admix. The Han Chinese and Malaysian Chinese Immunochip data sets were imputed separately as well following the Korean Immunochip protocol. We took 504 Asians (104 from the JPT population, 200 from the CHB population and 200 from the Southern Han Chinese (CHS) population) from 1000 Genomes Project data (1000 Genomes Project Project Phase 3 Integrated Release Version 5 Haplotypes) as the reference panel for imputations. All SNP names and strands for the three Immunochip data sets and the one out-of-study control data set were aligned with the Asian reference panel (\( n = 504 \)) before those four data sets were each imputed separately. This imputation strategy has been used by many earlier studies and has also been recommended as best practice by the eMERGE Network.

After imputation, we performed strict quality control on post-imputed SNPs. In addition to the quality control steps described above (Hardy-Weinberg equilibrium \( P > 0.0001 \) in controls, MAF > 0.5%), post-imputed SNPs were also required to have high imputation quality (Rsq > 0.7 for MAF ≥ 3% and Rsq > 0.9 for MAF < 3%) to be included for further analysis. To take into account imputation uncertainty, we used mach2dat and single-SNP post-imputation-based association tests and for conditional logistic regression analysis, with adjustment for population stratification. We used the first three principal components as covariates to correct for population stratification and potential batch effects (Supplementary Figs. 15 and 16). Additionally, as a complementary analysis, we used a newly developed genotype-conditional association test (GCAT) to confirm our principal-component analysis (PCA)-corrected associations, the results of which were very consistent (data not shown). We used MetaReg to perform the meta-analysis based on post-imputation associations for three Immunochip cohorts (KR, HC and MC), as well as for the combined KR data set (the merged dosage data set of KR Immunochip and KR GWAS controls), HC Immunochip and MC Immunochip. To include the highest quality SNPs in the follow-up association analysis, we used imputed SNPs with high imputation quality (Rsq > 0.7) in each of the separately imputed data sets (Immunochip and GWAS). We then merged the two imputation sets according to the stringent quality control criteria described above.

Finally, we analyzed SLE association in 152,918 post-imputation SNPs subjected to quality control and identified 20,213 associated SNPs (\( P_{\text{discovery-meta}} < 0.005 \), from which we successfully replicated 36 SLE loci with \( P_{\text{discovery-meta}} < 0.005 \) (Supplementary Table 10) and identified 16 new suggestive regions with \( P_{\text{discovery-meta}} < 5 \times 10^{-5} \) for follow-up replication (Supplementary Tables 26 and 27).

To test whether any systematic bias was introduced by this imputation procedure, we also performed an association analysis of the lead SNPs between the controls (Immunochip versus GWAS). We found no evidence of systematic bias introduced by imputation and thus consider the imputation results sound (Supplementary Table 28).

We performed conditional analysis for 20 known SLE-associated loci with genome-wide significance and ten new regions with genome-wide significance after replication in the largest cohort (KR). Conditional analysis was iterative, starting with the top SNP with that lowest \( P \) value as the first SNP to be conditioned on; all subsequent SNPs that were significant after conditioning were added to the regression model as covariates until no SNP with \( P < 5 \times 10^{-5} \) remained. To ensure that SNPs were truly independent, SNPs in high LD (\( r^2 > 0.3 \)) with the SNP being conditioned on) were filtered out before the next iteration, and only associated SNPs with \( P < 5 \times 10^{-5} \) entered conditional analysis.

Functional annotation of new loci. To localize candidate causal variants, we annotated each lead SNP along with its surrounding correlated SNPs (\( r^2 > 0.7 \) in Asian samples from the 1000 Genomes Project), as implemented in HaploReg on data obtained from Phase 1 of the 1000 Genomes Project and Ensembl. We surveyed allele-dependent gene expression regulation (eQTLs) by querying the blood eQTL database (which houses the experimental meta-analysis from gene expression experiments performed on non-transformed peripheral blood samples from 5,311 individuals of European descent and later replicated in 2,775 individuals) for cis- and trans-eQTLs (Supplementary Table 9). The functional significance of independent SNPs from new regions is shown in Supplementary Table 7, and we report eQTL results in Supplementary Table 9.

We annotated epigenetic regulatory features for all independent lead SNPs (and their correlated variants; \( r^2 > 0.8 \)) in our new regions using the HaploReg, GWAS3D and rSNPBase online tools. HaploReg provides functional annotations for binding motifs and epigenetic marks. GWAS3D aggregates epigenetic data from 16 cell types from multiple databases, including the ENCODE Project, and identifies multiple regulatory SNPs in high LD with the queried SNPs. Among the regulatory elements queried were enhancer marks (p300, H3K4me1 and H3K27ac), promoter regions, CTCF insulator marks and DNase I–hypersensitive sites (DHSs). ChromHMM was used to predict histone states and chromatin interactions. To understand distal regulatory relationships among the new loci, chromatin interactions between candidate loci were gathered from ChIA-PET and Hi-C data on eight cell lines (K562, NB4, GM12878, CD4+, T cells, H1–hESC, IMR90, RPPE1 and MCF-7), available through ENCODE. We reported data for lead SNPs with at least three ChIA-PET or Hi-C hits (Supplementary Fig. 5 and Supplementary Table 8). Additionally, rSNPBase provided putative functional SNPs with experimentally validated regulatory elements controlling transcriptional and post-transcriptional events.
Functional fine-mapping. To identify the set of variants most likely to house a functional variant, we used two Bayesian methods. The first one was based on a Bayesian regression to estimate each SNP’s Bayes factor and, thereafter, its posterior probability of association in the region. Second, we used the Probabilistic Identification of Causal SNPs (PICS) algorithm, which incorporates the underlying epigenetic information for those variants, to further narrow down the available SNPs within the Bayesian credible set.

Bayesian logistic regressions for each of the SNPs at the new imputation regions was implemented in the Bayes Factor (BF) library in R. Henceforth, we estimated the posterior probability for each SNP, as well as the proportion of the total Bayes factor explained by each variant. We formed 95–99% credible sets as the cumulative proportion of the Bayes factor to 0.1. To assess how much each of the effects could be explained by the credible sets, we annotated each candidate SNP with dbSNP functions (intron, missense, UTR, synonymous or intergenic), as well as epigenetic annotations (promoter, enhancer, DNAse hypersensitivity, bound protein motif disrupter disrupted, rSNP, LD proxy of rSNP (r² > 0.8), proximal regulation, distal regulation, microRNA regulation, RNA-binding protein–mediated regulation or eQTL).

We implemented the PICS method to identify the set of variants with probable functional effects. This method uses the epigenetic information at each locus and estimates the posterior probability of a SNP being causal given the strength of association and its linkage neighborhood, as well as regulatory element annotations.

Gene-gene interaction. To identify gene-gene interaction, we performed logistic regression with an interaction term between all pairs of lead SNPs using PLINK. Both BOOST3 and joint effects methods were used to screen for SNP-SNP interactions. We used a significance threshold of 1 × 10⁻⁴.

Network interactions. To investigate how our new loci interact with other genes, we curated network interactions from the Disease Association Protein-Protein Link Evaluator database (DAPPLE v2.0). We used a seed of all our new loci (flanking genes on both the left and right were also used for intergenic signals) and 20,000 within-degree-node permutations. We chose to simplify our network using the given number of potential interactions (Supplementary Fig. 12). The network represents all significant interactions between proteins that form a network.

Additionally, we confirmed network interactions using the aggregated database ConsensusPathDB28. ConsensusPathDB scores the confidence level of protein interactions on a scale between 0 and 1, and aggregates 11 pathway databases for GSEA. We chose interactions with a high confidence score (Intscore >0.9). Additionally, we plotted all possible high-confidence interactions for all new loci (Supplementary Figs. 9 and 10).

To investigate how our updated set of new SLE loci were related to each other and to previously established loci, we used a literature mining–based approach, implemented in IRIDESCENT29 (Supplementary Fig. 11). This approach identifies genes mentioned together in the same MEDLINE titles and/or abstracts (over 24 million currently) and weights their relevance on the basis of relative frequencies of gene mention and gene-gene co-mention.

Gene set enrichment analysis. To determine whether there were significant enrichments of our SLE (new and replicated) loci as compared to reported SLE loci in human and mouse ontologies, we performed GSEA using GREAT25 (Supplementary Table 19). To compare interacting pathways and the ontological properties of new versus published SLE genes, we used ConsensusPathDB28. Additionally, to identify and compare drug perturbation signatures for new and reported loci, we used the gene enrichment analysis software Enrichr25 (Supplementary Table 18).

To test whether there was bias in enrichment due to the choice of the Immunochip as a genotyping platform, we conducted 100 over-representation analysis tests using sets of 58 genes taken at random from the Immunochip gene set in ConsensusPathDB28. We computed the number of times any pathway or ontology category was observed in the 100 random sets (Supplementary Table 29).

Cell type–specific enrichment analysis. To identify enrichment in cell type–specific expression of new and replicated SLE loci (57 SNPs), we used a previously reported approach30,76 as follows. We used normalized expression data from 79 human cell types from GeneAtlas27 (curated by the Genomic Institute of the Novartis Research Foundation), as well as from 249 mouse cell types sorted by FACS and assayed at least three times from the Immunological Genome Project (ImmGen)17. Additionally, we used cell type–specific expression of the collection of 573 human cell samples from the FANTOM5 Project29.

In this analysis, we extracted genes from the regions where SNPs correlated with the lead SNPs (r² > 0.5; Table 1), spanning between recombination hotspots. We used the normalized cell type–specific expression profiles of the extracted genes to identify which cell types significantly express SLE candidate genes. Specificity P values were estimated on the basis of the permutation of ranked expression levels for each locus (10⁶ permutations) using SNPseq76 (Fig 4 and Supplementary Fig. 13). P values (blue bars) that passed the multiple-testing threshold (black line) indicate significant enrichment in SLE-associated loci. Threshold lines are dependent on the number of categories present in each database: that is, for the 1,751 GO categories possible, the significance threshold would be 2 × 10⁻⁵.

Explained heritability. We assessed the variance in liability (Vp) explained for each of our genome-wide significant SNPs using the liability threshold method7. We estimated Vp for new, reported and HLA loci separately. We used the weighted risk allele frequency and meta-analysis odds ratio for each variant to calculate the liability threshold for each genotype (Supplementary Table 24). We present values estimated using a prevalence estimate (K) of 0.0030653 following So and Sham7. To check the consistency of this heritability estimate, we also used the allele frequencies from each cohort, as well as the allele frequencies for the HapMap and 1000 Genomes Project CHB and JPT populations.

Sibling relative risk. We estimated the contribution of SLE susceptibility loci to the familiar relative risk (Supplementary Table 25), especially for the sibling relative risk (λs) under the multiplicative model, where λ is the overall sibling relative risk, assumed here to be ~30 (ref. 81), with the relative sibling risk from each locus (λ) given by

\[
\lambda = \frac{p^2 + q}{pr + q}\]

where p is the frequency of the risk allele (q = 1 − p) and r is the per-allele risk ratio82.

Weighted cumulative genomic risk score. To assess the effect of accumulation of risk variants between cases and controls, we estimated the wGRS for all individuals with high imputation quality (Rsq > 0.7). We weighted the number of risk variants by the natural logarithm of the meta-analysis odds ratio83 for all ten new loci, two HLA loci and 35 replicated loci from a total of 2,476 cases and 8,426 controls. Significant differences in wGRS were estimated using a logistic regression model including sex and the top three principal components as covariates (Supplementary Fig. 14). Differences between mean wGRS in cases and controls were estimated through a linear model.

Area under the curve. We estimated the predictive power of the wGRS for variants, as well as the marginal contribution of the new variants, by comparing the AUCs for the baseline model (including reported loci) and the expanded model (including reported and new loci) (Supplementary Fig. 14). AUC corrected for sex was estimated in R using the pROC library84. Confidence intervals for the AUC were estimated using the nonparametric DeLong method85.

Evidence for natural selection. To assess evidence for natural selection, we used HapMap 2 and Human Genome Diversity Project (HGDP) population data through Haplploter and the HGDP Selection Browser. For each of the ten
new loci, we looked for evidence of positive natural selection in the 1-Mb region around each gene. Haplotter uses three statistics: Hs (the integrated haplotype score), $F_{ST}$ (the fixation index of population differentiation) and the empirical $P$ value for the distribution of Tajima’s $D$ and Fay’s $H$ (ref. 86), whereas the HGDPS Selection Browser uses XP-EHH$^{87}$ (Cross-Population Extended Haplotype Homozygosity) to identify positive natural selection in addition to the Hs score. Evidence of natural selection was considered positive if the empirical $P$ value was <0.05 for the distribution of both Tajima’s $D$ and Fay’s $H$, and $-\log(\text{P value}) > 3$ for $F_{ST}$, $D$, Hs or XP-EHH, where $q$ is the empirical $P$ values rank ordering the summary statistic value (a given region divided by the total number of regions) (Supplementary Table 22).

Graphical display of the epigenetic landscape of the loci. For Supplementary Figure 4, plots were assembled similarly to ref. 32. Most data were downloaded from the UCSC Genome Browser and displayed using custom MATLAB code. ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) tracks for CD4$^+$ cells and GM12878 cells were downloaded from the Gene Expression Omnibus (GEO) under accession GSE47753 (ref. 88). DNase I hypersensitivity, ENCODE sequence classification, histone marks and binding data for transcription factors (to DNA) and RNA-binding proteins (to RNA) were all downloaded from the UCSC Genome Browser. ENCODE regulatory elements are color-coded according to their standard; other signals are shown in grayscale, with a darker shade of gray representing a higher signal. All tested SNPs are shown as bars with a height of $-\log(\text{P value})$ at the top. In the zoomed images, SNPs of interest are labeled.

58. Okada, Y. et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus erythematosus in Japanese. PLoS Genet. 8, e1002455 (2012).
59. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575 (2007).
60. Liu, E.Y., Li, M., Wang, W. & Li, Y. MacH-admix: genotype imputation for admixed populations. Genet. Epidemiol. 37, 25–37 (2013).
61. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119–124 (2012).
62. Liu, H. et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in IBD. Nat. Genet. 47, 267–271 (2015).
63. Verma, S.S. et al. Imputation and quality control steps for combining multiple genome-wide datasets. Front. Genet. 5, 370 (2014).
64. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. Annu. Rev. Genomics Hum. Genet. 10, 387–406 (2009).
65. Li, Y. & Abecasis, G.R. Mach 1.0: rapid haplotype reconstruction and missing genotype inference. Am. J. Hum. Genet. 79, 2290 (2006).
66. Song, M., Hao, W. & Storey, J.D. Testing for genetic associations in arbitrarily structured populations. Nat. Genet. 47, 550–554 (2015).
67. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. Nucleic Acids Res. 36, D422–D434 (2012).
68. Ward, L.D. & Kelis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 40, D930–D934 (2012).
69. Powell, J.C. et al. NGS tools comparison: metagene and METAL. Nucleic Acids Res. 42, D749–D755 (2014).
70. Liu, M.J., Wang, L.Y., Xia, Z., Sham, P.C. & Wang, J. GWA3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic Acids Res. 41, W150–W158 (2013).
71. Guo, L., Du, Y., Chang, S., Zhang, K. & Wang, J. sSNPbase: a database for curated regulatory SNPs. Nucleic Acids Res. 42, D1033–D1039 (2014).
72. Rouder, J.N. & Morey, R.D. Default Bayes factors for model selection in regression. Multivariate Behav. Res. 47, 877–903 (2012).
73. Wan, X. et al. BOOST: a fast approach to detecting gene-gene interactions in genome-wide case-control studies. Am. J. Hum. Genet. 87, 325–340 (2010).
74. Ueki, M. & Cordell, H.J. Improved statistics for genome-wide interaction analysis. PLoS Genet. 8, e1002625 (2012).
75. McLean, C.V. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 30, 2496–2497 (2012).
76. Slowikowski, K., Hu, X. & Raychaudhuri, S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. Bioinformatics 30, 496–497 (2014).
77. Su, A.I. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA 101, 6062–6067 (2004).
78. Hyatt, G. et al. Gene expression microarrays: glimpses of the immunological genome. Nat. Immunol. 7, 686–691 (2006).
79. Fantom Consortium and the RIKEN PMI and CLST (DGT). A promoter-mellar mammalian expression atlas. Nature 507, 462–470 (2014).
80. Risch, N. & Merikangas, K. The future of genetic studies of complex human diseases. Science 273, 1516–1517 (1996).
81. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat. Genet. 40, 204–210 (2008).
82. Zheng, W. et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum. Mol. Genet. 22, 2539–2550 (2013).
83. Hughes, T. et al. Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. Ann. Rheum. Dis. 71, 694–699 (2012).
84. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77 (2011).
85. DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845 (1988).
86. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. Science 338, 1532–1535 (2012).
87. Pickrell, J.K. et al. Genetic differentiation and recent positive selection in East Asians and African Americans. Genome Res. 22, 1529–1538 (2012).
88. International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN). et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat. Genet. 40, 204–210 (2008).
89. Zheng, W. et al. Common genetic determinants of breast-cancer risk in East Asian women: a collaborative study of 23 637 breast cancer cases and 25 579 controls. Hum. Mol. Genet. 22, 2539–2550 (2013).
90. Hughes, T. et al. Analysis of autosomal genes reveals gene-sex interactions and higher total genetic risk in men with systemic lupus erythematosus. Ann. Rheum. Dis. 71, 694–699 (2012).
91. Robin, X. et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 12, 77 (2011).
92. DeLong, E.R., DeLong, D.M. & Clarke-Pearson, D.L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 44, 837–845 (1988).
93. Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. A map of recent positive selection in the human genome. Science 338, 1532–1535 (2012).
94. Pickrell, J.K. et al. Genetic differentiation and recent positive selection in East Asians and African Americans. Genome Res. 22, 1529–1538 (2012).