Published in final edited form as:

Nat Genet. 2016 February 24; 48(3): 323–330. doi:10.1038/ng.3496.

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

Celi Sun[1,24], **Julio E. Molineros**[1,24], **Loren L. Looger**[2,24], **Xu-jie Zhou**[3,24], **Kwangwoo Kim**[4,24], **Yukinori Okada**[5,6], **Jianyang Ma**[7], **Yuan-yuan Qi**[3], **Xana Kim-Howard**[1], **Prasenjeet Motghare**[1], **Krishna Bhattarai**[1], **Adam Adler**[1], **So-Young Bang**[4], **Hye-Soon Lee**[4], **Tae-Hwan Kim**[4], **Young Mo Kang**[8], **Chang-Hee Suh**[9], **Won Tae Chung**[10], **Yong-Beom Park**[11], **Jung-Yoon Choe**[12], **Seung Cheol Shim**[13], **Yuta Kochi**[14], **Akari Suzuki**[14], **Michiaki Kubo**[15],

Address for Correspondence: Swapan K. Nath, Ph.D., ; Email: naths@omrf.org, Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, 1025 N.E. 13th Street, Oklahoma City, OK 73104, Phone: 405-271-7765 Fax: 405-271-4110 OR Sang-Cheol Bae, M.D., Ph.D., M.P.H., ; Email: scbae@hanyang.ac.kr, Hanyang University Hospital for Rheumatic Diseases, Seoul 133-792, Korea, Tel: 82-2-2290-9200, 9237, Fax: 82-2-2298-8231

[24]These authors contributed equally

URLs

GCTA v1.24: http://www.complextraitgenomics.com/software/gcta/
EIGENSTRAT: http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm
PLINK version 1.07: http://pngu.mgh.harvard.edu/~purcell/plink/
MACH-Admix: http://www.unc.edu/~yunmli/MaCH-Admix/
1000 Genomes Project: http://www.1000genomes.org/
Mach2dat: http://genome.sph.umich.edu/wiki/Mach2dat:_Association_with_MACH_output
METAL: http://www.sph.umich.edu/csg/abecasis/metal
HaploReg: http://www.broadinstitute.org/mammals/haploreg/haploreg.php
ENSEMBL: http://www.ensembl.org/
Blood eQTL: http://genenetwork.nl/bloodeqtlbrowser/
GWAS3D: http://jjwanglab.org/gwas3d
rSNPBase: http://rsnp.psych.ac.cn/
ENCODE Project: http://www.genome.gov/10005107
ChromHMM: http://compbio.mit.edu/ChromHMM/
DAPPLE v2.0: http://www.broadinstitute.org/mpg/dapple/dappleTMP.php
ConsensuspathDB: http://cpdb.molgen.mpg.de/
GREAT v2.0.2: http://bejerano.stanford.edu/great/public/html/
Enrichr: http://amp.pharm.mssm.edu/Enrichr/
GeneAtlas: http://biogps.gnf.org
ImmGen, http://www.immgen.org/
FANTOM5: http://fantom.gsc.riken.jp/5/
SNPSEA: http://www.broadinstitute.org/mpg/snpsea/
R version 3.0.0: http://www.r-project.org/
pROC: http://web.expasy.org/pROC/
Haplotter: http://haplotter.uchicago.edu/
HGDP Selection Browser: http://hgdp.uchicago.edu/cgibin/gbrowse/HGDP/
UCSC Genome Browser, http://genome.ucsc.edu;

AUTHORS CONTRIBUTIONS

S.K.N. and S-C.B. conceived and initiated the study. S.K.N. designed, coordinated and supervised the overall study. C.S., X-J.Z., P.M., K.B., A.A., and X.K-H prepared samples, performed genotyping, cleaned the data, combined various data sets and maintained the database. C.S., J.E.M., K.K. and Y. O. performed data imputation, association analysis, and various statistical analyses on the data. L.L.L., J.E.M., M.D. and J.D.W. performed the bioinformatic analysis. S-C.B., H.Z., K-H.C., X-J.Z., K.K., S-Y.B., H-S.L., T-H.K., Y.M.K., C-H.S., W.T.C., Y-B.P., J-Y.C., S.C.S., S-S.L., Y.J.K., B-G. H., ., Y. K., A. S., M. K., T.S., K.Y.,J. M., Y-Y. Q., K.M.K., J.B.H, and N, S. recruited and characterized SLE patients and controls, supplied the demographic and clinical data. C.S., J.E.M., X.K-H., K.K., S-C.B., L.L.L and S.K.N. drafted the manuscript. All authors approved the study, reviewed the manuscript, commented and helped revising the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Takayuki Sumida[16], **Kazuhiko Yamamoto**[14,17], **Shin-Seok Lee**[18], **Young Jin Kim**[19], **Bok-Ghee Han**[19], **Mikhail Dozmorov**[20], **Kenneth M. Kaufman**[21], **Jonathan D. Wren**[1], **John B. Harley**[21], **Nan Shen**[7,22], **Kek Heng Chua**[23], **Hong Zhang**[3], **Sang-Cheol Bae**[4], and **Swapan K. Nath**[1]

[1]Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA

[2]Howard Hughes Medical Institute, Janelia Research Campus, Ashburn, VA, USA

[3]Renal Division, Peking University First Hospital, Peking University Institute of Nephrology, Key Laboratory of Renal Disease, Ministry of Health of China, and Key Laboratory of Chronic Kidney Disease Prevention and Treatment (Peking University), Ministry of Education, Beijing, China

[4]Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul, Korea

[5]Department of Human Genetics and Disease Diversity, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo 113-8510, Japan

[6]Laboratory for Statistical Analysis, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan

[7]Joint Molecular Rheumatology Laboratory of the Institute of Health Sciences and Shanghai Renji Hospital, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, and Shanghai Jiaotong University School of Medicine, Shanghai 200025, China

[8]Kyungpook National University Hospital, Daegu, Korea

[9]Ajou University Hospital, Suwon, Korea

[10]Dong-A University Hospital, Department of Internal Medicine, Busan, Korea

[11]Department of Internal Medicine, Yonsei University College of Medicine, Seoul, Korea

[12]Daegu Catholic University Hospital, Daegu, Korea

[13]Daejeon Rheumatoid & Degenerative Arthritis Center, Chungnam National University Hospital, Daejeon, Korea

[14]Laboratory for Autoimmune Diseases, Center for Integrative Medical Sciences, RIKEN, Yokohama 230-0045, Japan

[15]Laboratory for Genotyping Development, RIKEN Center for Integrative Medical Sciences, Yokohama 230-0045, Japan

[16]Department of Internal Medicine, Faculty of Medicine, University of Tsukuba, Tsukuba 305-8575, Japan

[17]Department of Allergy and Rheumatology, Graduate School of Medicine, the University of Tokyo, Tokyo 113-0033, Japan

[18]Chonnam National University Hospital, Gwangju, Korea

[19]Korea National Institute of Health, Osong, Korea

[20]Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA

[21]Cincinnati Children's Hospital Medical Center and US Department of Veterans Affairs Medical Center, Cincinnati, OH, USA

[22]Center for Autoimmune Genomics and Etiology (CAGE), Cincinnati Children's Hospital Medical Center, Cincinnati Ohio 45229, USA

[23]Department of Biomedical Science, Faculty of Medicine, University of Malaya, 50603 Kuala Lumpur, Malaysia

Abstract

Systemic lupus erythematosus (SLE) has a strong but incompletely understood genetic architecture. We conducted an association study with replication in 4,492 SLE cases and 12,675 controls from six East-Asian cohorts, to identify novel and better localize known SLE susceptibility loci. We identified 10 novel loci as well as 20 known loci with genome-wide significance. Among the novel loci, the most significant was *GTF2IRD1-GTF2I* at 7q11.23 (rs73366469, $P_{meta}=3.75\times10^{-117}$, OR=2.38), followed by *DEF6*, *IL12B*, *TCF7*, *TERT, CD226*, *PCNXL3*, *RASGRP1, SYNGR1* and *SIGLEC6*. We localized the most likely functional variants for each locus by analyzing epigenetic marks and gene regulation data. Ten putative variants are known to alter *cis*- or *trans*-gene expression. Enrichment analysis highlights the importance of these loci in B- and T-cell biology. Together with previously known loci, the explained heritability of SLE increases to 24%. Novel loci share functional and ontological characteristics with previously reported loci, and are possible drug targets for SLE therapeutics.

SLE is a debilitating autoimmune disease (AID) 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 (*e.g.*, lupus nephritis)[1,2] compared to European-derived populations. SLE has a strong genetic component, *e.g.* a sibling risk ratio[3] (λ_s) of ~30, with ~40 susceptibility loci reported through candidate gene and genome-wide association studies (GWAS)[4–6]. However, only 8–15% of disease heritability[7,8] is accounted for, leaving many contributing loci unidentified. Since multiple susceptibility loci are shared among AIDs, and studying high-risk populations can facilitate novel risk locus identification, we performed high-density association analysis in East-Asians.

Our study was conducted in three stages (Fig. 1, Online Methods). First, we used ImmunoChip[9]-based association analysis in 2485 cases and 3947 controls from Koreans (KR), Han Chinese (HC) and Malaysian Chinese (MC), and identified 578 associated regions ($P<5\times10^{-3}$) (Supplementary Fig. 1, Supplementary Tables 1–3). To increase statistical power, we included 3669 out-of-study KR controls (Supplementary Table 1, Supplementary Fig. 2), and conducted imputation-based association analysis (Online Methods). Second, we followed up 16 novel loci with $P_{Discovery-meta}<5\times10^{-5}$ in three replication cohorts: one Japanese (JAP), and two independent Han Chinese from Beijing (BHC) and Shanghai (SHC). We identified 10 novel loci with genome-wide significance (GWS; $P_{meta}< 5\times10^{-8}$, Table 1, Figs. 2,3, Supplementary Fig. 3), and 6 novel suggestive loci (Supplementary Table 4). Third, we used a series of bioinformatic analyses including two recently developed Bayesian-based tests[10,11] (Online Methods) to identify the most likely

functional variants at each locus. Since the lead SNPs might not be functional, we examined SNPs in high linkage disequilibrium (LD; r^2>0.8). Variants were annotated using Encyclopedia of DNA Elements (ENCODE)[12] and Blood eQTL data[13]. We estimated the proportion of the heritability and sibling risk (λ_s) explained by novel and known SLE loci.

The strongest novel signal [$P_{meta}=3.75\times10^{-117}$, OR_{meta}(95% CI)=2.38 (2.22–2.56)] is at rs73366469 between two "general transcription factor" genes[14] *GTF2I* and *GTF2IRD1* (Supplementary Table 5). Surprisingly this signal is much stronger than human leukocyte antigen (HLA). Notably, rs117026326 within *GTF2I* (92kb from rs73366469) was recently identified as a major risk locus for primary Sjögren's syndrome (SS), another AID, in HC[15] and Southern Chinese[16]. Two recent SS GWAS[15,17] showed substantial overlap with SLE[18], emphasizing the validity and immune significance of this region. To confirm the veracity of this extraordinary 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 linked to rs73366469 ($r^2_{KR}=0.76$; $r^2_{BHC}=0.65$; $r^2_{SHC}=0.64$ in controls), making it difficult to separate their effects (Supplementary Table 6). Interestingly, conditional analysis on four SNPs showed that rs80346167 (*GTF2IRD1*) was independent in KR, supporting involvement of both genes. However, due to the strong correlation structure between variants, genotyping and fine-mapping at larger scale are required to further delineate this signal. ENCODE data indicate that high-LD SNPs rs7800325 (r^2 = 0.99) and in/del rs587608058 (r^2=0.81), ~1000bp from rs73366469, lie within conserved enhancers, active chromatin and transcription factor binding sites (TFBSs) 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 Tables 7,8, Supplementary Fig. 5).

The second strongest signal is at intronic rs10807150 (*DEF6*, $P_{meta}=6.06\times10^{-16}$) and correlated rs8205 (*ZNF76* promoter, r^2=1), a *cis*-eQTL altering expression of *ZNF76* and *DEF6* (Supplementary Tables 5,9). Nearby SNP rs4711414 (r^2=0.91) alters a highly conserved promoter/TFBS cluster (Supplementary Fig. 4b). The third strongest signal is near interleukin-12β (*IL12B*, rs2421184, $P_{meta}=4.67\times10^{-12}$), in a highly conserved enhancer (Supplementary Tables 5,7, Supplementary Fig. 4c).

Among the other novel signals, *TCF7* rs7726414 ($P_{meta}=1.13\times10^{-11}$) in the distal promoter is highly linked to rs6874758 (r^2=0.99), in a conserved enhancer (Supplementary Tables 5,7, Supplementary Fig. 4d). Nearby rs201806887 (r^2=0.79) alters a strong enhancer/TFBS cluster. The 5p15.33 signal is an oncogene[19] (*TERT*, intronic rs7726159, $P_{meta}=2.11\times10^{-11}$) tightly bound by RNA-binding proteins PABPC1 and SLBP (Supplementary Fig. 4e); high-LD rs7705526 (r^2=0.94) has been linked to chronic lymphocytic leukemia[20]. The *CD226* signal was explained by intronic rs1610555 ($P_{meta}=4.50\times10^{-11}$), linked ($r^2$= 0.74) to non-synonymous rs763361 (Supplementary Fig. 4f), associated with multiple AIDs. rs763361 is a *cis*-eQTL for *CD226* and also a *trans*-eQTL for *ACRBP* and *MAP3K7CL* (Supplementary Table 9). The signal at *PCNXL3* (rs2009453, $P_{meta}=9.61\times10^{-11}$) was 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 *RASGRP1* (rs12900339, $P_{meta}=4.73\times10^{-10}$) 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 (*SYNGR1,* $P_{meta}=5.73\times10^{-10}$), is in active chromatin (Supplementary Fig. 4i), and is a *cis*-eQTL of *SYNGR1* (Supplementary Table 9). Correlated SNP ($r^2=0.86$) rs909685 is associated with rheumatoid arthritis (RA) in Koreans[21]. Intronic rs2305772 (*SIGLEC6,* $P_{meta}=1.34\times10^{-9}$) 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 loci (Supplementary Table 10, Supplementary Fig. 6). Conditional analysis (Online Methods) at each locus identified secondary associations in 3 novel and 10 reported loci (Supplementary Table 11).

As expected, HLA association was replicated in all cohorts (Supplementary Table 10, Supplementary Fig. 7a). The strongest signal was at HLA Class II (rs113164910, $P_{Discovery\text{-}meta}=2.48\times10^{-37}$, OR=1.65), 14 kb 3′ of *HLA-DRA*. In order to further delineate the HLA signal, we imputed SNPs, classical HLA alleles and HLA amino acid residues in all three cohorts (Online Methods). The most significant association was identified at *HLA-DRB1* amino-acid position 13 ($P=9.5\times10^{-45}$) and its linked position 11 ($P = 7.37\times10^{-39}$), as shown in a recent HLA-fine-mapping study using a subset (~ 60%) of KR[22] (Supplementary Table 12). Our results also confirmed the reported associations of the two linked classical alleles, *HLA-DRB1**15:01 ($P=4.19\times10^{-29}$) and *HLA-DQB1**06:02 ($P=6.46\times10^{-26}$) (Supplementary Table 12; Supplementary Figure 7b). To investigate the secondary effect within and out of *HLA-DRB1*, we performed a conditional analysis. Consistent with the recent study[22], the associations of *HLA-DRB1* were almost explained by residues at amino-acid position 13 (and 11) with a primary effect and position 26 ($P=4.09\times10^{-17}$) as a secondary effect. After accounting for the effect of the *HLA-DRB1* locus (Online Methods), no new signals were detected. Thus the *DRB1* locus explained most of the MHC associations (Supplementary Table 12; Supplementary Figure 7c). Comparing SNP versus classical allele associations, we find that both association results co-locate the strongest effects towards the *HLA-DRB1* region (Supplementary Fig. 7b), as evidenced by *HLA-DRB1**15:01 and nearby rs113164910.

Additionally, we identified six novel “suggestive” loci ($1.9\times10^{-9}<P_{meta}<1.12\times10^{-5}$) with three missense variants (Supplementary Tables 4, 13). Although three of these loci (*ATG16L2-FCHSD2, MYNN-LRRC34* and *CCL22*) passed GWS, further replications are needed to confirm their association.

We replicated most of the previously reported genes with the same published or highly correlated SNPs (Supplementary Table 14). We also found four genes with novel uncorrelated SNPs shifting the association peaks in Asians (Supplemental Table 15). Of them, *ARHGAP31-TMEM39A-CD80* was of special interest: previously reported association signals from *TMEM39A* (rs1132200)[23] and *CD80* (rs6804441)[24] were now explained by a novel synonymous SNP in *ARHGAP31* (rs2305249, $P_{meta}=1.64\times10^{-9}$), a *cis*-eQTL of *B4GALT4* and *POGLUT1* (a *NOTCH1* signaling regulator[25]).

To identify the most likely functional variants within a locus, we used Bayesian-based analyses[10,11], eQTLs and epigenetic analyses (Online Methods, Supplementary Tables 7–9, 16). We found that lead SNPs in *GTF2I, IL12B, PCNXL3*, *SYNGR1, RASGRP1* and *SIGLEC6* had a high probability of being functional (Supplementary Table 17).

To explore biological functions and pathways related to SLE loci (novel and replicated), 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 novel and published loci. Moreover, GSEA with a drug target database[26] identified a set of 56 significantly enriched drugs (adjusted P-value <0.05, Supplementary Table 18), including SLE therapeutics[27] (cyclosporine, zinc acetate, hydrocortisone, methotrexate), that affected expression of the target loci. Of note was *GTF2I*, significantly enriched in drugs used for the treatment of leukemia (imatinib, $P_{adj}=1.82\times10^{-10}$) and lymphoma (cisplatin, $P_{adj}=2.68\times10^{-4}$). Immune system involvement was confirmed by enrichment analysis of SLE loci on mouse immune phenotypes, with significant enrichment in abnormal lymphocyte/leukocyte/immune cell physiology and abnormal cell-mediated/adaptive immunity (Supplementary Table 19).

To understand the relationship between our novel loci and known SLE loci, and to identify possible molecular mechanisms involved in SLE pathogenesis, we performed network interaction analsyis[28,29] (Online Methods). We found that SLE novel and replicated loci are connected directly and indirectly to each other through gene regulation, protein and biochemical interactions (Supplementary Fig. 9,10). Text-mining methods[30] confirmed that many of these loci have strong associations with one another in the literature, and show how the novel loci are related to the replicated loci (Supplementary Fig. 11). Within these relationships, we further identified sub-networks of molecules interacting with our novel loci in the context of known SLE genes (*i.e., TERT, IL12B*, *GTF2I, RELA, SRC* and *NFKB2* (Supplementary Fig. 12).

We identified only one non-synonymous variant (rs2305772, Pro246Ser/splice junction, *SIGLEC6*) in LD ($r^2\geq0.8$) with the novel SNPs (Supplementary Table 20), suggesting other variants likely contribute to SLE pathogenesis through epigenetic regulation, rather than protein structure/function alterations. Joint analysis of lead and correlated ($r^2>0.8$) SNPs indicated 13-fold enrichment in strong enhancers in K562 and up to 22-fold enrichment in DNase hypersensitivity in MCF-7 cells (Supplementary Table 21).

In six of the ten novel genes (*GTF2I, DEF6, CD226, PCNXL3*, *RASGRP1* and *SIGLEC6*), highly conserved, ancestral alleles were risk alleles. Except for *SIGLEC6*, all derived, protective alleles are major alleles in Asians and European-derived populations (CEU); in *SIGLEC6*, the derived, protective allele is major in only Europeans. Notably, derived risk alleles for *SYNGR1* occur at >80% in Asians (CHB+JPT), compared to CEU (~20%), 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 (novel and replicated) harbored genes expressed in distinct immune cell types[31] (Online Methods). We identified significant

($1\times10^{-9}<P<4\times10^{-4}$) cell type-specific expression of our loci in human B-cells, T-cells, natural killer cells and dendritic cells (Fig. 4, Supplementary Fig. 13a). This result was further strengthened by replication of homologous mouse genes in mouse cell lines, with significant enrichment in $CD19^+$ B-cells ($P=1.0\times10^{-5}$) and transitional B-cells ($P=1.0\times10^{-5}$) (Supplementary Fig. 13b). Thus, our results point to a strong (and conserved) effect of gene expression in B- and T-cells in SLE pathogenesis.

Six of the ten novel genes are also associated with other AIDs including celiac disease (CD), RA, T1D, and multiple sclerosis (MS) (Supplementary Table 23), suggesting pleiotropic effects. This pattern extended to suggestive *ATG16L2, PTPRC, UBAC2* and *RGS1*, which are reportedly associated with other AIDs.

Collectively, these novel and known SLE susceptibility variants (47 SNPs) explain 24% of total heritability of SLE in Asians (Supplementary Table 24). Among them, *HLA* explains 2%, and the 10 novel loci account for 6%. These loci also explain 24% of λ_s (Supplementary Table 25); novel loci explain 7%. To quantify the predictive effect of these variants, we estimated genetic risk through the weighted genetic risk score (wGRS). Novel risk alleles significantly ($P=6.58\times10^{-39}$) increased the wGRS area under the curve [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 10 novel SLE-predisposing loci, and validated association in 36 reported loci (often refining intervals). We pinpointed and annotated independently associated variants at each locus. Further analysis in additional populations and experimental validation in cultured and patient cell types (as previously performed[32–34]) will confirm 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.

Online Materials and 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 (QC) and preliminary association analysis to identify 578 regions with $P<5\times10^{-3}$. Then we increased the KR sample size with out-of-study controls and performed imputation-based meta-analysis to discover 16 novel regions with $P_{\text{Discovery-meta}}<5\times10^{-5}$. In the second stage, we followed up with these 16 novel regions, doing *in silico* replication on a Japanese (JAP) GWAS[59] data set and two independent replications on separate Beijing Han Chinese (BHC) and Shanghai Han Chinese (SHC) to identify 10 novel loci with $P_{\text{meta}}<5\times10^{-8}$. 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, Supplementary Note).

Imputation-based association, meta-analysis and conditional analysis

For the first stage of our study (Fig. 1), we performed single-SNP case-control association analysis based on QC'd ImmunoChip genotype data in each population. We calculated association P-values, standard error (SE), odds ratio (ORs) and 95% confidence intervals (95% CIs) using PLINK[60]. This identified 578 regions with $P<5\times10^{-3}$ in at least one Asian cohort for imputation (Supplementary Fig. 1, Supplementary Table 3). In order 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 $P<5\times10^{-3}$. Region size was defined by the length of the linkage disequilibrium (LD) region ($r^2>0.2$) with respect to the peak SNP. To avoid edge effects we extended a further 100 kilobases (kb) on each side for each region. The recursive algorithm to define imputed regions used the following steps:

- **a.** Find the peak SNP with minimal P value $\leq5\times10^{-3}$ in a region a(x,y) (the region starting with the whole chromosome (x=start position, y=end position)). If such a peak SNP exists, continue; otherwise, stop.
- **b.** Define imputation region d(u,v) = LD region ($r^2>0.2$ with peak SNP) ±100kb.
- **c.** If (x,u) exists, go to a(x,u) recursively (step a); if (v,y) exists, go to a(v,y) recursively (step a). Otherwise, stop.
- **d.** Collect all regions d(u,v) for final imputation.

For the second stage of our study (Fig. 1), we integrated additional GWAS data from KR out-of-study controls to increase both SNP density and statistical power. Since KR ImmunoChip and KR GWAS data sets are genotyped in two different platforms and their overlapping SNP number is less than the original SNP number from either KR ImmunoChip or KR GWAS data set, we imputed each set separately on its original number of real genotyping SNPs using MACH-Admix[61]. HC and MC ImmunoChip data sets were imputed separately as well following the KR IC protocol. We took 504 Asians (104 Japanese in Tokyo-JPT + 200 Han Chinese in Beijing-CHB + 200 Southern Han Chinese-CHS) from 1000 Genomes Project data (2013-05-02 1000G Phase 3 Integrated Release Version 5 Haplotypes) as the reference panel for imputations. All SNP names and strands for the three ImmunoChip and one out-of-study control datasets were aligned with the Asian reference panel (n=504) before those four datasets were imputed separately. This imputation strategy has been used by many earlier studies[62,63], and has also been recommended as a best practice by the eMERGE network[64].

After imputation, we performed strict QC on post-imputed SNPs. In addition to the QC steps described above ($P_{HWE}>0.0001$ 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. In order to take into account imputation uncertainty, we used mach2dat[65,66] for 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. Additionally, as a complementary analysis, we used a newly developed genotype-conditional association test (GCAT)[67] to confirm our PCA-corrected

associations, the results of which were very consistent (data not shown). We used METAL[68] 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 dataset (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 controls described above.

Finally we analyzed SLE-association in 152,918 post-imputation QC'd SNPs and identified 20,213 associated SNPs ($P_{Discovery\text{-}meta}<0.005$), from which we successfully replicated 36 SLE loci with $P_{Discovery\text{-}meta}<0.005$ (Supplementary Table 10) and identified 16 novel suggestive regions with $P_{Discovery\text{-}meta}<5\times10^{-5}$ for follow-up replication.

In order to test if any systematic bias was introduced by this imputation procedure, we also performed an association analysis of the lead SNPs between the controls (IC versus GWAS). We found no evidence of systematic bias introduced by the imputation and thus consider the imputation results sound (Supplementary Table 26).

We performed conditional analysis for 20 known SLE loci with genome-wide significance (GWS) and 10 novel regions with GWS after replication in the largest cohort (KR). Conditional analysis was iterative, starting with the top SNP with the lowest P-value as the first SNP to be conditioned upon; all subsequent significant SNPs after conditioning were added to the regression model as covariates until no SNP with $P<5\times10^{-5}$ remained. To ensure that SNPs were truly independent, SNPs in high LD ($r^2>0.3$ with the SNP being conditioned upon) were filtered out before the next iteration, and only the associated SNPs with $P<5\times10^{-5}$ entered conditional analysis.

Functional annotation of novel loci

In order 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[69] on data obtained from 1000 Genomes Phase 1, and ENSEMBL[70]. We surveyed allele-dependent gene expression regulation (*i.e.* expression quantitative trait loci, eQTLs) by querying the Blood eQTL[13] database (which houses the experimental meta-analysis from gene expression experiments performed on non-transformed peripheral blood samples of 5311 individuals of European descent and later replicated on 2775 individuals) for *cis*- and *trans*-eQTLs (Supplementary Table 9). The functional significance of independent SNPs from novel 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 novel regions using the Haploreg[69], GWAS3D[71], and rSNPBase[72] online tools. Haploreg[69] provides functional annotations for binding motifs and epigenetic marks. GWAS3D[71] 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, H3K27Ac), promoter regions, CTCF insulator marks and DNase hypersensitive sites (DHS). ChromHMM was used to predict histone states and chromatin interactions. In order to understand distal regulatory relationships among the novel loci, chromatin interactions between candidate loci were gathered from ChIA-PET and Hi-C data on 8 cell lines (K562, NB4, GM12878, $CD4^+$ T-cells, H1-hESC, IMR90, RWPE1, MCF-7), available through ENCODE. We reported data for lead SNPs with at least three ChIA-PET or Hi-C hits (Supplementary Table 8, Supplementary Fig. 5). Additionally, rSNPBase[72] provided putative functional SNPs with experimentally validated regulatory elements controlling transcriptional and post-transcriptional events.

Functional fine-mapping

In order 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 within the region[10]. Second, we used the Probabilistic Identification of Causal SNPs (PICS) algorithm[11], 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 novel imputation regions was implemented in the Bayes Factor (BF)[73] library in R. Henceforth, we estimated the posterior probability for each SNP, as well as the proportion of the total BF explained by each variant. We formed the 95%–99% credible sets as the cumulative proportion of the BF [10]. In order to assess how much of the effects could be explained by the credible sets, we annotated each candidate SNP with dbSNP functions (intron, missense, UTR, synonymous, intergenic), as well as epigenetic annotations (promoter, enhancer, DNase hypersensitivity, bound proteins, motifs, drivers disrupted, rSNP, LD-proxy of rSNP (r^2>0.8), proximal regulation, distal regulation, miRNA regulation, RNA-binding protein mediated regulation, eQTL).

We implemented the PICS method[11] 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 to be causal, given the strength of association, its linkage neighbourhood, as well as regulatory element annotations.

Gene-gene interaction

In order to identify gene-gene interaction, we performed logistic regression with an interaction term between all pairs of lead SNPs (Table 1) using PLINK. Both BOOST[74] and joint effects[75] methods were used to screen for SNP-SNP interactions. We used a significance threshold of 1×10^{-4}.

Network interactions

In order to investigate how our novel loci interact with other genes, we used curated network interactions using the Disease Association Protein-Protein Link Evaluator database (DAPPLE V2.0)[28]. We used a seed of all our novel loci (both left and right flanking genes were also used for intergenic signals) and 20,000 within-degree-node permutations. We chose to simplify our networks given the 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 ConsensusPathDB[29]. ConsensusPathDB scores the confidence level of protein interactions on a scale between 0 and 1, and aggregates 11 pathway databases for gene set enrichment analysis (GSEA). We chose interactions with high confidence score (Intscore >0.9). Additionally, we plotted all possible high-confidence interactions for all novel loci (Supplementary Figs. 9,10).

To investigate how our updated set of novel SLE loci were related to each other and to previously established loci, we used a literature mining-based approach, implemented in IRIDESCENT[30] (Supplementary Fig. 11). This approach identifies genes mentioned together in the same MEDLINE titles/abstracts (over 24 million currently) and weights their relevance based on relative frequencies of gene mention and gene-gene co-mention.

Gene set enrichment analysis (GSEA)

In order to identify if there were significant enrichments of our SLE (novel and replicated) loci as compared to reported SLE loci in human and mouse ontology, we performed gene set enrichment analysis using GREAT[76] (Supplementary Table 19). In order to compare interacting pathways and ontological properties of novel *versus* published SLE genes, we used ConsensusPathDB[29]. Additionally, in order to identify and compare drug perturbation signatures between novel and reported loci, we used the gene enrichment analysis software Enrichr[26] (Supplementary Table 18).

In order to test if there was bias in enrichment due to the choice of 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 ConsensusPathDB[29]. 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

In order to identify enrichment in cell type-specific expression of novel and replicated SLE loci (57 SNPs), we used a previously reported approach[31,77] described as follows. We used normalized expression data from 79 human cell types from GeneAtlas[78] (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)[79]. Additionally, we used cell-specific expression of the collection of 573 human cell samples from the FANTOM5[80] Project.

In this analysis, we extracted genes from the regions where SNPs correlated with the lead SNPs (Table 1; $r^2>0.5$), spanning between recombination hotspots. We used normalized cell-specific expression profiles of the extracted genes to identify which cell types significantly express SLE candidate genes. Specificity P-values were estimated based on the permutation of ranked expression levels for each locus (10^{10} permutations) using SNPSEA[77] (Fig. 4, Supplementary Fig. 13). P-values (blue bars) that passed the multiple testing threshold

(black line) show significant enrichment in SLE loci. Threshold lines are dependent on the number of categories present in each database, *i.e.*, for the 1751 GO categories possible, the significance threshold would be at 2×10^{-5}.

Explained heritability

We assessed the variance in liability (Vg) explained for each of our GWS SNPs using the liability threshold method[7]. We estimated Vg for novel, reported and HLA loci separately. We used the weighted risk allele frequency and meta-analysis OR 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 Sham[7]. In order to check the consistency of this heritability estimate, we also used the allele frequencies from each cohort, as well as the allele frequencies for HapMap/1000 Genome populations CHB and JPT.

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 (λ_s) under the multiplicative model[81]:

sibling relative risk

$$(\lambda_s)=\frac{ln(\lambda)}{ln(\lambda_0)},$$

where λ_0 is the overall sibling relative risk, assumed here to be ~30 (ref. [82]), with the relative sibling risk from each locus (λ) given by

$$\lambda=\frac{(pr^2+q)}{(pr+q)^2},$$

where p is the frequency of the risk allele (q=1−p) and r is the per-allele risk ratio[83].

Weighted cumulative genomic risk score

In order to assess the effect of accumulation of risk variants between cases and controls, we estimated the weighted cumulative genomic risk score (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 OR[84] for all 10 novel loci, 2 HLA loci and 35 replicated loci from a total of 2476 cases and 8426 controls. Significant differences in wGRS were estimated using a logistic regression model including gender 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 (AUC)

We estimated the predictive power of variants' wGRS, as well as the marginal contribution of the novel variants, by comparing AUC for the baseline model (including reported loci) *versus* the expanded model (including reported + novel loci) (Supplementary Fig. 14). AUC corrected for gender was estimated in R using the pROC library[85]. Confidence intervals for AUC were estimated using the nonparametric DeLong method[86].

Evidence for natural selection

To assess evidence for natural selection, we used HapMap2 and Human Genome Diversity Project (HGDP) population data through Haplotter and the HGDP Selection Browser. For each of the 10 novel genes we looked for evidence of positive natural selection in the 1 Mb region around each gene. Haplotter uses three statistics: iHS (Integrated Haplotype Score), F_{ST} (fixation index of population differentiation) and the empirical P-value for the distribution of Tajima's D and Fay's H[87], while the HGDP Selection Brower uses XP-EHH[88] (Cross Population Extended Haplotype Homozygosity) to identify positive natural selection in addition to iHS. Evidence of natural selection was considered positive if the empirical P-value<0.05 for the distribution of both Tajima's D and Fay's H, and $-\log(Q) > 3$ for Fst, D, iHS, or XP-EHH, where Q is the empirical P-values rank ordering the summary statistic value (a given region divided by total number of regions) (Supplementary Table 22).

Graphical display of the epigenetic landscape of the loci

For Supplementary Fig. 4, plots were assembled similarly to ref [33]. Most data were downloaded from the UCSC Genome Browser and displayed using custom MATLAB code. ATAC-Seq tracks for $CD4^+$ cells and GM12878 cells were downloaded from NCBI Gene Expression Omnibus accession number (GSE47753)[89]. DNase hypersensitivity, ENCODE sequence classification, histone marks and binding data for transcription factors (to the DNA) and RNA-binding proteins (to the RNA) were all downloaded from UCSC. ENCODE regulatory elements are color-coded according to their standard; other signals are shown in grey-scale, with dark representing higher signal. All tested SNPs are shown as bars of $-\log_{10}$ (P-value) height at the top. In the zoomed images, SNPs of interest are labeled.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

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 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 SLE patients were obtained from the Korean National Biobank at Wonkwang University Hospital that is supported by the Ministry of Health & Welfare, Republic of Korea.

This work was supported by grants from the US National Institutes of Health (AR060366, MD007909, AI103399, AI024717, AI083194, HG006828), the Department of Defense (PR094002), the U.S. Department of Veterans Affairs, the Research Fund of Beijing Municipal Science and Technology for the Outstanding PhD Program (20121000110), the National Natural Science Foundation of China (No. 81200524) and the High Impact Research MoE Grant UM.C/625/1/HIR/MoE/E000044-20001, Malaysia. This study was also supported by a grant from the

Korea Healthcare Technology R&D Project (HI12C1834, HI13C2124), Ministry for Health & Welfare, Republic of Korea.

References

1. Jakes RW, 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). 2012; 64:159–68. [PubMed: 22052624]
2. Danchenko N, Satia JA, Anthony MS. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus. 2006; 15:308–18. [PubMed: 16761508]
3. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2001; 2:802–9. [PubMed: 11526390]
4. Harley IT, Kaufman KM, Langefeld CD, Harley JB, Kelly JA. Genetic susceptibility to SLE: new insights from fine mapping and genome-wide association studies. Nat Rev Genet. 2009; 10:285–90. [PubMed: 19337289]
5. Boackle SA. Advances in lupus genetics. Curr Opin Rheumatol. 2013; 25:561–8. [PubMed: 23917156]
6. Yang W, et al. Meta-analysis followed by replication identifies loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as associated with systemic lupus erythematosus in Asians. Am J Hum Genet. 2013; 92:41–51. [PubMed: 23273568]
7. So HC, Gui AHS, Cherny SS, Sham PC. Evaluating the Heritability Explained by Known Susceptibility Variants: A Survey of Ten Complex Diseases. Genetic Epidemiology. 2011; 35:310–317. [PubMed: 21374718]
8. Gateva V, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nat Genet. 2009; 41:1228–33. [PubMed: 19838195]
9. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011; 13:101. [PubMed: 21345260]
10. Wellcome Trust Case Control, C. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet. 2012; 44:1294–301. [PubMed: 23104008]
11. Farh KK, et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature. 2015; 518:337–43. [PubMed: 25363779]
12. Consortium, E.P. et al. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57–74. [PubMed: 22955616]
13. Westra HJ, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013
14. Tantin D, Tussie-Luna MI, Roy AL, Sharp PA. Regulation of immunoglobulin promoter activity by TFII-I class transcription factors. J Biol Chem. 2004; 279:5460–9. [PubMed: 14645227]
15. Li Y, et al. A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjogren's syndrome at 7q11.23. Nat Genet. 2013; 45:1361–5. [PubMed: 24097066]
16. Zheng J, et al. The GTF2I rs117026326 polymorphism is associated with anti-SSA-positive primary Sjogren's syndrome. Rheumatology (Oxford). 2015; 54:562–4. [PubMed: 25480810]
17. Lessard CJ, et al. Variants at multiple loci implicated in both innate and adaptive immune responses are associated with Sjogren's syndrome. Nature Genetics. 2013; 45:1284. [PubMed: 24097067]
18. Perl A. Emerging new pathways of pathogenesis and targets for treatment in systemic lupus erythematosus and Sjogren's syndrome. Curr Opin Rheumatol. 2009; 21:443–7. [PubMed: 19584730]
19. Johnatty SE, et al. Evaluation of candidate stromal epithelial cross-talk genes identifies association between risk of serous ovarian cancer and TERT, a cancer susceptibility "hot-spot". PLoS Genet. 2010; 6:e1001016. [PubMed: 20628624]
20. Berndt SI, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat Genet. 2013; 45:868–76. [PubMed: 23770605]

21. Kim K, et al. High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. Ann Rheum Dis. 2014
22. Kim K, et al. The HLA-DRbeta1 amino acid positions 11-13-26 explain the majority of SLE-MHC associations. Nat Commun. 2014; 5:5902. [PubMed: 25533202]
23. Lessard CJ, et al. Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. Am J Hum Genet. 2012; 90:648–60. [PubMed: 22464253]
24. Yang W, et al. Meta-analysis Followed by Replication Identifies Loci in or near CDKN1B, TET3, CD80, DRAM1, and ARID5B as Associated with Systemic Lupus Erythematosus in Asians. Am J Hum Genet. 2013; 92:41–51. [PubMed: 23273568]
25. Chu Q, Liu L, Wang W. Overexpression of hCLP46 enhances Notch activation and regulates cell proliferation in a cell type-dependent manner. Cell Prolif. 2013; 46:254–62. [PubMed: 23692084]
26. Chen EY, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013; 14:128. [PubMed: 23586463]
27. Xiong W, Lahita RG. Pragmatic approaches to therapy for systemic lupus erythematosus. Nat Rev Rheumatol. 2014; 10:97–107. [PubMed: 24166241]
28. Trost B, Arsenault R, Griebel P, Napper S, Kusalik A. DAPPLE: a pipeline for the homology-based prediction of phosphorylation sites. Bioinformatics. 2013; 29:1693–5. [PubMed: 23658419]
29. Kamburov A, Stelzl U, Lehrach H, Herwig R. The ConsensusPathDB interaction database: 2013 update. Nucleic Acids Res. 2013; 41:D793–800. [PubMed: 23143270]
30. Wren JD, Bekeredjian R, Stewart JA, Shohet RV, Garner HR. Knowledge discovery by automated identification and ranking of implicit relationships. Bioinformatics. 2004; 20:389–98. [PubMed: 14960466]
31. Hu X, et al. Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. Am J Hum Genet. 2011; 89:496–506. [PubMed: 21963258]
32. Molineros JE, et al. Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production. PLoS Genet. 2013; 9:e1003222. [PubMed: 23441136]
33. Maiti AK, et al. Combined protein- and nucleic acid-level effects of rs1143679 (R77H), a lupus-predisposing variant within ITGAM. Hum Mol Genet. 2014; 23:4161–76. [PubMed: 24608226]
34. Guthridge JM, et al. Two functional lupus-associated BLK promoter variants control cell-type- and developmental-stage-specific transcription. Am J Hum Genet. 2014; 94:586–98. [PubMed: 24702955]
35. Vandeweyer G, Van der Aa N, Reyniers E, Kooy RF. The contribution of CLIP2 haploinsufficiency to the clinical manifestations of the Williams-Beuren syndrome. Am J Hum Genet. 2012; 90:1071–8. [PubMed: 22608712]
36. Howard ML, et al. Mutation of Gtf2ird1 from the Williams-Beuren syndrome critical region results in facial dysplasia, motor dysfunction, and altered vocalisations. Neurobiol Dis. 2012; 45:913–22. [PubMed: 22198572]
37. Antonell A, et al. Partial 7q11.23 deletions further implicate GTF2I and GTF2IRD1 as the main genes responsible for the Williams-Beuren syndrome neurocognitive profile. Journal of Medical Genetics. 2010; 47:312–320. [PubMed: 19897463]
38. Roy AL. Biochemistry and biology of the inducible multifunctional transcription factor TFII-I: 10 years later. Gene. 2012; 492:32–41. [PubMed: 22037610]
39. Malcolm T, Kam J, Pour PS, Sadowski I. Specific interaction of TFII-I with an upstream element on the HIV-1 LTR regulates induction of latent provirus. FEBS Lett. 2008; 582:3903–8. [PubMed: 18976654]
40. Gupta S, et al. T cell receptor engagement leads to the recruitment of IBP, a novel guanine nucleotide exchange factor, to the immunological synapse. J Biol Chem. 2003; 278:43541–9. [PubMed: 12923183]
41. Biswas PS, 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. 2012; 209:581–96. [PubMed: 22370718]

42. Noble JA, et al. A polymorphism in the TCF7 gene, C883A, is associated with type 1 diabetes. Diabetes. 2003; 52:1579–82. [PubMed: 12765974]
43. Klapper W, et al. Telomerase activity in B and T lymphocytes of patients with systemic lupus erythematosus. Ann Rheum Dis. 2004; 63:1681–3. [PubMed: 15547095]
44. Iguchi-Manaka A, et al. Accelerated tumor growth in mice deficient in DNAM-1 receptor. J Exp Med. 2008; 205:2959–64. [PubMed: 19029379]
45. Alcina A, et al. The autoimmune disease-associated KIF5A, CD226 and SH2B3 gene variants confer susceptibility for multiple sclerosis. Genes and Immunity. 2010; 11:439–445. [PubMed: 20508602]
46. Deshmukh HA, et al. Evaluation of 19 Autoimmune Disease-associated Loci with Rheumatoid Arthritis in a Colombian Population: Evidence for Replication and Gene-Gene Interaction. Journal of Rheumatology. 2011; 38:1866–1870. [PubMed: 21765104]
47. Hafler JP, et al. CD226 Gly307Ser association with multiple autoimmune diseases. Genes and Immunity. 2009; 10:5–10. [PubMed: 18971939]
48. Maiti AK, et al. Non-synonymous variant (Gly307Ser) in CD226 is associated with susceptibility to multiple autoimmune diseases. Rheumatology (Oxford). 2010; 49:1239–44. [PubMed: 20338887]
49. Qiu ZX, Zhang K, Qiu XS, Zhou M, Li WM. CD226 Gly307Ser association with multiple autoimmune diseases: A meta-analysis. Human Immunology. 2013; 74:249–255. [PubMed: 23073294]
50. Wieczorek S, et al. Novel association of the CD226 (DNAM-1) Gly307Ser polymorphism in Wegener's granulomatosis and confirmation for multiple sclerosis in German patients. Genes and Immunity. 2009; 10:591–595. [PubMed: 19536154]
51. Du Y, et al. Association of the CD226 single nucleotide polymorphism with systemic lupus erythematosus in the Chinese Han population. Tissue Antigens. 2011; 77:65–7. [PubMed: 20887380]
52. Stoeckman AK, et al. A distinct inflammatory gene expression profile in patients with psoriatic arthritis. Genes Immun. 2006; 7:583–91. [PubMed: 16971957]
53. Yasuda S, et al. Defective expression of Ras guanyl nucleotide-releasing protein 1 in a subset of patients with systemic lupus erythematosus. J Immunol. 2007; 179:4890–900. [PubMed: 17878389]
54. He CF, et al. TNIP1, SLC15A4, ETS1, RasGRP3 and IKZF1 are associated with clinical features of systemic lupus erythematosus in a Chinese Han population. Lupus. 2010; 19:1181–6. [PubMed: 20516000]
55. Iatropoulos P, et al. Association study and mutational screening of SYNGR1 as a candidate susceptibility gene for schizophrenia. Psychiatr Genet. 2009; 19:237–43. [PubMed: 19641478]
56. Liu JZ, et al. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat Genet. 2012; 44:1137–41. [PubMed: 22961000]
57. Gorski KS, et al. A set of genes selectively expressed in murine dendritic cells: utility of related cis-acting sequences for lentiviral gene transfer. Mol Immunol. 2003; 40:35–47. [PubMed: 12909129]
58. Patel N, et al. OB-BP1/Siglec-6. a leptin- and sialic acid-binding protein of the immunoglobulin superfamily. J Biol Chem. 1999; 274:22729–38. [PubMed: 10428856]
59. Okada Y, et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. PLoS Genet. 2012; 8:e1002455. [PubMed: 22291604]
60. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]
61. Liu EY, Li M, Wang W, Li Y. MaCH-admix: genotype imputation for admixed populations. Genet Epidemiol. 2013; 37:25–37. [PubMed: 23074066]
62. Jostins L, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature. 2012; 491:119–24. [PubMed: 23128233]
63. Liu H, et al. Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. Nat Genet. 2015; 47:267–71. [PubMed: 25642632]

64. Verma SS, et al. Imputation and quality control steps for combining multiple genome-wide datasets. Front Genet. 2014; 5:370. [PubMed: 25566314]
65. Li Y, Willer C, Sanna S, Abecasis G. Genotype imputation. Annu Rev Genomics Hum Genet. 2009; 10:387–406. [PubMed: 19715440]
66. Yang Li GA. Mach 1.0: Rapid Haplotype Reconstruction and Missing Genotype Inference. Am J Hum Genet. 2006; S79:2290.
67. Song M, Hao W, Storey JD. Testing for genetic associations in arbitrarily structured populations. Nat Genet. 2015
68. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–1. [PubMed: 20616382]
69. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic acids research. 2012; 40:D930–4. [PubMed: 22064851]
70. Flicek P, et al. Ensembl 2014. Nucleic Acids Res. 2014; 42:D749–55. [PubMed: 24316576]
71. Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: Detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. Nucleic Acids Res. 2013; 41:W150–8. [PubMed: 23723249]
72. Guo L, Du Y, Chang S, Zhang K, Wang J. rSNPBase: a database for curated regulatory SNPs. Nucleic Acids Res. 2014; 42:D1033–9. [PubMed: 24285297]
73. Rouder JN, Morey RD. Default Bayes Factors for Model Selection in Regression. Multivariate Behavioral Research. 2012; 47:877–903. [PubMed: 26735007]
74. Wan X, et al. BOOST: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. Am J Hum Genet. 2010; 87:325–40. [PubMed: 20817139]
75. Ueki M, Cordell HJ. Improved statistics for genome-wide interaction analysis. PLoS Genet. 2012; 8:e1002625. [PubMed: 22496670]
76. McLean CY, et al. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010; 28:495–501. [PubMed: 20436461]
77. Slowikowski K, Hu X, Raychaudhuri S. SNPsea: an algorithm to identify cell types, tissues and pathways affected by risk loci. Bioinformatics. 2014
78. Su AI, et al. A gene atlas of the mouse and human protein-encoding transcriptomes. Proc Natl Acad Sci U S A. 2004; 101:6062–7. [PubMed: 15075390]
79. Hyatt G, et al. Gene expression microarrays: glimpses of the immunological genome. Nat Immunol. 2006; 7:686–91. [PubMed: 16785882]
80. FANTOM Consortium and the RIKEN PMI and CLST (DGT) *et al.* A promoter-level mammalian expression atlas. Nature. 2014; 507:462–70. [PubMed: 24670764]
81. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science. 1996; 273:1516–7. [PubMed: 8801636]
82. International Consortium for Systemic Lupus Erythematosus, G. *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet. 2008; 40:204–10. [PubMed: 18204446]
83. 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. 2013; 22:2539–50. [PubMed: 23535825]
84. 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. 2012; 71:694–9. [PubMed: 22110124]
85. Robin X, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011; 12:77. [PubMed: 21414208]
86. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988; 44:837–45. [PubMed: 3203132]
87. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol. 2006; 4:e72. [PubMed: 16494531]

88. Pickrell JK, et al. Signals of recent positive selection in a worldwide sample of human populations. Genome Res. 2009; 19:826–37. [PubMed: 19307593]

89. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013; 10:1213–8. [PubMed: 24097267]

Box 1

Ten novel loci associated with SLE

***GTF2IRD1-GTF2I* (7q11.23)**: Our top signal was at rs73366469 in the intergenic region between critical "general transcription factor" (GTF) genes[34] *GTF2IRD1* and *GTF2I*. Both proteins are 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[35,36], as well as associated with SS. Deletion of this cytogenetic band reportedly alters craniofacial and neurocognitive characteristics[37]. Several studies also reported connections (especially for *GTF2I*) to transcriptional regulation induced in response to various signaling pathways, including immune response in both B-cells and T-cells[38,39].

***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-cells and T-cells[40]. DEF6 is implicated in autoimmunity through regulation of interferon regulatory factor 4 (*IRF4*) in interleukin-12 (IL-12) responsiveness[41].

***IL12B* (5q33.3)**: Interleukin 12B is a component of both IL12 (made in B-cells, macrophages, dendritic cells and neutrophils) and IL23 (macrophages and dendritic cells). Both interleukins are critical secreted signals in T-cell activation. Ustekinumab, a monoclonal antibody against IL12B, recognizes both IL12 and IL23; it is used in the treatment of psoriasis and is in testing for other AIDs.

***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 T1D risk[42]. A mouse *Tcf7* knockout showed reduced immunocompetence of T-cells in the periphery.

***TERT* (5p15.33)**: TERT (telomerase reverse transcriptase) plays a critical role in DNA replication and chromosomal stability, and is strongly associated with cancer[19]. Telomerase activity was dramatically upregulated in leukocytes from SLE patients, particularly in $CD19^{+}$ B-cells[43], and in untreated patients and patients with nephritis. Immunoglobin V(D)J recombination and telomere maintenance both function through non-homologous end joining, a core component of which is Ku70/80, first discovered as a lupus auto-antigen. 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[44]. It is a member of the Ig superfamily containing 2 Ig-like domains of the V-set, and is strongly associated with multiple AIDs[45–50]; previous associations with SLE[51] fell short of GWS.

***PCNXL3* (11q13.1)**: PCNXL3 (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 *PCNXL3* was one of the four most diagnostic genes for psoriatic arthritis (where it is down-regulated in symptomatic patients)[52].

***RASGRP1* (15q14)**: RAS guanyl releasing protein 1 (calcium and DAG-regulated) activates the Erk/MAP kinase cascade, which couples Ras to development, homeostasis and differentiation of T-cells and B-cells. *RASGRP1* was found to be down-regulated in symptomatic SLE patients[53]. The related gene *RASGRP3* has been associated with SLE and clinical features (discoid rash, malar rash and anti-nuclear antibodies) in a Han Chinese population[54].

***SYNGR1* (22q13.1)**: SYNGR1 (synaptogyrin1) has primary roles in neuronal synaptic transmission, is implicated in schizophrenia[55] and rheumatoid arthritis[21], and possibly primary biliary cirrhosis[56]. The related protein SYNGR2 is highly expressed in dendritic cells[57].

***SIGLEC6* (19q13.3)**: SIGLEC6 (Sialic acid binding Ig-like lectin 6) codes for a transmembrane receptor that binds leptin. SIGLEC6 mediates cell-cell adhesion by binding to glycans, and is expressed almost exclusively in the placenta and in B-cells[58].

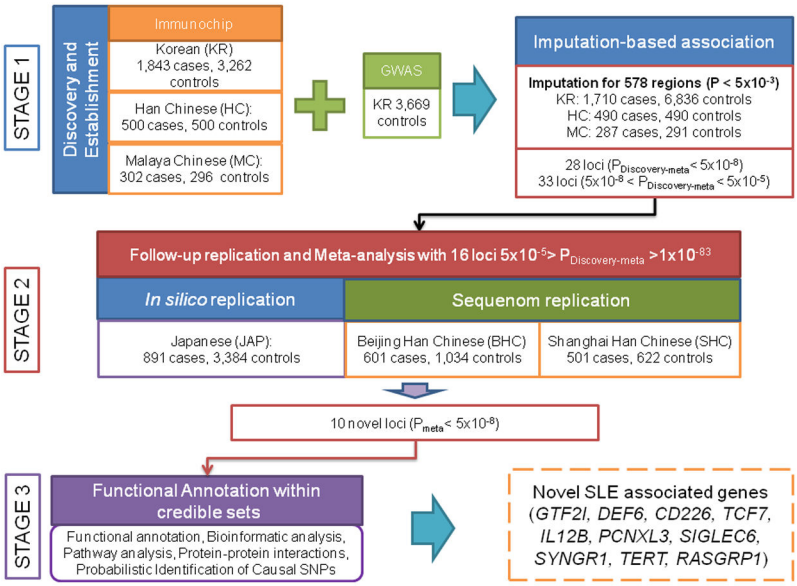

Figure 1.

Flowchart of our experimental design. This study followed three stages: in Stage 1, we genotyped three Asian cohorts of SLE patients and controls and identified 578 regions with $P<5\times10^{-3}$. Next we performed imputation-based fine-mapping, association tests and conditional analysis on the quality controlled data. We identified 16 statistically independent loci ($P<5\times10^{-5}$) for replication. In Stage 2, we performed an *in silico* replication of these 16 loci in an independent Japanese (JAP) cohort, and two independent Chinese cohorts from Shanghai (SHC) and Beijing (BHC). We identified novel regions represented by 10 replicated SNPs that passed the genome-wide significance threshold ($P < 5\times10^{-8}$). In Stage 3, we performed integrated functional and interaction analyses of SLE loci.

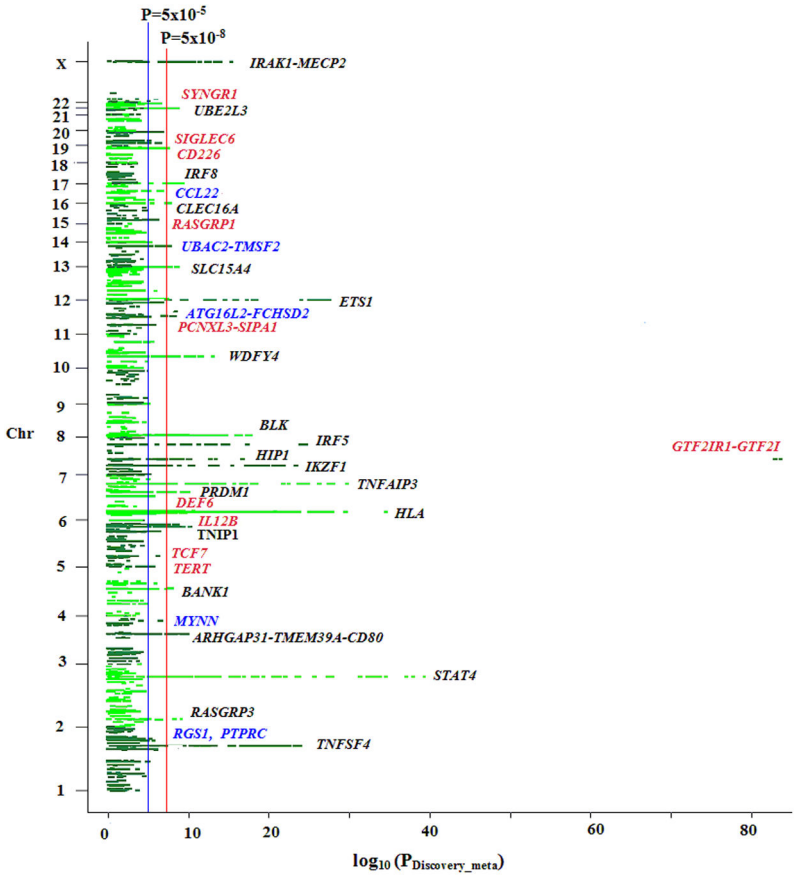

Figure 2.

Manhattan plot of the meta-analysis results using discovery sets. Novel significant loci are highlighted in red, “suggestive” loci are in blue and previously known SLE loci are in black.

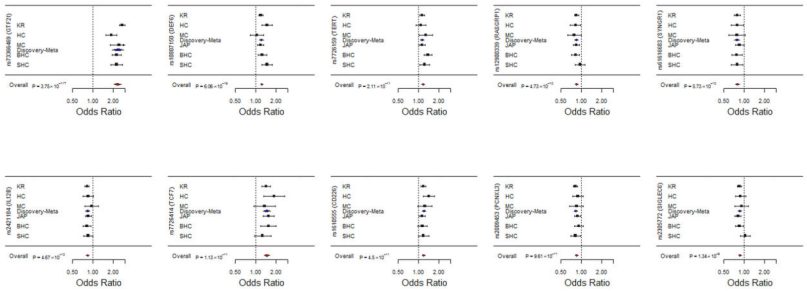

Fig. 3.

Meta-analysis of lead SNPs from 10 novel genes. We identified 10 novel loci in KR, HC and MC cohorts that were replicated in at least 2 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. KR: Korean; HC: Han Chinese; MC: Malaysian Chinese; JAP: Japanese; BHC: Beijing Han Chinese; SHC: Shanghai Han Chinese.

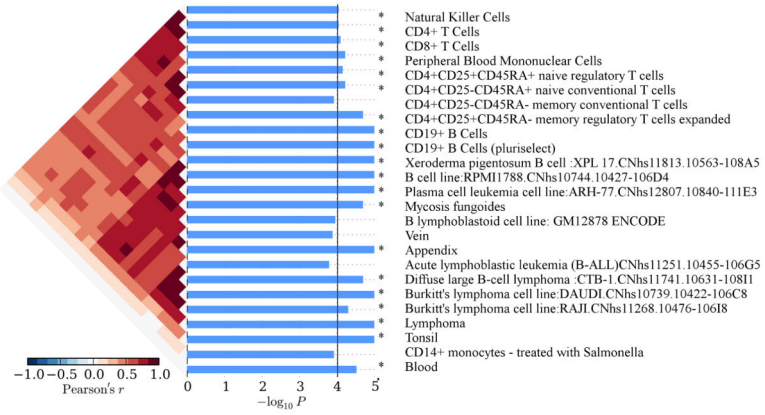

Fig. 4.

Cell-specific gene expression analysis of the novel and SLE loci. We estimated enrichment of our gene-set in a set of human (FANTOM5) cell lines. Overrepresented cell types have a high correlation (Pearson's correlation coefficient) of SLE-loci expression (dark red). P-values (blue bars) that passed the multiple testing threshold (black line) show significant enrichment in SLE loci.

Table 1

Meta-analysis results for novel and suggestive loci associated with SLE in Asian cohorts.

SNP (Alleles)	Gene	DISCOVERY COHORTS						REPLICATION COHORTS						OVERALL META ANALYSIS		
		KR 1710 cases, 6836 controls		HC 490 cases, 493 controls		MC 287 cases, 291 controls		JAP 891 cases, 3384 controls		SHC 501 cases, 622 controls		BHC 601 cases, 1034 controls		4480 cases, 12660 controls		
		MAF	P value	MAF	P value	MAF	P value	MAF	P value	MAF	P value	MAF	P value	P_{meta}	OR	CI
NOVEL SIGNALS																
rs73366469 (C/T)	*GTF2I*	0.29/0.13*	1.47×10^{-69}	0.36/0.23	1.71×10^{-9}	0.25/0.12	1.53×10^{-7}	NA	NA	0.30/0.16	1.17×10^{-16}	0.33/0.18	3.57×10^{-21}	3.75×10^{-117}	2.38	2.22–2.56
rs10807150 (C/T)	*DEF6*	0.36/0.32	9.21×10^{-6}	0.41/0.32	2.49×10^{-4}	0.37/0.36	5.81×10^{-1}	0.39/0.35	3.34×10^{-4}	0.41/0.32	$2.90\times10^{-}6$	0.38/0.33	1.36×10^{-3}	6.06×10^{-16}	1.25	1.19–1.32
rs2421184 (G/A)	*IL12B*	0.49/0.54	4.03×10^{-8}	0.49/0.53	1.54×10^{-1}	0.46/0.47	8.93×10^{-1}	0.47/0.51	1.35×10^{-3}	0.49/0.53	$4.69\times10^{-}2$	0.53/0.58	6.53×10^{-2}	4.67×10^{-12}	0.84	0.80–0.88
rs7726414 (T/C)	*TCF7*	0.07/0.05	1.20×10^{-4}	0.09/0.05	4.77×10^{-3}	0.13/0.10	8.17×10^{-2}	0.09/0.06	3.66×10^{-4}	0.10/0.08	$4.21\times10^{-}2$	0.09/0.06	1.19×10^{-2}	1.13×10^{-11}	1.40	1.27–1.54
rs7726159 (A/C)	*TERT*	0.40/0.37	2.10×10^{-4}	0.44/0.42	1.61×10^{-1}	0.48/0.42	3.70×10^{-2}	0.38/0.35	7.60×10^{-3}	0.46/0.41	$1.21\times10^{-}2$	0.49/0.41	6.92×10^{-6}	2.11×10^{-11}	1.21	1.14–1.28
rs1610555 (T/G)	*CD226*	0.43/0.39	2.19×10^{-5}	0.38/0.30	1.10×10^{-3}	0.37/0.32	7.16×10^{-2}	0.41/0.38	7.70×10^{-3}	0.38/0.34	$4.77\times10^{-}2$	0.34/0.31	3.86×10^{-2}	4.50×10^{-11}	1.19	1.13–1.26
rs2009453 (T/C)	*PCNXL3*	0.40/0.45	6.13×10^{-7}	0.42/0.45	2.08×10^{-1}	0.43/0.47	1.52×10^{-1}	0.37/0.40	2.56×10^{-3}	0.43/0.48	$2.43\times10^{-}2$	0.41/0.43	3.25×10^{-1}	9.61×10^{-11}	0.84	0.8–0.89
rs12900339 (G/A)	*RASGRP1*	0.40/0.44	5.85×10^{-5}	0.46/0.51	1.22×10^{-1}	0.46/0.52	3.32×10^{-2}	0.36/0.40	3.52×10^{-4}	0.51/0.52	$8.94\times10^{-}1$	0.45/0.50	9.00×10^{-3}	4.73×10^{-10}	1.18	1.12–1.24
rs61616683 (C/T)	*SYNGR1*	0.13/0.16*	4.50×10^{-4}	0.20/0.24	2.00×10^{-2}	0.19/0.23	7.00×10^{-2}	0.14/0.16	2.00×10^{-2}	0.20/0.24	$1.30\times10^{-}2$	0.18/0.22	4.00×10^{-3}	5.73×10^{-10}	0.79	0.73–0.85
rs2305772 (A/G)	*SIGLEC6*	0.42/0.46	2.33×10^{-6}	0.41/0.44	2.22×10^{-1}	0.39/0.41	5.17×10^{-1}	0.40/0.45	3.90×10^{-4}	0.42/0.41	$6.97\times10^{-}1$	0.42/0.46	3.40×10^{-2}	1.34×10^{-9}	0.86	0.81–0.9
SUGGESTIVE SIGNALS																
rs11235604 (T/C)	*ATG16L2-FCHSD2*	0.07/0.10	2.69×10^{-9}	0.09/0.1	5.19×10^{-1}	0.09/0.12	1.71×10^{-1}	0.07/0.09	2.45×10^{-2}	0.08/0.09	6.98×10^{-1}	0.09/0.10	5.87×10^{-1}	1.90×10^{-9}	0.76	0.69–0.84
rs10936599 (G/A)	*MYNN-LRRC34*	0.44/0.39	5.39×10^{-7}	0.49/0.46	1.15×10^{-1}	0.46/0.48	3.16×10^{-1}	0.39/0.36	1.49×10^{-2}	0.49/0.46	9.19×10^{-2}	0.48/0.46	1.27×10^{-1}	4.93×10^{-9}	0.87	0.82–0.91
rs223881 (C/T)	*CCL22*	0.46/0.49	9.31×10^{-5}	0.46/0.53	1.29×10^{-3}	0.51/0.59	7.23×10^{-3}	0.36/0.40	2.60×10^{-3}	0.54/0.53	9.08×10^{-1}	0.50/0.51	5.94×10^{-1}	2.06×10^{-8}	0.87	0.84–0.93
rs1885889 (A/G)	*UBAC2*	0.41/0.36	6.34×10^{-7}	0.40/0.37	1.38×10^{-1}	0.43/0.36	1.43×10^{-2}	0.43/0.41	2.25×10^{-1}	0.37/0.38	8.53×10^{-1}	0.38/0.37	4.47×10^{-1}	3.17×10^{-7}	1.14	1.08–1.2
rs7556469 (G/A)	*PTPRC*	0.12/0.15	2.33×10^{-6}	0.11/0.14	1.45×10^{-1}	0.11/0.10	4.94×10^{-1}	0.17/0.17	3.86×10^{-1}	0.12/0.14	1.59×10^{-1}	0.13/0.15	1.85×10^{-1}	1.95×10^{-6}	0.84	0.78–0.90
rs12022418 (C/A)	*RGS1*	0.32/0.36	1.53×10^{-4}	0.29/0.32	1.90×10^{-1}	0.31/0.38	1.26×10^{-2}	0.35/0.37	1.95×10^{-1}	0.32/0.35	9.56×10^{-2}	0.33/0.32	6.82×10^{-1}	1.12×10^{-5}	0.88	0.83–0.93

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

*Only from ImmunoChip controls.

***GTF2IRD1-GTF2I* (7q11.23)**: Our top signal was at rs73366469 in the intergenic region between critical "general transcription factor" (GTF) genes[14] *GTF2IRD1* and *GTF2I*. Both proteins are 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[35,36], as well as associated with SS. Deletion of this cytogenetic band reportedly alters craniofacial and neurocognitive characteristics[37]. Several studies also reported connections (especially for *GTF2I*) to transcriptional regulation induced in response to various signaling pathways, including immune response in both B-cells and T-cells[38,39].

***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-cells and T-cells[40]. DEF6 is implicated in autoimmunity through regulation of interferon regulatory factor 4 (*IRF4*) in interleukin-12 (IL-12) responsiveness[41].

***IL12B* (5q33.3)**: Interleukin 12B is a component of both IL12 (made in B-cells, macrophages, dendritic cells and neutrophils) and IL23 (macrophages and dendritic cells). Both interleukins are critical secreted signals in T-cell activation. Ustekinumab, a monoclonal antibody against IL12B, recognizes both IL12 and IL23; it is used in the treatment of psoriasis and is in testing for other AIDs.

***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 T1D risk[42]. A mouse *Tcf7* knockout showed reduced immunocompetence of T-cells in the periphery.

***TERT* (5p15.33)**: TERT (telomerase reverse transcriptase) plays a critical role in DNA replication and chromosomal stability, and is strongly associated with cancer[19]. Telomerase activity was dramatically upregulated in leukocytes from SLE patients, particularly in $CD19^{+}$ B-cells[43], and in untreated patients and patients with nephritis. Immunoglobin V(D)J recombination and telomere maintenance both function through non-homologous end joining, a core component of which is Ku70/80, first discovered as a lupus auto-antigen. 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[44]. It is a member of the Ig superfamily containing 2 Ig-like domains of the V-set, and is strongly associated with multiple AIDs[45–50]; previous associations with SLE[51] fell short of GWS.

***PCNXL3* (11q13.1)**: PCNXL3 (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 *PCNXL3* was one of the four most diagnostic genes for psoriatic arthritis (where it is down-regulated in symptomatic patients)[52].

***RASGRP1* (15q14)**: RAS guanyl releasing protein 1 (calcium and DAG-regulated) activates the Erk/MAP kinase cascade, which couples Ras to development, homeostasis and differentiation of T-cells and B-cells. *RASGRP1* was found to be down-regulated in symptomatic SLE patients[53]. The related gene *RASGRP3* has been associated with SLE and clinical features (discoid rash, malar rash and anti-nuclear antibodies) in a Han Chinese population[54].

***SYNGR1* (22q13.1)**: SYNGR1 (synaptogyrin1) has primary roles in neuronal synaptic transmission, is implicated in schizophrenia[55] and rheumatoid arthritis[21], and possibly primary biliary cirrhosis[56]. The related protein SYNGR2 is highly expressed in dendritic cells[57].

***SIGLEC6* (19q13.3)**: SIGLEC6 (Sialic acid binding Ig-like lectin 6) codes for a transmembrane receptor that binds leptin. SIGLEC6 mediates cell-cell adhesion by binding to glycans, and is expressed almost exclusively in the placenta and in B-cells[58].