HHS Public Access

Author manuscript

*Arthritis Rheumatol*. Author manuscript; available in PMC 2017 January 01.

Published in final edited form as:

*Arthritis Rheumatol*. 2016 January ; 68(1): 174–183. doi:10.1002/art.39403.

# Identification of a New Susceptibility Locus for Systemic Lupus Erythematosus on Chromosome 12 in Individuals of European Ancestry

**F. Yesim Demirci**[1], **Xingbin Wang**[1], **Jennifer A. Kelly**[2], **David L. Morris**[3], **M. Michael Barmada**[1], **Eleanor Feingold**[1], **Amy H. Kao**[4], **Kathy L. Sivils**[2], **Sasha Bernatsky**[5], **Christian Pineau**[5], **Ann Clarke**[6], **Rosalind Ramsey-Goldman**[7], **Timothy J. Vyse**[3], **Patrick M. Gaffney**[2], **Susan Manzi**[4], and **M. Ilyas Kamboh**[1]

[1]Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA

[2]Arthritis & Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA

[3]Department of Medical & Molecular Genetics, King's College London, Guy's Hospital, London SE1 9RT, UK

[4]Lupus Center of Excellence, Department of Medicine, Allegheny Health Network, Pittsburgh, PA 15224, USA

[5]Division of Rheumatology, Department of Medicine, McGill University, Montreal, QC H3G 1A4, Canada

[6]Division of Rheumatology, Department of Medicine, University of Calgary, Calgary, AB T2N 4Z6, Canada

[7]Division of Rheumatology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA

## Abstract

**Objective—**Genome-wide association studies (GWASs) in individuals of European ancestry identified a number of systemic lupus erythematosus (SLE) susceptibility loci using earlier versions of high-density genotyping platforms. Follow-up studies on suggestive GWAS regions using larger samples and more markers identified additional SLE loci in European-descent subjects. Here we report the results of a multi-stage study that we performed to identify novel SLE loci.

**Methods—**In Stage 1, we conducted a new GWAS of SLE in a North American case-control sample of European ancestry (*n*=1,166) genotyped on Affymetrix Genome-Wide Human SNP Array 6.0. In Stage 2, we further investigated top new suggestive GWAS hits by *in silico*

CORRESPONDENCE: F. Yesim Demirci, MD (fyd1@pitt.edu, phone: 412-383-7193, fax: 412-383-7844) or M. Ilyas Kamboh, PhD (kamboh@pitt.edu, phone: 412-624-3066, fax: 412-624-3020), Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA.

CONFLICT OF INTEREST STATEMENT: The authors have no conflict of interest to disclose.

evaluation and meta-analysis using an additional dataset of European-descent subjects (>2,500 individuals), followed by replication of top meta-analysis findings in another dataset of European-descent subjects (>10,000 individuals) in Stage 3.

**Results—**As expected, our GWAS revealed most significant associations at the major histocompatibility complex locus (6p21), which easily surpassed genome-wide significance threshold ($P<5\times10^{-8}$). Several other SLE signals/loci previously implicated in Caucasians and/or Asians were also supported in Stage 1 discovery sample and strongest signals were observed at 2q32/*STAT4* ($P=3.6\times10^{-7}$) and at 8p23/*BLK* ($P=8.1\times10^{-6}$). Stage 2 meta-analyses identified a new genome-wide significant SLE locus at 12q12 (meta $P=3.1\times10^{-8}$), which was replicated in Stage 3.

**Conclusion—**Our multi-stage study identified and replicated a new SLE locus that warrants further follow-up in additional studies. Publicly available databases suggest that this new SLE signal falls within a functionally relevant genomic region and near biologically important genes.

Systemic lupus erythematosus (SLE) is a prototypic multisystem autoimmune disease that is influenced by a complex interplay of heritable, hormonal, and environmental factors. Genetic basis of SLE was initially suggested by heritability estimates (~66%), familial clustering (sibling recurrence risk ratio: ~30), and twin studies (~10 times higher concordance rate in monozygotic vs. dizygotic twins) and subsequently supported by a growing number of susceptibility loci identified by association studies (1–5).

SLE is clinically highly heterogeneous and variable in disease course and prognosis. Underlying immunopathologic mechanisms also appear to be highly heterogeneous, involving various abnormalities in both innate and adaptive immune responses. Therefore, a large number of genetic factors are expected to influence SLE risk and several susceptibility loci have been identified to date through candidate gene and/or genome-wide association (GWA) studies (3–5).

GWA studies (GWASs) in European-descent subjects identified a number of new SLE loci and also confirmed some previously known loci, using earlier versions of high-density GWAS platforms. Three high-density GWASs of SLE were initially published in European-descent subjects (6–8) using the following platforms: Illumina HumanHap550 BeadChip (1311 cases, 1783 controls, and public data from 1557 controls), Illumina HumanHap300 BeadChip (720 cases, 475 controls, and public data from 1862 controls) and Affymetrix Human SNP Array 5.0 (431 cases and public data from 2155 controls). Follow-up studies on suggestive GWAS regions using larger samples and more markers identified additional SLE loci in European-derived populations (9–15). Although the data from initial GWASs have been extensively used for follow-up studies and additional analyses, no other ‘new’ high-density GWAS of SLE in European-derived populations (using new or extended samples genotyped on new platforms) has appeared in PubMed since 2008; however, a number of them were published from Asian populations (16–20) and the publication of a large European GWAS is underway (21, 22). Despite the tremendous progress, a substantial portion of genetic component of SLE still remains to be uncovered by additional studies/discoveries.

Given the heterogeneous nature of SLE and coverage differences between GWAS platforms, investigation of different and/or larger samples using different and/or higher density platforms is likely to increase the chance of discovering additional loci, as already exemplified by initial GWASs (6–8). Here we report the results of a multi-stage study that involved (1) a new GWAS of SLE in a North American case-control discovery sample of European ancestry ($n$=1,166) genotyped on Affymetrix Genome-Wide Human SNP Array 6.0 (Affy 6.0), (2) follow-up evaluation and meta-analysis of top new findings using an additional dataset of European-descent subjects (>2,500 individuals), and (3) replication of the most promising meta-analysis findings in another dataset of European-descent subjects (>10,000 individuals). Several previously reported SLE loci were supported in Stage 1 discovery sample, and Stage 2 meta-analyses implicated a new genome-wide significant locus at 12q12 (meta $P=3.1\times10^{-8}$), which was replicated in Stage 3.

## MATERIALS AND METHODS

### Stage 1 discovery sample for GWAS

A total of 1,166 European-descent subjects (676 SLE cases and 490 controls) were included in GWAS, after excluding subjects with cryptic relationship (one from each pair of duplicates or close relatives) from ~1,200 individuals initially genotyped. Participants were recruited at three sites [Pittsburgh ($n$=750), Chicago ($n$=204), and Montreal ($n$=212)] and all patients met 1982 or revised 1997 American College of Rheumatology classification criteria for SLE (23, 24) as determined by treating rheumatologist at each site. SLE cases were 18 years or older (mean age 45.1 ± 12.5 years; 97.3% women) and the controls were 21 years or older (mean age 48.9 ± 10.6 years; 100% women). Detailed information on these samples can be found elsewhere (25–28). All subjects provided written informed consent for genetic studies that were approved by the Institutional Review Board at each participating center.

### GWAS genotyping

DNA isolation was performed in the same lab (University of Pittsburgh Human Genetics Department) for all samples utilized in discovery stage for genotyping (at Expression Analysis, Durham, NC) on Affy 6.0 (containing probes for 906,600 SNPs). After applying the marker map corresponding to human genome build 37 (GRCh37/hg19), a total of 905,420 successfully mapped SNPs (excluding duplicates) were advanced into quality control (QC) filtering process prior to GWA analysis.

### GWAS QC filters

Samples with poor performance (15 with <95% average call rate across the array) and poor-quality markers (44,507 with <95% call rate across all samples) were removed from analysis. Markers with significant deviation from Hardy-Weinberg equilibrium in controls ($P\leq1\times10^{-6}$) and/or with low minor allele frequency (MAF<0.01) were also excluded ($n$=133,396). Population stratification analysis was conducted using 691,565 markers and a multi-dimensional scaling method implemented in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/). SNPs falling within genomic regions exhibiting abnormal linkage disequilibrium (LD) patterns and structural variations (chr 6: 24–38 Mb, chr 8: 7–13 Mb, chr 11: 42–58 Mb, and chr 17: 41–45 Mb) and those located on Y chromosome were excluded

from the calculation of principal components (PCs). First four components were determined to be relevant for controlling population substructure. Three samples were removed because they were extreme population outliers based on multi-dimensional scaling plots.

After applying above QC filters, a final set of 1,148 subjects (661 cases and 487 controls) and 727,517 markers were advanced into association analysis for SLE. While the entire QC-passed autosomal marker data ($n$=700,598) were assessed for associations at previously reported SLE genes/loci, only the results from 627,076 common (MAF≥0.05) autosomal markers were used to identify potentially new SLE signals/loci.

### GWAS analysis

Logistic regression analysis of the effects of genotypes on SLE risk was conducted under an additive model that included recruitment site, sex, age, and first 4 PCs as covariates. All analyses were performed in PLINK. Pairwise LD plots were created for selected markers by analyzing discovery sample data using Haploview (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview).

### Stage 2 dataset for follow-up investigation of suggestive GWAS hits (from Stage 1) and Meta-analyses of Stage 1 and 2 datasets for relevant markers

Data from an independent high-density GWAS of SLE in European-descent subjects (8) was used for follow-up investigation of new suggestive autosomal GWAS signals (from Stage 1) located outside the extended major histocompatibility complex (xMHC) region. This Stage 2 dataset included 431 SLE cases (97% women) and 2,155 database controls (50% women) successfully genotyped for 311,238 SNPs on Affy 5.0 (details can be found in original publication) (8). A total of 558 common markers were evaluated in this sample, which included index or proxy SNPs for new suggestive autosomal signals ($P$<0.001) that were present in this dataset. The SNAP tool (https://www.broadinstitute.org/mpg/snap/ldsearch.php) was utilized to determine proxy SNPs, using an $r^2$ threshold of 0.8 and 1000 genomes pilot and HapMap 3 data from CEU population as references. Initial meta-analysis of Stage 1 and 2 datasets for 558 SNPs identified 26 potentially new SLE loci harboring SNPs with meta $P$<0.001 and same direction of association in both datasets. Next, these 26 loci were more extensively examined in Stage 2 dataset by evaluating all relevant SNPs (567 common SNPs with $P$<0.05 in Stage 1 sample falling within/near potentially new signals). Data was available for 306 of 567 SNPs in Stage 2 dataset and thus final meta-analysis of Stage 1 and 2 datasets included 306 SNPs from 26 loci. Meta-analyses were performed using the METAL software (29).

### Stage 3 dataset for replication of findings from meta-analyses of Stage 1 and 2 datasets

A genome-wide significant locus identified by meta-analyses of Stage 1 and 2 datasets was further tested for replication in the data from a large European high-density GWAS of SLE (Stage 3 dataset) (21, 22). A total of 8 SNPs from this locus (with meta $P$<0.001 and same direction of association in Stage 1 and 2 samples) were evaluated in Stage 3 sample, which consisted of 4,036 SLE patients (90% women) and 6,959 controls (50% women; 5,699 from the NIH Health and Retirement Study) successfully genotyped for 644,674 markers on HumanOmni1-Quad BeadChip. This data was imputed to the density of 1000 genomes

project using IMPUTE V2.2.3 (30) and analyzed for association under an additive model computed by SNPTEST (30) with first 4 PCs used as covariates to account for European population substructure.

### In silico assessment of the regulatory potential of relevant variants

Relevant variants were evaluated for their regulatory potential by using RegulomeDB v1.1 (http://www.regulomedb.org/), a public database for human DNA features and regulatory elements in non-coding genomic regions. RegulomeDB is a powerful tool that uses various datasets (high-throughput experimental data from ENCODE project and other resources as well as computational and manual annotations) to score non-coding variants for their putative functional relevance. The RegulomeDB scoring scheme is provided in Table 1 footnote.

### In silico assessment of gene expression profiles and association patterns with expression quantitative trait loci (eQTLs) and methylation quantitative trait loci (mQTLs)

*In silico* assessment of expression profiles of genes of interest was performed using BioGPS gene annotation portal (http://biogps.org/#goto=welcome). The Genevar (GENe Expression VARiation) database v3.3.0 (http://www.sanger.ac.uk/resources/software/genevar/) and data from Multiple Tissue Human Expression Resource (MuTHER) (31–33) were utilized to visualize association patterns with eQTLs and mQTLs located within genomic region(s) of interest. In addition, an SQL database (ghs_probe_express030510) (34) compiling the results of genome-wide SNP associations with human monocyte expression traits (http://genecanvas.ecgene.net/uploads/ForReview/) was used to evaluate most relevant *cis* and *trans* eQTL effects of SNPs of interest.

### In silico enhancer enrichment analysis

HaploReg v2 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) is a functional annotation tool for non-coding variants on haplotype blocks (based on LD calculations on 1000 Genomes Phase 1 individuals; for our analysis we used the default settings that included EUR population and LD threshold of $r^2$=0.8) and it includes an extensive library of SNPs, motif instances, enhancer annotations, and eQTLs. HaploReg v2 also conducts enhancer enrichment analysis by comparing the coverage of all enhancers and strongest enhancers calculated for queried variants to that of all variants. When the coverage exceeds that of selected background set (for our analysis we selected all SNPs in 1KG CEU pilot), the enrichment is reported if nominally significant based on a binomial test.

## RESULTS

### GWAS of SLE in a North American discovery sample of European ancestry (Stage 1)

Quantile–quantile plot of observed vs. expected p-values under the null hypothesis is presented in Figure S1 for the GWAS of Stage 1 discovery sample. Manhattan plot summarizing association results for QC-passed autosomal common markers is presented in Figure S2, including over 1,000 markers with $P<0.001$. As expected, strongest associations were observed at the established MHC locus (6p21), which easily surpassed the stringent genome-wide significance threshold ($P<5\times10^{-8}$) in our discovery sample. About 200

markers from the xMHC region at 6p22–21 had $P<0.001$ and more than 600 had $P<0.05$. The most significant MHC SNP was rs2187668 in *HLA-DQA1* ($P=1.2\times10^{-9}$), followed by rs3129716 near *HLA-DQB1* ($P=4.2\times10^{-9}$), rs1150753 in *TNXB* ($P=5.6\times10^{-9}$), rs3131379 in *MSH5* ($P=7.6\times10^{-8}$), and rs9267531 in *CSNK2B* ($P=8.2\times10^{-8}$). Three of these SNPs (rs2187668, rs3129716, and rs3131379) showed the RegulomeDB score of "1f", indicating a strong regulatory potential.

Next we examined our Stage 1 data for 38 non-MHC autosomal SLE loci, where at least one common variant was previously shown to reach genome-wide significance in reported discovery and/or combined (discovery+replication) samples in GWASs or post-GWAS large studies of European-descent or Asian subjects (6–20, 35, 36) (see Table S1 footnote). Twenty-nine of these loci showed nominal associations ($P<0.05$) in our discovery sample, of which 10 were supported at the SNP level [by the same previously reported SNP and/or a proxy SNP ($r^2\geq0.8$ in 1000 genomes pilot and/or HapMap 3 data from CEU population)] and 19 at the locus level [by other SNP(s) located within the gene(s) of interest ± 50 kb]. The top ones were the 2q32.3 locus (*STAT4*/rs7582694, $P=3.6\times10^{-7}$) and the 8p23.1 locus (*BLK*/rs1478897, $P=8.1\times10^{-6}$, proxy for rs2248932), both of which were previously implicated in more than one high-density GWAS of European-descent subjects. The full list of nominally associated SNPs from 29 non-MHC autosomal SLE loci supported in our discovery sample is provided in Table S1. At the remaining 9 loci, we observed borderline associations [$P<0.10$ for some markers located within gene(s) of interest ± 50 kb] and/or nominal associations in extended regions [for some markers located within gene(s) of interest ± 100 kb] (data not shown).

## Follow-up investigation of new suggestive hits (from Stage 1) by using an additional dataset of European-descent subjects (Stage 2)

We next examined our GWAS data (SNPs with MAF≥0.05) for new suggestive autosomal signals and assessed them *in silico* in an independent GWAS data (Stage 2 dataset) (8). A total of 558 common markers (index or proxy SNPs for non-xMHC autosomal associations with $P<0.001$) had information available in Stage 2 dataset. Initial meta-analysis of Stage 1 and 2 datasets for 558 markers identified 26 new loci of interest, harboring SNPs with meta $P<0.001$ and same direction of association in both datasets. These 26 loci were further examined in Stage 2 dataset by evaluating all relevant GWAS SNPs residing within/near implicated loci/genes (567 common SNPs with $P<0.05$ in Stage 1 dataset). Information was available for 306 of 567 SNPs in Stage 2 dataset and final meta-analysis of Stage 1 and 2 datasets for these SNPs revealed a genome-wide significant new SLE locus at 12q12 (meta $P=3.1\times10^{-8}$ for the lead SNP) (Table 1, Figure 1).

## Replication of new genome-wide significant locus (from Stage 2) in another dataset of European-descent subjects (Stage 3)

We next sought *in silico* replication of new SLE locus identified by meta-analyses of Stage 1 and 2 datasets in an independent imputed GWAS data (Stage 3 dataset) (21, 22). A total of 8 SNPs from new 12q12 locus, with meta $P<0.001$ and same direction of association in Stage 1 and 2 samples (Figure 1-top panel), were further interrogated in Stage 3 sample. The genome-wide significant lead SNP (rs10506216/meta $P=3.1\times10^{-8}$) was replicated in Stage 3

dataset (rs10506216/$P$=2.3×10$^{-3}$) and three suggestive SNPs (rs1356422/meta $P$=3.5×10$^{-5}$, rs11181677/meta $P$=3.2×10$^{-4}$, and rs1914490/meta $P$=4.3×10$^{-4}$) also showed nominal significance in Stage 3 sample (rs1356422/$P$=1.1×10$^{-3}$, rs11181677/$P$=1.4×10$^{-3}$, and rs1914490/$P$=3.3×10$^{-2}$) (Table 1). The pairwise LD between these 4 SNPs in our discovery sample is illustrated in Figure 2.

## Functional annotation of the new locus

The novel 12q12 association signal falls within an intergenic region, adjacent to some large/long intergenic non-coding RNA (lincRNA) genes (Figure 1-bottom panel) including *LOC101927058* (http://www.ncbi.nlm.nih.gov/gene/101927058). Although *PRICKLE1* is the nearest protein-coding gene (~147 kb from the lead SNP), this region also harbors an immunologically important gene (*IRAK4;* ~1 Mb from the lead SNP) (Figure 1).

We first used RegulomeDB (http://www.regulomedb.org/) to assess the regulatory potential of 4 SNPs of interest at this new locus, which revealed no functionally relevant low scores for these SNPs (Table 1). Although no proxy with low score was identified for the lead SNP (rs10506216), the remaining 3 SNPs (rs1356422, rs11181677, and rs1914490) were in strong LD ($r^2$≥0.80) with 5 nearby potentially regulatory SNPs that were not analyzed in our study (rs868765, rs2897590, rs870972, rs12302566, and rs10785378; RegulomeDB scores ranging from "2b to 3a"). Two of these SNPs, rs870972 and rs12302566 (in strong LD with rs1356422 and rs11181677), were found to reside in a binding site or motif for STAT or IRF family of transcription factors (TFs) along with others.

We next used Genevar (http://www.sanger.ac.uk/resources/software/genevar/) to visualize eQTLs and mQTLs surrounding our SNPs of interest by examining the data from 856 Caucasian female twins with information on genetic variation, gene expression [tissues: adipose, lymphoblastoid cell lines (LCL) and skin], and DNA methylation (tissue: adipose) (31, 32). Figure S3 shows SNP-centric *cis*-eQTL analysis results for 4 SNPs of interest at 12q12, by displaying SNP-probe associations in a 2-Mb window (1-Mb on either side of each SNP) in 3 different tissues. Although the probes for *GXYLT1*, *YAF2*, *ZCRB1*, *PPHLN1*, *PRICKLE1*, *ADAMTS20*, and *PUS7L* were within targeted window for all 4 SNPs, those for *IRAK4* and downstream *PTK9* were included in analysis window for only rs1914490 due to their >1-Mb of distance from other 3 SNPs (see Figure 1 for gene locations). In adipose tissue, the most significant eQTL gene was *PRICKLE1* for all 4 SNPs (2.5×10$^{-4}$≤$P$≤5.0×10$^{-3}$), followed by *ZCRB1* (for 3 SNPs) and *IRAK4 & PPHLN1* (for rs1914490). In LCL, the most significant eQTL gene was *PPHLN1* (for rs1356422 and rs11181677, 3.4×10$^{-3}$≤$P$≤5.6×10$^{-3}$), followed by *ADAMTS20* (for rs10506216). Gene-centric *cis*-eQTL analysis (in a 2-Mb window around a given probe/gene) revealed that our SNPs and region of interest (GRCh37/hg19; chr12: 43130885–43214484, NCBI36/hg18; chr12: 41417152–41500751) overlapped with the most relevant cis-eQTL SNPs region for *PRICKLE1* in adipose tissue (Figure S4). In LCL, a similar overlap was observed for *PPHLN1* (for one probe) and our region of interest seemed to be also relevant for *IRAK4* expression (although it was only partly evaluated for this gene given its distant location) (Figure S5). SNP-centric *cis*-mQTL analysis [performed only in adipose tissue for common variants (MAF>0.05) located close to methylation sites/probes (probe ± 100 kb)] revealed

strong associations between our SNPs of interest at 12q12 and surrounding methylation sites/probes (Figure S6). Two probes (cg16758809 and cg22738642) were significantly associated with all 4 SNPs ($2.0\times10^{-10}\leq P\leq2.1\times10^{-4}$ and $2.2\times10^{-10}\leq P\leq1.8\times10^{-2}$, respectively).

Evaluation of 4 SNPs of interest at 12q12 in a database of genome-wide SNP associations (*cis* and *trans*) with monocyte expression in 1,490 European subjects (34) not only confirmed a *cis* effect on *PRICKLE1* expression (rs1356422/$P=3.5\times10^{-6}$ and rs11181677/$P=1.3\times10^{-6}$) but also revealed a *trans* effect of the lead SNP (rs10506216) on *PHRF1* expression at 11p15 ($P=1.3\times10^{-5}$).

Assessment of gene expression using BioGPS (http://biogps.org/#goto=welcome) (Figures S7–S8) revealed that although all genes of interest mentioned above are expressed in various cells/tissues including immune system, 3 of them (*ZCRB1*, *PPHLN1*, and *IRAK4)* exhibit a particularly high-level of expression in immune cells.

Finally, we also performed enhancer enrichment analysis using HaploReg v2 (http://www.broadinstitute.org/mammals/haploreg/haploreg.php) for 4 SNPs of interest (rs10506216, rs1356422, rs11181677, rs1914490) and 3 proxy SNPs (rs868765, rs2897590, rs870972) falling into RegulomeDB categories 1 or 2 (see above and Methods section) and the results for these 7 SNPs are shown in Figure S9. The most significant enrichment of strongest enhancers was detected in human embryonic stem cells (H1; $P=1.0\times10^{-6}$), B lymphoblastocytes (GM12878; $P=1.6\times10^{-4}$) and epidermal keratinocytes (NHEK; $P=3.7\times10^{-4}$).

## DISCUSSION

We conducted a new GWAS of SLE in a North American case-control sample of European ancestry and used the data from an independent GWAS of SLE in European-descent subjects (8) for follow-up evaluation of new suggestive autosomal signals, followed by meta-analyses that revealed a new genome-wide significant locus. This new locus was then replicated in imputed data from an independent European GWAS of SLE that is pending publication (21, 22).

### Replication of associations with previously reported autosomal SLE loci

Our GWAS further confirmed the MHC locus at 6p21 as the strongest SLE susceptibility locus ($P=1.2\times10^{-9}$ in our discovery sample), followed by 2q32/*STAT4* ($P=3.6\times10^{-7}$) and 8p23/*BLK* ($P=8.1\times10^{-6}$) (Figure S2). Overall, 30 of 39 previously reported autosomal SLE loci were supported in our discovery sample with $P<0.05$ (Table S1); 11 (*MHC region, STAT4, XKR6/BLK, IRF5/TNPO3, ITGAM, OLIG3/TNFAIP3, IL10, SNRPC/UHRF1BP1, IRF8, WDFY4, CREBL2/CDKN1B*) at the SNP level and 19 at the locus level (see Results for details). Of note, we cross-referenced our findings with only ‘selected genome-wide significant common SNPs’ reported at these 39 loci (not with all published significant SNPs), therefore the number of associations that we observed at the SNP level may be an underestimate. Our findings are not surprising given that several SLE loci reached genome-wide significance only after increasing the sample size by adding more controls from public

databases and/or after meta-analyses of discovery and replication samples and/or after performing follow-up studies by genotyping additional markers/samples (including the use of various ethnic groups for transancestral meta-analyses), all of which indicating small effect sizes for these loci. We considered nominal association(s) in our discovery sample as supporting evidence for previously reported loci. According to recommendations for publication of genetic associations with rheumatic diseases, it is acceptable to consider $P<0.05$ as significant in follow-up studies of established loci (37).

Our GWAS supported 11 SLE loci initially reported in Asian GWASs or follow-up studies (*RASGRP3, TET3, AFF1, HIP1, WDFY4, ARID5B, ETS1, CREBL2/CDKN1B, SLC15A4, ELF1*, and *PRKCB*) in our North American discovery sample, at the SNP (*WDFY4 and CREBL2/CDKN1B*) or locus level (Table S1). Although some of these loci have already been replicated in Caucasians (38), to our knowledge we are the first to report supporting evidence for *TET3, AFF1, HIP1, ARID5B, CREBL2/CDKN1B, ELF1*, and *PRKCB* in European-descent subjects.

As part of this study, we used some recently developed public databases including RegulomeDB, to assess the regulatory potential of SLE-relevant SNPs. A number of SNPs from established SLE loci with $P<0.05$ in our discovery sample showed low RegulomeDB scores, indicating a strong regulatory potential (Table S1).

In summary, our GWAS supported several, but not all, previously published SLE loci. Various factors may influence replication of reported loci, including GWAS platform coverage differences, study sample characteristics, study power, allele frequency/LD differences among various ethnic groups, allelic/genetic heterogeneity, effect size differences, gene-environment interactions and epigenetic factors.

## New SLE locus at 12q12

To follow-up on our new GWAS findings for SLE, we adopted a cost-effective multi-stage strategy by using available data from already performed studies; therefore, we could not evaluate all new suggestive loci and/or all relevant SNPs at these loci in follow-up analyses. Nevertheless, we identified and replicated a new genome-wide significant locus at **12q12** (Table 1). This new SLE signal falls within an intergenic region, located ~147 kb 5'-upstream of **PRICKLE1** [prickle homolog 1 (Drosophila)] and ~1 Mb 5'-upstream of **IRAK4** (interleukin-1 receptor-associated kinase 4) (Figure 1). This region also harbors some **lincRNA** genes that reside in closer vicinity to this SLE signal (Figure 1-bottom panel). Non-coding RNAs are emerging as essential regulators of various cellular processes including immune response (39, 40). Although these recently described lincRNAs do not seem to be represented in public gene expression databases (Genevar and BioGPS), *in silico* assessment of eQTL and mQTL associations and expression of relevant protein-coding genes (Figures S3–S8) suggest that this new locus probably functions as a distal regulatory element (i.e., enhancer) for the expression of neighboring genes in a tissue/cell type-specific manner, with indication of epigenetic modulation. The association of SLE-relevant SNPs with surrounding methylation sites is particularly strong at this new locus (Figure S6) and these sites may modulate chromatin accessibility and transcriptional potential of underlying DNA segment (31, 32). Recent studies suggest that inter-individual variation in DNA

methylation is genetically regulated (more variable and heritable for gene-body and intergenic methylation sites) and these regulatory variants usually function in a tissue-specific manner (31, 32). Moreover, our 'possible enhancer' hypothesis for this new intergenic locus has also been supported by the results of enhancer enrichment analysis performed in HaploReg v2 (Figure S9).

Based on available public data (Genevar and BioGPS), noteworthy genes that seem to be *cis*-regulated by this new locus include **PRICKLE1, PPHLN1, and IRAK4** (Figures S3–S5, S7–S8). *PRICKLE1* (http://www.ncbi.nlm.nih.gov/gene/144165) encodes a nuclear receptor that regulates the Wnt/beta-catenin signaling pathway. This pathway was (a) implicated in renal disease pathogenesis including lupus nephritis in mice, (b) shown to be activated in most autoimmune diseases, and (c) reported to play a critical role in senescence of bone marrow-mesenchymal stem cells from SLE patients (41–43). *PRICKLE1* mutations were linked to epilepsy and Autism Spectrum Disorders (44–46). *PPHLN1* (Periphilin 1; http://www.ncbi.nlm.nih.gov/gene/51535) appears to play a role in epithelial differentiation and epidermal integrity; however, it is most abundantly expressed in immune cells (Figure S7), thus awaiting further functional characterization. *IRAK4* (http://www.ncbi.nlm.nih.gov/gene/51135) is involved in NF-kappaB activation in Toll-like receptor and T-cell receptor signaling pathways and its mutations were reported to cause recurrent infections. *IRAK4* was shown to play an important role in Th17-mediated immunity and removal of autoreactive B cells (47, 48). IRAK1/4 inhibition was reported to affect plasmacytoid dendritic cell response to SLE serum (49) (*IRAK1* is an X-linked member of IRAK family previously implicated in SLE) (3–5).

A database of genome-wide SNP associations with monocyte expression (34) also suggested a possible *trans* effect of the lead SNP at 12q12 on *PHRF1* expression at 11p15. The *PHRF1-IRF7* region is among previously implicated SLE loci (3–5) and its interaction with the 12q12 locus warrants further investigation. Interestingly, three genomic regions reported to be major *trans*-acting regulators of multiple expression traits in monocytes are all located at 12q (12q13, 12q15, 12q24) (34). Although studies investigating *trans* eQTLs are relatively limited, a similar *trans* effect was also reported for another SLE locus (50).

According to RegulomeDB, this 12q12 regulatory region binds various TFs, including STAT and IRF family TFs that are essential for SLE as indicated by their interactions with type I interferon system that plays a central role in SLE and by direct implication of some of their members in SLE susceptibility.

Although public databases are extremely useful for preliminary functional evaluation of disease-associated variants/loci, they suffer from some limitations. Major limitation includes data/analyses being restricted to selected SNPs, selected probes (not representing all genes or all splice forms of a given gene), selected tissue/cell types, and selected methods (our region of interest was partially evaluated for its effect on *IRAK4* expression in Genevar-MuTHER due to a selected maximum window size). Therefore, an extensive functional analysis is necessary to determine the role of this new SLE locus in regulation of various genes in various cells/tissues involved in SLE pathogenesis. Follow-up deep sequencing is

also warranted to uncover all causative common/rare variants that may jointly contribute to SLE risk at this locus.

In conclusion, several previously published SLE loci were supported in our discovery sample and our multi-stage approach identified and replicated a new genome-wide significant locus at 12q12. Although recently developed public tools/databases significantly helped us to gain valuable insights into the regulatory potential of our region/SNPs of interest, they also emphasized the need for comprehensive follow-up functional studies to determine tissue/cell type-specific roles of implicated loci. Thus, this new 12q12 locus warrants further characterization using deep sequencing and extensive functional evaluation and is likely to shed further light on underlying disease mechanisms. Additional large studies are also necessary to examine association of this new locus with SLE subphenotypes and its interaction with hormonal and environmental factors. These factors could influence the effect size of this locus (given its association with DNA methylation and its predicted role in epigenetic regulation), depending on their enrichment in a given study sample.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

FUNDING: This work was supported by grants from the US National Institutes of Health [HL092397, HL088648, AR057028, AR046588, AR057338, HD066139, AR02318, AR30492, AR064464, AR30692, AR43274, AI63274, AR56360, and TR000150]; by a grant from the Lupus Foundation of America; and by grants from Wellcome Trust (Ref 085492) and Arthritis Research UK (Ref 19289).

## REFERENCES

1. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis and rheumatism. 2005; 52(4): 1138–1147. [PubMed: 15818688]
2. Deapen D, Escalante A, Weinrib L, Horwitz D, Bachman B, Roy-Burman P, et al. A revised estimate of twin concordance in systemic lupus erythematosus. Arthritis and rheumatism. 1992; 35(3):311–318. [PubMed: 1536669]
3. Guerra SG, Vyse TJ, Cunninghame Graham DS. The genetics of lupus: a functional perspective. Arthritis research & therapy. 2012; 14(3):211. [PubMed: 22640752]
4. Vaughn SE, Kottyan LC, Munroe ME, Harley JB. Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. Journal of leukocyte biology. 2012; 92(3): 577–591. [PubMed: 22753952]
5. Dai C, Deng Y, Quinlan A, Gaskin F, Tsao BP, Fu SM. Genetics of systemic lupus erythematosus: immune responses and end organ resistance to damage. Current opinion in immunology. 2014; 31:87–96. [PubMed: 25458999]
6. Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S, et al. Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. The New England journal of medicine. 2008; 358(9):900–909. [PubMed: 18204098]
7. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nature genetics. 2008; 40(2): 204–210. [PubMed: 18204446]

8. Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM, et al. Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40(9): 1059–1061. [PubMed: 19165918]
9. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS genetics. 2011; 7(10):e1002341. [PubMed: 22046141]
10. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X, et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. Nature genetics. 2009; 41(11):1228–1233. [PubMed: 19838195]
11. Graham RR, Hom G, Ortmann W, Behrens TW. Review of recent genome-wide association scans in lupus. Journal of internal medicine. 2009; 265(6):680–688. [PubMed: 19493061]
12. Jacob CO, Eisenstein M, Dinauer MC, Ming W, Liu Q, John S, et al. Lupus-associated causal mutation in neutrophil cytosolic factor 2 (NCF2) brings unique insights to the structure and function of NADPH oxidase. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(2):E59–E67. [PubMed: 22203994]
13. Lessard CJ, Adrianto I, Ice JA, Wiley GB, Kelly JA, Glenn SB, et al. Identification of IRF8, TMEM39A, and IKZF3-ZPBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. American journal of human genetics. 2012; 90(4):648–660. [PubMed: 22464253]
14. Lessard CJ, Adrianto I, Kelly JA, Kaufman KM, Grundahl KM, Adler A, et al. Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. American journal of human genetics. 2011; 88(1):83–91. [PubMed: 21194677]
15. Taylor KE, Chung SA, Graham RR, Ortmann WA, Lee AT, Langefeld CD, et al. Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes. PLoS genetics. 2011; 7(2):e1001311. [PubMed: 21379322]
16. Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z, et al. Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. Nature genetics. 2009; 41(11):1234–1237. [PubMed: 19838193]
17. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, Higasa K, et al. A genome-wide association study identified AFF1 as a susceptibility locus for systemic lupus eyrthematosus in Japanese. PLoS genetics. 2012; 8(1):e1002455. [PubMed: 22291604]
18. Yang J, Yang W, Hirankarn N, Ye DQ, Zhang Y, Pan HF, et al. ELF1 is associated with systemic lupus erythematosus in Asian populations. Human molecular genetics. 2011; 20(3):601–607. [PubMed: 21044949]
19. Yang W, Shen N, Ye DQ, Liu Q, Zhang Y, Qian XX, et al. Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. PLoS genetics. 2010; 6(2):e1000841. [PubMed: 20169177]
20. Yang W, Tang H, Zhang Y, Tang X, Zhang J, Sun L, 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. American journal of human genetics. 2013; 92(1):41–51. [PubMed: 23273568]
21. Bentham J, Vyse TJ. The development of genome-wide association studies and their application to complex diseases, including lupus. Lupus. 2013; 22(12):1205–1213. [PubMed: 24097992]
22. Bentham, J.; Morris, DL.; Alarcón-Riquelme, ME.; Anand, V.; Delgado-Vega, AM.; Fortin, PR., et al. Novel loci identified by the largest genome-wide association study of lupus to date; (Abstract). Presented at the 61st Annual Meeting of The American Society of Human Genetics; October 11-15, 2011; Montréal, Canada. 2011.
23. Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis and rheumatism. 1982; 25(11):1271–1277. [PubMed: 7138600]
24. Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis and rheumatism. 1997; 40(9):1725. [PubMed: 9324032]

25. Bernatsky S, Ramsey-Goldman R, Labrecque J, Joseph L, Boivin JF, Petri M, et al. Cancer risk in systemic lupus: an updated international multi-centre cohort study. Journal of autoimmunity. 2013; 42:130–135. [PubMed: 23410586]
26. Demirci FY, Dressen AS, Kammerer CM, Barmada MM, Kao AH, Ramsey-Goldman R, et al. Functional polymorphisms of the coagulation factor II gene (F2) and susceptibility to systemic lupus erythematosus. The Journal of rheumatology. 2011; 38(4):652–657. [PubMed: 21239755]
27. Pineau CA, Bernatsky S, Abrahamowicz M, Neville C, Karp I, Clarke AE. A comparison of damage accrual across different calendar periods in systemic lupus erythematosus patients. Lupus. 2006; 15(9):590–594. [PubMed: 17080914]
28. Rhew EY, Manzi SM, Dyer AR, Kao AH, Danchenko N, Barinas-Mitchell E, et al. Differences in subclinical cardiovascular disease between African American and Caucasian women with systemic lupus erythematosus. Translational research : the journal of laboratory and clinical medicine. 2009; 153(2):51–59. [PubMed: 19138649]
29. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics (Oxford, England). 2010; 26(17):2190–2191.
30. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nature reviews Genetics. 2010; 11(7):499–511.
31. Grundberg E, Meduri E, Sandling JK, Hedman AK, Keildson S, Buil A, et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. American journal of human genetics. 2013; 93(5):876–890. [PubMed: 24183450]
32. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. Nature genetics. 2012; 44(10):1084–1089. [PubMed: 22941192]
33. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, Stranger BE, et al. Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics (Oxford, England). 2010; 26(19):2474–2476.
34. Rotival M, Zeller T Fau - Wild PS, Wild Ps Fau - Maouche S, Maouche S Fau - Szymczak S, Szymczak S Fau - Schillert A, Schillert A Fau - Castagne R, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. (1553-7404 (Electronic)).
35. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, Sanchez E, et al. Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nature genetics. 2008; 40(2):211–216. [PubMed: 18204447]
36. Sheng YJ, Gao JP, Li J, Han JW, Xu Q, Hu WL, et al. Follow-up study identifies two novel susceptibility loci PRKCB and 8p11.21 for systemic lupus erythematosus. Rheumatology (Oxford, England). 2011; 50(4):682–688.
37. Plenge RM, Bridges SL Jr, Huizinga TW, Criswell LA, Gregersen PK. Recommendations for publication of genetic association studies in Arthritis & Rheumatism. Arthritis and rheumatism. 2011; 63(10):2839–2847. [PubMed: 21702018]
38. Wang C, Ahlford A, Jarvinen TM, Nordmark G, Eloranta ML, Gunnarsson I, et al. Genes identified in Asian SLE GWASs are also associated with SLE in Caucasian populations. European journal of human genetics : EJHG. 2013; 21(9):994–999. [PubMed: 23249952]
39. Carpenter S, Aiello D, Atianand MK, Ricci EP, Gandhi P, Hall LL, et al. A long noncoding RNA mediates both activation and repression of immune response genes. Science (New York, NY). 2013; 341(6147):789–792.
40. Pagani M, Rossetti G, Panzeri I, de Candia P, Bonnal RJ, Rossi RL, et al. Role of microRNAs and long-non-coding RNAs in CD4(+) T-cell differentiation. Immunological reviews. 2013; 253(1): 82–96. [PubMed: 23550640]
41. Gu Z, Tan W, Feng G, Meng Y, Shen B, Liu H, et al. Wnt/beta-catenin signaling mediates the senescence of bone marrow-mesenchymal stem cells from systemic lupus erythematosus patients through the p53/p21 pathway. Molecular and cellular biochemistry. 2013

42. Tuller T, Atar S, Ruppin E, Gurevich M, Achiron A. Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases. Genes and immunity. 2013; 14(2):67–82. [PubMed: 23190644]

43. Tveita AA, Rekvig OP. Alterations in Wnt pathway activity in mouse serum and kidneys during lupus development. Arthritis and rheumatism. 2011; 63(2):513–522. [PubMed: 21280006]

44. Bassuk AG, Wallace RH, Buhr A, Buller AR, Afawi Z, Shimojo M, et al. A homozygous mutation in human PRICKLE1 causes an autosomal-recessive progressive myoclonus epilepsy-ataxia syndrome. American journal of human genetics. 2008; 83(5):572–581. [PubMed: 18976727]

45. Cukier HN, Dueker ND, Slifer SH, Lee JM, Whitehead PL, Lalanne E, et al. Exome sequencing of extended families with autism reveals genes shared across neurodevelopmental and neuropsychiatric disorders. Molecular autism. 2014; 5(1):1. [PubMed: 24410847]

46. Tao H, Manak JR, Sowers L, Mei X, Kiyonari H, Abe T, et al. Mutations in prickle orthologs cause seizures in flies, mice, and humans. American journal of human genetics. 2011; 88(2):138–149. [PubMed: 21276947]

47. Isnardi I, Ng YS, Srdanovic I, Motaghedi R, Rudchenko S, von Bernuth H, et al. IRAK-4- and MyD88-dependent pathways are essential for the removal of developing autoreactive B cells in humans. Immunity. 2008; 29(5):746–757. [PubMed: 19006693]

48. Staschke KA, Dong S, Saha J, Zhao J, Brooks NA, Hepburn DL, et al. IRAK4 kinase activity is required for Th17 differentiation and Th17-mediated disease. Journal of immunology (Baltimore, Md : 1950). 2009; 183(1):568–577.

49. Chiang EY, Yu X, Grogan JL. Immune complex-mediated cell activation from systemic lupus erythematosus and rheumatoid arthritis patients elaborate different requirements for IRAK1/4 kinase activity across human cell types. Journal of immunology (Baltimore, Md : 1950). 2011; 186(2):1279–1288.

50. Westra HA-OX, Peters Mj Fau - Esko T, Esko T Fau - Yaghootkar H, Yaghootkar H Fau - Schurmann C, Schurmann C Fau - Kettunen J, Kettunen J Fau - Christiansen MW, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. (1546-1718 (Electronic)).

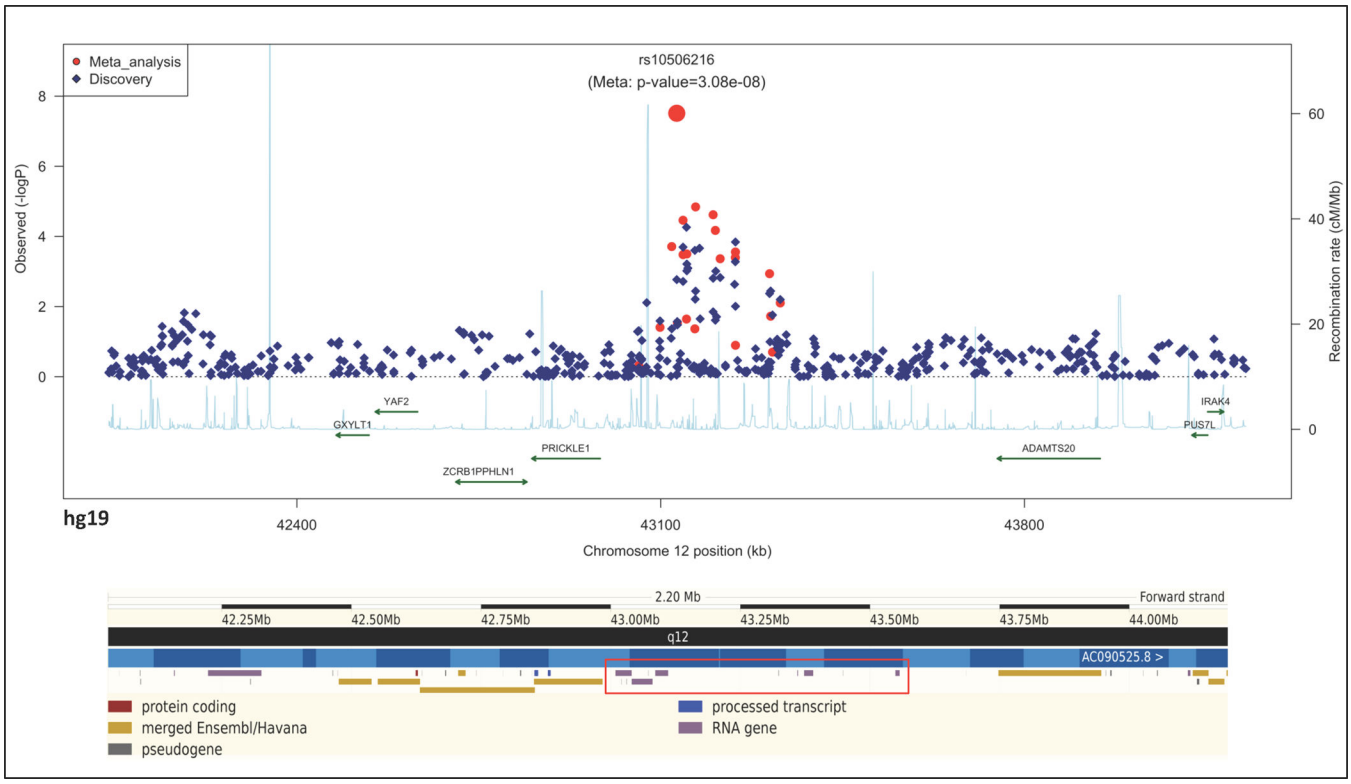


**Figure 1. Regional association plot on chromosome 12q12 (top panel) and the region overview of the corresponding chromosome segment in Ensembl genome browser (bottom panel)**

TOP PANEL: The associations observed in Stage 1 discovery sample are depicted as dark blue diamonds while the results from the meta-analysis of Stage 1 and Stage 2 samples are shown as red dots. The genes located in the region (based on the UCSC genome browser) and the recombination rates by position (light blue line) are also shown. The most relevant SNP with best meta *P* is labeled. BOTTOM PANEL: The red rectangle shows the large/long intergenic non-coding RNA (lincRNA) genes that reside adjacent to the SLE association signal in this region.

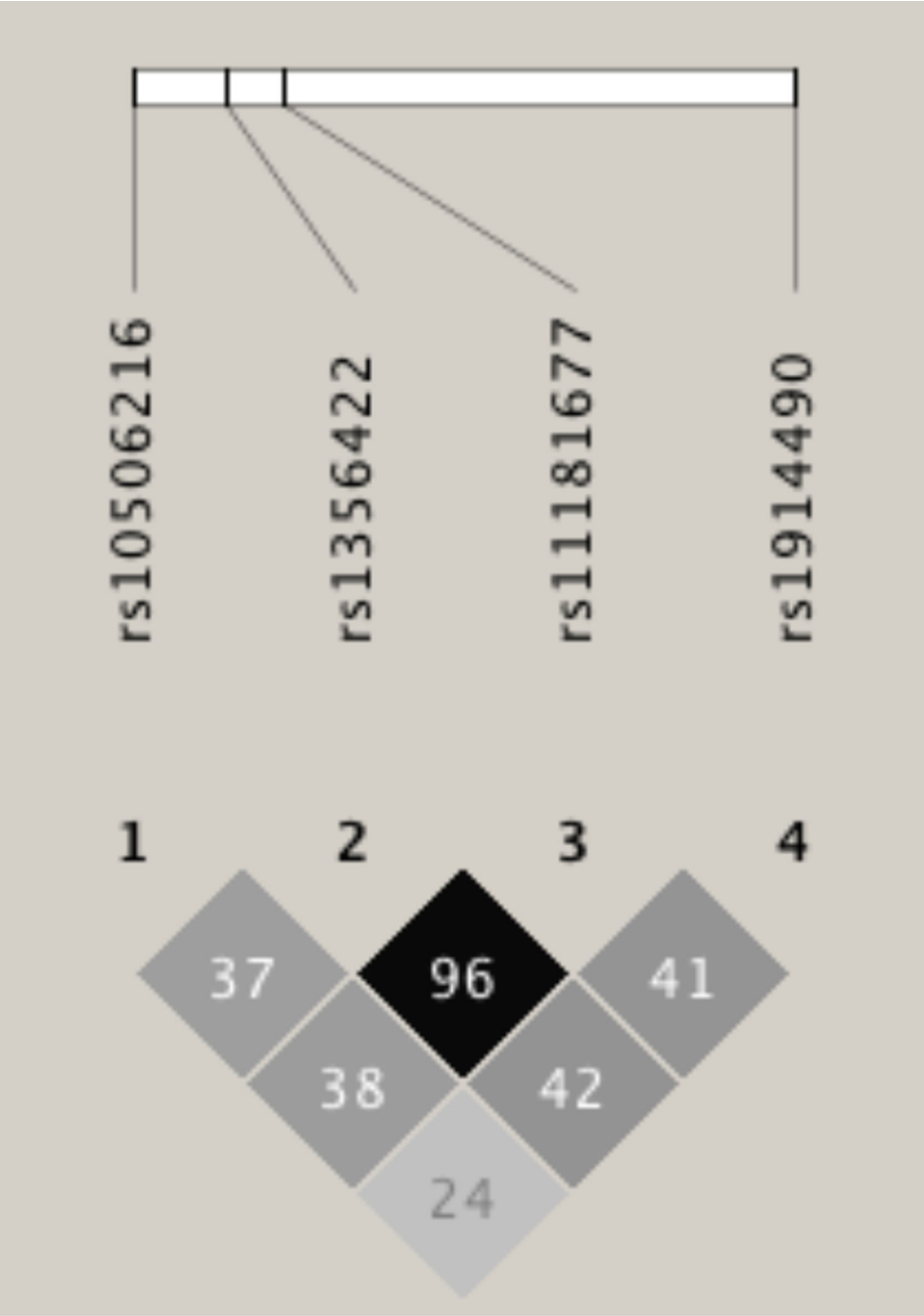


**Figure 2. The extent of pairwise LD between the SNPs of interest at the 12q12 locus**

The LD plot shows the pairwise $r^2 \times 100$ values between the SNPs of interest in Stage 1 discovery sample.

**Table 1**

New genome-wide significant SLE locus that was identified by Stage 1 and 2 analyses and then replicated in Stage 3 analysis

| Chr | SNP | Position in bp (hg19) | Allele | Stage 1 discovery set 661 cases & 487 controls | | | | Stage 2 data set 431 cases & 2,155 controls | | | | Stages 1&2 Meta *P* | Stage 3 replication data set 4,036 cases & 6,959 controls | | | | | Nearest Protein-coding Gene* | RegulomeDB score** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Case Freq | Control Freq | OR | *P* | Case Freq | Control Freq | OR | *P* | | Case Freq | Control Freq | OR | *P* | Imputed | | |
| 12q12 | rs10506216 | 43130885 | **T** | 0.109 | 0.079 | 1.701 | 1.71E-03 | 0.124 | 0.080 | 1.666 | 5.04E-06 | **3.08E-08** | 0.093 | 0.081 | 1.167 | **2.27E-03** | Yes | *PRICKLE1* | 6 |
| | rs1356422 | 43142996 | **G** | 0.239 | 0.184 | 1.570 | 2.02E-04 | 0.215 | 0.182 | 1.247 | 1.22E-02 | **3.46E-05** | 0.205 | 0.188 | 1.124 | **1.05E-03** | Yes | *PRICKLE1* | 5 |
| | rs11181677 | 43150285 | **C** | 0.236 | 0.183 | 1.520 | 6.13E-04 | 0.209 | 0.181 | 1.204 | 4.00E-02 | **3.17E-04** | 0.205 | 0.189 | 1.121 | **1.40E-03** | Yes | *PRICKLE1* | ND |
| | rs1914490 | 43214484 | G | 0.300 | 0.251 | 1.427 | 1.49E-03 | 0.268 | 0.238 | 1.186 | 3.47E-02 | **4.33E-04** | 0.270 | 0.257 | 1.072 | **3.29E-02** | Yes | *PRICKLE1* | 6 |

Only the SNPs with meta *P*<1.00E-03 and same direction of association in Stage 1 and 2 samples that had *P*<0.05 in Stage 3 sample are shown. **Bold** alleles: reverse strand alleles

*The nearest protein-coding gene within ± 500 kb (http://www.genome.ucsc.edu/).

**Please see the methods section and http://www.regulomedb.org/help;

ND: No Data.; RegulomeDB scoring scheme is divided into the following categories (1 through 6) based on available data resources: "1a: expression Quantitative Trait Locus (eQTL) + Transcription factor (TF) binding + matched TF motif + matched DNase Footprint + DNase peak, 1b: eQTL + TF binding + any motif + DNase Footprint + DNase peak, 1c: eQTL + TF binding + matched TF motif + DNase peak, 1d: eQTL + TF binding + any motif + DNase peak, 1e: eQTL + TF binding + matched TF motif, 1f: eQTL + TF binding / DNase peak, 2a: TF binding + matched TF motif + matched DNase Footprint + DNase peak, 2b: TF binding + any motif + DNase Footprint + DNase peak, 2c: TF binding + matched TF motif + DNase peak, 3a: TF binding + any motif + DNase peak, 3b: TF binding + matched TF motif, 4: TF binding + DNase peak, 5: TF binding or DNase peak, 6: other".