Published in final edited form as:
Nat Genet. 2009 November ; 41(11): 1228–1233. doi:10.1038/ng.468.

A large-scale replication study identifies *TNIP1, PRDM1, JAZF1, UHRF1BP1* and *IL10* as risk loci for systemic lupus erythematosus

Vesela Gateva[1], **Johanna K Sandling**[2], **Geoff Hom**[1], **Kimberly E Taylor**[3], **Sharon A Chung**[3], **Xin Sun**[1], **Ward Ortmann**[1], **Roman Kosoy**[4], **Ricardo C Ferreira**[1], **Gunnel Nordmark**[5], **Iva Gunnarsson**[6], **Elisabet Svenungsson**[6], **Leonid Padyukov**[6], **Gunnar Sturfelt**[7], **Andreas Jönsen**[7], **Anders A Bengtsson**[7], **Solbritt Rantapää-Dahlqvist**[8], **Emily C Baechler**[9], **Elizabeth E Brown**[10], **Graciela S Alarcón**[10], **Jeffrey C Edberg**[10], **Rosalind Ramsey-Goldman**[11], **Gerald McGwin Jr**[10], **John D Reveille**[12], **Luis M Vilá**[13], **Robert P Kimberly**[10], **Susan Manzi**[14], **Michelle A Petri**[15], **Annette Lee**[16], **Peter K Gregersen**[16], **Michael F Seldin**[4], **Lars Rönnblom**[5], **Lindsey A Criswell**[3], **Ann-Christine Syvänen**[2], **Timothy W Behrens**[1], and **Robert R Graham**[1]

[1] Immunology Biomarkers Group, Genentech, South San Francisco, California, USA

[2] Molecular Medicine, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

[3] Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, California, USA

[4] Rowe Program in Genetics, University of California at Davis, Davis, California, USA

[5] Section of Rheumatology, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

[6] Rheumatology Unit, Department of Medicine, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden

[7] Section of Rheumatology, Department of Clinical Sciences, Lund University Hospital, Lund, Sweden

[8] Department of Rheumatology, Umeå University Hospital, Umeå, Sweden

Correspondence should be addressed to R.R.G. (graham.robert@gene.com).

Note: Supplementary information is available on the Nature Genetics website.

AUTHOR CONTRIBUTIONS

V.G. and J.K.S. performed the primary statistical analyses and contributed to initial manuscript preparation; J.K.S managed DNA samples and performed genotyping. G.H. contributed to the statistical analyses and experimental design. K.E.T. and S.A.C. performed statistical analyses and contributed to manuscript preparation. X.S., W.O. and R.C.F. managed DNA samples and contributed to experimental design. G.N., I.G., E.S., L.P., G.S., A.J., A.A.B., S.R.-D., E.C.B, E.E.B., G.S.A., J.C.E., R.R.-G., G.M. Jr., J.D.R., L.M.V., R.P.K., S.M. and M.A.P. provided samples and phenotype information. A.L. managed samples and oversaw genotyping efforts. P.K.G. provided samples and contributed to the initial manuscript preparation. M.F.S. and R.K. contributed statistical analyses and contributed to the selection of the ancestry-informative markers. L.R., L.A.C. and A.-C.S. contributed samples, input into experimental design, data interpretation and initial manuscript preparation; A.-C.S. oversaw genotyping efforts. R.R.G. and T.W.B. contributed to experimental design and interpretation, statistical analyses and initial manuscript preparation. All authors contributed to the final paper.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/.

Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

[9] Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota, USA

[10] University of Alabama at Birmingham, Birmingham, Alabama, USA

[11] Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA

[12] University of Texas–Houston Health Science Center, Houston, Texas, USA

[13] University of Puerto Rico Medical Science Campus, San Juan, Puerto Rico

[14] University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania, USA

[15] Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

[16] The Feinstein Institute for Medical Research, North Shore–Long Island Jewish Health System, Manhasset, New York, USA

Abstract

Genome-wide association studies have recently identified at least 15 susceptibility loci for systemic lupus erythematosus (SLE). To confirm additional risk loci, we selected SNPs from 2,466 regions that showed nominal evidence of association to SLE ($P < 0.05$) in a genome-wide study and genotyped them in an independent sample of 1,963 cases and 4,329 controls. This replication effort identified five new SLE susceptibility loci ($P < 5 \times 10^{-8}$): *TNIP1* (odds ratio (OR) = 1.27), *PRDM1* (OR = 1.20), *JAZF1* (OR = 1.20), *UHRF1BP1* (OR = 1.17) and *IL10* (OR = 1.19). We identified 21 additional candidate loci with $P \leq 1 \times 10^{-5}$. A candidate screen of alleles previously associated with other autoimmune diseases suggested five loci ($P < 1 \times 10^{-3}$) that may contribute to SLE: *IFIH1*, *CFB*, *CLEC16A*, *IL12B* and *SH2B3*. These results expand the number of confirmed and candidate SLE susceptibility loci and implicate several key immunologic pathways in SLE pathogenesis.

Systemic lupus erythematosus (SLE) is a chronic inflammatory autoimmune disease characterized by the presence of antibodies to nuclear self-antigens. Many of the lupus autoantibodies recognize nucleic acids and nucleic acid binding proteins, which in turn activate Toll-like receptors, leading to the production of type I interferon[1]. Despite considerable clinical heterogeneity, SLE ranks among the most heritable of common autoimmune diseases, with a sibling risk ratio of ~30 (ref. [2]). Recent genome-wide association (GWA) and candidate gene studies have identified at least 15 common SLE risk alleles that achieve genome-wide significance ($P < 5 \times 10^{-8}$). These include genes encoding proteins important for adaptive immunity and the production of autoantibodies (HLA class II alleles, *BLK*, *PTPN22* and *BANK1*) and proteins with roles in innate immunity and interferon signaling (*ITGAM*, *TNFAIP3*, *STAT4* and *IRF5*)[3–10]. To identify additional risk loci, we performed a targeted replication study of SNPs from 2,466 loci that showed a nominal *P* value of <0.05 in a recent GWA[7] scan of 1,310 individuals with lupus (cases) and 7,859 controls. We also genotyped SNPs from 23 previously reported SLE risk loci, 42 SNPs implicated in other autoimmune diseases and over 7,000 ancestry-informative markers (Fig. 1).

We designed a custom SNP array (Illumina Infinium II) consisting of over 12,000 variants and genotyped two independent SLE case and control populations from the United States (1,129 SLE cases and 2,991 controls) and Sweden (834 SLE cases and 1,338 controls). Included among the US controls were 2,215 Alzheimer's disease case-control samples, which were judged to be acceptable as controls because the genetic factors underlying SLE and Alzheimer's disease are expected to be independent. We next applied data quality filters to remove poorly performing samples and SNPs, population outliers, duplicate samples and related individuals (see Online Methods). Following these quality control measures, we examined a final set of 10,848 SNPs (Fig. 1). Association statistics for 3,735 SNPs were calculated and corrected for population stratification using 7,113 ancestry-informative markers (see Online Methods).

We first examined 25 SNPs (from 23 loci) that were previously reported to be associated with SLE (Table 1 and Supplementary Table 1). We found further evidence of association for 21 of the variants ($P < 0.05$), including 9 loci that reached genome-wide significance ($P < 5 \times 10^{-8}$) in the current combined dataset. Among the loci with genome-wide significant results were *HLA-DRB1* (HLA*DR3 or DRB1*0301), *IRF5*, *TNFAIP3*, *BLK*, *STAT4*, *ITGAM*, *PTPN22*, *PHRF1* (also called *KIAA1542*) and *TNFSF4* (also called *OX40L*). The analysis also provided additional evidence of association for variants at nine loci for which a single previous study reported genome-wide levels of significance: *HLA-DRB1* (HLA*DR2 or DRB1*1501), *TNFAIP3* (rs6920220), *BANK1*, *ATG5*, *PTTG1*, *PXK*, *FCGR2A*, *UBE2L3* and *IRAK1-MECP2*.

An earlier candidate gene study[9] identified *MECP2* as a potential risk locus for SLE; however, in the current dataset, SNPs near *IRAK1*, a gene critical for Toll-like receptor 7 and 9 signaling and located within the identified region of linkage disequilibrium (LD) surrounding *MECP2*, showed the strongest evidence of association. Similar findings were recently reported[11], and further work will be required to determine the causal allele in the *IRAK1-MECP2* locus. We found additional evidence of association for three loci (*TYK2*, *ICA1* and *NMNAT2*) that had previously shown significant but not genome wide–level evidence for association[6,10]. For four previously implicated variants (*LYN*, *SCUBE1*, *TLR5* and *LY9*), no evidence of association was observed in the combined dataset.

To identify previously unknown SLE risk loci, we examined 3,188 SNPs from 2,446 distinct loci that showed evidence of association to SLE in our genome-wide dataset[7], which comprised 502,033 SNPs genotyped in 1,310 SLE cases and an expanded set of 7,859 controls. Using this dataset, we imputed over 2.1 million variants using Phase II HapMap CEU samples as a reference (see Online Methods) and generated a rank-ordered list of association statistics. Variants with $P < 0.05$ were selected for possible inclusion on the custom replication array. For efficient genotyping, we identified groups of correlated variants ($r^2 > 0.2$) and then carried out selection of at least two SNPs from each group where all SNPs had $P < 0.001$. For the remaining groups, the SNP with the lowest P value in the group was included. In the replication samples, we calculated the association statistics (see Online Methods) and observed a significant enrichment of the replication results relative to the expected null distribution (Fig. 2). Excluding previously reported SLE risk alleles, there were 134 loci with $P < 0.05$ (64 expected; $P = 2 \times 10^{-15}$) and 12 loci with $P < 0.001$ (1 expected; $P = 1 \times 10^{-9}$), suggesting the presence of true positive associations.

The replication study identified five new SLE risk loci with a combined P value that exceeded the genome-wide threshold for significance ($P < 5 \times 10^{-8}$): *TNIP1*, *PRDM1*, *JAZF1*, *UHRF1BP1* and *IL10* (Table 2 and Supplementary Table 2). These loci are discussed in more detail below.

A variant (rs7708392) on 5q33.1 that resides within an intron of *TNIP1* (encoding TNF-α-induced protein 3 (TNFAIP3)-interacting protein 1) was significantly associated with SLE in all three cohorts and had a combined $P = 3.8 \times 10^{-13}$ (Fig. 2). Variants near *TNIP1* were recently found to contribute to risk of psoriasis[12]; however, the SLE and psoriasis risk variants are separated by 21 kb and appear to have distinct genetic signals ($r^2 = 0.001$). TNIP1 and TNFAIP3 are interacting proteins[13], but the precise role of TNIP1 in regulating TNFAIP3 is unknown. The association of multiple distinct variants near *TNFAIP3* with

SLE[4,14], rheumatoid arthritis[15], psoriasis[12] and type 1 diabetes[16] suggests that this pathway has an important role in regulating autoimmunity.

A second confirmed risk variant (rs6568431, $P = 7.12 \times 10^{-10}$) was identified in an intergenic region between *PRDM1* (PR domain containing 1, with ZNF domain, also known as *BLIMP1*) and *ATG5* (APG5 autophagy 5-like). The signal at rs6568431 appears to be distinct from the previously reported[6] SLE risk allele within *ATG5* (rs2245214, Table 1), as rs6568431 has an $r^2 < 0.1$ with rs2245214, and rs2245214 remains significantly associated with SLE ($P < 1 \times 10^{-5}$) after conditional logistic regression incorporating rs6568431 (Fig. 2).

The promoter region of *JAZF1* (juxtaposed with another zinc finger gene 1) is a third newly confirmed SLE locus (rs849142, $P = 1.54 \times 10^{-9}$). Of note, this same variant was previously associated with risk of type 2 diabetes[17] and with height variation[18]. A separate prostate cancer risk allele near *JAZF1* (rs10486567)[19] showed no evidence for association in the current study.

A fourth newly identified SLE risk locus is defined by a nonsynonymous allele (R454Q) of *UHRF1BP1* (ICBP90 binding protein 1; rs11755393, $P = 2.22 \times 10^{-8}$). This allele encodes a nonconservative amino-acid change in a putative binding partner of UHRF1, a transcription and methylation factor linked to multiple pathways[20]. The *UHRF1BP1* risk allele is in a region of extended LD that encompasses multiple genes, including *SNRPC* (small nuclear ribonucleoprotein polypeptide C), which is part of a RNA processing complex often targeted by SLE autoantibodies.

The fifth newly identified SLE locus is *IL10* (interleukin-10; rs3024505, $P = 3.95 \times 10^{-8}$; Fig. 2). IL10 is an important immunoregulatory cytokine that functions to downregulate immune responses[21], and variation in *IL10* has inconsistently been reported to be associated with SLE[22]. The variant associated with SLE is identical to a SNP recently identified as contributing to risk of ulcerative colitis[23] and type 1 diabetes[24], suggesting the possibility of shared pathophysiology in the IL10 pathway across these disorders.

Using a significance threshold of $P < 1 \times 10^{-5}$ in the combined replication sample, we identified 21 additional SLE candidate risk loci (Table 2 and Supplementary Table 2). Less than one locus (0.01 loci, specifically) with $P < 1 \times 10^{-5}$ was expected under a null distribution for the meta-analysis ($P = 8 \times 10^{-77}$), suggesting that several of these loci are likely to be true positives for association to SLE. Notable candidate genes in this list include: (i) *IRF8* (interferon regulatory factor 8), which was implicated in a previous GWA study (GWAS)[4] and whose family members *IRF5* and *IRF7* are within confirmed SLE risk loci; (ii) TAOK3 (TAO kinase 3), a kinase expressed in lymphocytes, and the disease-associated variant is a missense allele (rs428073, N47S); (iii) *LYST* (lysosomal trafficking regulator), mutations of which cause Chediak-Higashi syndrome in humans, a complex disorder characterized by a lymphoproliferative disorder; and (iv) *IL12RB2* (interleukin 12 receptor, beta 2), a locus which includes *IL23R* and *SERPBP1* but appears distinct from the *IL23R* variants reported in inflammatory bowel disease, psoriasis and ankylosing spondylitis[25].

A noteworthy feature of recent GWAS is the large number of overlapping loci found to be shared between different complex diseases[26]. We tested 42 variants from 35 loci that were previously reported as autoimmune-disease risk alleles for association with SLE (Table 3 and Supplementary Table 3). No single locus had an unadjusted P value $< 5 \times 10^{-8}$; however, we found an enrichment of previously identified disease-associated alleles. From the 35 loci tested (42 total variants), there were 5 alleles with unadjusted $P < 0.0004$ (less than 1 expected by chance, $P = 4.4 \times 10^{-12}$) and with $P < 0.05$ after a Bonferroni correction for the 35 pre-specified loci. For each of the five variants, the SLE-associated allele matched a previously reported allele and had the same direction of effect (Table 3). We observed a highly significant association to SLE of a missense allele of *IFIH1* (rs1990760, $P = 3.3 \times 10^{-7}$) that has previously been associated with type 1 diabetes and Graves' disease[27,28]. We also observed an association of SLE with a missense allele (R32Q) of *CFB* (complement factor B, rs641153) that resides in the HLA class III region and is a validated risk allele for age-related macular degeneration[29]. This missense allele in *CFB* is not in significant LD with other HLA region variants associated with SLE (DR2/DR3) and remained significant ($P < 0.05$) after conditional logistic regression analyses that incorporated DR2 and DR3. The HLA is a complex genetic region, but it is noteworthy that the rs641153 SNP has a protective effect nearly identical to that of the reported age-related macular degeneration (AMD) risk allele[29]. Further validation of the five candidate disease alleles is required.

Using 26 SLE risk alleles (21 previously reported loci in Table 1 plus the 5 newly identified SLE loci), we performed several additional analyses. First, we performed pairwise interaction analysis with the previously confirmed loci, and, consistent with previous literature from SLE[6] and other complex diseases[30], we observed no evidence for non-additive interactions. Using conditional logistic regression analyses, we found no evidence for multiple independent alleles contributing to risk at any of the individual risk loci. We next estimated the percent of variance explained by each of the confirmed SLE risk alleles using previously described methods[30]. *HLA-DRB1* (HLA*DR3), *IRF5* and *STAT4* were each estimated to account for over 1% of the genetic variance, whereas the remaining loci each accounted for less than 1% of the variance. Together, the 26 SLE risk loci explain an estimated 8% of the total genetic susceptibility to SLE.

Targeted replication of GWAS results is an efficient study design to confirm additional risk loci[31]. However, there are few available data as to the probability of replicating results that fall short of accepted *P* value criteria for genome-wide significance. In the current study, all variants with $P < 0.05$ from the original GWAS were included for replication. The lower a locus' *P* value is in the GWAS, the higher is the probability of that locus reaching candidate or confirmed status in the replication meta-analysis (Fig. 3). Of note, no candidate or confirmed loci were obtained in the current study from the group of variants with a GWAS *P* value between 0.05 and 0.01, despite accounting for ~50% of all variants tested in the replication. These results may be useful in helping guide future targeted study designs, although clearly the size of the original GWAS population, the replication sample size, the disease architecture and the effect size of the candidate disease-associated variants need to be carefully considered in planning replication efforts.

These data provide further evidence that common variation in genes that function in the adaptive and innate arms of the immune system are important in establishing SLE risk. Although each of the identified alleles accounts for only a fraction of the overall genetic risk, these and other ongoing studies are providing new insight into the pathogenesis of lupus and are suggesting new targets and pathways for drug discovery and development.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the many affected individuals and physicians who contributed DNA samples and clinical data for this study; M.I. Kamboh and P. Davies for the use of Alzheimer's disease samples as controls in our study; B. Neale for assistance in the percent of genetic variance explained calculation; and S. Sanna and C. Willer for assistance in generating regional association plots. Genotyping of the Swedish samples by the 12K chips was performed using equipment of the SNP technology platform in Uppsala. We thank C. Enström and A.-C. Wiman for assistance with genotyping. Financial support was obtained from the Swedish Research Council for Medicine, the Knut and Alice Wallenberg Foundation the Swedish Rheumatism Association, the King Gustaf V 80th Birthday Foundation, COMBINE, and a Target Identification in Lupus (TIL) grant from the Alliance for Lupus Research, US. This work was supported in part by R01 AR44804, K24 AR02175, the Mary Kirkland Center for Lupus Research, RO1 AR43727 and Institute for Clinical and Translational Research UL1RR025005. These studies were performed in part in the General Clinical Research Center, Moffitt Hospital, University of California, San Francisco, with funds provided by the National Center for Research Resources, 5 M01 RR-00079, US Public Health Service.

References

1. Rönnblom L, Pascual V. The innate immune system in SLE: type I interferons and dendritic cells. Lupus 2008;17:394–399. [PubMed: 18490415]
2. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell 1996;85:311–318. [PubMed: 8616887]
3. Cunninghame Graham DS, et al. Polymorphism at the TNF superfamily gene *OX40L* confers susceptibility to systemic lupus erythematosus. Nat Genet 2008;40:83–89. [PubMed: 18059267]
4. Graham RR, et al. Genetic variants near *TNFAIP3* on 6q23 are associated with systemic lupus erythematosus. Nat Genet 2008;40:1059–1061. [PubMed: 19165918]
5. Graham RR, Hom G, Ortmann W, Behrens TW. Review of recent genome-wide association scans in lupus. J Intern Med 2009;265:680–688. [PubMed: 19493061]
6. Harley JB, 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;40:204–210. [PubMed: 18204446]
7. Hom G, et al. Association of systemic lupus erythematosus with *C8orf13-BLK* and *ITGAM-ITGAX*. N Engl J Med 2008;358:900–909. [PubMed: 18204098]
8. Kozyrev SV, et al. Functional variants in the B-cell gene *BANK1* are associated with systemic lupus erythematosus. Nat Genet 2008;40:211–216. erratum**40**, 484 (2004). [PubMed: 18204447]
9. Sawalha AH, et al. Common variants within *MECP2* confer risk of systemic lupus erythematosus. PLoS ONE 2008;3:e1727. [PubMed: 18320046]
10. Sigurdsson S, et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet 2005;76:528–537. [PubMed: 15657875]
11. Jacob CO, et al. Identification of *IRAK1* as a risk gene with critical role in the pathogenesis of systemic lupus erythematosus. Proc Natl Acad Sci USA 2009;106:6256–6261. [PubMed: 19329491]
12. Nair RP, et al. Genome-wide scan reveals association of psoriasis with IL-23 and NF-κB pathways. Nat Genet 2009;41:199–204. [PubMed: 19169254]
13. Heyninck K, Kreike MM, Beyaert R. Structure-function analysis of the A20-binding inhibitor of NF-κB activation, ABIN-1. FEBS Lett 2003;536:135–140. [PubMed: 12586352]
14. Musone SL, et al. Multiple polymorphisms in the *TNFAIP3* region are independently associated with systemic lupus erythematosus. Nat Genet 2008;40:1062–1064. [PubMed: 19165919]
15. Plenge RM, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–1482. [PubMed: 17982456]
16. Fung EY, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/*TNFAIP3* as a susceptibility locus. Genes Immun 2009;10:188–191. [PubMed: 19110536]
17. Zeggini E, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet 2008;40:638–645. [PubMed: 18372903]

18. Johansson A, et al. Common variants in the *JAZF1* gene associated with height identified by linkage and genome-wide association analysis. Hum Mol Genet 2009;18:373–380. [PubMed: 18952825]
19. Thomas G, et al. Multiple loci identified in a genome-wide association study of prostate cancer. Nat Genet 2008;40:310–315. [PubMed: 18264096]
20. Arita K, Ariyoshi M, Tochio H, Nakamura Y, Shirakawa M. Recognition of hemi-methylated DNA by the SRA protein UHRF1 by a base-flipping mechanism. Nature 2008;455:818–821. [PubMed: 18772891]
21. Diveu C, McGeachy MJ, Cua DJ. Cytokines that regulate autoimmunity. Curr Opin Immunol 2008;20:663–668. [PubMed: 18834938]
22. Nath SK, Harley JB, Lee YH. Polymorphisms of complement receptor 1 and interleukin-10 genes and systemic lupus erythematosus: a meta-analysis. Hum Genet 2005;118:225–234. [PubMed: 16133175]
23. Franke A, et al. Sequence variants in *IL10, ARPC2* and multiple other loci contribute to ulcerative colitis susceptibility. Nat Genet 2008;40:1319–1323. [PubMed: 18836448]
24. Barrett JC, et al. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–707.
25. Duerr RH, et al. A genome-wide association study identifies *IL23R* as an inflammatory bowel disease gene. Science 2006;314:1461–1463. [PubMed: 17068223]
26. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 2009;10:43–55. [PubMed: 19092835]
27. Smyth DJ, et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. Nat Genet 2006;38:617–619. [PubMed: 16699517]
28. Sutherland A, et al. Genomic polymorphism at the interferon-induced helicase (*IFIH1*) locus contributes to Graves' disease susceptibility. J Clin Endocrinol Metab 2007;92:3338–3341. [PubMed: 17535987]
29. Gold B, et al. Variation in factor B (*BF*) and complement component 2 (*C2*) genes is associated with age-related macular degeneration. Nat Genet 2006;38:458–462. [PubMed: 16518403]
30. Barrett JC, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955–962. [PubMed: 18587394]
31. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 2005;6:95–108. [PubMed: 15716906]
32. Awata T, et al. Association of type 1 diabetes with two loci on 12q13 and 16p13 and the influence of coexisting thyroid autoimmunity in Japanese. J Clin Endocrinol Metab 2009;94:231–235. [PubMed: 18940880]
33. Skinningsrud B, et al. Polymorphisms in *CLEC16A* and *CIITA* at 16p13 are associated with primary adrenal insufficiency. J Clin Endocrinol Metab 2008;93:3310–3317. [PubMed: 18593762]
34. Zoledziewska M, et al. Variation within the *CLEC16A* gene shows consistent disease association with both multiple sclerosis and type 1 diabetes in Sardinia. Genes Immun 2009;10:15–17. [PubMed: 18946483]
35. Fisher SA, et al. Genetic determinants of ulcerative colitis include the *ECM1* locus and five loci implicated in Crohn's disease. Nat Genet 2008;40:710–712. [PubMed: 18438406]
36. Hunt KA, et al. Newly identified genetic risk variants for celiac disease related to the immune response. Nat Genet 2008;40:395–402. [PubMed: 18311140]
37. Smyth DJ, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 2008;359:2767–2777. [PubMed: 19073967]

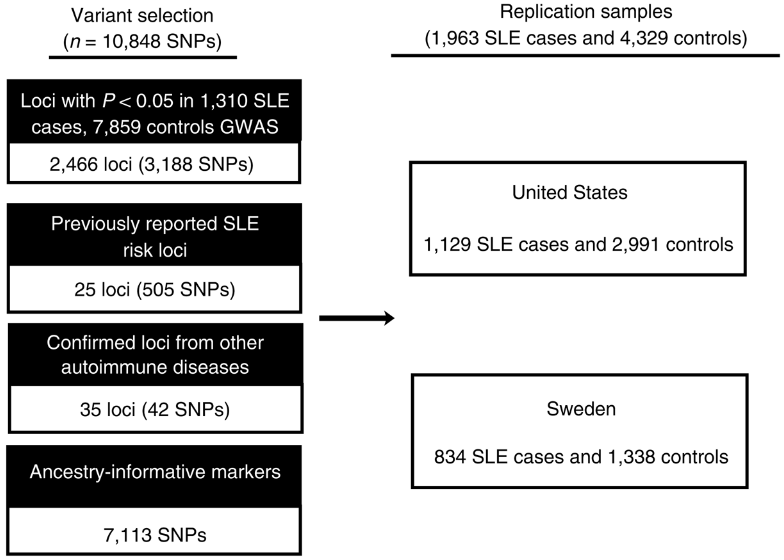


Figure 1.

Overview of experimental design. Variants were selected from loci with $P < 0.05$ in a genome-wide scan of 1,310 cases and 7,859 controls, previously reported SLE risk loci, confirmed loci from other autoimmune diseases and over 7,000 ancestry-informative markers, and these variants were incorporated into an Illumina custom SNP array. The array was genotyped in independent cases and controls from the United States and Sweden. 823 of the Swedish controls were genotyped using the Illumina 310K SNP array. Variants were analyzed as described in Online Methods.

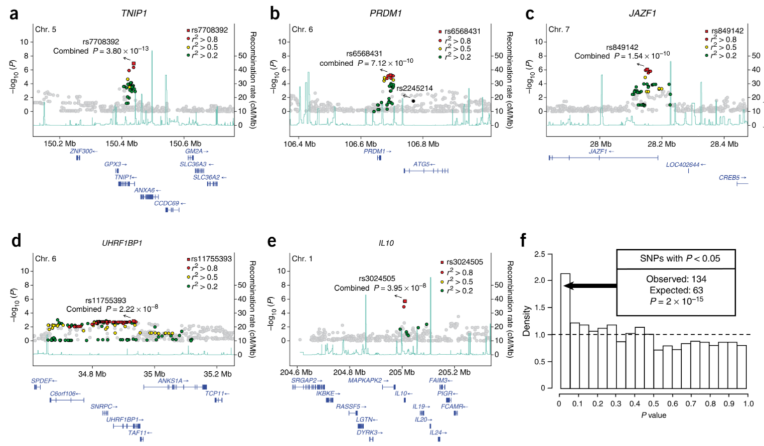
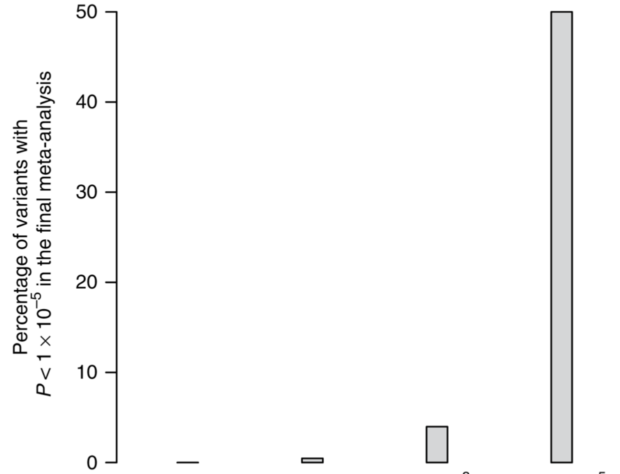


Figure 2.

Newly discovered genome-wide significant associations in SLE. (**a**–**e**) Association results from the GWA scan are plotted on the *y* axis versus genomic position on the indicated chromosome on the *x* axis within a 500-kb region surrounding the loci defined by (**a**) *TNIP1*, (**b**) *PRDM1*, (**c**) *JAZF1*, (**d**) *UHRF1BP1* and (**e**) IL10. The meta-analysis *P* value for the most strongly associated marker is indicated by a red square. *P* values from the genome scan are shaded to indicate LD to the genome-wide associated variant: red, $r^2 > 0.8$; yellow, $r^2 > 0.5$; green, $r^2 > 0.2$; gray, $r^2 < 0.2$. Along the bottom are the recombination rates from the CEU HapMap (light blue line) and the known human genes (blue). A previously reported and independent SLE risk locus at the nearby *ATG5* gene is indicated (**b**; rs2245214). (**f**) Histogram of *P* values of 1,256 independent SNPs ($r^2 < 0.1$ to any other SNP in the array) in the 1,963 case and 4,329 control replication samples. Under a null distribution, the expected density of results is indicated by the dashed line. A significant enrichment of results with $P < 0.05$ was observed.



P value in GWAS	0.05 – 0.01	0.01 – 0.001	1×10^{-3} – 1×10^{-5}	1×10^{-5} – 5×10^{-8}
n of independent variants	1,220	863	377	14
n of variants with meta $P < 1 \times 10^{-5}$	0	3	14	4
n of variants with meta $P < 5 \times 10^{-8}$	0	1	1	3

Figure 3.
Percentage of newly discovered variants reaching candidate ($P < 1 \times 10^{-5}$) and confirmed ($P < 5 \times 10^{-8}$) status in the meta-analysis stratified by the P value in the original GWAS.

Table 1

Replication results of previously reported SLE risk loci

SNP	Chr	Critical region	*P* GWAS	*P* US	*P* Sweden	*P* Combined	Gene of interest	Risk allele	Risk allele frequency	OR (95% CI)
Variants with $P < 5 \times 10^{-8}$ in the current dataset										
rs3135394[a]	6p21.32	32.027–32.874	7.8×10^{-22}	1.8×10^{-26}	8.3×10^{-21}	2.0×10^{-60}	*HLA-DRB1*[b]	G	0.10	1.98 (1.84–2.14)
rs7574865[a]	2q32.2	191.609–191.681	3.0×10^{-19}	6.4×10^{-16}	2.7×10^{-12}	1.4×10^{-41}	*STAT4*	T	0.23	1.57 (1.49–1.69)
rs2070197[a]	7q32.1	128.276–128.476	n.a.	1.4×10^{-16}	4.1×10^{-9}	5.8×10^{-24}	*IRF5*	C	0.11	1.88 (1.78–1.95)
rs11860650[a]	16p11.2	31.195–31.277	5.3×10^{-11}	1.8×10^{-5}	9.2×10^{-8}	1.9×10^{-20}	*ITGAM*	T	0.13	1.43 (1.32–1.54)
rs2736340	8p23.1	11.331–11.488	5.5×10^{-8}	4.6×10^{-9}	0.0028	7.9×10^{-17}	*BLK*	T	0.25	1.35 (1.27–1.43)
rs5029937[a]	6q23.3	138.174–138.284	1.0×10^{-4}	2.4×10^{-7}	3.1×10^{-5}	5.3×10^{-13}	*TNFAIP3*	T	0.03	1.71 (1.51–1.95)
rs2476601	1p13.2	113.963–114.251	3.3×10^{-5}	4.5×10^{-5}	1.5×10^{-5}	3.4×10^{-12}	*PTPN22*	A	0.10	1.35 (1.24–1.47)
rs4963128	11p15.5	0.485–0.664	0.0021	1.5×10^{-5}	8.7×10^{-4}	4.9×10^{-9}	*PHRF1*	C	0.67	1.20 (1.13–1.27)
rs2205960	1q25.1	171.454–171.523	9.5×10^{-6}	0.030	6.7×10^{-4}	6.3×10^{-9}	*TNFSF4*	T	0.23	1.22 (1.15–1.30)
Variants with a previous report of $P < 5 \times 10^{-8}$										
rs9271366[a]	6p21.32	32.446–32.695	0.0079	7.4×10^{-4}	8.3×10^{-5}	1.4×10^{-7}	*HLA-DRB1*[c]	G	0.16	1.26 (1.18–1.36)
rs6920220[a]	6q23.3	138.000–138.048	9.9×10^{-4}	5.2×10^{-4}	0.049	4.0×10^{-7}	*TNFAIP3*	A	0.21	1.17 (1.10–1.25)
rs2269368	Xq28	152.743–152.943	2.5×10^{-5}	n.a.	0.0049	7.5×10^{-7}	*IRAK1-MECP2*	T	0.14	1.11 (1.01–1.22)
rs2431099	5q33.3	159.813–159.821	1.5×10^{-5}	0.16	0.047	1.6×10^{-6}	*PTTG1*	G	0.52	1.15 (1.09–1.22)
rs5754217	22q11.2	20.240–20.315	0.0060	8.4×10^{-4}	0.018	2.3×10^{-6}	*UBE2L3*	T	0.19	1.20 (1.13–1.27)
rs2245214[a]	6q21	106.749–106.876	0.032	4.3×10^{-6}	0.35	1.2×10^{-5}	*ATG5*	G	0.37	1.15 (1.09–1.21)
rs10516487	4q24	102.930–103.134	0.097	0.091	0.0015	8.3×10^{-4}	*BANK1*	G	0.70	1.11 (1.04–1.18)
rs2176082[a]	3p14.3	58.214–58.443	0.010	0.012	0.0031	1.2×10^{-5}	*PXK*	A	0.28	1.17 (1.10–1.25)
rs1801274	1q23.3	159.724–159.746	4.1×10^{-4}	n.a.	n.a.	4.1×10^{-4}	*FCGR2A*	G	0.50	1.16 (1.09–1.20)
Variants with a previous report of $P > 5 \times 10^{-8}$										
rs280519[a]	19p13.2	10.387–10.430	7.1×10^{-4}	n.a.	0.036	7.4×10^{-5}	*TYK2*	A	0.48	1.13 (1.06–1.21)
rs10156091	7p21.3	8.134–8.154	0.095	0.0031	8.7×10^{-4}	6.5×10^{-4}	*ICA1*	T	0.10	1.16 (1.06–1.27)
rs2022013	1q25.3	181.538–181.670	0.26	2.05×10^{-5}	2.8×10^{-4}	0.0015	*NMNAT2*	T	0.60	1.09 (1.03–1.16)
rs7829816	8q12.1	56.985–57.025	0.49	0.76	0.19	0.17	*LYN*	A	0.79	1.05 (0.96–1.17)

SNP	Chr	Critical region	*P*				Gene of interest	Risk allele	Risk allele frequency	OR (95% CI)
			GWAS	US	Sweden	Combined				
rs2071725	22q13.2	41.908–41.970	0.63	0.34	0.29	0.30	*SCUBE1*	G	0.86	1.09 (0.98–1.20)
rs5744168[a]	1q41	n.a.	n.a.	1.00	0.40	0.67	*TLR5*	G	0.94	1.02 (0.94–1.12)
rs509749	1q23.3	158.993–159.067	0.64	0.94	0.93	0.76	*LY9*	G	0.96	1.01 (0.91–1.12)

Critical region here is defined as the minimal region containing variants with $r^2 > 0.4$ in the HapMap CEU population and is reported in HG18 coordinates (Mb). *P* values calculated from indicated case-control population (GWAS: 1,310 cases and 7,859 controls; US: 1,129 cases and 2,991 controls; Sweden: 834 cases and 1,338 controls; combined: 3,273 cases and 12,188 controls) and combined *P* values were calculated as described in Online Methods. Risk allele is reported relative to + reference strand. Risk allele frequency is the frequency in control chromosomes. OR is the combined odds ratio as described in Online Methods.

[a]Indicates markers that were imputed, as described in Online Methods, in the GWAS samples and directly genotyped in the replication samples.

[b]rs3135394 has an $r^2 = 0.87$ to the HLA*DR3 (DRB1*0301) allele.

[c]rs9271366 has an $r^2 = 0.97$ to the HLA*DR2 (DRB1*1501) allele. See Supplementary Table 1 for expanded summary statistics. n.a, not available due to failure to pass quality control measures (*TYK2*, *FCGR2A* and *IRAK1-MECP2*), or the specific variant was not present in the genome-wide array (*TLR5* and *IRF5*); however, rs2070197 (*IRF5* region) is in strong LD with rs10488631, which had a $P = 2 \times 10^{-11}$ in the genome scan.

Table 2

Newly discovered SLE risk loci in the combined dataset

			P							
SNP	**Chr.**	**Critical region**	**GWAS**	**US**	**Sweden**	**Combined**	**Gene of interest**	**Risk allele**	**Risk allele frequency**	**OR (95% CI)**
Genome-wide significant loci										
rs7708392[a]	5	150.419–150.441	4.5×10^{-7}	7.7×10^{-4}	1.2×10^{-5}	3.8×10^{-13}	*TNIP1*	C	0.24	1.27 (1.10–1.35)
rs6568431	6	106.675–106.705	6.1×10^{-6}	0.0016	0.0050	7.1×10^{-10}	*PRDM1*	A	0.38	1.20 (1.14–1.27)
rs849142[a]	7	28.108–28.223	4.5×10^{-7}	0.10	5.4×10^{-4}	1.5×10^{-9}	*JAZF1*	T	0.49	1.19 (1.13–1.26)
rs11755393[a]	6	34.658–35.090	0.0014	3.7×10^{-4}	5.1×10^{-4}	2.2×10^{-8}	*UHRF1BP1*	G	0.35	1.17 (1.10–1.24)
rs3024505	1	205.007–205.016	2.6×10^{-6}	0.062	1.8×10^{-4}	4.0×10^{-8}	*IL10*	A	0.16	1.19 (1.11–1.28)
Loci with combined *P* value < 1×10^{-6}										
rs10911363[a]	1	181.672–181.816	2.0×10^{-4}	1.5×10^{-5}	0.52	9.5×10^{-8}	*NCF2*	T	0.27	1.19 (1.12–1.26)
rs12444486[a]	16	84.548–84.576	3.5×10^{-5}	0.021	0.026	1.9×10^{-7}	*IRF8*	T	0.50	1.16 (1.10–1.23)
rs11013210[a]	10	23.181–23.337	1.6×10^{-5}	0.013	0.12	2.0×10^{-7}	*ARMC3*	T	0.21	1.18 (1.11–1.26)
rs1874791[a]	1	67.563–67.687	3.1×10^{-5}	0.012	0.11	3.4×10^{-7}	*IL12RB2*	A	0.18	1.18 (1.10–1.26)
rs9782955	1	233.893–234.107	6.4×10^{-6}	0.057	0.12	4.6×10^{-7}	*LYST*	C	0.74	1.18 (1.11–1.26)
rs7683537[a]	4	185.805–185.914	1.6×10^{-4}	0.11	0.0013	7.6×10^{-7}	*MLF1IP*	T	0.82	1.23 (1.14–1.33)
rs428073	12	117.706–117.315	1.7×10^{-5}	0.22	0.0079	7.7×10^{-7}	*TAOK3*	T	0.69	1.18 (1.11–1.26)
rs497273[a]	12	119.610–119.891	5.0×10^{-5}	0.068	0.021	8.2×10^{-7}	*SPPL3*	G	0.65	1.14 (1.08–1.21)
Loci with combined *P* value < 1×10^{-5}										
rs1861525	7	25.097–25.183	8.5×10^{-5}	0.16	0.0027	1.9×10^{-6}	*CYCS*	G	0.05	1.27 (1.12–1.45)
rs921916	7	50.193–50.205	4.8×10^{-4}	0.027	0.014	2.0×10^{-6}	*IKZF1*	C	0.18	1.15 (1.07–1.23)
rs7333671	13	73.177–73.198	2.2×10^{-4}	0.14	0.0027	2.2×10^{-6}	*KLF12*	G	0.08	1.22 (1.11–1.34)
rs12992463	2	22.312–22.464	2.1×10^{-5}	0.23	0.023	2.6×10^{-6}	–	A	0.50	1.12 (1.06–1.19)
rs12620999	2	237.616–237.770	1.6×10^{-5}	0.040	0.45	3.1×10^{-6}	*COPS8*	C	0.19	1.13 (1.06–1.21)
rs503425[a]	11	118.079–118.198	0.0012	3.3×10^{-4}	0.43	3.3×10^{-6}	*DDX6*	C	0.20	1.16 (1.08–1.24)
rs10742326[a]	11	34.733–34.809	1.4×10^{-4}	0.017	0.21	3.6×10^{-6}	*APIP*	G	0.59	1.14 (1.08–1.21)
rs4766921[a]	12	117.835–117.883	4.6×10^{-5}	n.a.	0.036	4.6×10^{-6}	*KIAA1853*	G	0.67	1.18 (1.09–1.27)

			P							
SNP	**Chr.**	**Critical region**	**GWAS**	**US**	**Sweden**	**Combined**	**Gene of interest**	**Risk allele**	**Risk allele frequency**	**OR (95% CI)**
rs11951576[a]	5	6.741–6.866	2.5×10^{-5}	0.42	0.014	4.6×10^{-6}	*POLS-SRD5A*	C	0.69	1.14 (1.08–1.22)
rs6438700	3	123.355–123.454	7.4×10^{-5}	0.23	0.020	5.5×10^{-6}	*CD86*	C	0.82	1.18 (1.09–1.27)
rs6486730[a]	12	127.830–127.840	8.2×10^{-5}	0.16	0.049	6.9×10^{-6}	*SLC15A4*	G	0.41	1.13 (1.07–1.19)
rs4748857[a]	10	23.529–23.654	2.2×10^{-4}	0.68	1.3×10^{-4}	6.9×10^{-6}	*C10orf67*	C	0.73	1.16 (1.09–1.24)
rs3914167[a]	5	39.426–39.454	1.8×10^{-4}	0.24	0.0081	7.6×10^{-6}	*DAB2-C9*	G	0.27	1.15 (1.09–1.23)

Samples, critical region, *P* values, risk alleles and ORs are as defined in the Table 1 legend.

[a] Indicates markers that were imputed, as described in Online Methods, from the GWAS samples and directly genotyped in the replication samples. See Supplementary Table 2 for expanded summary statistics.

Table 3

Candidate autoimmune loci with evidence of association to SLE

SNP	Gene	Chr	*P*				Combined corrected	Risk allele	Risk allele frequency	OR	Phenotype	References
			GWAS	US	Sweden	Combined						
rs1990760	*IFIH1*	2	3.2×10^{-5}	0.015	0.0039	3.34×10^{-7}	1.12×10^{-5}	T	0.60	1.17	T1D, Graves'	27,28
rs641153[a]	*CFB*	6	0.0079	n.a.	0.0011	1.4×10^{-4}	0.0049	G	0.91	1.30	AMD	29
rs12708716[a]	*CLEC16A*	16	0.15	1.3×10^{-4}	0.062	1.6×10^{-4}	0.0056	A	0.64	1.16	T1D, Addison's, MS	32–34
rs6887695[a]	*IL12B*	5	0.014	0.04	0.03	1.7×10^{-4}	0.0060	G	0.68	1.13	Psoriasis, IBD	12,35
rs17696736	*SH2B3*	12	0.0036	0.12	0.19	4.0×10^{-4}	0.014	T	0.50	1.08	T1D, Celiac, SLE	33,36,37

All alleles in the table either were identical to the reported variants or have $r^2 > 0.8$ to the reported variant and were the same risk allele with the same direction of effect. Samples, individual and combined *P* values, risk allele frequency and OR are as described in Table 1 legend. Combined-corrected *P* value is the Bonferroni-corrected *P* value for the 35 previously reported risk loci. Other autoimmunity associations: T1D, type 1 diabetes; AMD, age-related macular degeneration; MS, multiple sclerosis; IBD, inflammatory bowel disease. See Supplementary Table 3 for expanded summary statistics and a complete list of variants tested.

[a] Indicates markers that were imputed, as described in Online Methods, from the GWAS samples and directly genotyped in the replication samples.