

ASSOCIATION STUDIES ARTICLE

A plausibly causal functional lupus-associated risk variant in the *STAT1–STAT4* locus

Zubin H. Patel^{1,2,3}, Xiaoming Lu^{1,2}, Daniel Miller², Carmy R. Forney², Joshua Lee², Arthur Lynch², Connor Schroeder², Lois Parks², Albert F. Magnusen², Xiaoting Chen², Mario Pujato², Avery Maddox², Erin E. Zoller², Bahram Namjou^{2,36}, Hermine I. Brunner⁴, Michael Henrickson⁴, Jennifer L. Huggins⁴, Adrienne H. Williams⁵, Julie T. Ziegler⁵, Mary E. Comeau⁵, Miranda C. Marion⁵, Stuart B. Glenn⁶, Adam Adler⁶, Nan Shen^{2,7,36}, Swapan K. Nath⁶, Anne M. Stevens^{8,9}, Barry I. Freedman¹⁰, Bernardo A. Pons-Estel¹¹, Betty P. Tsao¹², Chaim O. Jacob¹³, Diane L. Kamen¹², Elizabeth E. Brown^{14,15}, Gary S. Gilkeson¹², Graciela S. Alarcón¹⁵, Javier Martin¹⁶, John D. Reveille¹⁷, Juan-Manuel Anaya¹⁸, Judith A. James^{6,19,20}, Kathy L. Sivils⁶, Lindsey A. Criswell²¹, Luis M. Vilá²², Michelle Petri²³, R. Hal Scofield^{6,19,24}, Robert P. Kimberly¹⁵, Jeffrey C. Edberg¹⁵, Rosalind Ramsey-Goldman²⁵, So-Young Bang²⁶, Hye-Soon Lee²⁶, Sang-Cheol Bae²⁶, Susan A. Boackle²⁷, Deborah Cunninghame Graham²⁸, Timothy J. Vyse²⁸, Joan T. Merrill^{6,19}, Timothy B. Niewold²⁹, Hannah C. Ainsworth⁵, Earl D. Silverman³⁰, Michael H. Weisman³¹, Daniel J. Wallace³¹, Prithvi Raj³², Joel M. Guthridge^{6,19,20}, Patrick M. Gaffney⁶, Jennifer A. Kelly⁶, Marta E. Alarcón-Riquelme^{33,34}, Carl D. Langefeld⁵, Edward K. Wakeland³², Kenneth M. Kaufman^{2,35,36,37}, Matthew T. Weirauch^{2,36,37,†}, John B. Harley^{2,35,36,†} and Leah C. Kottyan^{2,36,*,†}

¹Immunology Graduate Program, University of Cincinnati College of Medicine, Cincinnati, OH 45229, USA,

²Center for Autoimmune Genomics and Etiology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA, ³Medical Scientist Training Program, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA, ⁴Division of Rheumatology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA, ⁵Center for Public Health Genomics and the Department of Biostatistical Sciences, Wake Forest School of

[†]These authors contributed equally to this work and are joint Senior Authors.

Received: October 9, 2017. Revised: March 21, 2018. Accepted: April 13, 2018

© The Author(s) 2018. Published by Oxford University Press. All rights reserved. For permissions, please email: journals.permissions@oup.com

Medicine, Winston-Salem, NC 27157, USA, ⁶Arthritis and Clinical Immunology Research Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA, ⁷Shanghai Institute of Rheumatology, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai 200001, P.R. China, ⁸Center for Immunity and Immunotherapies, Seattle Children's Research Institute, Seattle, WA 98101, USA, ⁹Division of Rheumatology, Department of Pediatrics, University of Washington, Seattle, WA 98195, USA, ¹⁰Section on Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC 27157, USA, ¹¹Servicio de Reumatología, Sanatorio Parque, Rosario 2000, Santa Fe, Argentina, ¹²Division of Rheumatology and Immunology, Medical University of South Carolina, Charleston, SC 29425, USA, ¹³Department of Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA, ¹⁴Division of Molecular and Cellular Pathology, Department of Pathology and ¹⁵Department of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA, ¹⁶Instituto de Parasitología y Biomedicina Lopez-Neyra, CSIC, Granada 18001-18016, Spain, ¹⁷Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center at Houston, Houston, TX 77030, USA, ¹⁸Center for Autoimmune Diseases Research (CREA), Universidad del Rosario, Bogota 111711, Colombia, ¹⁹Department of Medicine and ²⁰Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104, USA, ²¹Department of Medicine, Rosalind Russell/Ephraim P Engleman Rheumatology Research Center, University of California San Francisco, San Francisco, CA 94143-0500, USA, ²²Division of Rheumatology, Department of Medicine, University of Puerto Rico Medical Sciences Campus, San Juan, PR 00936, USA, ²³Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21224, USA, ²⁴United States Department of Veterans Affairs Medical Center, Oklahoma City, OK 73104, USA, ²⁵Division of Rheumatology, Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA, ²⁶Department of Rheumatology, Hanyang University Hospital for Rheumatic Diseases, Seoul 04763, Korea, ²⁷Division of Rheumatology, University of Colorado School of Medicine, Aurora, CO 80045, USA, ²⁸Divisions of Genetics/Molecular Medicine and Immunology, King's College London, Guy's Hospital, London SE1 9RT, UK, ²⁹Division of Rheumatology, Department of Pathology, New York University, New York, NY 10016, USA, ³⁰Division of Rheumatology, The Hospital for Sick Children, Hospital for Sick Research Institute, University of Toronto, Toronto, ON M5G 1X8, Canada, ³¹Division of Rheumatology, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA, ³²Department of Immunology, University of Texas Southwestern Medical Center, Dallas, TX 75390, USA, ³³Unit of Chronic Inflammatory Diseases, Institute of Environmental Medicine, Karolinska Institutet, Stockholm 17167, Sweden, ³⁴Center for Genomics and Oncological Research, Pfizer-University of Granada-Junta de Andalucía, Parque Tecnológica de la Salud, Granada 18016, Spain, ³⁵United States Department of Veterans Affairs Medical Center, Cincinnati, OH 45220, USA, ³⁶Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH 45220, USA and ³⁷Divisions of Biomedical Informatics and Developmental Biology, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

*To whom correspondence should be addressed. Tel: 513-636-1316; Fax: 513-803-5246; Email: leah.kottyan@cchmc.org

Abstract

Systemic lupus erythematosus (SLE or lupus) (OMIM: 152700) is a chronic autoimmune disease with debilitating inflammation that affects multiple organ systems. The *STAT1-STAT4* locus is one of the first and most highly replicated genetic loci associated with lupus risk. We performed a fine-mapping study to identify plausible causal variants within the *STAT1-STAT4* locus associated with increased lupus disease risk. Using complementary frequentist and Bayesian approaches in trans-ancestral Discovery and Replication cohorts, we found one variant whose association with lupus risk is supported across ancestries in both the Discovery and Replication cohorts: rs11889341. In B cell lines from patients with lupus and healthy controls, the lupus risk allele of rs11889341 was associated with increased *STAT1* expression. We demonstrated that the transcription factor HMGA1, a member of the HMG transcription factor family with an AT-hook DNA-binding domain, has enriched binding to the risk allele compared with the non-risk allele of rs11889341. We identified a genotype-dependent repressive element in the DNA within the intron of *STAT4* surrounding rs11889341. Consistent with expression quantitative trait locus (eQTL) analysis, the lupus risk allele of rs11889341 decreased the activity of this putative repressor. Altogether, we present a plausible molecular mechanism for increased lupus risk at the *STAT1-STAT4* locus in which the risk allele of rs11889341, the most probable causal variant, leads to elevated *STAT1* expression in B cells due to decreased repressor activity mediated by increased binding of HMGA1.

Introduction

Systemic lupus erythematosus (SLE) or lupus is an autoimmune disorder characterized by multiple organ system pathology mediated by exaggerated antibody responses to self-antigens. The pathoetiology of lupus is postulated to be driven by environmental factors in the context of a genetic-risk background. Epidemiological studies support the genetic component of lupus risk. Specifically, the sibling risk ratio (λ_s) is 8–30, and monozygotic twins of individuals with lupus have a disease concordance rate of 20–60% (1–7). Many family-based and large cohort studies [e.g. candidate and genome-wide association studies (GWAS)] have identified tagging variants throughout the genome that contribute to lupus disease risk. These studies have identified over 85 different loci, including variants on the long arm of chromosome 2 spanning the *STAT1* and *STAT4* genes (8–10).

Unbiased genomic analyses have revealed an enrichment of activating chromatin marks and expressed gene products in B cells at lupus risk loci, including the *STAT1*–*STAT4* locus (11,12). B cells are critical cells in the development and pathogenesis of lupus. Patients with lupus have many autoantibodies produced from B cells that have eluded immunological tolerance. These autoantibodies are pathogenic in patients with lupus and result in immune complex deposition and inappropriate, self-directed inflammatory responses (13).

The *STAT1*–*STAT4* locus has been associated with increased lupus disease-risk in all major ancestries, although association in the African American ancestry is weaker ($P < 0.001$) and has never been reported at a genome-wide significant level (14–30). Most of these studies have used the ‘tag’ variants within the third intron of the *STAT4* gene (20) to establish a significantly higher minor allele frequency (MAF) in subjects with lupus compared to control subjects (14,18,19,21,22,24,31–38). Mechanistically, little is known about how genetic variation at this locus increases lupus disease risk. A recent study demonstrated that a haplotype of lupus-risk variants at this locus are associated with increased expression of both *STAT1* and *STAT4* in monocytes, and the haplotype containing these variants is the only disease-associated haplotype in the region (39).

In this study, we performed a fine-mapping analysis using frequentist and Bayesian statistical methods with an aim to identify a candidate causal variant by interrogating all common genetic variants (MAF > 0.01) within the *STAT1*–*STAT4* locus (GRCh37 chr2: 191,700,000–192,100,000) in multi-ethnic Discovery and Replication cohorts. These analyses revealed a single plausibly causal genetic variant located within intron 3 of *STAT4*, rs11889341. We established a *STAT1* cis-eQTL involving rs11889341 in Epstein-Barr Virus transformed lymphoblastoid cell lines (LCLs) generated from patients with lupus and healthy controls. We further identified allele-dependent differential binding of the HMGA1 transcription factor at the candidate causal variant. Finally, we demonstrated that the genomic region within intron 3 of *STAT4* containing this candidate causal variant represses gene expression in a genotype-dependent fashion, with the risk allele attenuating the repression. Together, this study provides a plausible biological mechanism through which a lupus risk variant at the *STAT1*–*STAT4* locus leads to increased *STAT1* expression in B cells.

Results

We genotyped 327 variants at the *STAT1*–*STAT4* locus in a multi-ancestral discovery cohort of 13577 individuals. This

cohort included individuals of European & European American, Asian & Asian American, African American, and Amerindian ancestry. We used data from The 1000 Genomes Project to impute an additional 485 variants with MAF > 0.01 and a genotyping rate > 90%. We used a total of 812 variants for model building and genetic analysis with the goal of identifying candidate causal variants to explain association at this locus with increased lupus disease risk.

First, we used a logistic regression model with the admixture estimates as covariates to identify variants highly associated with increased lupus disease risk. We identified genome-wide significant association in the European & European American, Asian & Asian American, and Amerindian cohorts, but did not see a statistically robust association of genetic variants in individuals of African American ancestry (Fig. 1, Supplementary Material, Table S1). To identify the number of genetic effects present within the locus, we performed a step-wise logistic regression analysis starting with rs11889341 as a covariate in the European, Asian, and Amerindian cohort. The genotype of rs11889341 was sufficient to account for virtually all of the lupus association at this locus (>10 orders of magnitude) in each of the ancestral cohorts (Supplementary Material, Figs S1 and S2, and Table S1). A small number of genetic variants in the region that were not initially associated with SLE risk demonstrated nominal significance ($10^{-4} < P < 10^{-2}$) after the first step-wise regression analysis. After conditioning on rs11889341, we performed further step-wise logistic regression by including variants with a residual association as an additional covariate. This did not result in significant decreases in the level of remaining association ($P < 0.01$). These analyses support a model with a single genetic effect tagged by the lupus-risk variant rs11889341 and a second minor, probably insignificant, genetic effect, which is likely complex in nature.

To complement the frequentist analysis, we performed a Bayesian analysis with the genotyped and imputed variants in the *STAT1*–*STAT4* region for each of the three cohorts with genome-wide significant lupus risk association at this locus. We identified a credible set of variants that account for 99% of the posterior probability in the *STAT1*–*STAT4* region (12, 15 and 5 genetic variants in the European & European American, Asian & Asian American, and Amerindian cohorts, respectively) (Fig. 2 and Supplementary Material, Table S2). In our genetic analysis, we identified a single major genetic effect in the European & European American, Asian & Asian American, and Amerindian samples, which is shared amongst the three ancestries. Based on the presence of a shared genetic effect, we inferred that the disease mechanism at *STAT1*–*STAT4* locus is likely common for all ancestries. Based on this assumption, we developed an ‘ancestry-informed credible set’ (AICS) by identifying variants shared amongst the three credible sets (Supplementary Material, Table S2). As expected, the variants in the AICS are in high linkage disequilibrium ($R^2 > 0.8$) and the variants with the most significant *P*-values have the highest Bayes factors (BFs). The AICS contains four lupus-risk genetic variants shared across the three ancestries with lupus association. We validated our assumption of a shared mechanism by performing a weighted trans-ancestral meta-analysis using METAL (40) and identified the four AICS variants as the four most significantly associated variants (Supplementary Material, Fig. S3 and Table S2).

We used an independent trans-ancestral Replication cohort of 7762 individuals to complement our initial genetic analysis. We identified a lupus-risk association of genetic variants in each of the Replication cohorts with genome-wide significant

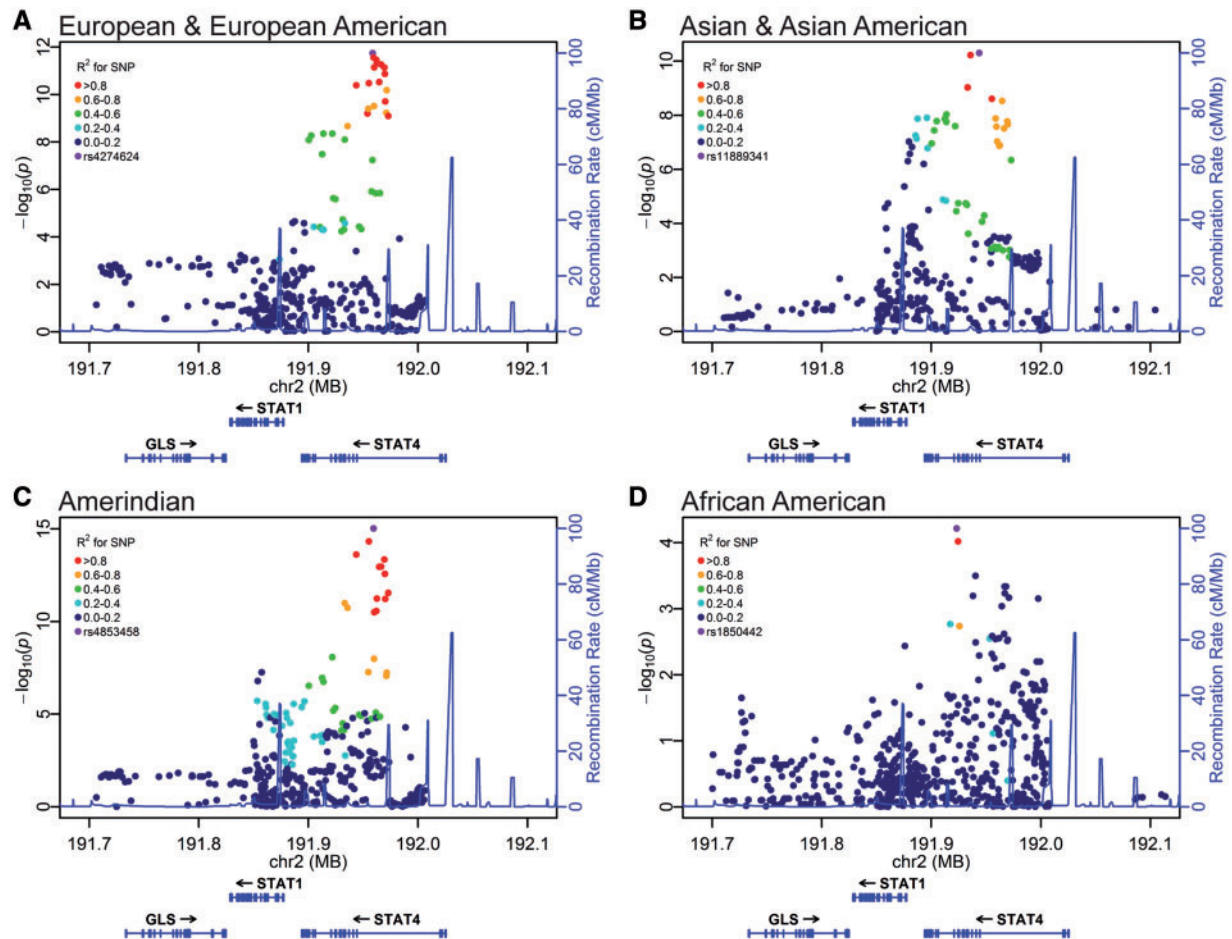


Figure 1. *STAT1-STAT4* variants show a genome-wide association in a multi-ethnic discovery cohort. Each variant is represented as a data point in the context of its genomic location and is colored on the basis of linkage disequilibrium with the most associated variant in each individual ancestral analysis (A, European & European American; B, Asian & Asian American; C, Amerindian; D, African American). Genomic position is provided using GRCh37 (hg19) coordinates. The variants were assessed in a logistic regression model using the admixture estimates as a covariate. Genome-wide significance was defined as $P < 5 \times 10^{-8}$.

associations in the European & European American and Asian & Asian American cohort (Supplementary Material, Fig. S4). Strikingly, the Bayesian analysis identified a single variant shared across the 99% credible sets: rs11889341 (Supplementary Material, Fig. S5 and Table S3). This variant is highly associated in all ancestries (Supplementary Material, Fig. S6). Based on our model of a shared mechanism across all ancestries, we performed another trans-ancestral meta-analysis using the association data from both the replication and discovery cohorts. The single variant shared across the 99 credible sets, rs11889341, is also the most strongly associated variant in our trans-ancestral meta-analysis (Fig. 3).

We then focused mechanistic analysis on this candidate causal variant. rs11889341 is located within the third intron of the *STAT4* gene in a genomic region with H3K4Me1 marks in multiple LCLs (Supplementary Material, Fig. S7), suggesting that the variant falls within a regulatory region in B cells (41). Using B cell lines developed from lupus patients and healthy subjects, we found an eQTL establishing that the risk allele of rs11889341 leads to increased *STAT1* expression in SLE cases (Fig. 4A, with subjects separated by case control status in Supplementary Material, Fig. S8). We observed a similar trend in the healthy controls, but this trend did not reach statistical significance. We did not observe genotype-dependent expression of the *STAT4* gene (Supplementary Material,

Fig. S8). Based on these results, we hypothesized that rs11889341 affects the gene expression of *STAT1* in B cells through genotype-dependent transcription factor binding.

We constructed a luciferase reporter with rs11889341 and the flanking genomic sequence inserted upstream of a minimal promoter. Using site-directed mutagenesis, we generated reporter vectors differing at only the genotype of the genetic variant. We observed a strong repression of the nanoluciferase reporter by the region containing the non-risk allele of rs11889341 compared with an empty vector construct (Fig. 4B). This repressor activity was significantly reduced for the construct containing the risk allele (Fig. 4B). These results suggest that the region containing the variant rs11889341 can act as strong repressor in B cells and that the risk allele at rs11889341 decreases repressor activity.

To identify transcription factors binding the rs11889341 variant in a genotype-dependent fashion, we performed DNA affinity precipitation assays (DAPA) followed by liquid chromatography-tandem mass spectrometry. This proteomic analysis identified the HMGA1 protein binding to both the risk and the non-risk alleles. Based on the DNA binding motif for HMGA1 and the 'DNA Scan' tool provided on the Cis-BP web server (42), we hypothesized that HMGA1 would bind more strongly to the risk allele (Supplementary Material, Fig. S9).

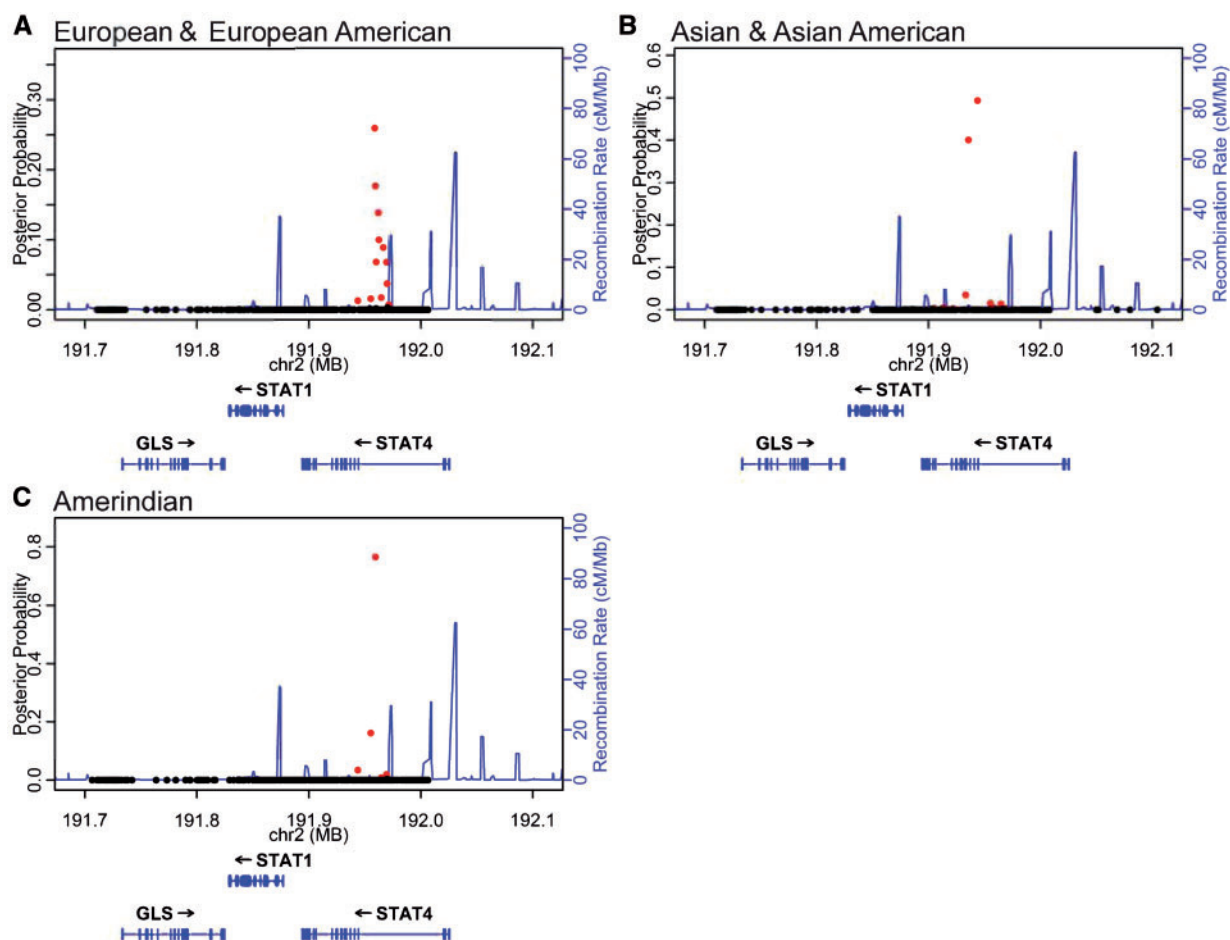


Figure 2. Discovery cohort Bayesian analysis identifies a small group of genetic variants in the *STAT4* gene that comprise the 99% credible set in multiple ancestral analyses. Each variant is represented as a data point in the context of its genomic location using genome build GRCh37 coordinates. Variants in red represent members of the 99% credible set for each ancestry (listed in [Supplementary Material, Table S2](#)). Variants with posterior probability greater than 0.01 are most likely to be causal.

To assess allelic binding of HMGA1 to the variant in B cells of patients with lupus, we performed chromatin immunoprecipitation (ChIP) followed by quantitative polymerase chain reaction (qPCR) in a B cell line that is heterozygous for rs11889341. These experiments revealed enhanced binding of HMGA1 to the risk allele of rs11889341 (>3-fold increased binding) ([Fig. 4C](#) and [Supplementary Material, Fig. S10](#)) in three independent experiments. To further connect the differential binding of HMGA1 to the decreased repressive activity seen in the luciferase assay, we scrambled the putative HMGA1 binding site in the luciferase reporter and found decreased reporter activity of the risk allele, as expected ([Supplementary Material, Fig. S11](#)). Taken together, our data suggest the variant rs11889341 alters the binding of the transcription factor HMGA1, explaining the decreased repression in a luciferase assay in LCLs and increased expression of the *STAT1* gene in LCLs derived from lupus patients.

Discussion

We undertook a fine-mapping study to identify plausible causal variants within the *STAT1-STAT4* locus in the context of increased lupus disease risk. Using complementary frequentist and Bayesian approaches in trans-ancestral Discovery and Replication cohorts, we found one variant whose association with lupus risk is supported across ancestries in both the

Discovery and Replication cohorts: rs11889341. In B cell lines from patients with lupus and healthy controls, the rs11889341 risk allele is associated with increased *STAT1* expression. The transcription factor HMGA1, which binds AT-rich DNA through an AT-hook, binds the risk allele of rs11889341 more strongly than the non-risk allele. We identified a genotype-dependent repressive element at the DNA surrounding rs11889341 within an intron of *STAT4*. Consistent with our eQTL analysis, the lupus risk allele of rs11889341 decreases the activity of this putative repressor, resulting in higher *STAT1* expression levels. Altogether, we present a plausible molecular mechanism for increased lupus risk at the *STAT1-STAT4* locus in which the risk allele of rs11889341 leads to increased *STAT1* expression in B cells due to decreased repressor activity mediated by increased HMGA1 binding.

Other studies have shown that HMGA1 can play a role in regions of the genome that repress gene expression. For example, HMGA1 is known to suppress CD4 and CD8 expression in T cells (43) and BRCA1 expression in carcinoma cell lines (44). Likewise, we herein propose a model in which HMGA1 binding contributes to a repressive element controlling the expression of *STAT1*. Based on previous studies, it is possible that the binding of HMGA1 to rs11889341 may be facilitated through interactions with other factors (45). In the mass spectrometry result from DAPA of rs11889341 ([Supplementary Material, Table S4](#)),

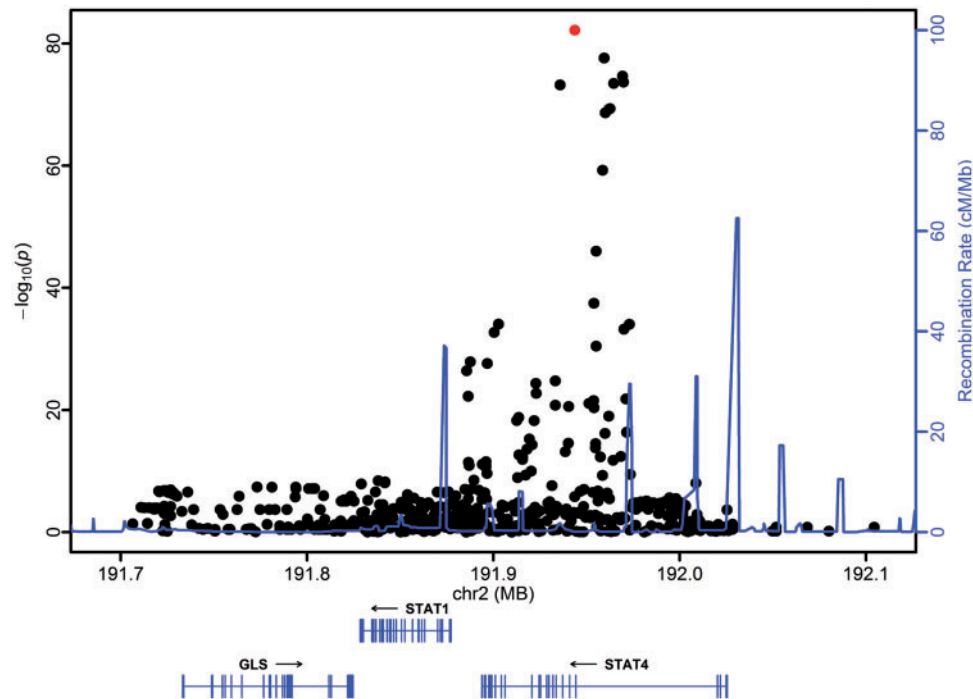


Figure 3. The single variant shared across the 99%-credible sets of the discovery and replication cohort also has the strongest association in a weighted trans-ancestral meta-analysis. A weighted meta-analysis was performed on the results of the logistic regression modelling of each cohort within the discovery and replication cohort. Each variant is represented as a data point in the context of its genomic location and is colored based on the variant's inclusion in the AICS analysis (shown in red). Genomic position is provided using GRCh37 coordinates. The only variant shared across the 99%- credible sets (red), rs11889341, shows the strongest association within a trans-ancestral meta-analysis.

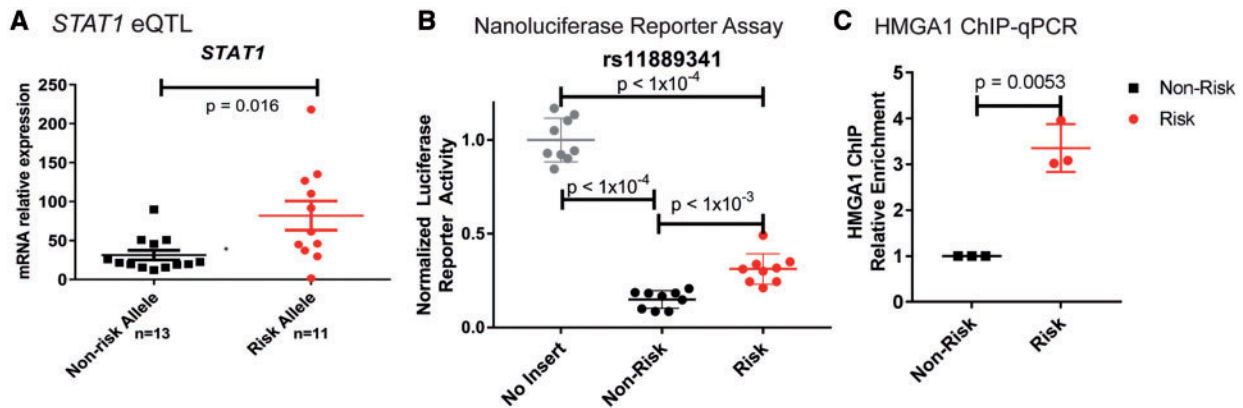


Figure 4. The lupus risk allele of rs11889341 increases HMGA1 binding and decreases repressor regulatory activity in a genotype-dependent manner. (A) *STAT1* mRNA levels were measured in LCLs from individuals with and without lupus who were homozygous for the risk or the non-risk allele (risk allele: C and non-risk allele: T). mRNA levels were normalized to a housekeeping gene, *HPRT1*. A total of 24 separate cell lines were assessed. Mean \pm SEM is shown. (B) GM12878 cell lines were transiently transfected with luciferase constructs generated by inserting the genomic region surrounding rs11889341 into a luciferase vector containing a minimal promoter. The risk and non-risk versions of the construct differed only at the rs11889341 variant (either risk allele or non-risk allele). Luciferase activity was measured 24 h post-transfection. About nine independent transfection experiments are represented with mean \pm SEM. Two-tailed one-way ANOVA with Holm-Sidak's multiple comparison test was used to estimate statistical significance. (C) The LCL GM12878, which is heterozygous for rs11889341, was used for ChIP-qPCR assessment of the differential binding of HMGA1 to the lupus risk and non-risk alleles. Cross-linked and sonicated chromatin was immunoprecipitated with an anti-HMGA1 antibody. Site-specific primers and probes specific to the rs11889341 risk and non-risk alleles were used for determining HMGA1 binding to immunoprecipitated DNA. Relative enrichment was calculated by normalizing to the non-risk allele.

we also found other proteins in the elution from both the risk and non-risk alleles, including PARP1. It has been shown that PARP1 can regulate the serine ADP-ribosylation of HMGA1, which may affect HMGA1 activity (46). Therefore, it is possible that these proteins might form a complex affecting *STAT1* gene expression.

STAT1 is a key transcription factor downstream of Type 1 and Type 2 interferon signaling (47). Type 1 interferon signaling is known to play a central role in lupus pathogenesis. Specifically, patients with lupus have higher levels of Type 1 interferon in their serum compared with healthy controls and higher levels of Type 1 interferon gene expression signatures

within their peripheral blood mononuclear cells (PBMCs) compared with PBMCs from healthy controls (as reviewed in 48 and 49). PBMCs from patients with lupus are more sensitive to Type 1 interferon compared with PBMCs derived from healthy controls (50,51). Moreover, in some rare cases, patients injected with Interferon- β develop a lupus-like disease characterized by the production of autoantibodies to nuclear antigens and multi-system pathology (52,53). Data from multiple studies support a key role for Type 2 Interferon in initiating autoreactive B-cell germinal centers and preceding Type 1 interferon signatures (54–56). Altogether, these data support a role of differential STAT1 expression in promoting lupus disease.

The involvement of STAT1 in lupus disease pathogenesis has been supported by other genetic studies. Data from our previous study of the ETS1 SLE genetic locus suggest that enhanced STAT1 binding to the risk allele of rs6590330 results in decreased ETS1 expression, which is associated with increased lupus risk (57), suggesting there might be an epistatic effect between rs11889341 and rs6590330 for lupus risk. In particular, it is possible that increased HMGA1 binding to rs11889341 results in increased STAT1 expression, which in turn leads to decreased ETS1 expression, disrupting B cell function, which may promote lupus development.

While we present one mechanism through which variants at the 2q32 lupus-risk locus increase STAT1 expression in B cells, this model is not mutually exclusive of other plausible etiological mechanisms in other cell types. For example, data from a recent study indicate that rs11889341 (the same variant identified as plausibly causal in our study) also affects STAT1 and STAT4 expression in monocytes (39). Additionally, a study by Kim-Hellmuth et al. (58) shows an association between the variant rs11889341 and STAT1 expression 6 h after LPS stimulation in monocytes from healthy subjects. These data suggest that in monocytes, rs11889341 alters STAT1 gene expression in response to immune stimuli (58). Further, while we demonstrate the functionality of rs11889341 in the context of nearby gene expression, other variants in the credible set might well have distinct effects in other cell types. Our analysis does not completely preclude other candidates, especially variants which are in tight LD with rs11889341. Genetic variants across this region have further been associated with a variety of other phenotypes with immune components (20,33,59,60). Future studies will establish whether or not there are shared and disease-specific mechanisms through which these variants increase disease risk (61). Additionally, numerous studies have revealed case-only associations of variants at this locus with the presentation and disease progression of lupus (15,48,62–66). We did not assess the genetic architecture of these sub-phenotypic associations. While the risk variants for lupus etiology are largely shared with risk variants associated with severity of lupus, it is plausible that both shared and distinct molecular mechanisms drive lupus risk and lupus disease severity.

Previous fine mapping studies of the STAT1–STAT4 region assessed a smaller number of variants (19) in a smaller trans-ancestral cohort (9923 individuals with and without lupus). As in our study, the previous fine-mapping analysis identified a group of variants in the third intron of STAT4 as most highly associated. Indeed, the functional variant identified in this study, rs11889341, is ‘tagged’ through strong linkage disequilibrium ($R^2 > 0.95$) by the previously most associated haplotype (19). Also consistent with the previous fine mapping study, we did not identify a lupus association at the STAT1–STAT4 locus in African American subjects. Although it is possible that this region is in fact not lupus associated in subjects of African ancestry, an

alternative explanation could be due to the frequency of the risk haplotype identified in African cohorts being substantially lower, as identified in the Raj et al. study (39). This could have resulted in less robust statistical power to find the lupus-risk association in the African cohort. Future well-powered studies aimed at elucidating the genetic etiology of lupus in individuals of African ancestry will be critical to further explain this genetic association.

In conclusion, we performed a genetic analysis in two independent trans-ancestral cohorts to identify rs11889341 as the variant most likely to be causal for lupus. Functional analyses revealed that the lupus-risk allele of rs11889341 enhances binding of HMGA1 and attenuates the repressor activity of the region surrounding the variant. We further established the minor allele of rs11889341 as increasing STAT1 expression in B cells from subjects with and without lupus. Altogether, this work provides a molecular mechanistic context for a rigorously established lupus risk locus.

Materials and Methods

Genotyping of genetic variants: discovery cohort

We used a large collection of samples from subjects with and without lupus from multiple ethnic groups (Supplementary Material, Table S5). These samples were from a collaborative study and were contributed by participating institutions in the United States, Latin America, Asia and Europe to be genotyped on the Illumina ImmunChip (67). Samples were genotyped on the custom-designed ImmunoChip Illumina Infinium Assay3 per manufacturer’s (Illumina) protocols, using the Illumina iScan scanner at: Oklahoma Medical Research Foundation (OMRF), University of Texas Southwestern (UTSW), HudsonAlpha Institute for Biotechnology (HA), and North Shore–LJ Health System’s Feinstein Institute for Medical Research (NSLIJ). A total of 327 common ($MAF > 0.01$) variants that met our quality control criteria spanning the STAT1–STAT4 locus (Supplementary Material, Table S5, spanning GRCh37 chr2: 191,700,000–192,100,000) were genotyped on this array. Subjects were grouped into four ethnic groups: European & European American (EU), African American (AA), Asian & Asian American (AS), and Amerindian (AI). All lupus patients met the American College of Rheumatology criteria for the classification of lupus (68) and were enrolled in this study through an informed consent process approved through the local Institutional Regulatory Boards.

Genotyping of genetic variants: replication cohort

In the Replication analysis, we genotyped 90 SNPs covering the STAT1–STAT4 region (Supplementary Material, Table S5), spanning GRCh37 chr2: 191,700,000–192,100,000, as part of a larger collaborative study, the Large Lupus Association Study 2 (LLAS2). The samples were collected from individuals in the United States, Asia, Europe, and Latin America. They were genotyped using the Illumina iSelect platform located at the Lupus Genetics Studies Unit at the OMRF. The subjects were grouped into the four ancestral groups given above. All lupus patients met the American College of Rheumatology criteria for the classification of lupus (68) and were enrolled in this study through an informed consent process approved through the local Institutional Regulatory Boards.

LLAS2 included genotyping of other SLE risk loci, and the analyses of those loci from this same collection, with and without SLE, have been published separately (8,57,69–78).

Genotyping sample quality control

Using standard Illumina genotyping procedures, we generated intensity data for all samples as reported previously (8,57,69,73–75,79). The individuals in the Discovery and Replication cohorts were unique. Some individuals of Asian ancestry that were called with the SLE ImmunoChip study samples were used for other published ImmunoChip studies of lupus risk (79). Samples were excluded if their call rates were <98% across SNPs that passed the other quality control filters. Duplicates and first-degree relatives, as defined by pihat greater than 0.4, were removed, retaining the sample with the highest call rate.

Ascertainment of population stratification

The ancestry of the subjects in this study were self-identified. Genetic outliers from each ethnic and/or racial group were removed from further analysis as determined by principal component (PC) analysis and admixture estimates, as described previously (72,80–81). A PC analysis of the remaining samples (after outlier removal) confirms that no sample has a PC1–3 that is more than 2 standard deviations outside of the mean. We used 347 ancestral informative markers (AIMs) from the same custom genotyping study that passed quality control in both EIGENSTRAT (81) and ADMIXMAP (82,83) to distinguish the four continental ancestral populations: Africans, Europeans, American Indians and East Asians, allowing identification of the substructure within the sample set (84,85). We utilized PCs from EIGENSTRAT outputs to identify outliers of each of the first three PCs for the individual population clusters through visual inspection [see Figure 1 of reference (72)]. Three PCs were used because they accounted for 95% of the eigenvalues. Because the four admixture estimate proportions sum to 1, any three of the four provide the full set of information, so only three proportions were necessary to include as covariates; the EA proportion was omitted from this analysis, as it corresponded to the largest ethnic segment of the combined population.

Statistical analysis: workflow

The analysis began by assessing the disease association of genotyped variants in each of the four ancestral Discovery cohorts individually, as published previously [see Figure 1 of reference (86) for visualization of the subjects before and after outlier removal]. We analyzed the genotyped, then imputed variants and built statistical models to account for the lupus-associated variability in each ancestry with genome-wide statistical association. In order to establish the number of independent genetic effects, we performed a conditional logistic regression analysis in which we included the genotype of each individual at rs11889341 as a covariate. Indeed, adjusting for any variant at this locus that is most highly associated with SLE in any ancestry is sufficient to remove residual association in all other variants in the locus. To generate the 99% credible-sets, we performed a Bayesian analysis (described below) and calculated the posterior probability for each variant. This analysis was repeated in our Replication cohort in ancestries that showed genome-wide significance in the Discovery cohort.

Statistical analysis: frequentist approach

We tested each variant for its association with lupus using logistic regression models that included three admixture proportion estimates as covariates, as implemented in PLINK v1.07

(87). The additive genetic model was assessed as the initially tested model of inheritance.

Using PLINK, step-wise logistic regression was performed to identify those genetic variants independently associated with the development of lupus. For these analyses, the allelic dosage(s) of specific genetic variant(s) were added to the logistic model as covariates in addition to the admixture estimates.

A trans-ancestral meta-analysis was performed using METAL (40). The P -value and odds ratio for each variant were included and the analysis was weighted for the number of individuals with data for each variant in an individual cohort. METAL performs a meta-analysis using P -values from genetic locus associations as input. The software allows the analysis to be weighted based on sample size and takes into account the direction and magnitude of the effect size (odds ratio) for the associations that are combined in the meta-analysis.

Statistical analysis: Bayesian approach

Using SNPTEST, we calculated the BF for each genetic variant. We calculated the probability of the genotype configuration at a genetic variant in cases and controls under the alternative hypothesis that the genetic variant is associated with disease status. Next, we divided this probability by the probability of the genotype configuration at that genetic variant in cases and controls under the null hypothesis that disease status is independent of genotype at that variant. These methods were performed as described previously [we used the previously introduced methods developed and implemented in references (75,88)]. We used three admixture estimates as covariates, as we did for the frequentist approach. Large values of the BF correlate to robust evidence for association, as small probabilities provide strong evidence in a frequentist approach. For well-powered studies, the BFs of relatively common variants are highly correlated with the frequentist-derived P -values (reviewed in 89). We used the additive model. The linear predictor is $\log(p_i/(1 - p_i)) = \mu + \beta G_i$, and the prior is $\mu \sim N(0, 1^2)$, $\beta \sim N(0, 0.2^2)$ [variables are defined in the supplemental note in reference (88) and https://mathgen.stats.ox.ac.uk/genetics_software/snptest/old/snptest.html].

To identify the variants most likely to be driving the statistical association, we calculated a posterior probability under the assumption that any of the variants within a single genetic effect could be causal and that only one of these variants is causal for each genetic effect. Regardless of whether the causal variants have been genotyped in this experiment, variants with a low posterior probability are unlikely to be causal (88).

We generated our credible set by calculating the posterior probability for association of each individual variant. The variants were ordered in descending order by posterior probability and the credible set was defined as the minimum set of variants with posterior probabilities summing to 0.99 or greater.

Imputation to composite 1000 genomes reference panel

To detect associated variants that were not directly genotyped, we imputed the STAT1–STAT4 region with IMPUTE2 using a composite imputation reference panel based on 1000 Genomes Project sequence data freezes from March 2012 (86,90). Imputed genotypes were included in the analysis if they had or exceeded a probability threshold of 0.9, an information measure of >0.4, and the same quality-control criteria thresholds described for the genotyped markers. The most likely genotype was used for variants passing quality controls in all analyses.

Identification of differential transcription factor binding

We used the Cis-BP web server to predict transcription factor binding affected by the alleles of rs11889341. Cis-BP contains transcription factor binding models collected from many sources, and allows users to analyze nucleotide sequences for predicted transcription factor binding. The server can also analyze regions containing variants to identify transcription factors whose binding may be altered by the variant (42).

eQTL analysis

RNA was extracted from 10 million cells from LCLs established from 14 lupus patients and 10 controls for eQTL analysis of STAT1 and 12 lupus patients and 12 controls for the assessment of STAT4 using the Qiagen RNA extraction kit. Cell lines for the expression quantitative trait locus (eQTL) study were selected to include patients and healthy controls with the homozygous risk and the homozygous non-risk genotype at the previously established lupus-associated variant rs11889341. The cell lines used for these experiments were derived from European subjects. About 200 ng of RNA was used to generate a cDNA library using Applied Biosystems High capacity RNA-cDNA Kit (Product # 4387406). STAT1, STAT4 and HPRT1 expression was assayed by qPCR using Taqman Probes [Applied Biosystems Assay # Hs01013996_m1 (STAT1), Hs01028017_m1 (STAT4) and Hs02800695_m1 (HPRT1)] spanning exons.

DNA affinity precipitation assay (DAPA)

Pairs of single stranded 5'-biotinylated 35 base oligonucleotides (obtained from IDT Inc., Coralville, Iowa, USA) were annealed to generate double-stranded probes. Nuclear lysates were prepared from an LCL (GM12878) using methods described in Miller *et al.* (91). Binding reactions were performed with biotinylated probes, cell lysate, binding buffer, binding enhancer, protease inhibitor, phosphatase inhibitor and 0.1 µg poly (dI-dC) along with protocols supplied with the µMACS Factor Finder Kit. Eluted probe-bound proteins were identified by nano liquid chromatography followed by tandem mass spectrometry (Nano-LC-MS/MS) analysis (92). The oligonucleotide sequences are provided in [Supplementary Material, Table S6](#).

Luciferase reporter assays

The 1500 bp genomic region containing rs11889341 was amplified from Jurkat genomic DNA using PCR. The allele of rs11889341 is in the middle of the genomic fragment. The primers used for these reactions introduced a flanking 15 bp sequence homologous to the vector pNL3.2 (Promega) at the 5' and 3' end ([Supplementary Material, Table S6](#)). The amplicon containing the genomic region and the homologous sequence was inserted into the pNL3.2 vector with a 5' HindIII and 3' NheI overhang using the Infusion HD Cloning Kit (Clontech, USA). The pNL3.2 vector contains a nano-luciferase gene driven by a minimal promoter. The vectors were sequenced to identify the allele present and a vector containing the other allele was generated using site-directed mutagenesis with the GeneArt® Site-Directed Mutagenesis System kit (Thermo Fischer Scientific, USA). The constructs were amplified in chemically competent DH5α cells using manufacturer provided instructions and were subsequently sequence-verified.

To perform the luciferase reporter assay, the pNL3.2 constructs containing the variants and the flanking regions and a

pGL3-control firefly luciferase construct were transfected into LCL GM12878 cells with the Neon transfection system using a single 1350 V pulse for 30 ms. Prior to transfection, the cells were seeded at 0.6×10^6 cells/ml and grown for 16 h in RPMI-1640 supplemented with 10% heat-inactivated FBS, 1× Anti-Anti (Gibco, Waltham, MA), 1 µg/µl of Normocin. For each reaction, 2×10^6 cells were transfected with 2.5 µg of the pNL3.2 nano-luciferase construct and 2.5 µg of the firefly luciferase construct. The cells were incubated for 24 h and luciferase expression was assessed using the One-Glo Ex reagent (Firefly-luciferase) and the NanoDLR Stop & Glo (Nano-luciferase).

ChIP-qPCR

Cross-linking of protein–chromatin complexes was achieved by incubating EBV-transformed cells in cross-linking solution [1% formaldehyde, 5 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 8.0, 10 mM sodium chloride (NaCl), 0.1 mM EDTA, 0.05 mM ethylene glycol tetraacetic acid (EGTA)] and shaking at room temperature for 10 min. Glycine was added to a final concentration of 0.125 M to quench the cross-linking. Cells were washed twice with ice-cold phosphate-buffered saline (PBS), resuspended in lysis buffer L1 (50 mM Hepes pH 8.0, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.25% Triton X-100 and 0.5% NP-40), and incubated for 10 min on ice. Protease and phosphatase inhibitors were added to all buffers. Nuclei were harvested after centrifugation at 5000 rpm for 10 min, resuspended in lysis buffer L2 (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 200 mM NaCl and 0.5 mM EGTA), and incubated at room temperature for 10 min. Nuclei were resuspended in sonication buffer (10 mM Tris, pH 8.0, 1 mM EDTA and 0.1% SDS) after centrifuging. A S220 focused-ultrasonicator (Covaris, Inc, Woburn, MA, USA) was used to shear genomic DNA (150–500 bp fragments) with 10% duty cycle, 175 peak power, 200 burst/cycle for 7 min. Sheared chromatin was precleared with 10 µl Dynabeads® Protein G (Life Technologies, Grand Island, NY, USA) at 4°C for 1 h. Antibody [anti-HMGA1 (ab4078), AbCam, San Francisco, CA, USA] was incubated with 20 µl Dynabeads® Protein G at room temperature for 1 h followed by washing with PBS once. The antibody-coated beads were incubated with sheared chromatin at 4°C overnight. A volume of 1% of sheared chromatin was used as input control. After immunoprecipitation, the beads were washed consecutively with low-salt wash buffer (0.1% SDS, 1% Triton X-100, 0.1% sodium deoxycholate, 1 mM EDTA, 50 mM Tris-HCl pH 8 and 150 mM NaCl) twice, high-salt wash buffer (as above with 500 mM NaCl) twice, LiCl wash buffer (0.5 M LiCl, 1% NP-40, 0.7% sodium deoxycholate, 1 mM EDTA and 50 mM Tris-HCl pH 8) twice and twice in 1 mM EDTA, 10 mM Tris-HCl pH 8. Purified chromatin fragments were eluted from the beads with elution buffer (340 mM NaCl, 1 mM EDTA and 10 mM Tris-HCl) and 1 mg/ml proteinase K, and incubated at 37°C for 1 h. DNA cross-links were reversed by incubating precipitates at 65°C for 5 h. DNA was purified by PureLink® PCR Micro Kit (Life Technologies, Grand Island, NY, USA) and resuspended in H₂O. DNA was then analyzed using qPCR with a single set of genotyping primers and differentially tagged fluorescent probes for the risk and non-risk allele of rs11889341. This qPCR was performed with a genotyping Taqman assay (Assay ID: C_26419582_10) using ABI 7500 with the VIC fluorophore for the non-risk allele and the FAM fluorophore for the risk allele). Being heterozygous at this variant provides a well-controlled comparison of the risk and non-risk haplotypes in the cell line studied. We normalized all of our ChIP-qPCR data against a 1% input control. The

experiments were done in the GM12878 cell line, which is heterozygous at the variant rs11889341. The crossing threshold (CT) value of each probe for the chromatin pulled down by anti-HMGA1 antibody was normalized to the CTs of each probe from the heterozygous cell DNA (input). Our quantification method is similar to that used in other previously published studies (57,93,94).

Supplementary Material

Supplementary Material is available at HMG online.

Acknowledgements

We thank Dr Artem Barski for his invaluable help and guidance in the development and analysis of the HMGA1 ChIP experiments. Mass spectrometry data were collected in the UC Proteomics Laboratory on the 5600 + TripleT of system funded in part through an NIH shared instrumentation grant (S10 RR027015-01; KD Greis-PI).

Conflict of Interest statement. None declared.

Funding

We are grateful for support from US Department of Veteran Affairs and Defense (BX001834, PR094002) and the National Institutes of Health (NIH) (R01AI024717, R01AI063274, R01AI082714, R01AI083194, U01AI130830, R01AR043274, R01AR043727, R01AR043814, R01AR051545, R01AR056360, R01AR057172, R01AR058959, R01AR060366, R01AR063124, R01AR065626, R01AR62277, R01AI024717, R01DK107502, GM103456, GM104938, HG006828, HG008666, K24AI078004, K24AR02318, K24AR002138, MD007909, P01AR49084, P30AR053483, P30AR055385, P30GM103510, P30AR070549, P30GM110766, P60AR053308, P60AR062755, P60AR064464, P60AR066464, R01AR44804, R01AR043727, R01AR069572, R01AR064820, R01NS099068, R21AI070304, R21HG008186 S10RR027015, TR000077, U01AI101934, U01HG006828, U01HG008666, U19AI082714, U54GM104938, UL1RR029882, UL1TR000004, UL1TR001417, ULTR000062, UL1TR000150, UL1TR000154, 1U54TR001353, 2U54MD007587, 4T32GM063483, 5T32GM105526). Support for the project was also provided by the Cincinnati Children's Research Foundation Endowed Scholar Award, Lupus Research Alliance "Novel Approaches" Award, Kirkland Scholar Award, National Basic Research Program of China (973 program) (2014CB541901), National Natural Science Foundation of China (No. 81230072; 81421001), grants from the State Key Laboratory of Oncogenes and Related Genes (No. 91-14-05), Key Research Program of the Chinese Academy of Sciences (KJZD-EW-L01-3), the Program of the Shanghai Commission of Science and Technology (No.12JC1406000; No. 12431900703), the Proyecto de Excelencia of the Junta de Andalucía (CTS2548), Arthritis Foundation, Alliance for Lupus Research "Target Identification in Lupus" grant, funds from the Spaulding Paolozzi Autoimmunity Center of Excellence, the Richard M Silver MD Endowment for Inflammation Research, the SmartState Center of Economic Excellence in Inflammation and Fibrosis research, and the Korea Healthcare technology R&D Project, Ministry for Health and Welfare, Republic of Korea (HI13C2124).

References

- Deafen, D., Escalante, A., Weinrib, L., Horwitz, D., Bachman, B., Roy-Burman, P., Walker, A. and Mack, T.M. (1992) A revised estimate of twin concordance in systemic lupus erythematosus. *Arthritis. Rheum.*, **35**, 311–318.
- Moser, K.L., Kelly, J.A., Lessard, C.J. and Harley, J.B. (2009) Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun.*, **10**, 373–379.
- Alarcón-Segovia, D., Alarcón-Riquelme, M.E., Cardiel, M.H., Caeiro, F., Massardo, L., Villa, A.R. and Pons-Estel, B.A. and on behalf of the Grupo Latinoamericano de Estudio del Lupus, E. (2005) Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1, 177 lupus patients from the GLADEL cohort. *Arthritis. Rheum.*, **52**, 1138–1147.
- Sestak, A.L., Shaver, T.S., Moser, K.L., Neas, B.R. and Harley, J.B. (1999) Familial aggregation of lupus and autoimmunity in an unusual multiplex pedigree. *J. Rheumatol.*, **26**, 1495–1499.
- Lawrence, J.S., Martins, C.L. and Drake, G.L. (1987) A family survey of lupus erythematosus. 1. Heritability. *J. Rheumatol.*, **14**, 913–921.
- Hochberg, M.C. (1987) The application of genetic epidemiology to systemic lupus erythematosus. *J. Rheumatol.*, **14**, 867–869.
- Block, S.R. (2006) A brief history of twins. *Lupus*, **15**, 61–64.
- Vaughn, S.E., Foley, C., Lu, X., Patel, Z.H., Zoller, E.E., Magnusen, A.F., Williams, A.H., Ziegler, J.T., Comeau, M.E. and Marion, M.C. (2015) Lupus risk variants in the PXX locus alter B-cell receptor internalization. *Front. Genet.*, **5**, 450.
- Vaughn, S.E., Kottyan, L.C., Munroe, M.E. and Harley, J.B. (2012) Genetic susceptibility to lupus: the biological basis of genetic risk found in B cell signaling pathways. *J. Leukoc. Biol.*, **92**, 577–591.
- Kottyan, L.C., Kelly, J.A. and Harley, J.B. (2015) Genetics of Lupus, Chapter # 127. In Hochberg, M.C. (ed), *Rheumatology*. Mosby/Elsevier, Philadelphia, PA, pp. 1045–1051.
- Hu, X., Kim, H., Stahl, E., Plenge, R., Daly, M. and Raychaudhuri, S. (2011) Integrating autoimmune risk loci with gene-expression data identifies specific pathogenic immune cell subsets. *Am. J. Hum. Genet.*, **89**, 496–506.
- Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S. and Raychaudhuri, S. (2013) Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.*, **45**, 124–130.
- Waldman, M. and Madaio, M.P. (2005) Pathogenic autoantibodies in lupus nephritis. *Lupus*, **14**, 19–24.
- Li, P., Cao, C., Luan, H., Li, C., Hu, C., Zhang, S., Zeng, X., Zhang, F., Zeng, C. and Li, Y. (2011) Association of genetic variations in the STAT4 and IRF7/KIAA1542 regions with systemic lupus erythematosus in a Northern Han Chinese population. *Hum. Immunol.*, **72**, 249–255.
- Taylor, K.E., Rimmers, E.F., Lee, A.T., Ortmann, W.A., Plenge, R.M., Tian, C., Chung, S.A., Nititham, J., Hom, G. and Kao, A.H. (2008) Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. *PLoS Genet.*, **4**, e1000084.
- Abelson, A.K., Delgado-Vega, A.M., Kozyrev, S.V., Sanchez, E., Velazquez-Cruz, R., Eriksson, N., Wojcik, J., Linga Reddy, M.V., Lima, G., D'Alfonso, S. et al. (2009) STAT4 associates with systemic lupus erythematosus through two independent effects that correlate with gene expression and act

- additively with IRF5 to increase risk. *Ann. Rheum. Dis.*, **68**, 1746–1753.
17. International Consortium for Systemic Lupus Erythematosus, G., Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D. et al. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.*, **40**, 204–210.
 18. Mirkazemi, S., Akbarian, M., Jamshidi, A.R., Mansouri, R., Ghoroghi, S., Salimi, Y., Tahmasebi, Z. and Mahmoudi, M. (2013) Association of STAT4 rs7574865 with susceptibility to systemic lupus erythematosus in Iranian population. *Inflammation*, **36**, 1548–1552.
 19. Namjou, B., Sestak, A.L., Armstrong, D.L., Zidovetzki, R., Kelly, J.A., Jacob, N., Ciobanu, V., Kaufman, K.M., Ojwang, J.O., Ziegler, J. et al. (2009) High-density genotyping of STAT4 reveals multiple haplotypic associations with systemic lupus erythematosus in different racial groups. *Arthritis Rheum.*, **60**, 1085–1095.
 20. Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.*, **357**, 977–986.
 21. Alarcon-Riquelme, M.E., Ziegler, J.T., Molineros, J., Howard, T.D., Moreno-Estrada, A., Sanchez-Rodriguez, E., Ainsworth, H.C., Ortiz-Tello, P., Comeau, M.E., Rasmussen, A. et al. (2016) Genome-wide association study in an amerindian ancestry population reveals novel systemic lupus erythematosus risk loci and the role of european admixture. *Arthritis. Rheumatol.*, **68**, 932–943.
 22. Demirci, F.Y., Wang, X., Kelly, J.A., Morris, D.L., Barmada, M.M., Feingold, E., Kao, A.H., Sivils, K.L., Bernatsky, S., Pineau, C. et al. (2016) Identification of a new susceptibility locus for systemic lupus erythematosus on chromosome 12 in individuals of European ancestry. *Arthritis. Rheumatol.*, **68**, 174–183.
 23. Sandling, J.K., Garnier, S., Sigurdsson, S., Wang, C., Nordmark, G., Gunnarsson, I., Svenungsson, E., Padyukov, L., Sturfelt, G., Jonsen, A. et al. (2011) A candidate gene study of the type I interferon pathway implicates IKBKE and IL8 as risk loci for SLE. *Eur. J. Hum. Genet.*, **19**, 479–484.
 24. Su, Y., Zhao, Y., Liu, X., Guo, J.P., Jiang, Q., Liu, X.Y., Zhang, F.C., Zheng, Y., Li, X.X., Song, H. et al. (2010) Variation in STAT4 is associated with systemic lupus erythematosus in Chinese Northern Han population. *Chin. Med. J.*, **123**, 3173–3177.
 25. Yuan, H., Feng, J.B., Pan, H.F., Qiu, L.X., Li, L.H., Zhang, N. and Ye, D.Q. (2010) A meta-analysis of the association of STAT4 polymorphism with systemic lupus erythematosus. *Mod. Rheumatol.*, **20**, 257–262.
 26. Yang, W., Shen, N., Ye, D.Q., Liu, Q., Zhang, Y., Qian, X.X., Hirankarn, N., Ying, D., Pan, H.F., Mok, C.C. et al. (2010) Genome-wide association study in Asian populations identifies variants in ETS1 and WDFY4 associated with systemic lupus erythematosus. *PLoS Genet.*, **6**, e1000841.
 27. Han, J.W., Zheng, H.F., Cui, Y., Sun, L.D., Ye, D.Q., Hu, Z., Xu, J.H., Cai, Z.M., Huang, W., Zhao, G.P. et al. (2009) Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat. Genet.*, **41**, 1234–1237.
 28. Hellquist, A., Sandling, J.K., Zucchelli, M., Koskenmies, S., Julkunen, H., D'Amato, M., Garnier, S., Syvanen, A.C. and Kere, J. (2010) Variation in STAT4 is associated with systemic lupus erythematosus in a Finnish family cohort. *Ann. Rheum. Dis.*, **69**, 883–886.
 29. Ji, J.D., Lee, W.J., Kong, K.A., Woo, J.H., Choi, S.J., Lee, Y.H. and Song, G.G. (2010) Association of STAT4 polymorphism with rheumatoid arthritis and systemic lupus erythematosus: a meta-analysis. *Mol. Biol. Rep.*, **37**, 141–147.
 30. Suarez-Gestal, M., Calaza, M., Endreffy, E., Pullmann, R., Ordi-Ros, J., Sebastiani, G.D., Ruzickova, S., Jose Santos, M., Papasteriades, C., Marchini, M. et al. (2009) Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. *Arthritis. Res. Ther.*, **11**, R69.
 31. Lessard, C.J., Sajuthi, S., Zhao, J., Kim, K., Ice, J.A., Li, H., Ainsworth, H., Rasmussen, A., Kelly, J.A., Marion, M. et al. (2016) Identification of a systemic lupus erythematosus risk locus spanning ATG16L2, FCHSD2, and P2RY2 in Koreans. *Arthritis. Rheumatol.*, **68**, 1197–1209.
 32. Kawasaki, A., Ito, I., Hikami, K., Ohashi, J., Hayashi, T., Goto, D., Matsumoto, I., Ito, S., Tsutsumi, A., Koga, M. et al. (2008) Role of STAT4 polymorphisms in systemic lupus erythematosus in a Japanese population: a case-control association study of the STAT1-STAT4 region. *Arthritis. Res. Ther.*, **10**, R113.
 33. Beltran Ramirez, O., Mendoza Rincon, J.F., Barbosa Cobos, R.E., Aleman Avila, I. and Ramirez Bello, J. (2016) STAT4 confers risk for rheumatoid arthritis and systemic lupus erythematosus in Mexican patients. *Immunol. Lett.*, **175**, 40–43.
 34. Ciccacci, C., Perricone, C., Ceccarelli, F., Rufini, S., Di Fusco, D., Alessandri, C., Spinelli, F.R., Cipriano, E., Novelli, G., Valesini, G. et al. (2014) A multilocus genetic study in a cohort of Italian SLE patients confirms the association with STAT4 gene and describes a new association with HCP5 gene. *PLoS One*, **9**, e111991.
 35. Piotrowski, P., Lianeri, M., Wudarski, M., Olesińska, M. and Jagodziński, P.P. (2012) Contribution of STAT4 gene single-nucleotide polymorphism to systemic lupus erythematosus in the Polish population. *Mol. Biol. Rep.*, **39**, 8861–8866.
 36. Liang, Y.L., Wu, H., Shen, X., Li, P.Q., Yang, X.Q., Liang, L., Tian, W.H., Zhang, L.F. and Xie, X.D. (2012) Association of STAT4 rs7574865 polymorphism with autoimmune diseases: a meta-analysis. *Mol. Biol. Rep.*, **39**, 8873–8882.
 37. Sánchez, E., Comeau, M.E., Freedman, B.I., Kelly, J.A., Kaufman, K.M., Langefeld, C.D., Brown, E.E., Alarcón, G.S., Kimberly, R.P., Edberg, J.C. et al. (2011) Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. *Arthritis. Rheum.*, **63**, 3493–3501.
 38. Luan, H., Li, P., Cao, C., Li, C., Hu, C., Zhang, S., Zeng, X., Zhang, F., Zeng, C. and Li, Y. (2012) A single-nucleotide polymorphism of the STAT4 gene is associated with systemic lupus erythematosus (SLE) in female Chinese population. *Rheumatol. Int.*, **32**, 1251–1255.
 39. Raj, P., Rai, E., Song, R., Khan, S., Wakeland, B.E., Viswanathan, K., Arana, C., Liang, C., Zhang, B., Dozmorov, I. et al. (2016) Regulatory polymorphisms modulate the expression of HLA class II molecules and promote autoimmunity. *Elife*, **5**
 40. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
 41. Grubert, F., Zaugg, J.B., Kasowski, M., Ursu, O., Spacek, D.V., Martin, A.R., Greenside, P., Srivas, R., Phanstiel, D.H., Pekowska, A. et al. (2015) Genetic control of chromatin states

- in humans involves local and distal chromosomal interactions. *Cell*, **162**, 1051–1065.
42. Weirauch, M.T., Yang, A., Albu, M., Cote, A.G., Montenegro-Montero, A., Drewe, P., Najafabadi, H.S., Lambert, S.A., Mann, I., Cook, K. et al. (2014) Determination and inference of eukaryotic transcription factor sequence specificity. *Cell*, **158**, 1431–1443.
 43. Xi, Y., Watanabe, S., Hino, Y., Sakamoto, C., Nakatsu, Y., Okada, S. and Nakao, M. (2012) Hmga1 is differentially expressed and mediates silencing of the CD4/CD8 loci in T cell lineages and leukemic cells. *Cancer Sci.*, **103**, 439–447.
 44. Baldassarre, G., Battista, S., Belletti, B., Thakur, S., Pentimalli, F., Trapasso, F., Fedele, M., Pierantoni, G., Croce, C.M. and Fusco, A. (2003) Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma. *Mol. Cell. Biol.*, **23**, 2225–2238.
 45. Panne, D. (2008) The enhanceosome. *Curr. Opin. Struct. Biol.*, **18**, 236–242.
 46. Bonfiglio, J.J., Fontana, P., Zhang, Q., Colby, T., Gibbs-Seymour, I., Atanassov, I., Bartlett, E., Zaja, R., Ahel, I. and Matic, I. (2017) Serine ADP-ribosylation depends on HPF1. *Mol. Cell*, **65**, 932–940. e936.
 47. Platanius, L.C. (2005) Mechanisms of type-I- and type-II-interferon-mediated signalling. *Nat. Rev. Immunol.*, **5**, 375–386.
 48. Kariuki, S.N., Kirou, K.A., MacDermott, E.J., Barillas-Arias, L., Crow, M.K. and Niewold, T.B. (2009) Cutting edge: autoimmune disease risk variant of STAT4 confers increased sensitivity to IFN- α in lupus patients in vivo. *J. Immunol.*, **182**, 34–38.
 49. Niewold, T.B. (2014) Type I interferon in human autoimmunity. *Front. Immunol.*, **5**, 306.
 50. Niewold, T.B., Hua, J., Lehman, T.J., Harley, J.B. and Crow, M.K. (2007) High serum IFN- α activity is a heritable risk factor for systemic lupus erythematosus. *Genes Immun.*, **8**, 492–502.
 51. Weckerle, C.E., Franek, B.S., Kelly, J.A., Kumabe, M., Mikolaitis, R.A., Green, S.L., Utset, T.O., Jolly, M., James, J.A., Harley, J.B. et al. (2011) Network analysis of associations between serum interferon- α activity, autoantibodies, and clinical features in systemic lupus erythematosus. *Arthritis Rheum.*, **63**, 1044–1053.
 52. Ronnblom, L.E., Alm, G.V. and Oberg, K.E. (1990) Possible induction of systemic lupus erythematosus by interferon- α treatment in a patient with a malignant carcinoid tumour. *J. Intern. Med.*, **227**, 207–210.
 53. Ronnblom, L.E., Alm, G.V. and Oberg, K.E. (1991) Autoimmunity after alpha-interferon therapy for malignant carcinoid tumors. *Ann. Intern. Med.*, **115**, 178–183.
 54. Munroe, M.E., Lu, R., Zhao, Y.D., Fife, D.A., Robertson, J.M., Guthridge, J.M., Niewold, T.B., Tsokos, G.C., Keith, M.P., Harley, J.B. et al. (2016) Altered type II interferon precedes autoantibody accrual and elevated type I interferon activity prior to systemic lupus erythematosus classification. *Ann. Rheum. Dis.*, **75**, 2014–2021.
 55. Jackson, S.W., Jacobs, H.M., Arkatkar, T., Dam, E.M., Scharping, N.E., Kolhatkar, N.S., Hou, B., Buckner, J.H. and Rawlings, D.J. (2016) B cell IFN- γ receptor signaling promotes autoimmune germinal centers via cell-intrinsic induction of BCL-6. *J. Exp. Med.*, **213**, 733–750.
 56. Jacob, C.O., van der Meide, P.H. and McDevitt, H.O. (1987) In vivo treatment of (NZB X NZW)F1 lupus-like nephritis with monoclonal antibody to gamma interferon. *J. Exp. Med.*, **166**, 798–803.
 57. Lu, X., Zoller, E.E., Weirauch, M.T., Wu, Z., Namjou, B., Williams, A.H., Ziegler, J.T., Comeau, M.E., Marion, M.C., Glenn, S.B. et al. (2015) Lupus risk variant increases pSTAT1 binding and decreases ETS1 expression. *Am. J. Hum. Genet.*, **96**, 731–739.
 58. Kim-Hellmuth, S., Bechheim, M., Putz, B., Mohammadi, P., Nedelec, Y., Giangreco, N., Becker, J., Kaiser, V., Fricker, N., Beier, E. et al. (2017) Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. *Nat. Commun.*, **8**, 266.
 59. Thompson, S.D., Sudman, M., Ramos, P.S., Marion, M.C., Ryan, M., Tsoras, M., Weiler, T., Wagner, M., Keddache, M., Haas, J.P. et al. (2010) The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANKPT1. *Arthritis. Rheum.*, **62**, 3265–3276.
 60. Zhernakova, A., Stahl, E.A., Trynka, G., Raychaudhuri, S., Festen, E.A., Franke, L., Westra, H.J., Fehrmann, R.S., Kurzeeman, F.A., Thomson, B. et al. (2011) Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.*, **7**, e1002004.
 61. Anaya, J.M. (2017) The autoimmune tautology. A summary of evidence. *Joint Bone Spine* **84**, 251–253.
 62. Bolin, K., Sandling, J.K., Zickert, A., Jonsen, A., Sjoball, C., Svenungsson, E., Bengtsson, A.A., Eloranta, M.L., Ronnblom, L., Syvanen, A.C. et al. (2013) Association of STAT4 polymorphism with severe renal insufficiency in lupus nephritis. *PLoS One*, **8**, e84450.
 63. Sanchez, E., Nadig, A., Richardson, B.C., Freedman, B.I., Kaufman, K.M., Kelly, J.A., Niewold, T.B., Kamen, D.L., Gilkeson, G.S., Ziegler, J.T. et al. (2011) Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus. *Ann. Rheum. Dis.*, **70**, 1752–1757.
 64. Goropevsek, A., Holcar, M., and Avcin, T. (2017) The role of STAT signaling pathways in the pathogenesis of systemic lupus erythematosus. *Clinical Rev. Allergy Immunol.*, **52**, 164–181.
 65. Chung, S.A., Taylor, K.E., Graham, R.R., Nititham, J., Lee, A.T., Ortmann, W.A., Jacob, C.O., Alarcon-Riquelme, M.E., Tsao, B.P., Harley, J.B. et al. (2011) Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet.*, **7**, e1001323.
 66. Svenungsson, E., Gustafsson, J., Leonard, D., Sandling, J., Gunnarsson, I., Nordmark, G., Jonsen, A., Bengtsson, A.A., Sturfelt, G., Rantapaa-Dahlqvist, S. et al. (2010) A STAT4 risk allele is associated with ischaemic cerebrovascular events and anti-phospholipid antibodies in systemic lupus erythematosus. *Ann. Rheum. Dis.*, **69**, 834–840.
 67. Cortes, A. and Brown, M.A. (2011) Promise and pitfalls of the ImmunoChip. *Arthritis. Res. Ther.*, **13**, 101.
 68. Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **40**, 1725.
 69. Lin, C.P., Adrianto, I., Lessard, C.J., Kelly, J.A., Kaufman, K.M., Guthridge, J.M., Freedman, B.I., Anaya, J.-M., Alarcón-Riquelme, M.E., Pons-Estel, B.A. et al. (2012) Role of MYH9 and APOL1 in African and non-African populations with lupus nephritis. *Genes Immun.*, **13**, 232–238.
 70. Nath, S.K., Han, S., Kim-Howard, X., Kelly, J.A., Viswanathan, P., Gilkeson, G.S., Chen, W., Zhu, C., McEver, R.P., Kimberly, R.P. et al. (2008) A nonsynonymous functional variant in

- integrin- α (M) (encoded by ITGAM) is associated with systemic lupus erythematosus. *Nat. Genet.*, **40**, 152–154.
71. Zhao, J., Wu, H., Khosravi, M., Cui, H., Qian, X., Kelly, J.A., Kaufman, K.M., Langefeld, C.D., Williams, A.H., Comeau, M.E. et al. (2011) Association of genetic variants in complement factor H and factor H-related genes with systemic lupus erythematosus susceptibility. *PLoS Genet.*, **7**, e1002079.
 72. Lessard, C.J., Adrianto, I., Kelly, J.A., Kaufman, K.M., Grundahl, K.M., Adler, A., Williams, A.H., Gallant, C.J., Anaya, J.-M., Bae, S.-C., Marta, E.A.-R.o.b.o.t.B., Networks, G. et al. (2011) Identification of a systemic lupus erythematosus susceptibility locus at 11p13 between PDHX and CD44 in a multiethnic study. *Am. J. Hum. Genet.*, **88**, 83–91.
 73. Namjou, B., Choi, C.-B., Harley, I.T.W., Alarcón-Riquelme, M.E., Kelly, J.A., Glenn, S.B., Ojwang, J.O., Adler, A., Kim, K., Gallant, C.J. et al. (2012) Evaluation of TRAF6 in a large multi-ancestral lupus cohort. *Arthritis. Rheum.*, **64**, 1960–1969.
 74. Namjou, B., Kim-Howard, X., Sun, C., Adler, A., Chung, S.A., Kaufman, K.M., Kelly, J.A., Glenn, S.B., Guthridge, J.M., Scofield, R.H. et al. (2013) PTPN22 association in systemic lupus erythematosus (SLE) with respect to individual ancestry and clinical sub-phenotypes. *PLoS One*, **8**, e69404.
 75. Kottyan, L.C., Zoller, E.E., Bene, J., Lu, X., Kelly, J.A., Rupert, A.M., Lessard, C.J., Vaughn, S.E., Marion, M., Weirauch, M.T. et al. (2015) The IRF5-TNPO3 association with systemic lupus erythematosus has two components that other autoimmune disorders variably share. *Hum Mol. Genet.*, **24**, 582–596.
 76. Sakurai, D., Zhao, J., Deng, Y., Kelly, J.A., Brown, E.E., Harley, J.B., Bae, S.-C., Alarcón-Riquelme, M.E., Edberg, J.C., Kimberly, R.P., Biolupus, networks, G. et al. (2013) Preferential binding to Elk-1 by SLE-associated IL10 risk allele upregulates IL10 expression. *PLoS Genet.*, **9**, e1003870.
 77. Deng, Y., Zhao, J., Sakurai, D., Kaufman, K.M., Edberg, J.C., Kimberly, R.P., Kamen, D.L., Gilkeson, G.S., Jacob, C.O., Scofield, R.H. et al. (2013) MicroRNA-3148 modulates allelic expression of toll-like receptor 7 variant associated with systemic lupus erythematosus. *PLoS Genet.*, **9**, e1003336.
 78. Kaufman, K.M., Zhao, J., Kelly, J.A., Hughes, T., Adler, A., Sanchez, E., Ojwang, J.O., Langefeld, C.D., Ziegler, J.T., Williams, A.H. et al. (2013) Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups. *Ann. Rheum. Dis.*, **72**, 437–444.
 79. Sun, C., Molineros, J.E., Looger, L.L., Zhou, X.J., Kim, K., Okada, Y., Ma, J., Qi, Y.Y., Kim-Howard, X., Motghare, P. et al. (2016) High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat. Genet.*, **48**, 323–330.
 80. McKeigue, P.M., Carpenter, J.R., Parra, E.J. and Shriver, M.D. (2000) Estimation of admixture and detection of linkage in admixed populations by a Bayesian approach: application to African-American populations. *Ann. Hum. Genet.*, **64**, 171–186.
 81. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
 82. Hoggart, C.J., Parra, E.J., Shriver, M.D., Bonilla, C., Kittles, R.A., Clayton, D.G. and McKeigue, P.M. (2003) Control of confounding of genetic associations in stratified populations. *Am. J. Hum. Genet.*, **72**, 1492–1504.
 83. Hoggart, C.J., Shriver, M.D., Kittles, R.A., Clayton, D.G. and McKeigue, P.M. (2004) Design and analysis of admixture mapping studies. *Am. J. Hum. Genet.*, **74**, 965–978.
 84. Smith, M.W., Patterson, N., Lautenberger, J.A., Truelove, A.L., McDonald, G.J., Waliszewska, A., Kessing, B.D., Malasky, M.J., Scafe, C., Le, E. et al. (2004) A high-density admixture map for disease gene discovery in african americans. *Am. J. Hum. Genet.*, **74**, 1001–1013.
 85. Halder, I., Shriver, M., Thomas, M., Fernandez, J.R. and Frudakis, T. (2008) A panel of ancestry informative markers for estimating individual biogeographical ancestry and admixture from four continents: utility and applications. *Hum. Mutat.*, **29**, 648–658.
 86. International HapMap, C., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Altshuler, D.M., Gibbs, R.A., Peltonen, L., Dermitzakis, E., Schaffner, S.F., Yu, F. et al. (2010) Integrating common and rare genetic variation in diverse human populations. *Nature*, **467**, 52–58.
 87. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 88. Maller, J.B., McVean, G., Byrnes, J., Vukcevic, D., Palin, K., Su, Z., Howson, J.M.M., Auton, A., Myers, S., Morris, A. et al. (2012) Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.*, **44**, 1294–1301.
 89. Stephens, M. and Balding, D.J. (2009) Bayesian statistical methods for genetic association studies. *Nat. Rev. Genet.*, **10**, 681–690.
 90. Marchini, J., Howie, B., Myers, S., McVean, G. and Donnelly, P. (2007) A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.*, **39**, 906–913.
 91. Miller, D.E., Patel, Z.H., Lu, X., Lynch, A.T., Weirauch, M.T., and Kottyan, L.C. (2016) Screening for functional non-coding genetic variants using electrophoretic mobility shift assay (EMSA) and DNA-affinity precipitation assay (DAPA). *J. Vis. Exp.*, doi: 10.3791/54093.
 92. Wijeratne, A.B., Manning, J.R., Schultz, J.E.J. and Greis, K.D. (2013) Quantitative phosphoproteomics using acetone-based peptide labeling: method evaluation and application to a cardiac ischemia/reperfusion model. *J. Proteome Res.*, **12**, 4268–4279.
 93. Lopez Rodriguez, M., Kaminska, D., Lappalainen, K., Pihlajamäki, J., Kaikkonen, M.U. and Laakso, M. (2017) Identification and characterization of a FOXA2-regulated transcriptional enhancer at a type 2 diabetes intronic locus that controls GCKR expression in liver cells. *Genome Med.*, **9**, 63.
 94. Chen, C., Cai, Q., He, W., Li, Z., Zhou, F., Liu, Z., Zhong, G., Chen, X., Zhao, Y., Dong, W. et al. (2016) An NKX3.1 binding site polymorphism in the l-plastin promoter leads to differential gene expression in human prostate cancer. *Int. J. Cancer*, **138**, 74–86.