Genetic analysis of the pathogenic molecular sub-phenotype interferon-alpha identifies multiple novel loci involved in systemic lupus erythematosus

SN Kariuki\textsuperscript{1,12}, Y Ghodke-Puranik\textsuperscript{2,12}, JM Dorschner\textsuperscript{2}, BS Chrabot\textsuperscript{3}, JA Kelly\textsuperscript{4}, BP Tsao\textsuperscript{5}, RP Kimberly\textsuperscript{6}, ME Alarcón-Riquelme\textsuperscript{4,7}, CO Jacob\textsuperscript{8}, LA Criswell\textsuperscript{9}, KL Sivils\textsuperscript{4}, CD Langefeld\textsuperscript{10}, JB Harley\textsuperscript{11}, AD Skol\textsuperscript{1} and TB Niewold\textsuperscript{2}

Systemic lupus erythematosus (SLE) is a chronic autoimmune disorder characterized by inflammation of multiple organ systems and dysregulated interferon responses. SLE is both genetically and phenotypically heterogeneous, greatly reducing the power of case-control studies in SLE. Elevated circulating interferon-alpha (IFN-\(\alpha\)) is a stable, heritable trait in SLE, which has been implicated in primary disease pathogenesis. About 40–50\% of patients have high IFN-\(\alpha\), and high levels correspond with clinical differences. To study genetic heterogeneity in SLE, we performed a case–case study comparing patients with high vs low IFN-\(\alpha\) in over 1550 SLE cases, including genome-wide association study and replication cohorts. In meta-analysis, the top associations in European ancestry were protein kinase, cyclic GMP-dependent, type I (\(PRKG1\)) \(r^2\beta^2733\) \((P_{\text{Meta}} = 2.75 \times 10^{-10})\) and purine nucleoside phosphorylase (\(PNP\)) \(rs1049564\) \((P_{\text{Meta}} = 1.24 \times 10^{-7})\). We also found evidence for cross-ancestral background associations with the ankyrin repeat domain 44 \((ANKRD44)\) and pleckstrin homology domain containing, family F member 2 gene \((PLEKHF2)\) loci. These loci have not been previously identified in case-control SLE genetic studies. Bioinformatic analyses implicated these loci functionally in dendritic cells and natural killer cells, both of which are involved in IFN-\(\alpha\) production in SLE. As case-control studies of heterogeneous diseases reach a limit of feasibility with respect to subject number and detectable effect size, the study of informative pathogenic sub-phenotypes becomes an attractive strategy for genetic discovery in complex disease.

\textsuperscript{1}Department of Human Genetics, University of Chicago, Chicago, IL, USA; \textsuperscript{2}Department of Immunology and Division of Rheumatology, Mayo Clinic, Rochester, MN, USA; \textsuperscript{3}Gwen Knapp Center for Lupus Research, University of Chicago, Chicago, IL, USA; \textsuperscript{4}Arthritis & Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA; \textsuperscript{5}Department of Medicine, University of California, Los Angeles, CA, USA; \textsuperscript{6}Department of Medicine, University of Alabama, Birmingham, AL, USA; \textsuperscript{7}GENYO Centre for Genomics and Oncological Research: Pfizer/University of Granada/Andalusian Regional Government, Granada, Spain; \textsuperscript{8}Department of Medicine, University of Southern California, Los Angeles, CA, USA; \textsuperscript{9}Rosalind Russell/Epstein P Engleman Rheumatology Research Center, University of California, San Francisco, CA, USA; \textsuperscript{10}Department of Biostatistical Sciences, Wake Forest University, Winston-Salem, NC, USA and \textsuperscript{11}Cincinnati Children’s Hospital Medical Center and Cincinnati VA Medical Center, Cincinnati, OH, USA.

Correspondence: Professor TB Niewold, Department of Immunology and Division of Rheumatology, Mayo Clinic, Guggenheim Building 3-42, 200 1st Street South West, Rochester, MN 55905, USA.

E-mail: Niewold.Timothy@mayo.edu

\textsuperscript{12}These authors contributed equally to this work.

Received 30 May 2014; revised 10 September 2014; accepted 11 September 2014; published online 23 October 2014
human IFN-α administered to humans as a therapy for chronic viral hepatitis and malignancy can induce de novo SLE in some cases. This IFN-α-induced SLE typically resolves after the IFN-α therapy is discontinued, which supports the idea that IFN-α is causal.\textsuperscript{18,19} Case-control genetic studies in SLE have demonstrated remarkable over-representation of genes involved in type I IFN signaling, production and response.\textsuperscript{11} We have shown that many of these SLE-risk loci in the IFN-α pathway are associated with increased IFN-α pathway activity in SLE patients,\textsuperscript{20–23} supporting the idea that these loci are gain-of-function in humans. High circulating levels of IFN-α correspond to particular clinical manifestations,\textsuperscript{15} and thus activation of this pathway contributes to both susceptibility and heterogeneity in SLE.\textsuperscript{24} We suspect that heterogeneity in the molecular pathogenesis of SLE between patients is a major factor in the unexplained heritability of the disease to date. In this study, we directly address this heterogeneity by mapping the causal IFN-α molecular trait, which allowed for detection of novel genetic variations underlying SLE disease pathogenesis. In addition, over-activity of the IFN-α pathway has been implicated in other autoimmune diseases such as Sjogren’s syndrome and inflammatory myositis,\textsuperscript{35,26} and it is possible that these IFN-related loci underlie some of the genetic architecture of these conditions as well.

RESULTS

SNPs associated with IFN-α in the discovery cohort

We generated serum IFN-α data (using reporter cell assay described in Materials and Methods section to detect functional IFN-α activity) in the SLE cases who were genotyped in the SLE Genetics (SLEGEN) consortium genome-wide association study (GWAS) study for whom there was a serum sample available ($n=400$).\textsuperscript{7} This group was used as our discovery cohort. Re-analyzing the GWAS data in a case–case analysis stratified by high vs low serum IFN-α, we found a number of strong associations (OR $>2.0$) with serum IFN-α activity. These included single nucleotide polymorphisms (SNPs) in the chromosome 7 open reading frame 57 (Cort57), protein kinase, cyclic GMP-dependent, type I (PRKG1), purine nucleoside phosphorylase (PNP), activating transcription factor 7 interacting protein (ATF7IP) and ankyrin repeat domain 44 (ANKRD44) loci (Supplementary Table 1, Figure 1a). Furthermore, a quantile–quantile plot of the observed $P$-values showed a deviation at the tail of the distribution from the null distribution (Figure 1b). We conducted a pathway analysis to identify canonical functional pathways that are enriched in the genes nearby these SNPs, and the results from this analysis are described later. This top 10 list did not share any SNPs or loci in common with the known case-control genetic associations with SLE, supporting the ability of this approach to detect novel associations. Many of the underlying genetic variations with SLE could impact particular pathways or subsets of this heterogeneous disease, and these genetic variants can be missed by large case-control SLE studies in which all patients are grouped together. We then planned to replicate all SNPs in the discovery phase with $P < 10^{-6}$ (323 SNPs). In this replication list, there was one SNP which has been previously reported as a case-control association with SLE (ITGAM, rs1143678, enrichment $P = 0.044$),\textsuperscript{7,27} and there were two loci on the replication list which had previously been associated with serum IFN-α levels in SLE patients (EFN5AS and ZKSCAN1/LAMTOR4, enrichment $P = 0.036$)\textsuperscript{28,29} (Supplementary Table 2).

Validation of GWAS candidates with serum IFN-α activity in an independent cohort

The 323 top SNPs which had a $P < 10^{-4}$ were validated in both European-American and African-American SLE cases using an independent multi-ethnic replication cohort of 1165 SLE cases (see Supplementary Table 3 for the characteristics of the replication cohort). We used logistic regression analysis to test SNPs from the discovery cohort for association with serum IFN-α, and European-American and African-American ancestral groups were analyzed separately. SNPs in the PRKG1 (rs7897633, rs7906944) and PNP (rs1049564) loci showed strong evidence for association (Table 1) in the European-American patients. In meta-analysis, both PRKG1 rs7897633 and PNP rs1049564 were associated with serum IFN-α in European ancestry with $P$-values that exceeded a Bonferroni correction for multiple comparisons ($P < 1.71 \times 10^{-7}$, Table 1). Thus, the novel loci identified in the current study achieve genome-wide significance in the overall meta-analysis of discovery and replication sets, but none of the loci have been formally replicated. Table 2 shows a list of top SNPs in African-Americans. No significant SNP–SNP interactions were detected. Haplotype analysis was performed when evidence for association was observed for two nearby SNPs, but none of the haplotype models were superior to the individual SNP models of association. For the SNPs which demonstrated evidence for association in both European-American and African-American ancestral backgrounds, those with homogenous effects by Breslow–Day testing were analyzed in meta-analysis assuming a fixed-effect model. The two SNPs included in this cross-ancestral background meta-analysis were ANKRD44 rs4850410 (OR = 0.64; 95%CI (0.48–0.84); $P_{\text{Meta}} = 1.3 \times 10^{-6}$) and pleckstrin homology domain containing, family F member 2 gene (PLEKHF2) rs297573 (OR = 0.70; 95%CI (0.50–0.98); $P_{\text{Meta}} = 1.2 \times 10^{-4}$).

![Figure 1](image-url) Figure 1. Top signals of association with increased serum IFN-α activity in SLE cases in the discovery phase. (a) Manhattan plot shows top association signals by chromosome. (b) Quantile–quantile Plot showing association of SLE GWAS SNPs with serum IFN-α.
Table 1. List of top replicated SNPs associated with IFN-α in European-Americans

| Chromosome | Locus | SNP     | SNP type | Associated allele/minor allele | Odds ratio (95% CI) | P-discovery | P-replication | P_Meta |
|------------|-------|---------|----------|-------------------------------|---------------------|-------------|---------------|---------|
| 10         | PRKG1 | rs7897633 | Intron | C                             | 0.59 (0.44–0.78)    | 1.07E–05     | 2.96E–04      | 2.75E–08 |
| 14         | PNP   | rs1049564 | Missense | T                             | 2.08 (1.34–3.21)    | 1.32E–05     | 9.88E–04      | 1.24E–07 |
| 6          | DLL   | rs1028488 | Intergenic | A                             | 0.51 (0.38–0.70)    | 8.50E–04      | 3.12E–05      | 2.21E–07 |
| 4          | GXR1C | rs6850699 | Intergenic | A                             | 0.64 (0.50–0.83)    | 4.75E–04      | 5.88E–04      | 1.81E–06 |
| 1          | CHIA  | rs7411387 | Intron | C                             | 1.61 (1.24–2.11)    | 1.23E–03      | 3.80E–04      | 3.07E–06 |
| 11         | TMRRSS | rs3934007 | Intergenic | T                             | 1.55 (1.19–2.00)    | 4.86E–04      | 9.98E–04      | 3.12E–06 |
| 2          | ANKRDD4 | rs1429411 | Intron | C                             | 1.55 (1.20–2.00)    | 9.56E–04      | 8.4E–04       | 5.04E–06 |

Abbreviations: CI, confidence interval; IFN-α, interferon-alpha; SNP, single nucleotide polymorphism.

Table 2. List of top SNPs associated with serum IFN-α in African-Americans

| Chromosome | Locus | SNP     | SNP type | Associated allele/minor allele | Odds ratio (95% CI) | P-value |
|------------|-------|---------|----------|-------------------------------|---------------------|---------|
| 10         | NRG3  | rs1649949 | Intron | C                             | 1.60 (1.20–2.15)    | 1.37E–03 |
| 2          | ANKRDD4 | rs4850410 | Intron | C                             | 0.70 (0.50–0.95)    | 2.47E–02 |
| 5          | LOC729506 | rs1666793 | Intergenic | C                             | 1.61 (1.10–2.12)    | 1.10E–02 |
| 8          | ASPH  | rs7812327 | Intron | C                             | 1.45 (0.94–2.02)    | 2.90E–02 |
| 20         | PLCB4 | rs2299676 | Intron | C                             | 1.81 (0.93–3.51)    | 1.58E–02 |
| 5          | FGFI8 | rs7711912 | Near 3′ | A                             | 1.60 (1.10–2.15)    | 1.37E–03 |
| 16         | RBFOX1 | rs4608354 | Intron | C                             | 1.45 (0.94–2.02)    | 2.90E–02 |
| 8          | PLEKHF2 | rs2975737 | Near 3′ | C                             | 1.60 (1.10–2.15)    | 1.37E–03 |
| 12         | KCNAM | rs5286654 | Near 3′ | C                             | 1.60 (1.10–2.15)    | 1.37E–03 |

Abbreviations: CI, confidence interval; IFN-α, interferon-alpha; SNP, single nucleotide polymorphism.

Table 3. Top canonical pathways from IFN-α associated SNPs in initial discovery GWAS data

| Canonical Pathways | Ratio | P-value |
|--------------------|-------|---------|
| Axonal guidance signaling | 0.02 | 5.02E–04 |
| Synaptic long term depression | 0.03 | 4.58E–03 |
| Xanthine and xanthosine salvage | 0.11 | 7.65E–03 |
| Dopamine–DARPP32 feedback in CAMP signaling | 0.03 | 8.37E–03 |
| Guanine and guanosine salvage I | 0.11 | 1.52E–02 |
| Adenine and adenosine salvage | 0.11 | 1.52E–02 |
| Antiproliferative role of TOB in T-cell signaling | 0.08 | 1.67E–02 |
| Cellular effects of sildenafil (viagra) | 0.03 | 1.73E–02 |
| Caveolar-mediated endocytosis signaling | 0.04 | 1.91E–02 |
| Cardiac 3′-adrenergic signaling | 0.03 | 2.01E–02 |

Abbreviations: cAMP, cyclic AMP; GWAS, genome-wide association study; IFN-α, interferon-alpha; SNP, single nucleotide polymorphism. Ratio and P-values are calculated as described in the Materials and methods section.

Association of GWAS candidates with autoantibody subsets in the replication cohort

Because the presence of particular autoantibodies has been strongly associated with high IFN-α in SLE, we also tested the SNPs which were replicated from the GWAS study for association with SLE autoantibodies. Supplementary Tables 4 and 5 show the autoantibody associations observed in different ancestral groups in replication cohort. These include some of the top SNP associations with autoantibodies observed in different ancestral groups (Supplementary Figure 2).

Canonical pathway analysis of GWAS candidate SNPs

A pathway analysis of the networks enriched among the top SNPs in the discovery cohort was generated through the use of Ingenuity Pathway Analysis (IPA; Ingenuity Systems, www.ingenuity.com). All SNPs from the discovery cohort with P < 10−4 were included. The top canonical pathways related to IFN-α associated SNPs which pass a Benjamini–Hochberg false discovery rate of 0.05 are shown in Table 3. There was prominent representation of pathways associated with neural signaling and transmission and purine metabolism, and T-cell signaling was also highlighted. Some of the key molecules defining these pathways were also some of the top validated serum IFN-α-associated loci in our replication cohort, such as PNP and PRKG1. Networks enriched in our study included those with various cellular functions such as cell morphology, cellular assembly and organization (PRKG1), cellular development and cell-mediated immune response (PNP), (Supplementary Table 6).

Genome-scale integrated analysis of gene networks in tissues (GIANT)

Because the top loci identified in this study were not classical type 1 IFN pathway genes, we used the GIANT software to query potential relevance of the gene products encoded by these loci in various immune cell subsets. Figure 2 shows the networks produced by the GIANT algorithm when the top hits from our study are used as the input data in the various immune cell subset
Networks with the highest density were observed in dendritic cells and natural killer (NK) cells, and low density networks were seen in T and B lymphocytes (Table 4). Similarly, the top associations with serum IFN-α generally demonstrated the greatest network strength in plasmacytoid dendritic cells and NK cells. These data support biological relevance of the transcripts in dendritic cells, which have been implicated as the major IFN-α producing cell type in SLE, and NK cells, which have been reported to have a critical cooperative role with dendritic cells in the production of IFN-α. In addition, when examining the other molecules functionally implicated in these networks, a number of SLE-associated molecules are observed in the network diagrams, including IL12, TLR7 and the JAK/STAT pathways.

Figure 2. Tissue-specific analysis of gene networks in different immune cells. Networks demonstrate relationships between PNP, PRKG1, ANKRD44 and PLEKHF2 to other molecules in immune cells. Edges with weight (relative confidence) 0.4 are shown. Each network diagram represents a different immune cell type as follows: a, B lymphocyte; b, dendritic cell; c, monocyte; d, neutrophil; e, NK cell; f, T lymphocyte.
**DISCUSSION**

SLE is a highly heterogeneous disease, hence it is likely that certain genetic factors will be related to particular disease phenotypes and pathogenic pathways.\(^1,12,32\) and that genetic associations will not be shared between all SLE patients. We suspect this is a major factor in the unexplained heritability of the disease to date. Directly studying this heterogeneity by mapping a causal molecular trait in this study greatly enhanced our power to detect novel genetic variations underlying SLE disease pathogenesis. The top loci in our study have not been previously reported in other case-control studies of SLE, and were not top loci in the initial case-control analysis of the GWAS data set we used in this study.\(^7\) Thus, our alternative strategy was capable of finding genetic variants associated with disease that are not readily apparent in case-control designs, supporting a complexity in genetic architecture that will require molecular subphenotyping to fully delineate.

**Table 4.** Network density and network strength analysis for tissue-specific gene networks in different immune cells

| Cells           | Network density | Network strength |
|-----------------|-----------------|------------------|
|                 | ANKRD44 | PNP  | PRKG1 | PLEKHF2 |
| B lymphocyte    | 0.12   | 1.1  | 5.5   |         |
| Dendritic cell  | 0.58   | 9.3  | 27.4  |         |
| Monocyte        | 0.18   | 0.4  | 3.7   | 3.4     |
| Neutrophil      | 0.13   | 1.4  | 1.9   | 8.7     |
| T lymphocyte    | 0.09   | 0.4  | 1.4   | 7.3     |
| NK cell         | 0.52   | 11.7 | 1.4   | 17.0    |

Abbreviation: NK cell, natural killer cell. Networks generated by the GIANT software program for each immune cell type. Network density and strength calculated as described in the Methods. Density is calculated for the overall network in the cell, and strength is calculated for each of the loci entered in the analysis.

and idiopathic thrombocytopenic purpura.\(^38\) The SNP we find associated is a common coding-change variant, which does not cause complete deficiency, and whether this variant results in some change in enzyme function is not currently known. In *silico* bioinformatic analysis using Polymorphism Phenotyping 2 (PolyPhen2) and Sorting Intolerant From Tolerant (SIFT) predicts this SNP as non-damaging, but an effect on enzyme activity would still be possible. There was strong representation of the purine metabolic pathway in our canonical pathway analysis, and PNP was the key molecule associated with this pathway. Some rare, highly penetrant variants in genes involved in nucleic acid metabolism have been associated with SLE, such as *TREX1* and *DNA1EL3.*\(^39-41\) Given this precedent, *PNP* is a fascinating genetic association with SLE.

The non-European ancestral backgrounds studied were smaller, and did not allow for strong independent significance. Our discovery set was exclusively of European ancestry, and thus variants specific to other ancestral backgrounds could not have been discovered. Despite these limitations, we observed some interesting evidence for associations, which were of similar effect in both European-American and African-American ancestral backgrounds. Both *ANKRD44* (rs4850410; \(P_{\text{Meta}} = 1.3 \times 10^{-5}\)) and *PLEKHF2* (rs297573; \(P_{\text{Meta}} = 1.2 \times 10^{-5}\)) were associated with IFN-\(\alpha\) in African-American and European ancestral backgrounds. *PLEKHF2* is an endosome-associated protein responsible for modulating the structure and function of endosomes, as well as the endocytotic process.\(^42\) *PLEKHF2* can increase the activity of caspase 12, and a role in ER-related apoptotic pathway has been suggested.\(^43\) ENCODE ChiP-seq data demonstrate that rs297573 SNP downstream of *PLEKHF2*, which was associated with IFN-\(\alpha\) in our study, resides in NFkB transcription factor binding site. ANKRD44 has not been extensively studied, but it binds to the catalytic subunit of protein phosphatase 6\(^44\) which has a role in cell-cycle progression.

Our initial discovery cohort showed association of two SNPs in the *C7orf57* locus in chromosome 7 with serum IFN-\(\alpha\) activity; however, this locus failed to replicate. One of the possible reasons for lack of replication could be that this locus was related to some unique feature of the discovery cohort that was not present in the replication cohort. One previous GWAS study of amyotrophic lateral sclerosis, which used a very similar Illumina genotyping platform, found evidence for association between these two SNPs and amyotrophic lateral sclerosis, which then failed to replicate in an additional independent replication cohort.\(^45\) It is possible that some peculiarity of the earlier Illumina genotyping platform made it more likely for these SNPs to be spuriously associated, although this locus was not associated in the original SLEGEN GWAS case-control study.\(^7\) We used an entirely different reverse transcriptase PCR-based genotyping method for our replication cohort to eliminate potential platform-related biases.

As referenced above, discovery methods followed by replication in non-European ancestral backgrounds would be an important next step to this work. It is likely that some polymorphisms will be ancestry specific, and will not be evident until a discovery strategy is used in that particular ancestral background. This would be especially important for African-Americans who have a higher incidence of SLE and more severe clinical manifestations.\(^1\) African-American SLE cases also have higher levels of serum IFN-\(\alpha\) activity,\(^15\) which could be one factor related to the increased incidence and severity of the disease. Our findings could have pharmacogenomic implications, as therapeutics targeting the IFN-\(\alpha\) pathway are currently in development for SLE. Knowledge of the functional genetic factors underlying IFN-\(\alpha\) dysregulation in a given patient could be useful in individualizing therapy with these agents.
We followed up IFN-α from all patients in both cohorts included in this study, and the study was summarized in Supplementary Table 3. Informed consent was obtained from all patients in both cohorts included in this study, and the study was approved by the institutional review boards at the respective institutions. We incorporated 238 Hispanic/Native-American and 450 African-Americans. We incorporated 238 Hispanic/Native-American and 450 African-Americans. We incorporated 238 Hispanic/Native-American and 450 African-Americans. We incorporated 238 Hispanic/Native-American and 450 African-Americans.

DNA using the Fluidigm Biomark microfluidic qPCR system (Fluidigm Corp., South San Francisco, CA, USA). All DNA samples were pre-amplified using the SNP-type primers from the genotyping assays, according to the manufacturer’s protocol. PCR data were analyzed using the BioMark SNP Genotyping Analysis software version 3 (Fluidigm Corp) to obtain genotype calls. Scatter plots were all reviewed individually for quality, and SNPs that deviated significantly from the expected Hardy–Weinberg proportions (P < 0.001) or with < 95% genotyping success were excluded from the analysis.

Measurement of serum IFN-α activity

Enzyme-linked immunosorbent assay methods for the measurement of type I IFN in human sera have been complicated by low sensitivity and specificity.47 We used a well-documented sensitive and reproducible reporter cell assay to generate IFN-α activity data from patient sera.48–50 The reporter cells in this assay (WISH cells, ATCC# CCL-25, Manassas, VA, USA) measures the ability of patient sera to cause IFN-induced gene expression. These cells are an epithelial-derived cell line that are highly responsive to reporter cells in this assay (WISH cells, ATCC# CCL-25, Manassas, VA, USA). Serum from healthy unrelated controls (n = 200) were tested to establish a normal value for the assay. Results from patient samples are expressed as the number of standard deviation (s.d.) above the mean of healthy unrelated control sera. The sum of the number of s.d. above healthy controls for the three transcripts is used as the quantitative output from the assay, representing a serum IFN-α activity score. This assay has been extremely informative in SLE and other autoimmune diseases.48–50

Statistical analysis

Control for population structure. To account for potential differences in admixture or population structure within self-reported ancestral backgrounds in the discovery and replication cohort, we performed a PCA using the GWAS SNPs and the independent ancestry-informative markers SNPs, respectively. PCA in the discovery cohort was carried out on all of the GWAS SNPs that passed quality control thresholds. This cohort is composed of SLE cases with self-reported European ancestry, and as shown in the PCA plots (Figure 3a), there were no major population outliers. As expected, cases with varying proportions of Northern- and Southern-European ancestry were included in the study, and some cases clustered with the Ashkenazi Hap Map reference population, suggesting Jewish ancestry (Figure 3b). PCA of the ancestry-informative markers genotyped in the replication cohort revealed that the PC 1 obtained in this analysis provided a strong separation between subjects of self-reported African-American ancestry and the non-African ancestral backgrounds, while PCs 2 and 3 provided separation between the PCA-defined Asian-, Hispanic- and European-American ancestry (Figures 3c and d). Self-reported Hispanic- or Native-American ancestry subjects were largely overlapping in this analysis, and are considered together in this analysis. Association analyses were not performed in the Hispanic- or Native-American (n = 238) and Asian-American cases (n = 40), due to the small number of cases. These subjects were included in the PCA only of the ancestry-informative markers to assist with the determination of population structure. Correction for population structure within the discovery and replication cohort was done using the first three PCs as covariates in the logistic regression association analyses. PCA analysis was performed using Cluster 3.0 software.51

Association analyses. Logistic regression analysis was used to detect associations between the SNPs and serum IFN-α in both stages of the study. IFN-α activity was studied as a categorical trait because the trait distribution is highly skewed, such that log transformation does not result in a normal distribution and the highly skewed data did not allow for linear modeling in a quantiative trait locus analysis. We used a binning strategy that has been highly informative in previous large scale studies and multivariate analyses of the serum IFN trait in SLE.51–53 In which subjects with a value > 2 s.d. above the mean of healthy controls are binned as high IFN-α, and the rest are binned as low IFN-α. Using this binning strategy prevents high outlying values from exerting an inordinate amount of influence in the model. In the discovery cohort, 88 were categorized as high IFN-α, and 322 were categorized as low IFN-α. Logistic regression analysis was carried out using PLINK v.1.07 software.51 The first three PCs from the PCA of the GWAS SNPs were used as covariates in the logistic regression to control for population structure in the discovery cohort. In the replication cohort, each self-reported ancestral background was analyzed separately, and the first three PCs were included as covariates to correct for population structure and admixture. Regression analysis was also performed to detect any potential associations between the presence of particular autoantibodies and SNPs in the replication cohort, because autoantibodies have been associated with high IFN-α in SLE patients.15 In the replication cohort, we used the Benjamini–Hochberg procedure to control the false discovery rate at 0.05, and the SNPs which passed this threshold were considered for meta-analysis. The P-value threshold used for significance in the overall meta-analysis corrects for the number of SNPs which were analyzed for association in the initial GWAS discovery analysis, controlling the family-wise error rate at the 0.05 level. For SNPs that demonstrated a homogenous effect across the discovery and replication sets by Breslow–Day testing, meta-analysis was performed using the weighted Z-score method52 using R 2.11.1 statistical analysis software from www.projectmidas.org for statistical correction. For SNP×ancestral background interactions, we applied a Bonferroni correction to the meta-analysis results using the number of SNPs that passed quality control in the discovery GWAS (n = 291 943), resulting in a threshold P-value for this study of P = 1.71 × 10−7. In the cross-ancestral background analysis, SNPs that demonstrated a homogenous effect across both ancestral backgrounds were meta-analyzed using the same weighted Z-score method,52 assuming a fixed-effect model. Enrichment P-values were calculated using a Fisher’s exact test with the following parameters: for the SNP-wise calculation for the SLE-associated ITGAM SNP, the number of possible combinations was calculated using the same weighted Z-score method,52 assuming a fixed-effect model. Enrichment P-values were calculated using a Fisher’s exact test with the following parameters: for the SNP-wise calculation for the SLE-associated ITGAM SNP, the number of possible combinations was calculated using the same weighted Z-score method,52 assuming a fixed-effect model. Enrichment P-values were calculated using a Fisher’s exact test with the following parameters: for the SNP-wise calculation for the SLE-associated ITGAM SNP, the number of possible combinations was calculated using the same weighted Z-score method,52 assuming a fixed-effect model.

Canonical pathway analysis. From the initial discovery GWAS data, IFN-α associated SNPs (n = 323) with a P-value ≤ 1 × 10−4 were analyzed further using IPA to identify the top canonical pathways related to IFN-α associated SNPs. SNPs were attributed to the nearby gene, and the genes were then compared with curated functional attribution lists organized by
canonical pathway function. The magnitude of over-representation of a particular canonical pathway in the gene list from our study was calculated as the ratio of the number of molecules from our data set that map to the pathway divided by the total number of reference molecules in that pathway in the IPA database (a list of genes belonging to major canonical pathways is curated in IPA based on published literature). Statistical significance was determined using the Fisher’s exact test, comparing the observed ratio of genes in a particular pathway to the null expectation (that the genes would assort proportionally across all IPA pathways), to estimate the probability that the observed over-representation of the particular pathway would arise by chance.

Genome-scale integrated analysis of gene networks in tissues (GIANT). The top genes from the replication cohort were queried using the GIANT software program to determine likely functional relationships of these genes in various types of immune cells. GIANT is a public, web-based software program that uses tissue-specific gene expression databases to predict tissue-specific gene interactions (http://giant.princeton.edu/about/). About 145 tissues/cell types are available to be queried, including major immune cell subsets. The software generates functional networks based on the genes queried via the integration of thousands of publicly available gene expression datasets, sequence data, transcription factor binding sites and protein–protein interaction data to generate gene-association matrix. Bayesian weights derived from the gold-standard tissue-specific datasets are then applied, and networks are generated for each tissue queried, which illustrate the most probable functional relationships between the queried genes and other molecules in that particular tissue. Network relationship confidence (edge weight) was set at a minimum of 0.4 for our analyses. After the networks were generated, we calculated overall network density and network strength of each of our study genes in each immune cell subset network. Network density ($D$) was defined as a ratio of the number of edges ($E$) to the number of possible edges, given by the binomial coefficient $N \choose 2$, giving $D = 2E/N (N − 1)$; where $N$ = number of nodes. In these weighted networks, we calculated strength as the sum of a node’s edge weights.

Prediction of the impact of coding-change SNPs
Prediction of consequences on protein structure and/or function of non-synonymous SNPs were evaluated using the prediction programs SIFT (http://sift.bii.a-star.edu.sg/) and PolyPhen (http://genetics.bwh.harvard.edu/pph/). These two programs use algorithms to determine the likelihood that a particular coding-change polymorphism impacts protein-folding based upon local protein structure as well as the particular amino acid substitution.

**CONFLICT OF INTEREST**
The authors declare no conflict of interest.
ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institutes of Health under award numbers: R01 AR060861 (NTB), UL1 RR024999 (NTB), K08 A0083790 (NTB), R3P A0071651 (NTB), R01 AR62277 (HUB), R372471 (HUB), AR842460 (HUB), P01 A0083194 (HUB), P20 RR020143 (HUB), R01 DE018209 (HUB), R01 AR033274 (SKL), R01 CA141700 (ARME), R01 AR085621 (ARME), R01 AR43814 (TBP), P60 AR53308 (CLA), R01 AR44804 (CLA) and UL1 TR000004. Additional support was granted from the HHMI Gilliam Fellowship for Advanced Study (KSN), Alliance for Lupus Research (NTB, SKL, CLA and JCO), Merit Award from the US Department of Veterans Affairs (HUB), Lupus Research Institute (NTB), the Arthritis National Research Foundation Eng Tan Scholar Award (NTB), and the Lupus Foundation of Minnesota (NTB and SKL). Mayo Clinic Foundation (NTB), Kirkland Scholar Award (CLA) and the Wake Forest University Health Sciences Center for Public Health Genomics (LCO). We also thank the Instituto de Salud Carlos III (PI12/02558), partly supported with FEDER funds from the EU and the RNP Network BIOLUPUS financed by the European Science Foundation. The funders had no role in the study design, data collection and analysis, decision to publish or preparation of the manuscript.

REFERENCES

1. Harby JB, Kelly JA, Kaufman KM. Unraveling the genetics of systemic lupus erythematosus. Springer Semin Immunopathol 2006; 28: 119–130.
2. Ko K, Koldobskaya Y, Rosenzweig E, Niewold TB. Activation of the interferon pathway is dependent upon autoantibodies in African-American SLE patients, but not in European-American SLE patients. Front Immunol 2013; 4: 309.
3. Sanchez E, Comeau ME, Freedman BL, Kelly JA, Kaufman KM, Langefeld CD et al. Identification of novel genetic susceptibility loci in African American lupus patients in a candidate gene association study. Arthritis Rheum 2011; 63: 3493–3501.
4. Ko K, Franek BS, Marion M, Kaufman KM, Langefeld CD, Harley JB et al. Genetic ancestry, serum interferon-alpha activity, and autoantibodies in systemic lupus erythematosus. J Rheumatol 2012; 39: 1238–1240.
5. Niewold TB, Hua J, Lehman TJ, Harley JB, Crow MK. High serum IFN-alpha activity is a heritable risk factor for systemic lupus erythematosus. Genes Immun 2007; 8: 492–502.
6. Graham RR, Kozeyv SV, Baechler EC, Reddy MV, Plenge RM, Bauer JW et al. A common haplootype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 2006; 38: 550–555.
7. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. Nat Genet 2008; 40: 204–210.
8. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007; 357: 977–986.
9. Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X et al. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci in systemic lupus erythematosus. J Rheumatol 2011; 38: 1288–1293.
10. Rullo OJ, Tsoo BP. Recent insights into the genetic basis of systemic lupus erythematosus. Ann Rheum Dis 2013; 72(Suppl 2): i56–i61.
11. Ghodke-Puranik Y, Niewold TB. Genetics of the type I interferon pathway in systemic lupus erythematosus. Int J Clin Rheumatol 2013; 8: 657–669.
12. Kim K, Brown EE, Choi CB, Alarcon-Riquelme ME, Kelly JA, Glenn SB et al. Variation in the ICAM1-ICAM4-ICAM5 locus is associated with systemic lupus erythematosus susceptibility in multiple ancestries. Ann Rheum Dis 2012; 71: 1809–1814.
13. Blanco P, Palacuka AK, Gill M, Pascual V, Banchereau J. Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. Science 2001; 294: 1540–1543.
14. Hooks JJ, Moutsopoulos HM, Geis SA, Stahl NJ, Decker JL, Notkins AL. Immune interferon in the circulation of patients with autoimmune disease. N Engl J Med 1979; 301: 5–8.
15. Weckerle CE, Franek BS, Kelly JA, Kumabe M, Mikolaitis RA, Green SL et al. Network analysis of associations between serum interferon-alpha activity, autoantibodies, and clinical features in systemic lupus erythematosus. Arthritis Rheumatol 2011; 63: 1044–1053.
16. Patil M, Singh S, Tesfayeone H, Dedrick R, Fry K, Lal P et al. Longitudinal expression of type I interferon responsive genes in systemic lupus erythematosus. Lupus 2013; 19: 980–989.
17. Niewold TB, Clark DN, Salloum R, Poole BD. Interferon alpha in systemic lupus erythematosus. J Biomed Biotechnol 2010; 2010: 948364.
18. Niewold TB. Interferon alpha-induced lupus: proof of principle. J Clin Rheumatol 2008; 14: 131–132.
42 Lin WJ, Yang CY, Lin YC, Tsai MC, Yang CW, Tung CY et al. Phafin2 modulates the structure and function of endosomes by a Rab5-dependent mechanism. Biochem Biophys Res Commun 2010; 391: 1043–1048.

43 Li C, Liu Q, Li N, Chen W, Wang L, Wang Y et al. EAPF/Phafin-2, a novel endoplasmic reticulum-associated protein, facilitates TNF-alpha-triggered cellular apoptosis through endoplasmic reticulum-mitochondrial apoptotic pathway. J Mol Med (Berl) 2008; 86: 471–484.

44 Stefansson B, Ohama T, Daugherty AE, Brautigan DL. Protein phosphatase 6 regulatory subunits composed of ankyrin repeat domains. Biochemistry 2008; 47: 1442–1451.

45 Chio A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai SL et al. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. Hum Mol Genet 2009; 18: 1524–1532.

46 Kosoy R, Nassir R, Tian C, White PA, Butler LM, Silva G et al. Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 2009; 30: 69–78.

47 Jabs WJ, Hennig C, Zawatzky R, Kirchner H. Failure to detect antiviral activity in serum and plasma of healthy individuals displaying high activity in ELISA for IFN-alpha and IFN-beta. J Interferon Cytokine Res 1999; 19: 463–469.

48 Hua J, Kirou K, Lee C, Crow MK. Functional assay of type I interferon in systemic lupus erythematosus plasma and association with anti-RNA binding protein autoantibodies. Arthritis Rheum 2006; 54: 1906–1916.

49 Feng X, Han D, Kilaru BK, Franek BS, Niewold TB, Reder AT. Inhibition of interferon-beta responses in multiple sclerosis immune cells associated with high-dose statins. Arch Neurol 2012; 69: 1303–1309.

50 Yeung KY, Ruzzo WL. Principal component analysis for clustering gene expression data. Bioinformatics 2001; 17: 763–774.

51 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.

52 Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher’s approach. J Evol Biol 2005; 18: 1368–1373.

Supplementary Information accompanies this paper on Genes and Immunity website (http://www.nature.com/gene)