Objective. The overexpression of interferon (IFN)–inducible genes is a prominent feature of systemic lupus erythematosus (SLE); it serves as a marker for active and more severe disease, and is also observed in other autoimmune and inflammatory conditions. This study was undertaken to investigate the genetic variations responsible for sustained activation of IFN-responsive genes in SLE.

Methods. We systematically evaluated association of SLE with a total of 1,754 IFN pathway–related genes, including IFN-inducible genes known to be differentially expressed in SLE patients and their direct regulators. We used a 3-stage study design in which 2 cohorts (total of 939 SLE cases and 3,398 controls) were analyzed independently and jointly for association with SLE, and the results were adjusted for the number of comparisons.

Results. A total of 15,166 single-nucleotide polymorphisms (SNPs) passed all quality control filters; 305 of these SNPs demonstrated replicated association with SLE in both cohorts. Nine variants were further genotyped for confirmation in an average of 1,316 independent SLE cases and 3,215 independent controls. Association with SLE was confirmed for several genes, including those for the transmembrane receptor CD44.
(CD44 [rs507230]; \( P = 3.98 \times 10^{-12} \)), the cytokine pleiotrophin (PTN [rs919581]; \( P = 5.38 \times 10^{-4} \)), the heat-shock protein DnaJ (DNAJ1 [rs10971259]; \( P = 6.31 \times 10^{-5} \)), and the nuclear import protein karyopherin α1 (KPN1 [rs6810306]; \( P = 4.91 \times 10^{-5} \)).

Conclusion. This study expands the number of candidate genes that have been shown to be associated with SLE and highlights potential of pathway-based approaches for gene discovery. Identification of the causal alleles will help elucidate the molecular mechanisms responsible for activation of the IFN system in SLE.

Systemic lupus erythematosus (SLE) (MIM no. 152700) is a chronic and severe systemic autoimmune disease characterized by the production of high-titer autoantibodies directed against native DNA and a wide variety of other cellular constituents. The prevalence of SLE in the US is estimated to be between 0.05% and 0.1% of the population, and disproportionately affects women and African Americans (1). SLE susceptibility is strongly influenced by genetic factors (2–7), and associations of SLE with \( \sim 38 \) loci have thus far been convincingly established. Clustering of some genetic associations identified to date appears to fall into at least 3 major pathways, including immune complex processing, lymphocyte signaling, and interferon (IFN) pathways (8).

Numerous studies have clearly demonstrated that dysregulation of the IFN system occurs in SLE and closely related autoimmune phenotypes, including Sjögren’s syndrome, psoriasis, and others (9). Genome-wide transcriptional profiling in SLE has shown that many patients overexpress IFN-inducible genes (10–19). This overexpression of IFN-inducible genes, known as the “IFN signature,” is a marker for active and severe disease. Dysregulation of IFN responses also correlates with several clinical and laboratory criteria, and is present in virtually all pediatric cases (9). Furthermore, some individuals treated with IFNα later develop antinuclear antibodies or even SLE (20). High serum IFNα activity, consistent with overexpression of IFN-inducible genes, is a heritable trait in families with SLE (21). Sustained overproduction of IFN activates dendritic cells, autoactive T cells, autoactive B cells, and cytotoxic effector cells. Thus, many of the immunologic disturbances observed in SLE, such as peripheral tolerance breakdown, nuclear autoantibody production, immune complex formation, and systemic tissue damage, may be explained at least in part by an impaired IFN system (22).

The role of IFNs in the homeostasis of the immune system and their observed dysregulation in patients with SLE makes any gene in this system a potential candidate for SLE susceptibility. To date, association analyses have established the gene for IFN regulatory factor 5 (IRF-5) and a few others related to IFN pathways (e.g., STAT4, SPP1, and TREX1) as risk factors for SLE (8,23). It is likely that there are additional, as-yet- unidentified, IFN-related genes important in SLE, and it is crucial to investigate for their contributions to the disease.

In this study, we utilized a pathway-centric approach to perform the first comprehensive genetic association analysis of genes known to constitute the IFN signature, their direct regulators, and all other known IFN pathway genes based on literature and database searches. Independent discovery (stage 1) and replication (stage 2) data sets consisted of both the observed and the imputed IFN-related single-nucleotide polymorphisms (SNPs) from the genome-wide association studies (GWAS) by Harley et al (3) and Graham et al (2), respectively. We performed single-locus tests of association and admixture adjustments, and adjusted our results for the number of comparisons. We then confirmed the top findings in a third confirmation cohort (stage 3). In addition, we performed 2-way interaction tests of association and applied a more novel approach, alternating decision trees (ADTrees), to test the predictive ability of these polymorphisms and their potential higher-order architecture. We report novel SLE risk loci with confirmed evidence for association in all of the cohorts studied.

PATIENTS AND METHODS

Criteria for definition of IFN pathway-related genes. Set 1. We compiled 2 lists with different sets of IFN-related genes. Set 1 consisted of all genes that have been reported as differentially expressed IFN-inducible genes in SLE, plus all known IFN genes. This list was compiled from gene expression profiling studies that demonstrated an IFN signature in SLE patients (10–19), or by National Center for Biotechnology Information (NCBI) (www.ncbi.nlm.nih.gov/gene) and Ingenuity Pathway Analysis (IPA) (www.ingenuity.com) searches for genes (and pseudogenes) with “IFN” in the gene or protein name or alias.

Set 2. Set 2 comprised the full set of direct regulators of the differentially expressed IFN-inducible genes compiled from the literature and included in set 1. We used IPA to identify all of the regulators (cytokines, transporters, kinases, peptidases, phosphatases, growth factors, ion channels, nuclear receptors, transmembrane receptors, G protein–coupled receptors, transcription and translation regulators) that are upstream of, and have direct interactions with, the above genes from the literature.

Association analysis in the discovery cohort (stage 1). As described by Harley et al (3), a GWAS using 317,501 SNPs was performed by the International Consortium for Systemic
Lupus Erythematosus Genetics (SLEGEN; www.slegen.org) in 706 Caucasian women with SLE and 2,317 controls genotyped on the Illumina Infinium HumanHap300 BeadChip. Genotypes from these subjects were imputed using the program IMPUTE (24) version 0.5 for SNPs not genotyped or poorly genotyped. Imputation was performed using high-quality genotype data from the SLEGEN GWAS (3) and phased HapMap Phase II (NCBI B35 assembly) genotype data from 60 CEU HapMap founders. We considered SNPs that mapped within 50 kb upstream and 10 kb downstream of each IFN-related gene. We used SNPs that met the following quality criteria: 1) No statistically significant differences in the proportions of missing genotype data between cases and controls (i.e., \( P > 0.05 \)); 2) overall missing genotype data \(<10\%\); 3) Hardy-Weinberg expectations (HWE) \( P > 0.01 \) in controls and \( P > 0.0001 \) in cases; and 4) minor allele frequencies (MAFs) in controls within a 95% or 99.99% confidence interval for ethnicity-matched HapMap MAFs, for genotyped and imputed SNPs, respectively. SNPs that were retained had an estimated MAF of \( >0.01 \) in the control samples, an information score of \( >0.50 \), and a confidence score of \( >0.90 \). Imputed SNPs were analyzed using SNPTEST (24). We report the lowest \( P \) value among the additive, dominant, and recessive models; however, since these tests can be affected by low genotype counts, at least 30 homozygotes for the minor allele were required in order for the recessive or additive models to be considered. All genetic models were defined relative to the minor allele. To account for potential population stratification, we performed principal component analysis using all SNPs, as described (3). After adjustment for 4 principal components, the genome-wide inflation factor was \( \lambda = 1.05 \), indicating minimal inflation of the test statistics.

**Association analysis in the replication cohort (stage 2).** Replication studies were carried out on an independent set of subjects from the University of Minnesota SLE Cohort using GWAS data as described by Graham et al (2). Genotypes from 412 cases and 1,081 controls were imputed using IMPUTE version 0.5 for SNPs not genotyped or poorly genotyped in the Minnesota GWAS. Imputation was performed using high-quality genotype data from the Minnesota GWAS (Affymetrix Genome-Wide Human SNP Array 5.0 platform) and phased HapMap Phase II (NCBI B35 assembly) genotype data from 60 CEU HapMap founders. SNPs that were retained had an estimated MAF of \( >0.01 \) in the control samples, an information score of \( >0.50 \), and a confidence score of \( >0.90 \). Imputed SNPs were analyzed using SNPTEST. We used SNPs that met the same quality criteria as described above, and report the \( P \) value chosen as described above. After adjustment for 4 principal components, the genome-wide inflation factor in this analysis was \( \lambda = 1.05 \). Only those replications that showed consistency of the risk allele in both the discovery and replication studies are reported.

**Joint analysis of the discovery and replication cohorts (stages 1 and 2).** We combined the genotypic and imputed data from both the discovery (stage 1) (3) and replication (stage 2) (2) GWAS data sets and performed a joint analysis. A total of 179 SLE cases were duplicates or first-degree relatives between the studies and were removed from the replication set, bringing the number of Minnesota cases in the joint analysis to 233. We used SNPs that met the same quality criteria as described above, and report the \( P \) value chosen as described above. These analyses were adjusted for 4 principal components to account for admixture. The genome-wide inflation factor in the joint analysis was \( \lambda = 1.08 \). The reported \( P \) values for the tests of association are not adjusted for the number of comparisons. However, we corrected for multiple comparisons by applying a false discovery rate (FDR) (25) to all SNPs in the joint analysis that passed quality control within set 1 (2,169) and set 2 (12,997), and declare statistical significance only for those that met an FDR-adjusted threshold of significance (\( P < 0.05 \)) within each set.

**Meta-analysis with the confirmation cohort (stage 3).** We genotyped most of the SNPs that replicated in an additional, independent Caucasian confirmation cohort. This confirmation cohort consisted of 474 cases and 539 controls from the Lupus Family Registry and Repository (LFRR) (26), 739 cases from the PROFILE study (27), 902 cases and 214 controls taken from the UK, and 2,729 out-of-study controls. The UK sample was collected by one of the authors (TJV) and includes 200 controls taken from the 1958 British Birth Cohort. These SLE probands from the UK met the American College of Rheumatology criteria for SLE (28). Written consent was obtained from all participants. In the UK, ethics approval was obtained from Multi-Centre Research Ethics Committee. The out-of-study controls were available from the following studies at the dbGaP database (http://www.ncbi.nlm.nih.gov/gap): the FUSION study (Finland–United States Investigation of Non-insulin-dependent diabetes mellitus Genetics), the National Institute of Diabetes and Digestive and Kidney Diseases Inflammatory Bowel Disease Genetics Consortium Crohn’s Disease GWAS, the Cancer Genetic Markers of Susceptibility (CGEMS) Prostate Cancer GWAS—Stage 1: Prostate, Lung, Colon and Ovarian, the CGEMS Breast Cancer GWAS—Stage 1: Nurses’ Health Study, and the National Institute of Neurological Disorders and Stroke Parkinson’s Disease study. We excluded duplicates and relatives between these studies, and used all GWAS SNPs to compute principal components and exclude all outliers.

The following SNPs were genotyped in the LFRR samples: rs10124051, rs1880791, rs2285210, rs2613310, rs507230, rs6810306, rs749701, and rs755690. All of these SNPs plus rs10971259, rs366078, and rs4659444 were genotyped in the PROFILE collection. Only SNPs rs10124051, rs10971259, and rs919581 were genotyped in the UK samples. In the LFRR and PROFILE samples SNPs were genotyped using TaqMan Pre-Designed SNP Genotyping Assays, according to the protocol recommended by the manufacturer (Applied Biosystems). The UK samples were genotyped on a custom Illumina array. All SNPs had \(<5\%\) missing genotype data overall and HWE at \( P > 0.05 \) in their genotyping cohorts. The out-of-study controls had \(<14\%\) missing genotypes and HWE at \( P > 0.01 \). We performed a weighted Z score meta-analysis as implemented in METAL (www.sph.umich.edu/csg/abecasis/metal) between stages 1, 2, and 3 SNPs. METAL combines study-specific \( P \) values and direction of effect independently of \( \beta \) estimates, and then converts them into signed Z statistics that are summed with weights proportional to the square root of the sample size for each data set.

**Two-locus interaction analysis.** Using the discovery cohort (stage 1), we computed all 2-locus interaction analyses among all SNPs that had an individual locus \( P \) value of \(<0.2\). Specifically, we computed a logistic regression model with each SNP modeled under an additive genetic model and the interaction as the centered crossproduct of the SNPs under the
additive model. This analysis included 6,324 SNPs with \( P < 0.2 \) identified in the discovery cohort analysis. These SNPs met the same quality criteria defined above. In order to reduce false-positive interactions due to low MAFs, we rejected all of the pairs for which the expected number of individuals in our data was <10 for minor allele homozygotes. In addition, we excluded all SNP pairs with a linkage disequilibrium (LD) measure of \( r^2 > 0.2 \). The interactions were adjusted for 4 principal components to account for admixture. We then performed these analyses in the replication cohort. Although no pairs remained significant after correction for the number of comparisons, several interactions passed all quality control criteria and were replicated.

**ADTrees.** We also used an ADTrees approach (29,30) to identify the variants that best distinguish case versus control status and the multilocus relationships among the variants. This analysis included the 6,324 SNPs used for the 2-locus interaction analysis described above. In order to minimize biased results, we only used autosomal non-HLA-region SNPs that met the quality criteria defined above. We used Weka 3 software (www.cs.waikato.ac.nz/ml/weka/index.html) to create 100 bootstrap samples (random sampling with replacement) from the discovery cohort data set, and an ADTree was constructed for each sample. Common structural elements across the 100 trees were recorded, and the SNPs involved in each common element that appeared in at least 5% of the trees were removed from the data set. Structural elements are defined as paths through a tree that include the root and a leaf node. In the event that no structural element appeared in at least 5% of the trees, the criteria were reduced to paths through the tree that include only the root node and only paths appearing in at least 10% of the trees. Analysis was repeated until no common structural elements met the threshold of inclusion. The resulting paths were ranked by order in which they were found, and then by number of trees in which they appeared.

**SNP functionality evaluation.** LD between SNPs was assessed using Haploview 4.1 (www.broadinstitute.org/mpg/haploview) with HapMap CEU data (Release 24/Phase II) (hapmap.ncbi.nlm.nih.gov). SNP function was evaluated with the UCSC genome browser (genome.ucsc.edu) March 2006 assembly. SNP effects on expression levels were examined using the SCAN database (scan.bsd.uchicago.edu).

## RESULTS

**Findings of the association analysis.** We compiled a list of 323 IFN-inducible genes identified through gene expression profiling experiments in SLE and 136 additional IFN pathway genes for set 1. An additional 1,761 genes identified as direct regulators of the IFN-inducible genes in set 1 were included in set 2. Thus, a total of 2,220 genes were selected for this targeted IFN pathway study (Table 1).

We next searched for SNPs that mapped within 50 kb upstream and 10 kb downstream of each IFN-related gene in sets 1 and 2 and that were present in the discovery cohort GWAS data set (3) (stage 1). Set 1 had 4,759 SNPs and set 2 had 26,265 SNPs, both genotyped and imputed, that were then tested for association with SLE. We repeated this process using data from our replication cohort data set (2) (stage 2) and selected 4,680 SNPs in set 1 and 25,196 SNPs in set 2, also genotyped and imputed, to test for association with SLE. A total of 2,169 SNPs in set 1 and 12,997 SNPs in set 2 passed all quality control filters in both data sets. Of these, 56 SNPs within 26 loci in set 1 and 249 SNPs within 119 loci in set 2 resulted in \( P < 0.05 \) in both the discovery and replication cohorts (stages 1 and 2) (Table 1). It is noteworthy that we observed an enrichment of results above the null distribution: from the 2,169 SNPs tested in set 1 and the 12,997 tested in set 2, we observed 325 SNPs in set 1 and 1,608 SNPs in set 2 with \( P \) values of <0.05, well above the 109 SNPs expected in set 1 and 650 SNPs expected in set 2 by chance. The unadjusted \( P \) values from the joint analysis (\( P_{\text{joint}} \)) are reported herein, and, unless explicitly noted, only variants that survived an FDR correction (25) for the number of comparisons in the joint analysis (\( P_{\text{FDR}} \)) of stages 1 and 2 are reported.

Finally, we attempted to confirm 11 of the most significant novel SNPs observed in the joint analysis in an independent confirmation cohort (stage 3). A meta-analysis of data from the joint analysis of stages 1 and 2 and the confirmation cohort (stage 3) was then computed.

Examination of the 561 loci with evidence for association in both stage 1 and stage 2 revealed both known and novel SLE risk genes. Not surprisingly, the most significant SNPs were shown to reside in genes in the HLA region (\( CFB \) [rs1270942]; \( P_{\text{joint}} = 5.29 \times 10^{-25} \), \( P_{\text{FDR}} = 1.11 \times 10^{-21} \)). Other genes with well-
established SLE associations included IRF5-TNPO3 (rs10488631; $P_{\text{joint}} = 1.45 \times 10^{-6}$, $P_{\text{FDR}} = 1.17 \times 10^{-16}$), ITGAM-ITGAX (rs10661710; $P_{\text{joint}} = 1.17 \times 10^{-7}$, $P_{\text{FDR}} = 1.43 \times 10^{-2}$), another SNP in this locus in LD with the former ($r^2 = 0.8$) remained associated in the meta-analysis between the stage 1, 2, and 3 cohorts ($P = 6.31 \times 10^{-3}$). Differential expression of this gene in the peripheral blood of SLE patients versus controls has been reported by Baechler et al. (2010).

Two other loci in this set, RS6540A1 (rs64693444; $P_{\text{joint}} = 1.14 \times 10^{-3}$, $P_{\text{FDR}} = 8.85 \times 10^{-2}$) and IRF8 (rs366078; $P_{\text{joint}} = 2.41 \times 10^{-4}$, $P_{\text{FDR}} = 2.75 \times 10^{-2}$), showed associations that were further confirmed in a joint analysis between stages 1, 2, and 2 additional cases from stage 3 ($P = 2.14 \times 10^{-3}$ and $P = 9.63 \times 10^{-3}$, respectively).

The loci in set 2 (Table 3) showed stronger associations with SLE than those in set 1. The most significant variant found in set 2 lies in the spleen tyrosine kinase gene (SYK [rs2613310]; $P_{\text{joint}} = 1.53 \times 10^{-5}$, $P_{\text{FDR}} = 3.11 \times 10^{-3}$), but it did not remain associated in the meta-analysis of all cohorts (Table 4).

The next strongest effect identified resides in the gene for pleiotrophin (PTN), where 1 of 3 replicated variants passed an FDR adjustment (rs919581; $P_{\text{joint}} = 5.86 \times 10^{-5}$, $P_{\text{FDR}} = 1.12 \times 10^{-2}$). This SNP was genotyped on the UK samples from the confirmation (stage 3) cohort, and continued to show evidence for association in the meta-analysis between these samples and the original cohorts ($P = 5.38 \times 10^{-3}$). Nevertheless, this variant was not in LD ($r^2 < 0.2$) with any known functional SNP in the gene. PTN has direct interactions.

### Table 3. Most significant replicated SNPs in set 2*

<table>
<thead>
<tr>
<th>Marker</th>
<th>Chr.</th>
<th>Position, Mb</th>
<th>Locus</th>
<th>MA</th>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6810306</td>
<td>3</td>
<td>123 731</td>
<td>KPN1A1</td>
<td>G</td>
<td>1.25 \times 10^{-5} (D)</td>
<td>7.85 \times 10^{-3} (D)</td>
</tr>
<tr>
<td>rs919581</td>
<td>7</td>
<td>136 631</td>
<td>PTN</td>
<td>G</td>
<td>1.94 \times 10^{-3} (R)</td>
<td>6.66 \times 10^{-3} (R)</td>
</tr>
<tr>
<td>rs2613310</td>
<td>9</td>
<td>92 670</td>
<td>SYK</td>
<td>C</td>
<td>1.34 \times 10^{-4} (A)</td>
<td>5.49 \times 10^{-4} (D)</td>
</tr>
<tr>
<td>rs507230</td>
<td>11</td>
<td>35 086</td>
<td>CD44</td>
<td>G</td>
<td>1.25 \times 10^{-2} (R)</td>
<td>8.55 \times 10^{-2} (R)</td>
</tr>
<tr>
<td>rs755900</td>
<td>19</td>
<td>43 859</td>
<td>ACTN4</td>
<td>A</td>
<td>1.60 \times 10^{-3} (R)</td>
<td>3.87 \times 10^{-3} (R)</td>
</tr>
<tr>
<td>rs749701</td>
<td>19</td>
<td>43 882</td>
<td>ACTN4</td>
<td>T</td>
<td>2.40 \times 10^{-4} (R)</td>
<td>6.51 \times 10^{-5} (R)</td>
</tr>
<tr>
<td>rs575421</td>
<td>22</td>
<td>20 270</td>
<td>UBE2L3</td>
<td>T</td>
<td>1.22 \times 10^{-4} (A)</td>
<td>3.72 \times 10^{-4} (R)</td>
</tr>
<tr>
<td>rs2298429</td>
<td>22</td>
<td>20 313</td>
<td>UBE2L3</td>
<td>T</td>
<td>3.65 \times 10^{-3} (A)</td>
<td>3.72 \times 10^{-3} (R)</td>
</tr>
</tbody>
</table>

* MHC region not included. The reported $P$ values are not adjusted for the number of comparisons, but all markers met an FDR-adjusted threshold of significance in the joint analysis, as described in Patients and Methods. The smallest $P$ values are presented. See Table 2 for definitions.

† Imputed in the stage 2 cohort.
with several reported IFN-inducible molecules, i.e., proteasome subunit β type 10 (15), apolipoprotein B messenger RNA editing enzyme, catalytic polypeptide–like 3G (10), cystatin S (10), agrin (9,11), plexin B2 (10), and vascular endothelial growth factor A (10).

CD44 is important in lymphocyte activation and homing. One replicated variant 31.2 kb upstream the CD44 antigen gene, located in a large copy number variant (CNV), survived an FDR adjustment (rs507230; $P_{\text{FDR}} = 2.39 \times 10^{-5}$) (Table 3). Furthermore, this SNP showed a significant association in the meta-analysis of all cohorts ($P = 3.98 \times 10^{-12}$) (Table 4). This variant did not show LD ($r^2 < 0.2$) with any known functional SNP in CD44. This receptor has direct interactions with 4 reported IFN-inducible molecules (10): CD9, matrix metalloproteinase 9, versican, and erythrocyte membrane protein band 4.1–like 3. Furthermore, secreted phosphoprotein 1 (also known as osteopontin), a ligand for CD44, has previously been established as an SLE risk locus involved in IFN pathways (8,33).

A variant in the karyopherin α1 (KPN1) region showed association in the stage 1 and stage 2 analyses (rs6810306; $P_{\text{FDR}} = 2.68 \times 10^{-5}$) (Table 3), and its association was subsequently confirmed in the meta-analysis ($P = 4.91 \times 10^{-5}$) (Table 4). This variant locates 15 kb upstream of KPN1, in PARP9, in a 120-kb region of very high LD ($r^2 > 0.8$) comprising both genes. Karyopherin α1 is a direct regulator of the IFN signature genes STAT1 (10,11,18) and STAT2 (10,18).

UBE2L3, the gene for ubiquitin-conjugating enzyme E2L3, is another associated locus. However, we chose not to attempt to confirm it since the most significant variant identified in the present study has already been established as being associated with SLE (3,34).

**Findings of the interaction and ADTrees analyses.** The 2-locus interaction analysis revealed replicated interactions not due to LD between pairs of SNPs (Table 5). The most interesting pairs included serpin peptidase inhibitor, clade D, member 1 (SERPIND1; also known as heparin cofactor II [HCF2]) with β-parvin (PARVB), and Fyn-related kinase (FRK) with receptor-type protein tyrosine phosphatase D (PTPRD). An intronic variant in SERPIND1 showed interaction with a variant 37 kb upstream of PARVB, the gene for actin-binding protein, a focal adhesion protein. Both molecules are involved in cellular growth and proliferation. The interacting SNPs in FRK and PTPRD are also intronic. Interestingly, the latter lies in a CpG island and a region of CNV. None of these pairs are known to date to have any direct biologic interaction, and none have been previously associated with SLE. Although these interactions did not remain significant after adjustment for the number of combinations tested, they passed all quality control criteria and were replicated.

After exclusion of the HLA region and chromosome X, analyses using the ADTrees approach identified SNPs in the IRF5-TNPO3 and ITGAM loci as the best discriminators between case status and control status, thus confirming the association results; rs10488631 and rs4728142, in the IRF5 region, were found in 16% and 8% of the bootstrap samples, respectively, and
rs9888739 and rs9937837, in ITGAM, were present in 10% and 8% of the bootstrap samples, respectively. Three novel loci were chosen in 6% of all bootstrap data sets: rs11605818 at ATG16L2-FCHSD2, rs11655550 at MED1-CRKRS, and rs2850724 in NFATC1 were present in the same structural feature in 6% of the bootstrap samples, suggesting a potential interaction in the discovery cohort data.

**DISCUSSION**

The discovery that a signature of coordinately overexpressed IFN-inducible genes is prominent in a substantial proportion of patients with SLE has fueled interest in the IFN pathway as a potential target for therapeutic intervention. This molecular signature is correlated with increased disease activity and specific clinical manifestations, such as low complement levels, high levels of anti–double-stranded DNA, elevated erythrocyte sedimentation rates, and increased renal complications (35). The goal of this study was to identify the genetic variation that leads to dysregulation of the IFN-related pathways and genes, including IFN-inducible genes and their direct regulators. This essentially Bayesian approach of selecting candidate genes based on prior knowledge serves to increase the reliability and likelihood of finding genes that are truly associated with disease (36–38). To our knowledge, this study represents the most comprehensive IFN pathway–based genetic analysis to date. Using independent cohorts for discovery and replication, we have evaluated a total of 1,754 genes. Eight genes were confirmed as being associated with SLE.

The associations between SLE and IFN-related genes that were most significant overall were observed with SNPs located in regions that have been previously reported and firmly established as risk factors (HLA, IRF5-TNPO3, ITGAM-ITGAX, STAT4, and TNFAIP3).

Two additional genes with previous evidence for association include IRF8 (34) and the ubiquitin enzyme gene UBE2L3 (3). The majority of additional loci we report represent novel genetic associations with SLE, underscoring the power of this candidate pathway approach.

The strongest novel genetic effect locates in CD44. CD44 is an integral cell membrane glycoprotein that is important in cell–cell interactions. As a key regulator of many molecules, including IFN/H9253 and Lck, CD44 has important roles in lymphocyte activation, recirculation and homing, hematopoiesis, tumor metastasis, and inflammation. Additionally T cells from SLE patients have been shown to display an increased and abnormal distribution of CD44 (39). Kaufman et al (40) genotyped 4 SNPs within CD44 in 13 African American families and found no evidence of association. However, given the number of SNPs and families tested, that study was likely underpowered to detect the effect that we report herein.

Another significant association signal was identified in the heparin-binding PTN gene. PTN is a developmentally regulated cytokine with fibrinolytic, antiapoptotic, mitogenic, transforming, angiogenic, and chemotactic activities (41). It has recently been shown to induce the expression of inflammatory cytokines, including tumor necrosis factor α, interleukin-1β, and interleukin-6, in quiescent human peripheral blood mononuclear cells (42). Other than being a regulator of several of the IFN signature genes, no link between PTN and SLE has been established to date. Nevertheless, its expression is up-regulated in experimental autoimmune encephalomyelitis (43).

We have also confirmed an association with DNAJ. DnaJ is a heat-shock protein that assists the chaperone Hsp70 in protein translation, folding, unfolding, translocation, and degradation. As a stress response protein, DnaJ is involved in repair and removal of damaged proteins, and is therefore important for main-

<table>
<thead>
<tr>
<th>SNP pair</th>
<th>Stage 1</th>
<th>Stage 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1544155 (FRK; Chr. 6, 116,453 kb) and rs1021720 (PTPRD; Chr. 9, 8,391 kb)</td>
<td>P = 3.40 × 10^-3</td>
<td>OR (95% CI) 9.50 × 10^-3 1.47 (1.10–1.97)</td>
</tr>
<tr>
<td>rs1050373 (SERPIND1; Chr. 22, 19,460 kb) and rs2073082 (PARVB; Chr. 22, 42,685 kb)†</td>
<td>P = 3.00 × 10^-4</td>
<td>OR (95% CI) 1.50 (1.21–1.86) 1.46 (1.19–1.79)</td>
</tr>
</tbody>
</table>

* See Table 2 for definitions. † Stage 1 r^2 = 0.00038.
taining cell viability. Heat-shock proteins are potential targets of an autoimmune response and have been implicated in the induction and propagation of autoimmunity in several diseases, including rheumatoid arthritis and type 1 diabetes mellitus (44). Experimental evidence suggests that improper protein folding may promote autoimmunity (45). Thus, DnaJ and related proteins have potentially important but poorly understood roles in autoimmune diseases.

Karyopherin α1 binds recombination-activating gene 1, a lymphoid-specific recombinase essential for V(D)J recombination, and influences its subnuclear localization, hence controlling the generation of immunoglobulins and T cell receptors (46). It has also been shown to bind activated STAT-1 and IRF-5 proteins and transport them to the nucleus (47,48).

Even though our objective was to perform a comprehensive analysis of all IFN-related genes, we cannot exclude the possibility that strong associations were missed due to the genomic coverage of the genotyping arrays or the a priori selection of specific genes reported in the literature to be IFN-related or interactors. We therefore could have missed some unknown interactions. In addition, the dysregulation of IFN pathway genes is not a uniform feature among all patients with lupus, and as such we would expect to detect moderate genetic effects that would occur in probably half of our cases. Nonetheless, replication of the identified novel effects in a second cohort, correction for multiple comparisons, and confirmation in a third cohort increase our confidence in the robustness of these associations.

In addition to the conventional statistical approaches, we used a data mining approach, ADTrees, to try to corroborate the association results and identify novel variants that best discriminate between cases and controls, as well as to confirm and reveal potential interactions between genetic variants. The ADTrees analysis validated the association results, and replication analyses are under way to confirm the 2-locus interactions that were observed.

In summary, we have identified multiple IFN pathway-related genes that show confirmatory evidence for association with SLE. For the majority of loci, this is the first report of a genetic association with SLE with confirmation. Taken together, these new data expand the growing list of genes that show association with SLE and emphasize the genetic contribution of dysregulated IFN pathways. Understanding how these genetic factors might contribute to pathogenesis should ultimately lead to important opportunities for developing therapeutic targets to control the active IFN signature seen in patients with SLE.

ACKNOWLEDGMENTS

We are extremely grateful to Miranda Marion, Joshua Grab, and Lingyi Lu for assistance with analysis of the data, and to Debbie McDuffie for assistance with the genotyping assays.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Ramos had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.


REFERENCES


ture in peripheral blood cells of patients with severe lupus. Proc Natl Acad Sci U S A 2003;100:2610–5.