nature genetics

A common haplotype of interferon regulatory factor 5 (*IRF5*) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus

Robert R Graham[1,9], Sergey V Kozyrev[2,9], Emily C Baechler[3,9], MV Prasad Linga Reddy[2,9], Robert M Plenge[1], Jason W Bauer[3], Ward A Ortmann[3], Thearith Koeuth[3], Ma Francisca González Escribano[4], the Argentine and Spanish Collaborative Groups, Bernardo Pons-Estel[5], Michelle Petri[6], Mark Daly[1], Peter K Gregersen[7], Javier Martín[8], David Altshuler[1], Timothy W Behrens[3] & Marta E Alarcón-Riquelme[2]

Systemic lupus erythematosus (SLE) is a complex autoimmune disease[1] characterized by activation of the type I interferon (IFN) pathway[2–4]. Here we convincingly replicate association of the IFN regulatory factor 5 (*IRF5*) rs2004640 T allele with SLE[5] in four independent case-control cohorts ($P = 4.4 \times 10^{-16}$) and by family-based transmission disequilibrium test analysis ($P = 0.0006$). The rs2004640 T allele creates a 5′ donor splice site in an alternate exon 1 of *IRF5*, allowing expression of several unique IRF5 isoforms. We also identify an independent *cis*-acting variant associated with elevated expression of *IRF5* and linked to the exon 1B splice site. Haplotypes carrying the variant associated with elevated expression and lacking the exon 1B donor site do not confer risk of SLE. Thus, a common *IRF5* haplotype driving elevated expression of multiple unique isoforms of IRF5 is an important genetic risk factor for SLE, establishing a causal role for type I IFN pathway genes in human autoimmunity.

Type I IFN is a central mediator of viral immunity[6], and recent studies have shown that many individuals with SLE strongly overexpress IFN-responsive genes in blood cells[4,7,8]. However, it is not known whether the IFN expression signature is a general biomarker of a dysregulated immune system or rather if it reflects primary genetic variation causal to the pathogenesis of human SLE. A recent study reported an allelic association between interferon regulatory factor 5 (*IRF5*) and SLE, with the best signal at SNP rs2004640 (ref. 5). IRF5 is a member of a family of transcription factors that controls inflammatory and immune responses[9]. It is critical for the production of the pro-inflammatory cytokines TNF-α, IL-12 and IL-6 following Toll-like receptor (TLR) signaling, as determined by knockout mouse studies[10], and is also important for transactivation of type I IFN and IFN-responsive genes[11,12].

In an effort to replicate the *IRF5* association with SLE, we genotyped four sets of SLE cases and controls from the US, Spain, Sweden and Argentina (total of 1,661 cases and 2,508 controls) and assessed association of the *IRF5* rs2004640 T allele using a standard case-control study design. We found a significant enrichment of the T allele in cases compared with matched controls (overall, 60.4% in cases versus 51.5% in controls; $P = 4.4 \times 10^{-16}$; **Table 1**). The frequency of the T allele was lower in the Argentine sample, possibly owing to the mixed ethnicity of the individuals studied. Notably, in a subset of 470 cases from the US for which family members were available, a family-based association ruled out the possibility that stratification could explain the results ($P = 0.0006$, **Table 2**).

When we examined all the available case-control data[13,14] (four independent cohorts described here, plus the previously published Swedish and Finnish cohorts[5]), we observed robust and consistent association of the rs2004640 T allele with SLE, with individual odds ratios (OR) ranging between 1.31 and 1.84 (**Table 1**). In the combined analysis, the best estimate of risk for the rs2004640 SNP T allele is OR = 1.47 (1.36–1.60), with an overall $P = 4.2 \times 10^{-21}$ (**Table 1**). We found a single copy of the rs2004640 T allele in 45% of cases, which conferred modest risk (pooled OR = 1.27, $P = 0.0031$), whereas the 38% of cases homozygous for the T allele were at a greater risk for SLE (pooled OR = 2.01, $P = 3.7 \times 10^{-14}$; **Supplementary Table 1** online). These data are most consistent with a general model of inheritance in which individuals carrying one copy of rs2004640 T allele are at increased risk for SLE, and individuals carrying two copies

[1]Program in Medical and Population Genetics, Broad Institute of Harvard and the Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA, and the Center for Human Genetics Research and Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [2]Department of Genetics and Pathology, Unit of Medical Genetics, Rudbcek Laboratory, Uppsala University, Uppsala 75185, Sweden. [3]Center for Immunology, University of Minnesota Medical School, Minneapolis, Minnesota 55455, USA. [4]Servicio de Inmunología, Hospital Virgen del Rocio, Sevilla 41013, Spain. [5]Sanatorio Parque, Rosario 2000, Argentina. [6]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [7]The Feinstein Institute for Medical Research, North Shore-Long Island Jewish Health System, Manhasset, New York 11030, USA. [8]Instituto de Biomedicina López-Neyra, Consejo Superior de Investigación Científica (CSIC), Granada 18100, Spain. [9]These authors contributed equally to this work. Correspondence should be addressed to M.E.A.-R. (marta.alarcon@genpat.uu.se) or T.W.B. (behre001@umn.edu).

Received 6 December 2005; accepted 14 March 2006; published online 16 April 2006; doi:10.1038/ng1782

Table 1 Case-control association analysis of the *IRF5* rs2004640 T allele with SLE

		N	No. of T alleles	Freq. of T alleles	No. of G alleles	Freq. of G alleles	OR (95% c.i.)	χ^2	*P*	Pooled OR[a]	Pooled *P*
Argentina	Cases	284	309	0.54	259	0.46	1.52 (1.20–1.93)	12.8	0.00035		
	Controls	279	245	0.44	313	0.56					
Spain	Cases	444	559	0.63	329	0.37	1.42 (1.18–1.71)	14.3	0.00016		
	Controls	541	589	0.54	493	0.46					
										1.45 (1.32–1.58)	4.4×10^{-16}
Sweden-1	Cases	208	260	0.63	156	0.38	1.31 (1.01–1.71)	4.1	0.04268		
	Controls	254	284	0.56	224	0.44					
USA	Cases	725	879	0.61	571	0.39	1.47 (1.29–1.67)	34.8	3.6×10^{-9}		
	Controls	1434	1467	0.51	1401	0.49					
Sweden-2[b]	Cases	480	595	0.62	365	0.38	1.51 (1.21–1.87)	13.8	0.0002		
	Controls	256	266	0.52	246	0.48					
										1.59 (1.31–1.94)	7.1×10^{-7}
Finland[b]	Cases	109	137	0.63	81	0.37	1.84 (1.27–2.66)	10.3	0.00133		
	Controls	121	116	0.48	126	0.52					
Combined analysis	Cases	2250	2739	0.61	1761	0.39					
	Controls	2885	2967	0.51	2803	0.49				1.47 (1.36–1.60)	4.2×10^{-21}

N, Number of individuals. c.i., confidence interval. 'Number of alleles' refers to number of alleles of rs2004640. [a]Mantel-Haenszel test of pooled odds ratios and 95% confidence intervals. [b]Data from ref. 5.

have an even higher relative risk. The relative risks for these genotypes are independent, separate parameters. Multiplicative, additive, dominant and recessive models were rejected at *P* values less than 0.032, 0.0008, 0.0015 and 1×10^{-13}, respectively (see URLs section of Methods). Thus, the evidence for association of the T allele of rs2004640 is highly significant, well surpassing even correction for testing all common variants in the human genome.

Table 2 TDT analysis of *IRF5* SNPs and haplotypes with SLE in 467 US pedigrees of European ancestry

Marker	Allele	Frequency[a]	T[b]	U[b]	T/U[b]	χ^2	Nominal *P*[c]	Permuted *P*[d]
rs729302	A	0.69	199	130	1.53	14.5	0.0001	0.0007
rs2004640	T	0.57	219	153	1.43	11.7	0.0006	0.0028
rs752637	C	0.68	199	161	1.24	4.0	0.0450	0.2999
rs2280714	T	0.72	153	127	1.20	2.4	0.1202	0.6248

Haplotype[e]	Frequency	T	U	T/U	χ^2	Nominal *P*	Permuted *P*
ATCT	0.54	192	139	1.38	8.5	0.0035	0.0192
CGTC	0.15	69	101	0.69	5.7	0.0167	0.1093
AGTC	0.13	88	77	1.14	0.7	0.4047	1.0000
CGCT	0.10	52	79	0.66	5.3	0.0216	0.1424
CGTT	0.04	18	29	0.60	2.9	0.0897	0.5105
CTCT	0.02	16	11	1.54	1.2	0.2645	0.9048
XTXT[f]	0.56	208	149	1.39	9.5	0.0021	–
XGXT	0.14	70	108	0.65	8.2	0.0042	–
XGXC	0.28	157	177	0.88	1.3	0.2542	–

[a]Frequency in parental chromosomes. [b]Transmitted and untransmitted chromosomes, and the transmission ratio (T/U). [c]*P* value, uncorrected for multiple tests. [d]*P* value from 1,000,000 random iterations of the genotype data (see Methods). [e]Haplotype consisting of markers rs729302, rs2004640, rs752637, rs2280714. [f]Haplotypes carrying the indicated allele of rs2004640 and rs2280714.

Given these convincing data for association of *IRF5* with SLE risk, we next investigated the potential functional consequences of the rs2004640 T allele. Examination of the genomic sequence of *IRF5* showed that the rs2004640 T allele is located 2 bp downstream of the intron-exon border of exon 1B, creating a consensus GT donor splice site (**Fig. 1a**). This observation led us to ask if rs2004640 influenced splicing of *IRF5*. The splicing of *IRF5* is known to be highly complex[15]. *IRF5* transcripts are initiated at one of three promoters, giving rise to transcripts containing exon 1A, exon 1B or exon 1C (**Fig. 1a**). Transcripts initiated at exon 1A and exon 1B are constitutively expressed in plasmacytoid dendritic cells and B cells, whereas transcripts bearing exon 1C are inducible by type I IFNs[15]. In addition, multiple *IRF5* isoforms are initiated at each promoter, with nine previously identified isoforms (V1–V9, **Fig. 1a**)[15].

We first examined whether rs2004640 affected expression of *IRF5* transcripts bearing exon 1B. We isolated peripheral blood mononuclear cells from individuals carrying GG, GT or TT rs2004640 genotypes and synthesized first-strand cDNA using RNA from these cells. Using specific primers to detect transcripts associated with each of the three exon 1 variants, we found that individuals with SLE and controls homozygous for the G allele expressed *IRF5* isoforms containing exon 1A and exon 1C, but not exon 1B (**Fig. 1b**). In contrast, individuals homozygous or heterozygous for the T allele expressed transcripts containing exon 1B as well as transcripts containing exons 1A and 1C. Using TaqMan quantitative PCR assays, we detected exon 1B transcripts only in individuals carrying the rs2004640 T allele (**Fig. 1c**). In all samples studied to date (66 total individuals; rs2004640 genotypes: GG, $n = 17$; GT, $n = 28$; TT, $n = 21$), transcripts containing exon 1A were more abundant than the other mRNA classes. From this data, we conclude that only individuals with the rs2004640 T allele express the multiple isoforms of *IRF5* initiated at exon 1B.

We successfully cloned two previously unidentified isoforms of *IRF5* from the peripheral blood mRNA of rs2004640 heterozygote donors:

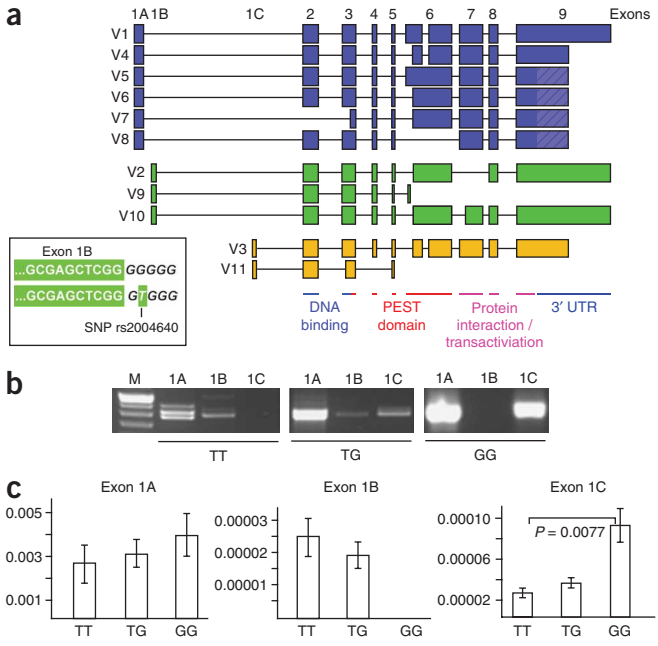

Figure 1 The rs2004640 T allele is associated with expression of variant exon 1B *IRF5* mRNA transcripts. (**a**) mRNA isoforms of *IRF5*. Three sets of isoforms derive from three alternative promoters in the *IRF5* 5′ region. The location of the exons encoding the DNA-binding, PEST and protein interaction domains, as well as the 3′ untranslated region, are annotated. Protein translation begins 10 bp from the 5′ end of exon 2 at a consensus ATG. Shown in the box is the location of the rs2004640 SNP, 2 bp downstream of exon 1B. Two polyadenylation sites are present in the *IRF5* 3′ UTR, and the lengths of the 3′ UTRs for V5, V6, V7 and V8 are unknown. Exon/intron structure is not shown to scale. (**b**) RT-PCR analysis of total RNA purified from human PBMCs. Representative results for individuals with defined rs2004640 genotypes are shown. Exon 1B transcripts are found only in individuals carrying the rs2004640 T allele. The identities of most of the bands observed in these reactions correlate with known isoforms of the gene, confirmed by sequencing. (**c**) Summary of TaqMan real-time quantitative RT-PCR analysis for transcripts associated with exons 1A, 1B and 1C. Each bar represents mean ± s.e.m. expression level. $n = 8$ SLE cases for each genotype. Similar data were obtained for normal controls (data not shown). ΔCt values were calculated from duplicate samples normalized to human β2-microglobulin and were converted to linear scale.

V10, which includes exon 1B and has an in-frame deletion of 30 nt at the beginning of exon 7 and a predicted protein ten residues shorter than V2, and V11, a transcript derived from exon 1C, with a 28-nt deletion of exon 3, predicted to encode a truncated protein translated from an alternate reading frame (**Fig. 1a**). Notably, several of these isoforms, including isoforms initiated at exon 1B, contain splicing variation in and around exon 6, which encodes part of an extended PEST domain, a motif highly enriched for proline, glutamic acid, serine and threonine, that is associated with control of protein stability. Several unique and constitutively expressed *IRF5* isoforms are initiated at exon 1B, and these isoforms may influence the function of IRF5 or the transcriptional profile of IRF5 target genes.

While this work was ongoing, others identified a common variant near *IRF5* as one of the polymorphisms most strongly associated with gene expression variation in a genome-wide scan[16,17]. Specifically, a variant ~5 kb downstream of *IRF5* (rs2280714 T allele) was identified as a *cis*-acting determinant of *IRF5* expression (or in strong linkage disequilibrium (LD) with a *cis*-acting determinant) by testing the 1 million HapMap phase I variants for association to gene expression in Epstein-Barr virus (EBV)-transformed B cells[17].

The observation that a common variant is strongly associated with *IRF5* expression led us to ask whether elevated expression of *IRF5* might be associated with the exon 1B splice site (**Fig. 2**). First, we determined the relationship between rs2004640 and rs2280714 in 30 independent Centre d'Etude du Polymorphisme Human (CEPH) trios from the HapMap project. *D′* for the two SNPs is 0.96; that is, nearly all copies of the splice site rs2004640 T allele are on haplotypes bearing the rs2280714 T allele. However, r^2 for these SNPs is 0.68, as the downstream rs2280714 T allele is also found on haplotypes that lack the splice site rs2004640 T allele (see also **Table 2** and **Fig. 2e**). Although these two SNPs are strongly linked, the fact that the

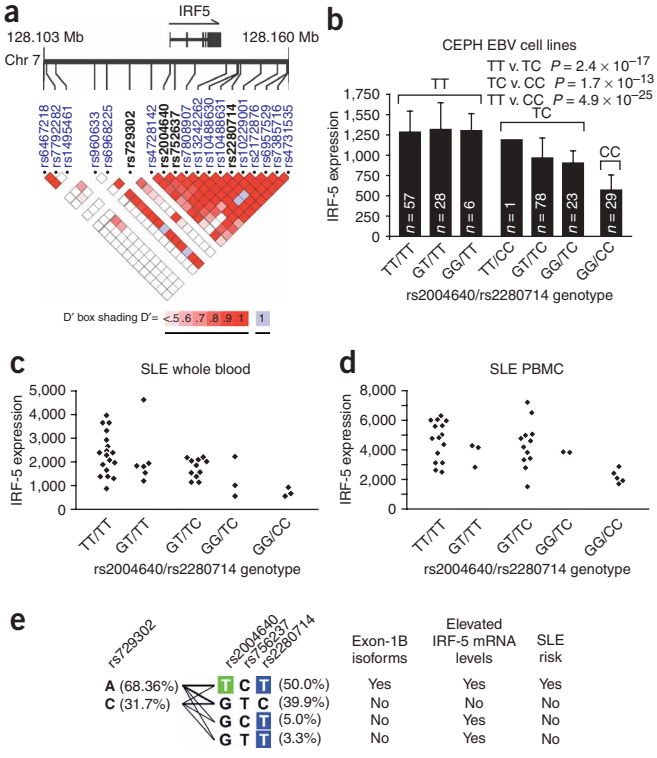

Figure 2 The SLE risk haplotype is associated with elevated expression of *IRF5*. (**a**) rs2004640 was typed in 90 HapMap CEPH (CEU) DNA samples, and the haplotype structure of the region was analyzed using Haploview and Locusview. (**b**) Levels of *IRF5* mRNA, as determined by microarray, were compared between EBV-transformed cell lines from CEPH individuals typed for rs2004640 and rs2280714. Identical findings were observed when only CEPH founders were examined. (**c**,**d**) Levels of *IRF5* mRNA transcripts, measured by Affymetrix microarrays, were compared in the whole blood ($n = 41$) and PBMCs ($n = 37$) of two sets of independent SLE cases. *IRF5* mRNA transcript levels were compared by rs2280714 genotype in whole-blood SLE samples: TT versus TC, $P = 0.01$; TC versus CC, $P = 0.0006$; TT versus CC, $P = 0.000002$. A similar analysis was performed for the SLE PBMC samples: TT versus TC, $P = 0.66$; TC versus CC, $P = 0.0002$; TT versus CC, $P = 0.000004$. (**e**) The SLE risk haplotype contains both the rs2004640 T allele (green) and the rs2280714 T allele (blue). Shown are haplotype frequencies of CEPH founders, as determined by Haploview. rs729302 is located 5′ of the haplotype marked by rs2004640, rs752637 and rs2280714. Haplotypes that contain the rs2280714 T allele together with the rs2004640 G allele do not confer risk for SLE (see **Table 2**).

3′ rs2280714 T allele can be observed in the absence of the upstream splice site SNP allowed us to determine which variant is the best predictor of *IRF5* expression and also SLE risk.

We examined the association between IRF5 expression and the two SNPs using expression data from EBV-transformed B cells of CEPH family members and from peripheral blood cells of two independent sets of SLE cases. We first genotyped rs2004640 and rs2280714 in 233 CEPH individuals previously used for a genome-wide survey of determinants of gene expression[16] and examined these variants for association with *IRF5* expression. We found that the T alleles of both rs2004640 and rs2280714 were associated with higher levels of *IRF5* mRNA expression (**Fig. 2b**). However, the rs2280714 T allele was a better predictor of *IRF5* overexpression in 92 unrelated individuals than was the rs2004640 T allele ($P = 2 \times 10^{-16}$ versus $P = 5.3 \times 10^{-11}$, respectively) and in the full data set of 233 individuals, consisting of 14 extended pedigrees and 38 unrelated individuals (**Fig. 2b**). We observed similar findings in the peripheral blood cells of two independent groups of SLE cases (**Fig. 2c,d**). On the basis of these data, we can reject the hypothesis that the splice site rs2004640 SNP is the *cis*-acting variant controlling expression, as rs2280714 remains significantly associated with *IRF5* expression ($P = 4.7 \times 10^{-7}$) after logistic regression conditional on rs2004640, whereas rs2004640 no longer remains significant after controlling for rs2280714.

Using the recently released phase II HapMap genotype data (~5 million SNPs across the genome), we next tested all available variants (including rs2004640 and rs2280714) within 100 kb of *IRF5* for association with *IRF5* expression in EBV-transformed B cells from 42 unrelated individuals from the HapMap CEPH (CEU) population[16,18]. We found that rs2280714 and four polymorphisms that are perfect proxies of rs2280714 ($r^2 = 1.0$) are the most strongly associated with *IRF5* gene expression ($P = 1.0 \times 10^{-10}$, **Supplementary Table 2** online). Given that these variants are well downstream of *IRF5*, and that they do not lie in a recognizable regulatory region, we hypothesize that there may be additional genetic variation in tight LD with rs2280714 that drives the expression phenotype.

We next asked whether overexpression of *IRF5* (rs2280714), the presence of exon 1B–initiated *IRF5* isoforms (rs2004640) or both are associated with SLE. The fact that ~14% of *IRF5* haplotypes are associated with overexpression but lack the exon 1B splice site allows the opportunity to test whether the allele associated with overexpression (rs2280714) is independently associated with SLE (**Table 2**). Indeed, in 470 SLE pedigrees, only haplotypes bearing the exon 1B splice site (rs2004640 T allele) show overtransmission using the transmission disequilibrium test[19] (208:149 T:U, $P = 0.0021$; **Table 2**). Haplotypes associated with overexpression of *IRF5* (rs2280714 T allele) but lacking the exon 1B splice site show no evidence of association with increased risk of SLE (rs2004640/rs2280714 'G/T' haplotype; 70:108 T:U; **Table 2**). Supporting the family-based analysis, we found no difference in the frequency of the rs2004640/rs2280714 'G/T' haplotype between SLE cases ($n = 1{,}358$, 13%) and controls ($n = 2{,}278$, 15%; $P = 0.98$; **Supplementary Table 3** online). Additionally, rs2280714 was not significantly associated with SLE in the case-control analysis after logistic regression conditional on rs2004640 ($P = 0.22$). We therefore conclude that overexpression of *IRF5* in the absence of the exon 1B splice site does not confer risk of SLE.

How might the exon 1B *IRF5* isoforms cause increased risk of SLE? Type I IFN and TLR signaling are known to induce the phosphorylation of IRF5, leading to the transactivation of downstream genes[9], including the type I IFNs themselves[11,12,20,21]. Multiple *IRF5* isoforms are initiated at exon 1B, and they differ in both the coding and noncoding portions of the transcript. Thus, individuals with exon 1B transcripts may express IRF5 protein isoforms with different stability, affinity for transcriptional cofactors and/or activation properties. Transfection studies have shown that the IRF5 isoforms differ in their ability to transactivate IFN genes[15]. For instance, the exon 1B transcript V2 stimulates production of IFN-β but not IFN-α1 or IFN-α2 (ref. 15), suggesting that differential expression of IRF5 isoforms may influence the pattern of IFN gene induction directly. Notably, the reference *IRF5* genomic sequence for chimpanzee contains the exon 1B splice site (rs2004640 T allele), suggesting that the T allele is ancestral and that the function of exon 1B isoforms is evolutionarily conserved.

In summary, these data confirm the association of *IRF5* with SLE and identify an *IRF5* risk haplotype that contains both a splice donor site that allows expression of multiple *IRF5* isoforms containing exon 1B and a separate genetic effect associated with elevated *IRF5* expression. The *IRF5* exon 1B isoforms are strongly linked to elevated expression of *IRF5* and to risk of SLE; however, overexpression of *IRF5* in the absence of exon 1B isoforms does not confer risk. These data provide new mechanistic insights into SLE pathogenesis and should serve to focus future research on the genetic basis of SLE.

METHODS

Clinical samples. A collection of American SLE families of European descent consisting of 187 sib-pair and 223 trio pedigrees was recruited at the University of Minnesota. An additional 63 trios from the National Institute of Arthritis and Musculoskeletal and Skin Disease (NIAMS)-sponsored Lupus Multiplex Registry at Oklahoma Medical Research Foundation (OMRF) were included in the analysis. The overall US family cohort comprised 681 individuals with SLE and 824 other family members. We genotyped 459 probands from the US family collection, 266 cases from the Hopkins Lupus Cohort, 41 controls from Minnesota and 1,393 controls of European ancestry from the New York Health Project[22] collection for the case-control analysis.

Three additional SLE case-control cohorts were studied. A cohort of 444 individuals with SLE and 541 control individuals was collected in several clinics in the Andalucia region of southern Spain. All individuals are of Spanish European ancestry. A second cohort of 284 SLE cases and 279 matched controls was collected through a multicenter collaboration in Argentina. Individuals are of European (72.5%) and mixed (20%) ancestry. Six percent were of Amerindian ($n = 1$), Asian ($n = 2$) or unknown ancestry ($n = 22$). A third set of 208 ethnic European Swedish SLE cases and 254 controls from the Stockholm-Uppsala area were studied (no overlap with the previously published cases[5]). All cases fulfilled the revised American College of Rheumatology criteria for SLE[23]. These studies were approved by the Human Subject Institutional Review Boards at each institution or the ethics committees from each country. Informed consent was obtained from all subjects.

Genotyping. Four polymorphisms from *IRF5* (rs729302, rs2004640, rs752637 and rs2280714) were genotyped in the 470 families by primer extension of multiplex products with detection by matrix-assisted laser desorption ionization–time of flight mass spectroscopy using a Sequenom platform. The average genotype completeness for the four assays was 98.3%. The genotyping consensus error rate was 0.7% (nine errors in mendelian inheritance from 1,288 parent-offspring transmissions; all errors were zeroed out, and no single marker had more than three errors).

For the US case-control studies, rs2004640 was typed by TaqMan for the Johns Hopkins cases and for the US controls, and by Sequenom for all other samples. Genotyping of rs2004640 was performed separately for the Spanish, Swedish and Argentine cases and controls. Briefly, these three sets were genotyped at the Rudbeck Laboratory at Uppsala University using the TaqMan Assay-on-Demand (ABI) for rs2004640. The average genotype completeness was 99% for the Swedish samples, 98% for the Argentine samples and 86% for Spanish samples.

rs2280714 was typed for the case-control studies on both the Sequenom platform and using a TaqMan assay (Rudbeck Laboratory). Over 1,100

individuals were typed on both platforms with 98.2% concordance of results. The following samples were not typed for rs2280714: 63 OMRF trios, 96 Spanish SLE cases, 126 Swedish cases and 161 Swedish controls. Hardy-Weinberg equilibrium (HWE) *P* values for rs2004640 and rs2280714 for each population are presented in **Supplementary Table 4** online. Primer and probe sequences are provided in **Supplementary Table 5** online.

Family-based association analysis. The transmission disequilibrium test (TDT) was performed using Haploview v3.2 under default settings. Haploview v3.2 examines the transmission patterns of all complete trios within each pedigree. To assess the statistical significance of the results, the transmitted or untransmitted status of each genotype and haplotype was randomly permutated for 1,000,000 iterations and the best χ^2 value generated for each permutated dataset was recorded. The number of times the permutated χ^2 value exceeded the nominal χ^2 value was divided by the number of iterations (here, 1,000,000) to generate the permutated *P* value. The pedigree disequilibrium test (PDT) was performed as described[24].

Case-control analysis. We used χ^2 analysis to evaluate the significance of differences in genotype and allele frequencies in the case-control samples. The allele frequencies for cases and controls were used to calculate the odds ratio (OR) and the 95% confidence interval using Woolf's method ($\ln(OR) - 1.96(1/A + 1/B + 1/C + 1/D)^{1/2}$). The χ^2 value was calculated from the 2×2 contingency tables, and *P* values were determined using one degree of freedom. For the case-control haplotype analysis, Haploview v3.2 was used to generate haplotype frequencies and calculate the significance of associations. In the case-control haplotype analysis, only the 1,358 cases and 2,278 controls for which complete genotype data at both rs2004640 and rs2280714 were available were used in the analysis.

Meta-analysis. Results for rs2004640 in SLE cases collected in Argentina, Spain, Sweden and the US were combined using the Mantel-Haenszel meta-analysis of the ORs[13,14], and these data were subsequently combined in a separate analysis with the published results of the association of rs2004640 with SLE in Finnish and Swedish collections[5].

Determination and quantification of IRF5 UTR-specific transcripts. Total RNA from individuals with SLE carrying the various genotypes was purified from peripheral blood mononuclear cells (PBMCs) with TRIZOL Reagent (Invitrogen). We reverse-transcribed 2 μg of total RNA with 2 units (U) of MultiScribe transcriptase in PCR buffer II containing 5 mM $MgCl_2$, 1 mM dNTPs, 0.4 U of RNase inhibitor and 2.5 μM random hexamers (all results were confirmed using oligo-dT–primed cDNA). All reagents were from Applied Biosystems. Synthesis was performed at 42 °C for 45 min, and the reaction was terminated at 95 °C for 5 min.

IRF5 isoforms with distinct 5′-UTRs were quantified by real-time PCR with TaqMan on an ABI PRISM 7700 Sequence Detector (Applied Biosystems) with SDS 1.9.1 software. Primers used to distinguish PCR products with different UTRs are shown in **Supplementary Table 5**. We used a common reverse primer lying in exon 3 as well as a common TaqMan probe. We performed 45 or 52 cycles of two-step PCR (95 °C for 15 s and 65 °C for 1 min) in a buffer containing 1.5 mM $MgCl_2$, 200 μM of each of dNTP, 0.5 U of Platinum Taq polymerase (Invitrogen), primer-probe mix and cDNA. Expression levels were normalized using human β2-microglobulin with commercial primer-probe mix (Applied Biosystems).

Standard PCR amplification of diverse isoforms of *IRF5* was performed with the same forward primers as for the TaqMan assay with a reverse primer designed to allow amplification of all transcripts containing exon 8 (primers shown in **Supplementary Table 5**). Cycle conditions were 95 °C for 3 min, followed by 40 cycles of 95 °C for 15 s, 60 °C for 15 s and 72 °C for 1.5 min. PCR was performed in 25 μl reaction volume with 0.5 U of Platinum Taq polymerase (Invitrogen) in the buffer supplied by the manufacturer. PCR products were electrophoresed on a 1.5% agarose gel.

The statistical analysis of isoform expression was performed using a two-tailed *t*-test included in GraphPad Software.

Cloning and sequencing of IRF5 isoforms. To isolate new isoforms, we conducted RT-PCR, using as a template total RNA isolated from human PBMCs of two individuals with SLE carrying the TG allele of rs2004640. We used the same forward primers matching exon 1 used for the TaqMan RT-PCR assays and a common reverse primer lying in the last exon (**Supplementary Table 5**). PCR products were analyzed by gel electrophoresis and individual bands were excised and purified. Sequencing was performed using the BigDye reaction at the Uppsala Genome Center. Two new transcripts, which we called V10 and V11, were identified and deposited in GenBank. All primer and probe sequences for isoform expression analysis and sequencing are provided in **Supplementary Table 5**.

IRF5 expression analysis. Two SNPs in the *IRF5* region (rs2004640 and rs2280714) were genotyped using the Sequenom platform described above in 30 CEPH trios (CEU, 90 individuals) from the International HapMap Project[18], and the data was integrated into the Phase II data (HapMap data release #19) for 100 kb flanking *IRF5*. In addition, three SNPs (rs726302, rs2004640 and rs2280714) were genotyped in the 233 CEPH individuals (14 extended pedigrees, including 21 trios that are part of the HapMap CEU samples, and 38 unrelated individuals) described in ref. 16, using Sequenom. Linear regression (performed with the R statistical package) was used to test the significance of association of genetic variants to *IRF5* expression levels using publicly available gene expression data (Gene Expression Omnibus (GEO) accession number GSE1485, *IRF5* probe 205469_s_at)[16] in the 233 CEPH individuals, subdivided by (i) 42 unrelated founders included in the HapMap CEPH (CEU) population, (ii) 92 unrelated individuals and (iii) all 233 individuals. Gene expression data were also obtained from the PBMCs of 37 individuals with SLE (Affymetrix U95A chips, *IRF5* probe set 36465_at)[4] and from PaxGene RNA from whole blood of 41 individuals with SLE of European descent (Affymetrix U133A chips, *IRF5* probe set 205469_s_at).

Accession codes. Sequences for isoforms V10 and V11 have been deposited in GenBank under accession numbers DQ277633 and DQ277634, respectively.

URLs. Genetic model testing: http://pngu.mgh.harvard.edu/~purcell/model/model.html. Haploview v3.2: http://www.broad.mit.edu/mpg/haploview/. GraphPad Software: http://www.graphpad.com.

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

ACKNOWLEDGMENTS

We thank H. Yin and the members of the Uppsala Genome Center for help with sequencing and S. Young, J. Novitzke, L. Kyle, K. Espe, C. Gillett, P. Gaffney and K. Moser for assistance. The Swedish samples were obtained from the Lupus cohort at Karolinska University Hospital, Solna. This work was supported by grants from the Swedish Research Council (M.E.A.-R.), the Clas Groschinski Foundation (M.E.A.-R.), the Marcus Borsgtröms Foundation (M.E.A.-R.), the Torsten and Ragnar Söderberg's Foundation (M.E.A.-R.), the Swedish Association against Rheumatism (M.E.A.-R.), the Alliance for Lupus Research (M.E.A.-R., T.W.B.), the Lupus Research Institute (T.W.B.), the Mary Kirkland Center for Lupus Research (T.W.B.), the Lupus Foundation of Minnesota (T.W.B.), the National Institute of Allergy and Infectious Diseases of the US National Institutes of Health (NIH) (T.W.B.), the National Institute of Arthritis, Musculoskeletal and Skin Diseases of the NIH (R.R.G., E.C.B., T.W.B.) and the Ministerio de Ciencia y Tecnología (SAF2003-3460; J.M.). The Johns Hopkins Lupus Cohort is supported by NIH grant AR 43727, and the Johns Hopkins General Clinical Research Center is supported by NIH M01-RR-00052.

AUTHORS' CONTRIBUTIONS

T.W.B. and M.E.A.-R. codirected this project. B.P.-E. is the coordinator of the Argentine Collaborative Group.

The Argentine and Spanish Collaborative Groups comprise the following individuals. In Argentina, coordinators at each center were Hugo R. Scherbarth, Servicio de Reumatología, Hospital Interzonal General de Agudos 'Dr. Oscar Alende', Mar del Plata; Susana Gamron, Servicio de Reumatología de la UHMI 1, Hospital Nacional de Clínicas, Universidad Nacional de Córdoba, Córdoba; Alberto Allievi, Hospital General de Agudos Dr. Juán A. Fernandez, Buenos Aires; Simon A. Palatnik, Facultad de Ciencias Medicas, Universidad Nacional de Rosario y Hospital Provincial del Centenario, Rosario; Alejandro Alvarellos, Servicio de Reumatología, Hospital Privado, Centro Medico de Córdoba, Córdoba; Sergio Paira, Sección de Reumatología, Hospital José M. Cullen, Santa Fe; Cesar Caprarulo, Hospital Felipe Heras, Concordia, Entre Ríos; Carolina Guillerón, Departamento de Inmunología, Instituto de Investigaciones Médicas 'Alfredo Lanari', Buenos Aires; Luis J. Catoggio, Sección Reumatología, Servicio de

Clínica Médica, Hospital Italiano de Buenos Aires y Fundación Dr. Pedro M. Catoggio para el Progreso de la Reumatología, Buenos Aires; Cristina Prigione, Servicio de Reumatología, Hospital Provincial de Rosario, Rosario; Guillermo A. Berbotto, Servicio de Reumatología, Hospital Escuela Eva Perón, Granadero Baigorria, Rosario; Mercedes A. García, Servicio de Reumatología, Hospital Interzonal General de Agudos General San Martín, La Plata; Carlos E. Perandones, Centro de Educación Médica e Investigaciones Clínicas, Buenos Aires; and Cristina G. Battagliotti, Hospital de Niños Orlando Alassia, Santa Fe.

In Argentina, participants were Pilar C. Marino, Estela L. Motta Servicio de Reumatología, Hospital Interzonal General de Agudos 'Dr. Oscar Alende', Mar del Plata; Cristina Drenkard, Emilia Menso Servicio de Reumatología de la UHMI 1, Hospital Nacional de Clínicas, Universidad Nacional de Córdoba, Córdoba; Guillermo A. Tate, Organización Médica de Investigación, Buenos Aires; Jose L. Presas, Hospital General de Agudos Dr. Juán A. Fernandez, Buenos Aires; Marcelo Abdala and Mariela Bearzotti, Facultad de Ciencias Medicas, Universidad Nacional de Rosario y Hospital Provincial del Centenario, Rosario; Francisco Caeiro, Ana Bertoli Servicio de Reumatología, Hospital Privado, Centro Medico de Córdoba, Córdoba; Susana Roverano, Hospital José M. Cullen, Santa Fe; Cesar E. Graf and Estela Bertero, Hospital San Martín, Paraná, Concordia, Entre Ríos; Griselda Buchanan, Hospital Felipe Heras, Concordia, Entre Ríos; Sebastian Grimaudo, Jorge Manni Departamento de Inmunología, Instituto de Investigaciones Médicas 'Alfredo Lanari', Buenos Aires; Enrique R. Soriano, Carlos D. Santos Sección Reumatología, Servicio de Clínica Medica, Hospital Italiano de Buenos Aires y Fundación Dr. Pedro M. Catoggio para el Progreso de la Reumatología, Buenos Aires; Fernando A. Ramos, Sandra M. Navarro Servicio de Reumatología, Hospital Provincial de Rosario, Rosario; Marisa Jorfen and Elisa J. Romero, Servicio de Reumatología Hospital Escuela Eva Perón, Granadero Baigorria, Rosario; Juan C. Marcos and Ana I. Marcos, Servicio de Reumatología, Hospital Interzonal General de Agudos General San Martín, La Plata; and Alicia Eimon, Centro de Educación Médica e Investigaciones Clínicas, Buenos Aires.

In Spain, participants were Norberto Ortego, Servicio Medicina Interna, Hospital Clínico San Cecilio, Granada; Enrique de Ramón, Servicio de Medicina Interna, Hospital Carlos Haya, Malaga; Juan Jiménez-Alonso, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Julio Sánchez-Román, Servico de Medicina Interna, Hospital Virgen del Rocio, Sevilla; and Miguel Angel López-Nevot, Servicio de Inmunología, Hospital Virgen de las Nieves, Granada.

COMPETING INTERESTS STATEMENT

The authors declare competing financial interests (see the *Nature Genetics* website for details).

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

1. Alarcon-Segovia, D. *et al.* 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 (2005).
2. Blanco, P., Palucka, A.K., Gill, M., Pascual, V. & Banchereau, J. Induction of dendritic cell differentiation by IFN-alpha in systemic lupus erythematosus. *Science* **294**, 1540–1543 (2001).
3. Ronnblom, L. & Alm, G.V. A pivotal role for the natural interferon alpha-producing cells (plasmacytoid dendritic cells) in the pathogenesis of lupus. *J. Exp. Med.* **194**, F59–F63 (2001).
4. Baechler, E.C. *et al.* Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc. Natl. Acad. Sci. USA* **100**, 2610–2615 (2003).
5. Sigurdsson, S. *et al.* Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Genet.* **76**, 528–537 (2005).
6. Isaacs, A. & Lindenmann, J. Virus interference. 1 The interferon. *Proc. R. Soc. Lond. B* **147**, 258–273 (1957).
7. Bennett, L. *et al.* Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J. Exp. Med.* **197**, 711–723 (2003).
8. Kirou, K.A. *et al.* Coordinate over-expression of interferon-alpha-induced genes in systemic lupus erythematosus. *Arthritis Rheum.* **50**, 3958–3967 (2004).
9. Honda, K., Yanai, H., Takaoka, A. & Taniguchi, T. Regulation of the type I IFN induction: a current view. *Int. Immunol.* **17**, 1367–1378 (2005).
10. Takaoka, A. *et al.* Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature* **434**, 243–249 (2005).
11. Barnes, B.J., Moore, P.A. & Pitha, P.M. Virus-specific activation of a novel interferon regulatory factor, IRF-5, results in the induction of distinct interferon alpha genes. *J. Biol. Chem.* **276**, 23382–23390 (2001).
12. Barnes, B.J. *et al.* Global and distinct targets of IRF-5 and IRF-7 during innate response to viral infection. *J. Biol. Chem.* **279**, 45194–45207 (2004).
13. Lohmueller, K.E., Pearce, C.L., Pike, M., Lander, E.S. & Hirschhorn, J.N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat. Genet.* **33**, 177–182 (2003).
14. Woolson, R.F. & Bean, J.A. Mantel-Haenszel statistics and direct standardization. *Stat. Med.* **1**, 37–39 (1982).
15. Mancl, M.E. *et al.* Two discrete promoters regulate the alternatively spliced human interferon regulatory factor-5 isoforms. Multiple isoforms with distinct cell type-specific expression, localization, regulation, and function. *J. Biol. Chem.* **280**, 21078–21090 (2005).
16. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
17. Cheung, V.G. *et al.* Mapping determinants of human gene expression by regional and genome-wide association. *Nature* **437**, 1365–1369 (2005).
18. Altshuler, D. *et al.* A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
19. Spielman, R.S., McGinnis, R.E. & Evans, W.J. Transmission test for linkage disequilibrium: the insulin gene region and insulin dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* **52**, 506–516 (1993).
20. Barnes, B.J., Field, A.E. & Pitha-Rowe, P.M. Virus-induced heterodimer formation between IRF-5 and IRF-7 modulates assembly of the IFNA enhanceosome *in vivo* and transcriptional activity of IFNA genes. *J. Biol. Chem.* **278**, 16630–16641 (2003).
21. Barnes, B.J., Kellum, M.J., Field, A.E. & Pitha, P.M. Multiple regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes. *Mol. Cell. Biol.* **22**, 5721–5740 (2002).
22. Mitchell, M.K., Gregersen, P.K., Johnson, S., Parsons, R. & Vlahov, D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. *J. Urban Health* **81**, 301–310 (2004).
23. Hochberg, M.C. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.* **40**, 1725 (1997).
24. Martin, E.R., Monks, S.A., Warren, L.L. & Kaplan, N.L. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* **67**, 146–154 (2000).