Contribution of IKBKE and IFIH1 gene variants to SLE susceptibility


Linköping University Post Print

N.B.: When citing this work, cite the original article.

Original Publication:


The type I interferon system genes *IKBKE* and *IFIH1* are associated with the risk of systemic lupus erythematosus (SLE). To identify the sequence variants that are relevant for the disease association, we resequenced the genes *IKBKE* and *IFIH1*. Eighty-six single-nucleotide variants (SNVs) with potentially functional effect or differences in allele frequencies between patients and controls determined by sequencing were further genotyped in 1140 SLE patients and 2060 controls. In addition, 108 imputed sequence variants in *IKBKE* and *IFIH1* were included in the association analysis. Ten *IKBKE* SNVs and three *IFIH1* SNVs were associated with SLE. The SNVs rs1539241 and rs12142086 tagged two independent association signals in *IKBKE*, and the haplotype carrying their risk alleles showed an odds ratio of 1.68 (P-value = 1.0 × 10^{-7}). The risk allele of rs12142086 affects the binding of splicing factor 1 in vitro and could thus influence its transcriptional regulatory function. Two independent association signals were also detected in *IFIH1*, which were tagged by a low-frequency SNV rs78456138 and a missense SNV rs3747517. Their joint effect is protective against SLE (odds ratio = 0.56; P-value = 6.6 × 10^{-7}). In conclusion, we have identified new SLE-associated sequence variants in *IKBKE* and *IFIH1*, and proposed functional hypotheses for the association signals.

Genes and Immunity (2013) 14, 217–222; doi:10.1038/gene.2013.9; published online 28 March 2013

**Keywords:** *IKBKE; IFIH1; resequencing; association study; systemic lupus erythematosus*

---

**INTRODUCTION**

Systemic lupus erythematosus (SLE) is a clinically heterogeneous disease, which affects multiple organ systems and causes significant morbidity and mortality. About 8–15% of the genetic component of SLE is accounted for by DNA sequence variants identified by association studies in European populations. These findings have underscored the crucial role of the innate immune system, and particularly of the type I interferon (IFN) system in SLE. Genes involved in the type I IFN system such as STAT4, IRF5, TNFAIP3, TNFSF4 and TNIP1 have been replicated as SLE susceptibility genes in multiple ethnic groups. Previously, we have provided evidence of association for two additional type I IFN system genes, namely inhibitor of kappa light polypeptide enhancer in B cells kinase epsilon (*IKBKE*) and IFN induced with helicase C domain 1 (*IFIH1*). *IKBKE* encodes IKKε, which belongs to the IkB kinase family. Its expression is restricted to lymphoid tissues and peripheral blood leukocytes. IKKε is activated downstream of Toll-like receptors or cytosolic RNA/DNA sensors, and mediates phosphorylation and activation of the transcription factors, for example, IFN regulatory factors 3 and 7, or activation of the nuclear factor of kappa B (NF-κB) transcription factor system. Both signalling cascades can trigger the expression of type I IFN and other pro-inflammatory cytokines. Recently, *IKBKE* was also found to be associated with rheumatoid arthritis and antibody-positive primary Sjögren’s syndrome.

IFIH1, also known as melanoma differentiation-associated protein 5, is a cytosolic RNA sensor belonging to the retinoic acid-inducible gene I-like receptor family. IFIH1 is expressed in most cell types at low levels, and its expression can be largely enhanced by exposure to IFNs or viruses. When recognizing viral nucleic acids, IFIH1 triggers activation of IFN regulatory factor 3 and NF-κB via mitochondrial antiviral signalling protein, leading to antiviral and pro-apoptotic responses. Genetic association studies have demonstrated an important role for *IFIH1* in autoimmune diseases such as type I diabetes, rheumatoid arthritis, multiple sclerosis, Graves’ disease and psoriasis.

With the aim of identifying functional sequence variants of relevance for SLE in *IKBKE* and *IFIH1*, we resequenced the genes including exonic and intronic regions in pools of SLE patients and controls from Sweden. The identified sequence variants together with imputed variants were subsequently tested for association with SLE in a large Swedish case–control cohort. With this strategy, our study explored the variation spectrum of *IKBKE* and *IFIH1*, and elucidated their contribution to the association to SLE.
RESULTS
Resequencing and variant identification
The IKBKE and IFIH1 genes, including exonic, intronic and untranslated regions were amplified by long-range PCRs in a pool of 100 SLE patients or of 100 controls, and then sequenced for comprehensive identification of potentially functional sequence variants for SLE. In total, 91 high-quality single-nucleotide variants (SNVs) in IKBKE and 138 high-quality SNVs in IFIH1 were identified. About 70% of these SNVs were already archived in the Short Genetic Variations database dbSNP v137, whereas 70 novel sequence variants were discovered. Of the identified SNVs, 10 were located in exonic regions, 4 in untranslated regions, one at a splice site and the remaining ones were intronic variants.

SLE association analysis
To limit the number of SNVs for genotyping but still maintain our discovery potential, we selected SNVs found by resequencing based on their potentially functional effect and allele frequencies in patients and controls. Three SNVs (rs1539243 and rs17433930 in IKBKE, and rs1990760 in IFIH1) associated with SLE and one SNV (rs35732034 in IKBKE) within each gene (Supplementary Table 3). For IKBKE, the two SNVs at the S’-gene region, rs11117858 and rs1539241, had an r² value of 0.60 and a D’ value of 1.00. The SNV in exon 8 rs2297545, and the four SNVs spanning intron 13 and intron 15, rs11118087, rs12409804, rs12142086 and rs11118132 also showed a high correlation (r² = 0.45–1.00; D’ = 0.93–1.00). For IFIH1, strong LD was observed between the SNVs rs3747517 and rs1990760 (r² = 0.65; D’ = 1.00), but not for the SNV rs78456138 (r² = 0.04–0.06; D’ = 1.00).

Independence test and analysis for the joint effect of risk alleles
To dissect the relationship between the associated SNVs and seek for independent association signals, we calculated pairwise linkage disequilibrium (LD) values between the SLE-associated SNVs for both genes (Supplementary Table 2). For IKBKE, the two SNVs at the S’-gene region, rs11117858 and rs1539241, had an r² value of 0.60 and a D’ value of 1.00. The SNV in exon 8 rs2297545, and the four SNVs spanning intron 13 and intron 15, rs11118087, rs12409804, rs12142086 and rs11118132 also showed a high correlation (r² = 0.45–1.00; D’ = 0.93–1.00). For IFIH1, strong LD was observed between the SNVs rs3747517 and rs1990760 (r² = 0.65; D’ = 1.00), but not for the SNV rs78456138 (r² = 0.04–0.06; D’ = 1.00).

The results from pairwise conditional logistic regression analysis were in accordance with the correlation patterns among SNVs within each gene (Supplementary Table 3). For IKBKE, when conditioning on the most significantly associated SNV rs12142086, the association signals from the other SNVs disappeared, except for the two SNVs at the gene S’-region; whereas at least marginal significant association P-values remained for rs12142086 when conditioning on the other SNVs, except for the SNV rs12409804, which is in high LD with rs12142086. In addition, when conditioning on the two SNVs at the gene 5’-region, rs11117858 and rs1539241, an r² value of 0.60 and a D’ value of 1.00.

Table 1. SLE-associated SNVs in IKBKE and IFIH1

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNV</th>
<th>Position (GRCh37/hg19)</th>
<th>Annotation</th>
<th>Minor/major allele</th>
<th>MAF in patients</th>
<th>MAF in controls</th>
<th>Unadjusted P-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKBKE</td>
<td>rs11117858</td>
<td>chr1:2066464274</td>
<td>Intron 2</td>
<td>G/T</td>
<td>0.25</td>
<td>0.22</td>
<td>0.023</td>
<td>1.15 (1.02–1.30)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs1539241</td>
<td>chr1:206647450</td>
<td>Intron 3</td>
<td>A/G</td>
<td>0.17</td>
<td>0.15</td>
<td>0.0094</td>
<td>1.21 (1.05–1.34)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs2297545</td>
<td>chr1:206651107</td>
<td>Exon 8</td>
<td>A/G</td>
<td>0.22</td>
<td>0.19</td>
<td>5.5 × 10⁻⁴</td>
<td>1.26 (1.10–1.43)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs17433909</td>
<td>chr1:206651725</td>
<td>Intron 9</td>
<td>T/C</td>
<td>0.049</td>
<td>0.037</td>
<td>0.032</td>
<td>1.33 (1.03–1.71)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs11118087</td>
<td>chr1:206656577</td>
<td>Intron 13</td>
<td>A/G</td>
<td>0.29</td>
<td>0.25</td>
<td>0.0029</td>
<td>1.19 (1.06–1.34)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs11118092</td>
<td>chr1:206656770</td>
<td>Intron 13</td>
<td>T/C</td>
<td>0.34</td>
<td>0.37</td>
<td>0.033</td>
<td>0.89 (0.79–0.99)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs12409804</td>
<td>chr1:206657181</td>
<td>Intron 13</td>
<td>A/T</td>
<td>0.24</td>
<td>0.21</td>
<td>0.0016</td>
<td>1.22 (1.08–1.38)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs12142086</td>
<td>chr1:206657802</td>
<td>Intron 15</td>
<td>C/T</td>
<td>0.25</td>
<td>0.20</td>
<td>2.1 × 10⁻⁴</td>
<td>1.27 (1.12–1.43)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs11118132</td>
<td>chr1:206659360</td>
<td>Exon 21</td>
<td>C/G</td>
<td>0.36</td>
<td>0.33</td>
<td>0.040</td>
<td>1.12 (1.01–1.25)</td>
</tr>
<tr>
<td>IKBKE</td>
<td>rs3748022</td>
<td>chr1:206669465</td>
<td>Missense</td>
<td>T/C</td>
<td>0.24</td>
<td>0.21</td>
<td>0.0068</td>
<td>1.19 (1.05–1.35)</td>
</tr>
<tr>
<td>IFIH1</td>
<td>rs78456138</td>
<td>chr2:163132346</td>
<td>Intron 11</td>
<td>T/C</td>
<td>0.014</td>
<td>0.024</td>
<td>0.0059</td>
<td>0.56 (0.37–0.85)</td>
</tr>
<tr>
<td>IFIH1</td>
<td>rs3747517</td>
<td>chr2:163128824</td>
<td>Exon 13</td>
<td>T/C</td>
<td>0.26</td>
<td>0.29</td>
<td>0.025</td>
<td>0.87 (0.78–0.98)</td>
</tr>
<tr>
<td>IFIH1</td>
<td>rs1990760</td>
<td>chr2:163124051</td>
<td>Exon 15</td>
<td>C/T</td>
<td>0.36</td>
<td>0.39</td>
<td>0.028</td>
<td>0.89 (0.80–0.99)</td>
</tr>
</tbody>
</table>

Abbreviations: Ala, alanine; Arg, arginine; chr, chromosome; CI, confidence interval; His, histidine; IFIH1, interferon induced with helicase C domain 1; IKBKE, inhibitor of kappa light polypeptide gene enhancer in B cells kinase epsilon; Leu, leucine; MAF, minor allele frequency; OR, odds ratio; Pro, proline; SLE, systemic lupus erythematosus; SNV, single-nucleotide variation; Thr, threonine. *Alleles called based on the forward strand. **Ppatients = 1140; Pcontrols = 2060. **Genotypes imputed based on a reference panel for sequence variants from the 1000 Genomes Phase I (build 37, release March 2012).
and rs1539241, most of the association signals from the SNVs located in the other gene regions persisted. Conditioning on the other SNVs, significant associations from these two SNVs at the gene 5′-region were retained in most cases; whereas they lost their significant P-values when conditioned on each other. These observations indicate that there are two independent association signals for the IKBKE SNVs in our study. One originates from the 5′-region of the gene, which is more preferably tagged by the SNV rs1539241, and the other from the remaining gene regions, which is most likely to be accounted for by the SNV rs12142086. Similarly, two independent association signals were also revealed for IFIH1. The first one was from the low-frequency SNV rs78456138, and the second one from the two common SNVs rs3747517 and rs1990760 located in the exonic regions of IFIH1.

Further analysis was conducted to investigate the possible joint effect of the independent risk or protective alleles in either gene (Table 2). For IKBKE, the haplotype composed of the risk alleles of rs1539241 and rs12142086 had a frequency of about 5% in our sample collection, and conferred a disease risk as high as 1.68 (P-value = 1.0 × 10^{-5}). In contrast, the haplotype composed of their respective protective alleles, which was the most abundant haplotype in the population, showed a protective effect against SLE (odds ratio = 0.82, P-value = 9.3 × 10^{-4}). The association signals from these two haplotypes were also retained after permutations. For IFIH1, the association of the haplotype composed of the minor alleles of the SNVs rs78456138 and rs3747517 was comparable to that of rs78456138 alone (odds ratio = 0.56), which indicated that the T allele of the low-frequency SNV rs78456138 appeared to dominate the protective effect; whereas the most abundant haplotype of these two SNVs render a weak risk for SLE (odds ratio = 1.14, P-value = 0.026). Robust association was only observed for the minor allele haplotype after permutations.

To investigate the SLE association for the rare SNV alleles (MAFs < 0.01), we also examined the combination of all rare alleles in IKBKE or IFIH1 using the sequence kernel association test, which largely increases the statistical power for detecting potential effect of rare variants. No significant association was observed for the combinations of all rare SNV alleles with MAFs < 0.01 (data not shown).

Functional implications for associated SNVs

Considering that the associated SNVs may affect interactions between DNA and protein factors, we compared the 30-bp flanking sequences of these SNVs with canonical protein binding motifs in databases. The major allele T of rs12142086, which was

**Table 2.** Haplotype association analysis of rs1539241 and rs12142086 in IKBKE, and rs78456138 and rs3747517 in IFIH1.

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>Frequency in patients</th>
<th>Frequency in controls</th>
<th>Unadjusted P-value</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IKBKE</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1539241A–rs12142086C</td>
<td>0.082</td>
<td>0.051</td>
<td>1.0 × 10^{-5}</td>
<td>1.68 (1.35–2.05)</td>
</tr>
<tr>
<td>rs1539241G–rs12142086G</td>
<td>0.66</td>
<td>0.70</td>
<td>9.3 × 10^{-4}</td>
<td>0.82 (0.75–0.93)</td>
</tr>
<tr>
<td>rs1539241A–rs12142086T</td>
<td>0.16</td>
<td>0.15</td>
<td>0.12</td>
<td>1.12 (0.94–1.25)</td>
</tr>
<tr>
<td>rs1539241G–rs12142086T</td>
<td>0.090</td>
<td>0.096</td>
<td>0.92</td>
<td>0.99 (0.77–1.11)</td>
</tr>
<tr>
<td>IFIH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs78456138T–rs3747517T</td>
<td>0.013</td>
<td>0.024</td>
<td>6.6 × 10^{-3}</td>
<td>0.56 (0.35–0.82)</td>
</tr>
<tr>
<td>rs78456138C–rs3747517T</td>
<td>0.74</td>
<td>0.71</td>
<td>0.026</td>
<td>1.14 (1.03–1.31)</td>
</tr>
<tr>
<td>rs78456138C–rs3747517T</td>
<td>0.25</td>
<td>0.27</td>
<td>0.16</td>
<td>0.92 (0.80–1.12)</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; IFIH1, interferon induced with helicase C domain 1; IKBKE, inhibitor of kappa light polypeptide gene enhancer in B cells kinase epsilon; OR, odds ratio. *n_{patients} = 1140; n_{controls} = 2060.

**Figure 1.** The IKBKE SNV rs12142086 affects the binding of SF1 with DNA. (A) The SNV rs12142086 is predicted to be located in the binding motif of SF1. Introns are represented by grey horizontal lines and exons by vertical bars with numbers. The position of rs12142086 is highlighted with a star. In the logo of the canonical binding motifs for SF1, the height of the letter stack at each nucleotide position signifies the information content of that position. A nucleotide position that is perfectly conserved contains two bits of information, whereas the position with all four nucleotides occurring equally contains no information. In each letter stack, the height of an individual letter is proportional to the probability of occurrence of the corresponding nucleotide. (B) DNA EMSA for the SNV rs12142086, which indicates allelic differences in binding with proteins in the nuclear extract. The following four reactions (lanes 1–4, from left to right) were performed for both alleles: 1, biotinylated probes only; 2, biotinylated probes and nuclear extract; 3, biotinylated probes, nuclear extract and the unlabelled probes as competitor in 500-fold excess; 4, biotinylated probes, nuclear extract and the unlabelled probes as competitor in 500-fold excess. Arrows indicate the positions of the bands with allelic differences in binding. (C) Gel super-shift assays for SF1-antibody complexes with DNA probes for the SNV rs12142086. The following five reactions (lanes 1–5, from left to right) were performed for both alleles: 1, biotinylated probes only; 2, biotinylated probes and nuclear extract; 3, biotinylated probes, nuclear extract and 6 μg SF1 antibodies; 4, biotinylated probes, nuclear extract and 12 μg SF1 antibodies; 5, biotinylated probes and 6 μg SF1 antibodies. Arrows indicate the positions of the super-shifted bands.
the strongest associated SNV in IKBKE, was found to be located at a highly stringent nucleotide in the binding motif of splicing factor 1 (SF1; P-value = 4.8 × 10⁻⁶, expected number of false positives (E-value) = 8.0 × 10⁻³; Figure 1a). Evidence was provided experimentally by electrophoretic mobility shift assays (EMSSAs), which showed that the major allele (T) of rs12142086 had a stronger binding potential with certain protein factors in nuclear extract compared with the minor allele (C; Figure 1b). Subsequently, we demonstrated that SF1 was involved in the allelic differential binding by super-shift assays using antibodies against SF1 (Figure 1c). SF1 has been reported to have regulatory roles in transcription.27,28 As SF1 is also known to facilitate the recognition of splice sites and the assembly of the early spliceosomal complex with pre-mRNA,29 we also compared the sense- and antisense-RNA sequences bearing the major (U/A) or minor (C/G) allele of rs12142086 using gel shift assays. The sense-RNA sequence bound more prominently to protein factors in the nuclear extract than the antisense-RNA sequence (Supplementary Figure 2a). However, we did not observe any differential binding of protein factors to the U and C allele of rs12142086, or any super-shift signal from a possible interaction with SF1 antibodies (Supplementary Figure 2b).

**DISCUSSION**

Prompted by the rapid development of sequencing technologies, deep resequencing has become an increasingly popular approach for studying disease candidate genes, especially when integrated with association analysis for identifying causal mutations, which can account for previously known association signals. In this study, we explored the spectrum of sequence variants in IKBKE and IFIH1, and identified 12 SNVs as novel risk factors for SLE. In addition, we confirmed the association of one SNV in IFIH1, which has been reported previously. The association signals of all 13 SLE-associated SNVs we identified are supported by statistical power of over 60% (Supplementary Figure 1). Most of them are common in the population (MAFs = 0.15–0.39 after genotyping) and have modest effects.

Two independent association signals were detected from IKBKE, which can be tagged by the SNVs rs1539241 and rs12142086, respectively. Association analysis of the haplotype composed of the risk alleles of these two SNVs reveals a potential synergic effect. We further examined an extended genomic region adjacent to IKBKE based on the data from the HapMap project (Supplementary Figure 3a) and found that IKBKE is located in a relatively isolated LD block. The LD between either of the two tagging SNVs in IKBKE and SNVs located in the neighbouring genes is relatively low (r² < 0.2), supporting the potential functional importance of IKBKE in SLE. The SNV rs12142086 is of particular interest, as our in vitro assays demonstrated that its minor allele may influence the transcriptional regulatory function of SF1 by impairing its binding motif. In addition, we showed that rs12142086 was unlikely to affect the binding of SF1 at the RNA level. Indeed, the pre-mRNA sequence bearing rs12142086 lacks the canonical binding motif of SF1, as well as the polypyrimidine tract recognized by the U2 snRNP auxiliary factor 65 kDa subunit (U2AF65), which functions as a cofactor with SF1 in the spliceosomal complex.30

The case with IFIH1 is more complicated. The minor alleles of the two missense SNVs, rs3747517 and rs1990760, are both predicted to have benign effects on the protein. The impact of rs1990760 has been extensively investigated. Its risk allele is associated with a lower IFN-α level, but enhanced IFN-induced gene expression in the SLE patients who carry anti-dsDNA antibodies.31 In addition, this allele is associated with a higher expression level of the IFIH1 gene itself.32 IFIH1 is located in a high LD region that is shared by the genes fibroblast activation protein alpha (FAP) and grancalcin EF-hand calcium binding protein (GCA) (Supplementary Figure 3b). FAP is an important factor in tumour development and the wound-healing processes.32 It is inducible by tumour necrosis factor-α,33 and overexpression of FAP has been reported in rheumatoid arthritis34 and Crohn’s disease.35 The expression of GCA is restricted to neutrophils and monocytes,36 in which GCA participates in leucocyte-specific host defence.37 Nevertheless, none of the other variants in this LD block provide a more plausible functional hypothesis than the two missense SNVs in IFIH1, rs3747517 and rs1990760.

No clear functional hypotheses could be proposed for the two intronic SNVs (rs11117858 and rs1539241) at the 5’-region of IKBKE or the intronic SNV (rs78456138) in IFIH1, which tag independent association signals for their respective genes in this study. Particularly, the low-frequency SNV rs78456138, which gives the strongest association signal in IFIH1 in our study, may tag other independent functional variants in the high LD region.

Resequencing of targeted genomic regions in pools of individuals is a promising strategy for detecting novel rare alleles.38 However, a proportion of the SNVs with differences in allele frequencies between patients and controls according to resequencing were not replicated in our follow-up association analysis of a larger number of samples. Despite that no clear difference was observed in the distribution of the local read coverage depth for the SNVs between the SLE patients and the controls (Supplementary Figure 4), the different pooling strategies for patients and controls might be one source of bias for accurate estimation of allele frequencies. The stochastic experimental factors during PCR amplification and second generation sequencing may also contribute to the inconsistent findings between the two stages. In addition, the coverage depth along the entire surveyed regions was not perfectly consistent, and the low local coverage depth at certain amplicons could have affected the identification of sequence variants (Supplementary Figure 4). The imbalance of local coverage depth may be due to the imperfect pooling of equal amounts of amplicons, or due to the properties of the DNA sequences, which may have caused bioinformatic mis-alignment. Improved methods for target selection that avoid PCR, as well as increased sequencing accuracy and capacity, together with large-scale barcoding of individual samples are likely to facilitate direct association analysis through resequencing in the future. The accuracy in estimating allele frequencies based on the counts of sequence reads may be improved by increasing the number of subjects and the depth during the sequencing stage.39 However, considering the cost for sequencing, association analysis based on genotyping in large populations is still indispensable for genetic studies. In particular thanks to the recent large-scale projects on human genetic variation, imputation has become a powerful tool for comprehensive investigation of sequence variants based on a limited number of genotyped loci. Indeed, in our study many single-nucleotide polymorphisms located in regions of low coverage depth were retrieved using imputation. In conclusion, we have identified new sequence variants in IKBKE and IFIH1 that are associated with SLE, dissected the contribution of the associated variants to the risk of SLE, and provided functional hypothesis to one independent disease association signal in IKBKE. More exhaustive investigation of sequence variants, especially low-frequency and rare variants in these two genes within a larger cohort, or with subjects of different origins may help to pinpoint any unidentified functional variants. Systematic functional studies of the disease-associated variants are required for a better understanding of the mechanism underlying their association with the risk of SLE.

**SUBJECTS AND METHODS**

**Subjects**

A total of 100 Swedish SLE patients and 100 controls for which there was sufficient amount of DNA available were selected for targeted sequencing. Genotyping was performed in an extended cohort of 1140 SLE patients.
and 2060 controls collected in Sweden. All patients fulfilled at least four of the classification criteria for SLE as defined by the American College of Rheumatology. The study was approved by the ethics committees of all involved institutions. Informed consents were obtained from all SLE patients and controls.

Sequencing

Primers for 14 long-range PCRs (2.5–11.7 kb) were designed to cover the gene regions for both IKBKE (27.4 kb) and IFIH1 (46.2 kb; Supplementary Table 4). An equal amount of purified genomic DNA from each of the 100 SLE patients were pooled before being used as PCR template, and subsequently an equal amount of PCR product obtained with each primer pair from both IKBKE and IFIH1 were combined. Alternatively, purified genomic DNA from each of the 100 controls was used as PCR template, after which an equal amount of PCR products amplified from each individual with each primer pair were combined for both genes. Sample DNA concentrations were quantified using the NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA) and the Quant-IT PicoGreen dsDNA Assay (Invitrogen, Eugene, OR, USA). PCRs were performed using Phusion High-Fidelity DNA polymerase (Finnzymes, Vantaa, Finland). Details of the PCR conditions are summarized in Supplementary Table 4.

Sequencing libraries were prepared from the pooled PCR products from the SLE patients or the controls using a standard protocol for genomic DNA preparation from Illumina (Illumina, San Diego, CA, USA) and reagents from the NEBNext DNA Sample Preparation kit (New England BioLabs, Ipswich, MA, USA). Each of the two libraries were sequenced individually in a single lane of a flow cell using a 76-cycle (for the library of SLE patients) or 36-cycle (for the library of controls) single-end run on a Genome Analyzer Ix (GA Ix) following the manufacturer’s instructions (Illumina). About 9.2 and 12.5 M high-quality sequence reads were obtained from the libraries prepared from the SLE patients and the controls, respectively.

Sequence data analysis and variant calling

Sequence reads were aligned to their target regions in the human reference genome (GRCh37/hg19), and potential SNVs were called using the software Maq v.0.7.1, with manual inspection of the sequence reads at called positions. For IKBKE and IFIH1, 65% and 60% of the reads aligned to the reference sequences, respectively. For both genes, >80% of the amplified regions were covered by at least 10 sequence reads. The median coverage depth for the reads was >2000 from SLE patients and >1000 from the controls, which corresponded to >20- and >10-fold coverage per individual. For further analysis, we retained SNVs that were located at high-quality base positions (quality score ≥ 28) and that were supported by at least 10 reads for the minor allele. In addition, SNVs for which the number of reads for the minor alleles was less than twice the reads for the background were removed. The MAFs of the called SNVs were estimated by counting the number of reads, which carried either allele. Any SNV with a MAF < 0.05, corresponding to 1 out of 200 chromosomes, was also filtered out. Putative SNVs were classified as novel or known by referencing to the Short Genetic Variations database dbSNP v137 (http://www.ncbi.nlm.nih.gov/projects/SNP/). Annotations for all identified SNVs were retrieved from the SeattleSeq Annotation server (http://gvs.gs.washington.edu/SeattleSeqAnnotation).

Genotyping

SNVs were selected based on the following criteria: non-intronic SNVs (n = 15); SNVs called exclusively in SLE patients or in controls (n = 144); and SNVs occurring in both sample groups with an allele frequency difference ≥ 0.03 (n = 19). A total of 86 SNVs in IKBKE and IFIH1 were included for genotyping with custom GoldenGate Genotyping assays using VeraCode microbeads on the Illumina BeadXpress system (Illumina). A subset of 72 SNVs passed the following quality filters applied to exclude poor genotyping assays from the downstream analysis: SNVs or samples with > 10% missing data (n = 13) and SNVs with genotypes, which deviated from Hardy–Weinberg equilibrium (P-value < 0.001, n = 1) in controls.

Imputation

Imputation was performed based on the genotyped 72 SNVs, which passed quality filters using the software IMPUTE2, with sequence variants from the 1000 Genomes Phase I (build 37, release March 2012) as reference. We kept imputed sequence variants with information scores > 0.8, and imputed genotypes with confidence probability > 0.9. The imputed sequence variants with > 10% missing genotypes in all individuals and the ones deviating from Hardy–Weinberg equilibrium (P-value < 0.001) in controls were excluded from downstream analysis.

Statistical analysis

Using the software PLINK,44 allelic association analysis for individual SNVs was performed by comparing the allele counts in cases and controls with Fisher’s exact tests, with 109 permutations to control for the rare SNVs and multiple testing involved in this study. D2- and r2-based pairwise LD values between associated SNVs were calculated using the genotype data from the controls. Independence of association signals between SNVs was tested with conditional logistic regression analysis under an additive model. Haplotype association analysis was performed using logistic regression analysis, also with 109 permutations, all using PLINK. For the rare SNVs (MAFs < 0.01), we examined their joint effects in IKBKE or IFIH1 using the R package sequence kernel association test (SKAT) v0.6356 with default parameters. Statistical power was estimated using the software Quant v1.2.4 (http://hydra.usc.edu/gxe/) based on the total number of genotyped SLE patients and controls, with a log-additive model, a two-sided type I error rate of 0.05 and a Swedish SLE prevalence of 0.068.45

Functional prediction

Potential binding sites for protein factors were predicted by comparing the sequences around each SNV, including 10-bp upstream and downstream of the variation site, with known motifs in the JASPAR CORE 2009 database46 and the TRANSFAC database 7.047 using the software Motif Comparison Tool (TOMTOM) v4.7.0.48 A match with a P-value < 1 x 10–5 and an expected number of false positives (E-value) < 0.05 was called as significant. Effects of the non-synonymous SNVs were predicted using the software Structure SNP (StSNP).49

Electrophoretic mobility shift assay

DNA and RNA EMSA were performed according to the instructions from the LightShift Chemiluminescent DNA/RNA EMSA Kit (Thermo Scientific Pierce Protein Research, Rockford, IL, USA). Four pairs of 31-nt long single-stranded DNA oligonucleotides (Integrated DNA Technologies, Leuven, Belgium) with the T/A or C/G allele of the IKBKE SNV rs12142086 positioned in the centre, and either S– or biotinylated or unlabelled, were annealed to generate double-stranded DNA probes. The S– biotinylated single-stranded RNA probes (Integrated DNA Technologies) shared base compositions with the single-stranded DNA oligonucleotides, corresponding to the forward and reverse strand bearing either the U/C or A/G allele. Nuclear extract was prepared from human peripheral blood mononuclear cells from a healthy blood donor using the NE-PCR Nuclear and Cytoplasmic Extraction Reagents Kit (Thermo Scientific Pierce Protein Research). For super-shift assays, nuclear extract was incubated with S1 antibody (sc-48792 X) (Santa Cruz Biotechnology, Dallas, TX, USA) before the binding reactions.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Knut and Alice Wallenberg Foundation (2011.0073 to A-CS and LR), the Swedish Research Council for Medicine and Health (A0280001 to A-CS and A0258801 to LR) and the Swedish Research Council for Science and Technology (90559401 to A-CS). The SNP&SEQ Technology Platform, which is part of the genomics platform of Science for Life L fterat in Uppsala, is supported by Uppsala University, Uppsala University Hospital and by the Swedish Research Council for Research Infrastructures (80576801 and 70374401). We also acknowledge the support of the following: the Swedish Heart–Lung Foundation, the Stockholm County Council and Karolinska Institutet (ALF), the King Gustaf V’s 80th Birthday Fund, the Swedish Rheumatism Association, the COMBINE, the Ragnar Söderbergs Foundation, the Foundation in memory of Claus G roscinsky and the Swedish Society of Medicine. We thank research nurse Rezvan Kiani from the Rheumatology clinic at Uppsala University Hospital for collecting the patient blood samples. We also thank Dr Sofia Nordman, Torbjörn Öst and the SNP&SEQ Technology Platform for help with sequencing and genotyping, and Anders Lundmark for help with sequence data analysis.
REFERENCES


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