ORIGINAL ARTICLE

Contribution of *IKBKE* and *IFIH1* gene variants to SLE susceptibility

C Wang[1,9], A Ahlford[1,9,10], N Laxman[1,11], G Nordmark[2], M-L Eloranta[2], I Gunnarsson[3], E Svenungsson[3], L Padyukov[3], G Sturfelt[4], A Jönsen[4], AA Bengtsson[4], L Truedsson[5], S Rantapää-Dahlqvist[6], C Sjöwall[7], JK Sandling[1,8], L Rönnblom[2] and A-C Syvänen[1]

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 able to account 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 $= 1.0 \times 10^{-5}$). 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 $= 6.6 \times 10^{-3}$). In conclusion, we have identified new SLE-associated sequence variants in *IKBKE* and *IFIH1*, and proposed functional hypotheses for the association signals.

INTRODUCTION

Systemic lupus erythematosus (SLE) is a clinically heterogeneous disease, which affects multiple organ systems and causes significant morbidity and mortality.[1] About 8–15% of the genetic component of SLE is accounted for by DNA sequence variants identified by association studies in European populations.[2] These findings have underscored the crucial role of the innate immune system, and particularly of the type I interferon (IFN) system in SLE.[3,4] Genes involved in the type I IFN system such as *STAT4*, *IRF5*, *TNFAIP3*, *TNFSF4* and *TNIP1*,[5–8] have been replicated as SLE susceptibility genes in multiple ethnical groups. Previously, we have provided evidence of association for two additional type I IFN system genes, namely inhibitor of kappa light polypeptide gene enhancer in B cells kinase epsilon (*IKBKE*) and IFN induced with helicase C domain 1 (*IFIH1*).[2,9]

IKBKE encodes IKKε, which belongs to the IκB kinase family. Its expression is restricted to lymphoid tissues and peripheral blood leukocytes.[10,11] 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.[12–14] 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[15] and antibody-positive primary Sjögren's syndrome.[16]

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.[17,18] When recognizing viral nucleic acids, IFIH1 triggers activation of IFN regulatory factor 3, 7 and NF-κB via mitochondrial antiviral signalling protein, leading to antiviral and pro-apoptotic responses.[19] Genetic association studies have demonstrated an important role for *IFIH1* in autoimmune diseases such as type I diabetes,[20] rheumatoid arthritis,[21] multiple sclerosis,[22] Graves' disease[23] and psoriasis.[24]

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.

[1]Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden; [2]Department of Medical Sciences, Rheumatology, Uppsala University, Uppsala, Sweden; [3]Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden; [4]Section of Rheumatology, Department of Clinical Sciences, Lund University, Lund, Sweden; [5]Section of MIG, Department of Laboratory Medicine, Lund University, Lund, Sweden; [6]Department of Public Health and Clinical Medicine, Rheumatology, Umeå University, Umeå, Sweden; [7]Rheumatology/AIR, Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden and [8]Genetics of Complex Traits in Humans, Department of Human Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. Correspondence: Professor A-C Syvänen, Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Entrance 70, 3rd Floor, Research Department II, University Hospital, Uppsala SE-751 85, Sweden.
E-mail: ann-christine.syvanen@medsci.uu.se
[9]These authors contributed equally to this work.
[10]Current address: Science for Life Laboratory, Department of Biochemistry and Biophysics, Molecular Diagnostics, Stockholm University, Sweden.
[11]Current address: Department of Medical Sciences, Metabolic Bone Diseases, Uppsala University, Sweden.
Received 30 October 2012; revised 13 February 2013; accepted 25 February 2013; published online 28 March 2013

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*,[9] and rs1990760 in *IFIH1*[2,25]) associated with SLE and one SNV (rs35732034 in *IFIH1*[20]) associated with type I diabetes in previous studies were also added to the genotyping panel. In total, 86 SNVs were genotyped in an extended cohort of 1140 SLE patients and 2060 controls from Sweden. In addition, 108 high-quality sequence variants that were imputed based on our genotyped SNVs and data from the 1000 Genomes Phase I (build 37, release March 2012) were included in the association analysis. In all, 10 SNVs in *IKBKE* and 3 SNVs in *IFIH1* showed significant association signals with SLE, of which the associations of 12 SNVs have not been reported before (Table 1 and Supplementary Table 1). The minor alleles of nine of the associated SNVs in *IKBKE* conferred a moderate risk for SLE (odds ratio = 1.12–1.33). The most profound *P*-values were from the SNVs rs12142086 (*P*-value = 2.1×10^{-4}) and rs2297545 (*P*-value = 5.5×10^{-4}), which are located in intron 13 and exon 8 of *IKBKE*, respectively. Their association signals were retained after permutations. For *IFIH1*, the

minor alleles of the associated SNVs and particularly the low-frequency SNV rs78456138 appeared to confer protection against SLE. This SNV is located in intron 11 of *IFIH1*, and its T allele (minor allele frequency (MAF) = 0.01–0.02) showed an odds ratio as low as 0.56 (*P*-value = 0.0059). The association of the previously reported common SLE SNV rs1990760 was also replicated. Although we could not detect a significant association signal from the SNV rs17433930 in *IKBKE*, the effect of the minor allele was in the same direction as reported in our previous study.[9] None of the novel SNVs in *IKBKE* or *IFIH1* identified by resequencing in this study or the diabetes-associated rare SNV rs35732034 in *IFIH1* (MAFs = 0.0005–0.01 after genotyping) gave any significant association signal, which might be due to the limited statistical power (Supplementary Figure 1).

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 5′-gene region, rs11117858 and rs1539241, had an r^2 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 ($r^2 = 0.45$–1.00; $D' = 0.93$–1.00). For *IFIH1*, strong LD was observed between the SNVs rs3747517 and rs1990760 ($r^2 = 0.65$; $D' = 1.00$), but not for the SNV rs78456138 ($r^2 = 0.04$–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 5′-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

Table 1. SLE-associated SNVs in *IKBKE* and *IFIH1*

Gene	SNV	Position (GRCh37/hg19)	Annotation	Minor/major allele[a]	MAF in patients[b]	MAF in controls[b]	Unadjusted P-value	OR (95% CI)
IKBKE	rs11117858	chr1:206646274	Intron 2	G/T	0.25	0.22	0.023	1.15 (1.02–1.30)
IKBKE	rs1539241	chr1:206647450	Intron 3	A/G	0.17	0.15	0.0094	1.21 (1.05–1.39)
IKBKE	rs2297545	chr1:206651107	Exon 8 Synonymous ACG = > ACA	A/G	0.22	0.19	5.5×10^{-4}	1.26 (1.10–1.43)
IKBKE	rs17433909	chr1:206651725	Intron 9	T/C	0.049	0.037	0.032	1.33 (1.03–1.71)
IKBKE	rs11118087	chr1:206656577	Intron 13	A/G	0.29	0.25	0.0029	1.19 (1.06–1.34)
IKBKE	rs11118092	chr1:206656770	Intron 13	T/C	0.34	0.37	0.033	0.89 (0.79–0.99)
IKBKE	rs12409804[c]	chr1:206657181	Intron 13	A/T	0.24	0.21	0.0016	1.22 (1.08–1.38)
IKBKE	rs12142086	chr1:206657802	Intron 13	C/T	0.25	0.20	2.1×10^{-4}	1.27 (1.12–1.43)
IKBKE	rs11118132	chr1:206659360	Intron 15	C/G	0.36	0.33	0.040	1.12 (1.01–1.25)
IKBKE	rs3748022	chr1:206669465	Exon 21 Missense Pro = > Leu	T/C	0.24	0.21	0.0068	1.19 (1.05–1.35)
IFIH1	rs78456138	chr2:163132346	Intron 11	T/C	0.014	0.024	0.0059	0.56 (0.37–0.85)
IFIH1	rs3747517	chr2:163128824	Exon 13 Missense His = > Arg	T/C	0.26	0.29	0.025	0.87 (0.78–0.98)
IFIH1	rs1990760	chr2:163124051	Exon 15 Missense Thr = > Ala	C/T	0.36	0.39	0.028	0.89 (0.80–0.99)

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. [a]Alleles called based on the forward strand. [b]$n_{\text{patient}} = 1140$; $n_{\text{control}} = 2060$. [c]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*.

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

Haplotype	Frequency in patients[a]	Frequency in controls[a]	Unadjusted P-value	OR (95% CI)
IKBKE				
rs1539241A–rs12142086C	0.082	0.051	1.0×10^{-5}	1.68 (1.35–2.05)
rs1539241G–rs12142086T	0.66	0.70	9.3×10^{-4}	0.82 (0.75–0.93)
rs1539241G–rs12142086C	0.16	0.15	0.12	1.12 (0.94–1.25)
rs1539241A–rs12142086T	0.090	0.096	0.92	0.99 (0.77–1.11)
IFIH1				
rs78456138T–rs3747517T	0.013	0.024	6.6×10^{-3}	0.56 (0.35–0.82)
rs78456138C–rs3747517C	0.74	0.71	0.026	1.14 (1.03–1.31)
rs78456138C–rs3747517T	0.25	0.27	0.16	0.92 (0.80–1.02)

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. [a]$n_{patient} = 1140$; $n_{control} = 2060$.

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.[26] 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
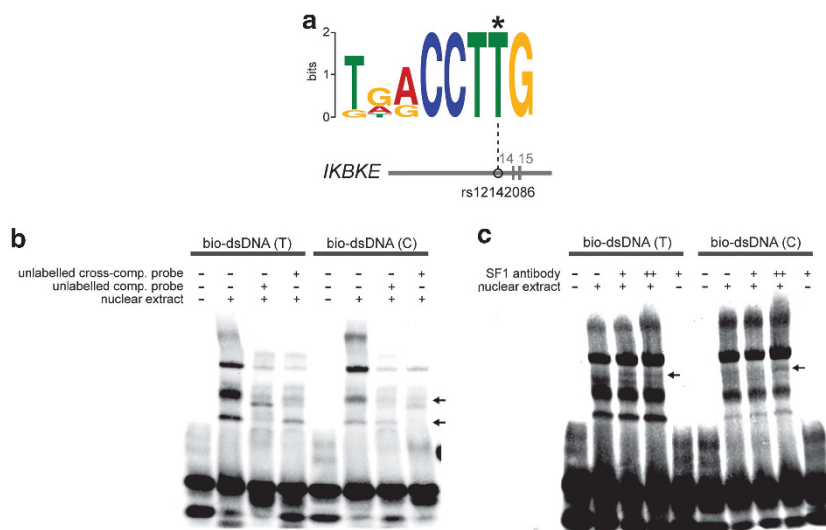

Figure 1. The *IKBKE* SNV rs12142086 affects the binding of SF1 with DNA. (**a**) The SNV rs12142086 is predicted to be located in the binding motif of SF1. Introns are represented by grey horizontal lines and exons by vertical bars with numbers. The position of rs12142086 is highlighted with a star. 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 for the other allele of the polymorphism as cross-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 $= 4.8 \times 10^{-6}$, expected number of false positives (E-value) $= 8.0 \times 10^{-3}$; Figure 1a). Evidence was provided experimentally by electrophoretic mobility shift assays (EMSAs), 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 ($r^2 < 0.2$), 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.[25] *IFIH1* is located in a high LD region that is shared by the genes fibroblast activation protein alpha (*FAP*) and granalcin 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 arthritis[34] and Crohn's disease.[35] The expression of *GCA* is restricted to neutrophils and monocytes,[36] in which GCA participates in leukocyte-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.[40,41] 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 IIx (GA IIx) 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,[42] 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 <0.005, 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,[43] 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 10^6 permutations to control for the rare SNVs and multiple testing involved in this study. D'- and r^2-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 10^6 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) v0.63[26] with default parameters. Statistical power was estimated using the software Quanto 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 database[46] and the TRANSFAC database 7.0[47] using the software Motif Comparison Tool (TOMTOM) v4.7.0.[48] A match with a P-value $< 1 \times 10^{-3}$ 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 5'- biotinylated or unlabelled, were annealed to generate double-stranded DNA probes. The 5'-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 the A/G allele. Nuclear extract was prepared from human peripheral blood mononuclear cells from a healthy blood donor using the NE-PER Nuclear and Cytoplasmic Extraction Reagents Kit (Thermo Scientific Pierce Protein Research). For super-shift assays, nuclear extract was incubated with SF1 antibody (sc-48792 X) (Santa Cruz Biotechnology, Dallas, TX, USA) before the binding reactions.

REFERENCES

1 Hewagama A, Richardson B. The genetics and epigenetics of autoimmune diseases. *J Autoimmun* 2009; **33**: 3–11.

2 Gateva V, Sandling JK, Hom G, Taylor KE, Chung SA, Sun X *et al.* A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet* 2009; **41**: 1228–1233.

3 Ronnblom L, Alm GV, Eloranta ML. The type I interferon system in the development of lupus. *Semin Immunol* 2011; **23**: 113–121.

4 Bronson PG, Chaivorapol C, Ortmann W, Behrens TW, Graham RR. The genetics of type I interferon in systemic lupus erythematosus. *Curr Opin Immunol* 2012; **24**: 530–537.

5 Hom G, Graham RR, Modrek B, Taylor KE, Ortmann W, Garnier S *et al.* Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N Engl J Med* 2008; **358**: 900–909.

6 Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, Moser KL *et al.* Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet* 2008; **40**: 204–210.

7 Graham RR, Cotsapas C, Davies L, Hackett R, Lessard CJ, Leon JM *et al.* Genetic variants near TNFAIP3 on 6q23 are associated with systemic lupus erythematosus. *Nat Genet* 2008; **40**: 1059–1061.

8 Han JW, Zheng HF, Cui Y, Sun LD, Ye DQ, Hu Z *et al.* Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus. *Nat Genet* 2009; **41**: 1234–1237.

9 Sandling JK, Garnier S, Sigurdsson S, Wang C, Nordmark G, Gunnarsson I *et al.* A candidate gene study of the type I interferon pathway implicates IKBKE and IL8 as risk loci for SLE. *Eur J Hum Genet* 2011; **19**: 479–484.

10 Shimada T, Kawai T, Takeda K, Matsumoto M, Inoue J, Tatsumi Y *et al.* IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases. *Int Immunol* 1999; **11**: 1357–1362.

11 Peters RT, Liao SM, Maniatis T. IKKepsilon is part of a novel PMA-inducible IkappaB kinase complex. *Mol Cell* 2000; **5**: 513–522.

12 Sharma S, tenOever BR, Grandvaux N, Zhou GP, Lin R, Hiscott J. Triggering the interferon antiviral response through an IKK-related pathway. *Science* 2003; **300**: 1148–1151.

13 Fitzgerald KA, McWhirter SM, Faia KL, Rowe DC, Latz E, Golenbock DT *et al.* IKKepsilon and TBK1 are essential components of the IRF3 signaling pathway. *Nat Immunol* 2003; **4**: 491–496.

14 Mattioli I, Geng H, Sebald A, Hodel M, Bucher C, Kracht M *et al.* Inducible phosphorylation of NF-kappa B p65 at serine 468 by T cell costimulation is mediated by IKK epsilon. *J Biol Chem* 2006; **281**: 6175–6183.

15 Dieguez-Gonzalez R, Akar S, Calaza M, Perez-Pampin E, Costas J, Torres M *et al.* Genetic variation in the nuclear factor kappaB pathway in relation to susceptibility to rheumatoid arthritis. *Ann Rheum Dis* 2009; **68**: 579–583.

16 Nordmark G, Kristjansdottir G, Theander E, Appel S, Eriksson P, Vasaitis L *et al.* Association of EBF1, FAM167A(C8orf13)-BLK and TNFSF4 gene variants with primary Sjogren's syndrome. *Genes Immun* 2011; **12**: 100–109.

17 Kang DC, Gopalkrishnan RV, Lin L, Randolph A, Valerie K, Pestka S *et al.* Expression analysis and genomic characterization of human melanoma differentiation associated gene-5, mda-5: a novel type I interferon-responsive apoptosis-inducing gene. *Oncogene* 2004; **23**: 1789–1800.

18 Yoneyama M, Kikuchi M, Natsukawa T, Shinobu N, Imaizumi T, Miyagishi M *et al.* The RNA helicase RIG-I has an essential function in double-stranded RNA-induced innate antiviral responses. *Nat Immunol* 2004; **5**: 730–737.

19 Barral PM, Sarkar D, Su ZZ, Barber GN, DeSalle R, Racaniello VR *et al.* Functions of the cytoplasmic RNA sensors RIG-I and MDA-5: key regulators of innate immunity. *Pharmacol Ther* 2009; **124**: 219–234.

20 Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ *et al.* A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 2006; **38**: 617–619.

21 Martinez A, Varade J, Lamas JR, Fernandez-Arquero M, Jover JA, de la Concha EG *et al.* Association of the IFIH1-GCA-KCNH7 chromosomal region with rheumatoid arthritis. *Ann Rheum Dis* 2008; **67**: 137–138.

22 Enevold C, Oturai AB, Sorensen PS, Ryder LP, Koch-Henriksen N, Bendtzen K. Multiple sclerosis and polymorphisms of innate pattern recognition receptors TLR1-10, NOD1-2, DDX58, and IFIH1. *J Neuroimmunol* 2009; **212**: 125–131.

23 Sutherland A, Davies J, Owen CJ, Vaikkakara S, Walker C, Cheetham TD *et al.* Genomic polymorphism at the interferon-induced helicase (IFIH1) locus contributes to Graves' disease susceptibility. *J Clin Endocrinol Metab* 2007; **92**: 3338–3341.

24 Li Y, Liao W, Cargill M, Chang M, Matsunami N, Feng BJ *et al.* Carriers of rare missense variants in IFIH1 are protected from psoriasis. *J Invest Dermatol* 2010; **130**: 2768–2772.

25 Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L *et al.* Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genet* 2011; **7**: e1002341.

26 Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* 2011; **89**: 82–93.

27 Zhang D, Paley AJ, Childs G. The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. *J Biol Chem* 1998; **273**: 18086–18091.

28 Goldstrohm AC, Albrecht TR, Sune C, Bedford MT, Garcia-Blanco MA. The transcription elongation factor CA150 interacts with RNA polymerase II and the pre-mRNA splicing factor SF1. *Mol Cell Biol* 2001; **21**: 7617–7628.

29 Liu Z, Luyten I, Bottomley MJ, Messias AC, Houngninou-Molango S, Sprangers R *et al.* Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 2001; **294**: 1098–1102.

30 Gupta A, Jenkins JL, Kielkopf CL. RNA induces conformational changes in the SF1/U2AF65 splicing factor complex. *J Mol Biol* 2011; **405**: 1128–1138.

31 Robinson T, Kariuki SN, Franek BS, Kumabe M, Kumar AA, Badaracco M *et al.* Autoimmune disease risk variant of IFIH1 is associated with increased sensitivity to IFN-alpha and serologic autoimmunity in lupus patients. *J Immunol* 2011; **187**: 1298–1303.

32 Aertgeerts K, Levin I, Shi L, Snell GP, Jennings A, Prasad GS *et al.* Structural and kinetic analysis of the substrate specificity of human fibroblast activation protein alpha. *J Biol Chem* 2005; **280**: 19441–19444.

33 Brokopp CE, Schoenauer R, Richards P, Bauer S, Lohmann C, Emmert MY *et al.* Fibroblast activation protein is induced by inflammation and degrades type I collagen in thin-cap fibroatheromata. *Eur Heart J* 2011; **32**: 2713–2722.

34 Bauer S, Jendro MC, Wadle A, Kleber S, Stenner F, Dinser R *et al.* Fibroblast activation protein is expressed by rheumatoid myofibroblast-like synoviocytes. *Arthritis Res Ther* 2006; **8**: R171.

35 Rovedatti L, Di Sabatino A, Knowles CH, Sengupta N, Biancheri P, Corazza GR *et al.* Fibroblast activation protein expression in Crohn's disease strictures. *Inflamm Bowel Dis* 2011; **17**: 1251–1253.

36 Boyhan A, Casimir CM, French JK, Teahan CG, Segal AW. Molecular cloning and characterization of grancalcin, a novel EF-hand calcium-binding protein abundant in neutrophils and monocytes. *J Biol Chem* 1992; **267**: 2928–2933.

37 Liu F, Shinomiya H, Kirikae T, Hirata H, Asano Y. Characterization of murine grancalcin specifically expressed in leukocytes and its possible role in host defense against bacterial infection. *Biosci Biotechnol Biochem* 2004; **68**: 894–902.

38 Prabhu S, Pe'er I. Overlapping pools for high-throughput targeted resequencing. *Genome Res* 2009; **19**: 1254–1261.

39 Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR. Low-coverage sequencing: implications for design of complex trait association studies. *Genome Res* 2011; **21**: 940–951.

40 Tan EM, Cohen AS, Fries JF, Masi AT, McShane DJ, Rothfield NF *et al.* The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1982; **25**: 1271–1277.

41 Hochberg MC. Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum* 1997; **40**: 1725.

42 Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 2008; **18**: 1851–1858.

43 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009; **5**: e1000529.

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

45 Stahl-Hallengren C, Jonsen A, Nived O, Sturfelt G. Incidence studies of systemic lupus erythematosus in Southern Sweden: increasing age, decreasing frequency of renal manifestations and good prognosis. *J Rheumatol* 2000; **27**: 685–691.

46 Portales-Casamar E, Thongjuea S, Kwon AT, Arenillas D, Zhao X, Valen E *et al.* JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2010; **38**: D105–D110.

47 Wingender E, Dietze P, Karas H, Knuppel R. TRANSFAC: a database on transcription factors and their DNA binding sites. *Nucleic Acids Res* 1996; **24**: 238–241.

48 Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. *Genome Biol* 2007; **8**: R24.

49 Uzun A, Leslin CM, Abyzov A, Ilyin V. Structure SNP (StSNP): a web server for mapping and modeling nsSNPs on protein structures with linkage to metabolic pathways. *Nucleic Acids Res* 2007; **35**: W384–W392.

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