Concise report

Replication of GWAS-identified systemic lupus erythematosus susceptibility genes affirms B-cell receptor pathway signalling and strengthens the role of IRF5 in disease susceptibility in a Northern European population

Tiina M. Järvinen1,2,3,4,5, Anna Hellquist1, Marco Zucchelli1, Sari Koškenmies2,3,4, Jaana Panelius2, Taina Hasan6, Heikki Julkunen7, Mauro D’Amato1 and Juha Kere1,3,4,5,8

Abstract

Objective. A large number of genes, including several not previously implicated in SLE susceptibility, have recently been identified or confirmed by genome-wide association studies (GWAS). In this study, we sought to replicate some of these results in Finnish SLE patients (n = 275) and control individuals (n = 356).

Methods. We genotyped 32 single nucleotide polymorphisms (SNPs) in 12 of the best-supported GWAS-identified SLE genes and loci. We further investigated gene-gene interactions between the loci included in the study.

Results. The strongest evidence of association was found at the IRF5-TNPO3 locus, with the most significant P-value being 2.0 × 10⁻⁷ and an odds ratio of 1.95 (95% CI 1.51, 2.50). Association between SLE and TNFAIP3, FAM167A-BLK, BANK1 and KIAA1542 was also confirmed, although at a lower significance level and contribution to individual risk. No significant association was found with 1q25.1, PXK, ATG5, ICA1, XKRI6, LYN and SCUBE1. Furthermore, no significant gene–gene interactions were detected.

Conclusion. Replication of previous GWAS findings across diverse populations is of importance to validate these associations and to get a better understanding of potential genetic heterogeneity between populations in SLE susceptibility. Our results attest the importance of B-cell receptor pathway and IFN signalling in SLE pathogenesis.

Key words: B-cell receptor pathway, epistasis, Finnish population, genome-wide association study replication, interferon regulatory factor 5–transportin 3

Introduction

SLE is a complex autoimmune disease, characterized by production of pathogenic autoantibodies against nuclear antigens due to a breakdown in self-tolerance. The pathogenesis is associated with the formation of ICs, followed by tissue inflammation in multiple organs, such as the skin, joints, heart and kidneys. SLE is an unusually heterogeneous disease and its clinical classification is based on criteria set by the ACR [1].

The genetic component in SLE is strong, with familial aggregation studies showing a sibling risk ratio (r(s)) of 8–29 [2]. Twin studies further support this view with a 10 times...
higher disease concordance in monozygotic twins (24–56%) compared with dizygotic twins (2–9%) [2].

Following the introduction of genome-wide association studies (GWAS) in SLE, a number of susceptibility genes and loci have been identified and consistently replicated in the past couple of years. To date, there are more than 30 identified loci that show robust association with SLE (supplementary table 1, available as supplementary data at *Rheumatology* Online), and several others for which confirmatory studies are needed through replication in different populations [2]. Interestingly, most of these SLE-predisposing genes appear to be involved in similar and/or related biological pathways, including the processing of ICs, Type 1 IFN production and immune signal transduction [2]. A few other genes, on the contrary, have no assigned function or obvious role in the immune system, and thus represent ideal candidates to reveal novel disease mechanisms. In the present study, we set up to confirm the pathogenetic role of 12 GWAS-identified SLE risk loci in the Finnish population. Single nucleotide polymorphisms (SNPs) at these loci were tested for association with disease and for potential gene–gene interactions (epistasis) in a well-characterized case–control cohort.

**Materials and methods**

**Patients and controls**

SLE patients (*n* = 275) and unmatched controls (*n* = 356) included in this study have been described in detail in previous publications [3, 4] and their demographic and clinical characteristics are reported in supplementary table 2 (available as supplementary data at *Rheumatology* Online).

All patients met the ACR classification criteria for SLE [1] and gave a written informed consent in accordance with the tenets of the Declaration of Helsinki. The study was reviewed and approved by local ethical committees (University of Helsinki and Helsinki University Central Hospital, Helsinki; University of Tampere and Tampere University Hospital, Tampere, Finland; and Karolinska Institutet, Stockholm, Sweden).

**SNP selection**

Relevant SNPs at SLE risk loci were selected based on findings from recent GWAS [5–8]. Priority was given to genes with strong evidence of association, including *BANK1, TNFAIP3, IRF5-TNPO3, FAM167A-BLK, KIAA1542, Tq25.1* and *PXR*. A number of genes that showed suggestive association [6] were also included in the study (*ATG5, ICA1, LYN, XKR6* and *SCUBE1*). At each locus, the SNPs providing the best evidence of association in GWAS, together with those already known to be of functional relevance or to strongly associate with SLE, were selected (supplementary table 3, available as supplementary data at *Rheumatology* Online). The SNP rs10954213 in *IRF5*, as well as a number of SNPs in *STAT4* and *ITGAM-ITGAX* have been previously studied in the Finnish population [9–11].

**Genotyping**

For SNP genotyping, the Sequenom iPLEX Gold chemistry (Sequenom Inc., San Diego, CA, USA) [11] was applied, and the CGGGG insertion/deletion (indel) polymorphism at the *IRF5-TNPO3* locus was genotyped as described previously [12]. The success rate for all markers was >90%, and no significant deviation from Hardy–Weinberg equilibrium was detected in control individuals. The PedCheck program was used to detect Mendelian inconsistencies in the family material and two families were excluded from further analysis owing to Mendelian inheritance errors.

**Statistical analyses**

A total of 275 SLE patients and 356 controls were included in the statistical analyses. Statistical power was calculated based on reported minor allele frequencies [5–8, 13, 14] at a significance level of 0.05, using the online Power for Association with Error program v. 1.2. The Haploview software v. 4.0 was used to explore linkage disequilibrium (LD) define haplotype block structure (D’) and perform an association test at each locus. Two-tailed odds ratios (ORs) with their corresponding 95% CIs were calculated using GraphPad Prism v. 4.03. Since the present study was a replication of previous findings, uncorrected *P* < 0.05 was considered statistically significant and reported.

Gene–gene interaction [9] was studied using a multiple logistic regression under a codominant or additive model by adding an interaction term between the genotypes from all SNPs investigated, as well as previously studied SNPs in *IRF5, STAT4* and *ITGAM-ITGAX* [10–12]. The significance threshold was set to a *P* < 3 × 10^{-5} by standard Bonferroni correction for multiple testing. Analyses were performed with the R software v. 2.6.2.

The *IRF5-TNPO3* locus was chosen for meta-analysis to highlight our results effectively to support results from previous studies [6, 7, 9, 13, 15, 16], despite the small sample size. Genotype frequencies were estimated based on published allele frequencies and meta-analysis was performed using PLINK with data set as a confounding factor.

**Results**

Altogether, 31 SNPs and 1 indel polymorphism in 12 SLE susceptibility genes were genotyped and tested for association with disease by a case–control-based study design in the Finnish population. The salient results of this analysis are reported in Table 1, whilst more comprehensive information is given in supplementary table 4 (available as supplementary data at *Rheumatology* Online). The *IRF5-TNPO3* locus resulted in the strongest association with SLE, with several SNPs showing highly significant *P*-values ranging between 0.01 and 2.0 × 10^{-7}. Meta-analysis demonstrated that our results conform to published data at this locus (supplementary figure 1, available as supplementary data at *Rheumatology* Online). More modest association was seen with SNPs in *TNFAIP3*.
and FAM167A-BLK, with P-values in the range of 0.02–4.5 \times 10^{-3} and 0.01–1.5 \times 10^{-3}, respectively. The highest ORs were, however, observed for several SNPs in TNFAIP3, but with larger CIs when compared with the other genes. Borderline association was also detected for BANK1 and KIAA1542, while no association was detected for 1q25.1, PXK, ATG5, ICA1, XKR6, LYN and SCUBE1 (supplementary table 4, available as supplementary data at Rheumatology Online).

We next tested whether some of the observed associations were due to the specific contribution of certain haplotypes. LD block structure (supplementary figure 2, available as supplementary data at Rheumatology Online) was utilized to explore haplotypes at each locus and determine, if they were significantly associated with risk of, or protection from SLE (Table 2). Neither stronger significance nor additional information beyond single-marker analysis was observed, because most of the detected haplotype associations are being driven by individual-associated SNPs. This result is not surprising, given the strong LD occurring at most of the studied loci, and the fact that a larger number of study subjects would likely be needed to effectively test different haplotypes at each locus with sufficient power.

In summary, we replicated successfully 5 out of 12 known SLE risk loci in the Finnish population, all of which have been found to significantly associate with SLE in one or several GWAS [5–8]. We were unable to reproduce associations with 1q25.1, ATG5, ICA1, XKR6, LYN, PXK and SCUBE1, of which only 1q25.1 and PXK have reached significant association in one previous GWAS [6].

Given that none of the GWAS or subsequent follow-up studies has exhibited strong evidence of gene–gene interaction, a likely consequence of the stringent thresholds set for multiple testing, we evaluated potential interaction effects between studied SNPs. However, no evidence of epistasis was detected through this analysis (data not shown). Although we have shown an interaction between IRF5 and TYK2 genes in this same study population [9], it might be that we lack power to detect epistasis of small effect size variants identified by GWAS. Statistical power calculations showed that we had >97% power to detect association to IRF5-TNPO3 with an OR of \geq 1.58 and >84% power in TNFAIP3 (OR \geq 2.00), while other loci had lower power (~73%).

**Discussion**

In the present study, we set out to replicate a number of GWAS-identified SLE candidate genes in Finnish SLE patients and controls. SNPs at the IRF5-TNPO3 locus showed a strong association with SLE, while the association with SNPs in FAM167A-BLK, TNFAIP3, BANK1 and KIAA1542 was more moderate. Furthermore, none of the analysed SNPs in 1q25.1, PXK, ATG5, ICA1, XKR6, LYN and SCUBE1 showed a significant association. These results are in line with the reported P-values and ORs, as well as with the number of studies supporting each gene (supplementary table 1, available as supplementary data at Rheumatology Online). The strong LD at these genetic regions (supplementary figure 2, available as supplementary data at Rheumatology Online) makes it difficult to pinpoint the true causative SNPs responsible for the observed associations with SLE. This was evident also from our haplotype analysis, where statistical significance did not improve after taking into account specific combination of risk alleles. However, the IRF5-TNPO3 locus contains several variants suggested to be causal [2], including a promoter CGGGGG indel that was proposed to explain the association signal from all previously identified causal variants [13]. Interestingly, haplotype analysis of the IRF5-TNPO3 locus identified risk and protective haplotypes, which contain, respectively, also the risk (4×CGGGG) or protective (3×CGGGG) indel alleles (Table 2).

Concrete evidence of association with SLE was also obtained for the FAM167A-BLK locus, although it was not as strong as the association with IRF5-TNPO3 locus. This may be explained by the modest power for detecting association for this gene. One regulatory variant (rs13277113) at this locus, correlating with opposite effects on BLK and FAM167A mRNA expression, has been reported [7] and did show significant association also in the present study.

TNFAIP3, BANK1, PXK, KIAA1542 and 1q25.1 have reached genome-wide significance levels in a single GWAS (supplementary table 1, available as supplementary data at Rheumatology Online), and have been replicated in independent follow-up studies [15, 17]. The strongest evidence of a predisposing role was previously obtained for alleles at the TNFAIP3 locus [5, 14], and our results are consistent with this observation (Table 1), although low minor allele frequencies warrant some caution when interpreting the data. Interestingly, association with TNFAIP3 was recently confirmed in a Chinese Han population, thus further strengthening its global role in SLE susceptibility [18]. Only borderline significant association was obtained with SNPs in BANK1 and KIAA1542, whereas no significant signal was detected for PXK and 1q25.1. Two additional non-replications for PXK have been reported in Hong Kong Chinese [18, 19], suggesting a different effect between populations. However, minor allele frequencies differed considerably in our Finnish control individuals compared with other European populations analysed [6, 17] and, given the relatively low power to detect association for these loci, a Type 2 error cannot be excluded.

The genes ATG5, ICA1, XKR6, LYN and SCUBE1, for which we could not detect any significant signal, have shown only suggestive associations in a GWAS [6]. Out of these genes, an independent replication supports ATG5 [15, 18], ICA1 [15] and LYN [20]. Also for these genes, however, the power to detect association in our sample was very modest, and control frequencies differed from those previously reported for other European populations [6].

To date, two comprehensive replication studies of GWAS-identified SLE susceptibility genes have been
attempted in study populations of European descent [15, 17]. Exploration of multiple other populations is needed to verify the contribution of single loci to disease pathogenesis. It is known, for example, that STAT4 and BANK1 are global risk factors with similar effect size across populations, whereas ITGAM, PXK and LYN show discrepancy between populations [2]. In conclusion, the association of BANK1 controlling the activation of B cells, and BLK influencing their tolerance provides further evidence on the role of B-cell receptor signalling pathway in disease pathogenesis [2]. The association of upstream regulators of Type 1 IFN response

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP ID (major/ minor allele)</th>
<th>Risk allele</th>
<th>Case frequency</th>
<th>Control frequency</th>
<th>P-value (uncorrected) OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BANK1</td>
<td>rs10516487 (G/A)</td>
<td>G</td>
<td>0.73</td>
<td>0.68</td>
<td>0.04</td>
</tr>
<tr>
<td>FAM167A-BLK</td>
<td>rs2736340 (C/T)</td>
<td>T</td>
<td>0.30</td>
<td>0.24</td>
<td>0.01</td>
</tr>
<tr>
<td>FAM167A-BLK</td>
<td>rs13277113 (G/A)</td>
<td>A</td>
<td>0.30</td>
<td>0.23</td>
<td>0.01</td>
</tr>
<tr>
<td>FAM167A-BLK</td>
<td>rs4840568 (G/A)</td>
<td>A</td>
<td>0.31</td>
<td>0.24</td>
<td>4.00 x 10^{-3}</td>
</tr>
<tr>
<td>FAM167A-BLK</td>
<td>rs2618476 (T/C)</td>
<td>C</td>
<td>0.31</td>
<td>0.23</td>
<td>1.50 x 10^{-3}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs729302 (A/C)</td>
<td>A</td>
<td>0.76</td>
<td>0.62</td>
<td>2.06 x 10^{-7}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>CGGGG (3x4/C2)</td>
<td></td>
<td>0.58</td>
<td>0.43</td>
<td>3.67 x 10^{-7}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs3807306 (T/G)</td>
<td>T</td>
<td>0.62</td>
<td>0.50</td>
<td>2.37 x 10^{-5}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs2070197 (T/C)</td>
<td>C</td>
<td>0.24</td>
<td>0.15</td>
<td>2.00 x 10^{-4}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs10488631 (T/C)</td>
<td>C</td>
<td>0.23</td>
<td>0.15</td>
<td>2.00 x 10^{-4}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs2280714 (T/C)</td>
<td>T</td>
<td>0.78</td>
<td>0.71</td>
<td>0.01</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs12539741 (C/T)</td>
<td>T</td>
<td>0.24</td>
<td>0.15</td>
<td>1.00 x 10^{-4}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs10279821 (C/T)</td>
<td>C</td>
<td>0.79</td>
<td>0.72</td>
<td>3.10 x 10^{-3}</td>
</tr>
<tr>
<td>IRF5-TNPO3</td>
<td>rs12537284 (G/A)</td>
<td>A</td>
<td>0.23</td>
<td>0.17</td>
<td>8.50 x 10^{-3}</td>
</tr>
<tr>
<td>KIAA1542</td>
<td>rs4963128 (C/T)</td>
<td>C</td>
<td>0.74</td>
<td>0.68</td>
<td>0.02</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>rs10498197 (T/G)</td>
<td>G</td>
<td>0.04</td>
<td>0.01</td>
<td>6.60 x 10^{-3}</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>rs5029939 (C/G)</td>
<td>G</td>
<td>0.04</td>
<td>0.02</td>
<td>7.90 x 10^{-3}</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>rs2230926 (T/G)</td>
<td>G</td>
<td>0.04</td>
<td>0.02</td>
<td>4.50 x 10^{-3}</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>rs7749323 (G/A)</td>
<td>A</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Haplotype structures based on LD are shown in supplementary figure 2, available as supplementary data at Rheumatology Online. *SNP no. is as referred to in supplementary table 3, available as supplementary data at Rheumatology Online.
pathway, \textit{IRF5-TNPO3} and \textit{TNFAIP3}, highlights yet again the importance of this regulatory mechanism in the outbreak of SLE [2]. Our findings are only partially concordant with previous results, possibly due to insufficient power to detect associations or true population differences in allele frequencies. These findings underline once more the importance of assessing genetic variants in different populations, even within Europe, to conclusively define the genetic architecture of SLE and the magnitude of the effects of specific risk alleles in different populations.

\textbf{Rheumatology key messages}

- The number of SLE susceptibility genes has expanded to more than 30 in recent years.
- Non-European cohorts are needed to validate the worldwide role of recently identified predisposing genetic variants.
- Our results support the role of B-cell receptor pathway and IFN signalling in disease pathogenesis.

\textbf{Acknowledgements}

This study was completed before the untimely death of Professor Ulpu Saarialho-Kere, whose contribution to the study was substantial. We thank Professors Anna-Elina Lehesjoki and Albert de la Chapelle for providing control samples, and Dr Elisabet Einarsdóttir for statistical advice.

\textbf{Funding:} This study was supported by Finska Läkaresällskapet, Academy of Finland, Sigrid Jusélius Foundation, Helsinki University Central Hospital Research Fund (TYH 2008248), Finland, and the Swedish Research Council, Sweden. Personal support was provided by Helsinki Biomedical Graduate School LERU PhD Program in Biomedicine, Maire Lisko Foundation, Biomedicum Helsinki Foundation and University of Helsinki Funds to T.M.J., and Medical Research Fund of Tampere University Hospital to T.H., Finland.

\textbf{Disclosure statement:} The authors have declared no conflicts of interest.

\textbf{Supplementary data}

Supplementary data are available at \textit{Rheumatology} Online.

\textbf{References}