Mini Review

Role of *IRF5*, *STAT4* and *BLK* polymorphisms for the genetic predisposition to systemic lupus erythematosus in Japanese

Naoyuki Tsuchiya *, Aya Kawasaki, and Ikue Ito

Molecular and Genetic Epidemiology Laboratory, Doctoral Program in Life System Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Tsukuba, Japan

Recent large-scale association studies revealed new susceptibility genes to systemic lupus erythematosus (SLE) including *IRF5*, *STAT4* and *BLK*. Association of these genes have been quickly replicated by many studies in multiple populations. In this minireview, we discuss our recent studies on the association of these genes with SLE in Japanese. Although association of these genes was replicated, notable differences were observed between Caucasian and Japanese populations.

In Japanese, *IRF5* risk haplotype in the Caucasians carrying three functional polymorphisms (a single nucleotide polymorphism [SNP] at exon 1B splice site, 10 amino acid insertion/deletion, a SNP at poly A signal) is almost absent. However, another intron 1 SNP, which was not described in the Caucasians, was significantly associated with SLE in Japanese.

On the other hand, both the *STAT4* intronic SNPs and *C8orf13-BLK* intergenic SNPs associated in Caucasians were similarly associated with SLE in Japanese. Moreover, because of higher population frequencies of the risk genotypes and higher odds ratios, the contribution of these genes appeared to be greater in the Japanese than in the Caucasians.

Although association of these genes with SLE is established, the molecular mechanisms of the association remain largely unknown. In addition, further studies are required to develop applications of the genetics information for clinical use. Genetics finally began to reveal new and reliable clues to gain insight into the pathogenesis of highly complicated disorders such as SLE.

Rec.12/19/2008, Acc.1/22/2009, pp190-197

*Correspondence should be addressed to:*
Naoyuki Tsuchiya, MD, PhD, Doctoral Program in Life System Medical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba 305-8575, Japan. Phone/FAX.+81-29-853-3071, e-mail: tsuchiya@md.tsukuba.ac.jp

**Key words** *BLK, IRF5, polymorphism, STAT4, systemic lupus erythematosus*

Introduction

Epidemiological data strongly implicate significant contribution of genetic background in the development of systemic lupus erythematosus (SLE). *HLA-DRB1, C4, FCGR2A, 2B, 3A and 3B* polymorphisms have been repeatedly associated with susceptibility to SLE. However, a number of other susceptibility
genes remain to be discovered.

Since 2005, a number of exciting new discoveries have been made in the field of lupus genetics. A candidate pathway approach focusing on type I interferon (IFN) pathway identified association of interferon regulatory factor 5 (IRF5) polymorphisms with SLE\(^6\). Subsequently, signal transducer and activator of transcription 4 (STAT4), identified as a susceptibility gene to rheumatoid arthritis (RA), has been shown to be associated with SLE at the same time\(^8\).

In parallel with these discoveries, genome-wide association studies revealed new susceptibility genes such as B lymphoid tyrosine kinase (BLK), integrin alpha M (ITGAM) and tumor necrosis factor alpha-induced protein 3 (TNFAIP3), in addition to confirming the strong association of IRF5 and STAT4\(^{10}\). Importantly, association of these genes with SLE has been replicated in most of the replication studies.

In this minireview, we discuss our findings on the association of IRF5\(^9\), STAT4\(^{10}\) and BLK\(^{11}\) polymorphisms with SLE in Japanese, focusing on the similarities and differences as compared with Caucasians.

**IRF5**

Type I IFN family includes at least 13 IFN\(_\alpha\), as well as IFN\(_\beta\), IFN\(_\kappa\), IFN\(_\tau\) and IFN\(_\omega\). Type I IFN has been heavily implicated in SLE and autoimmune diseases (reviewed in 12). Serum IFN\(_\alpha\) has been reported to be elevated in SLE. Treatment of viral hepatitis with type I IFN sometimes induces antinuclear antibodies and symptoms of autoimmune diseases. Furthermore, mRNA expression profiling on peripheral blood cells from SLE unanimously reported overexpression of genes induced by type I IFN (“interferon signature”). These studies established crucial role of type I IFN in the pathogenesis of SLE.

These lines of evidence led the researchers to examine the contribution of type I IFN pathway-related gene polymorphisms to the genetic susceptibility to SLE. The initial study in a Swedish population identified an association of a single nucleotide polymorphism (SNP) rs2004640 in intron 1 of IRF5 with SLE\(^9\). Subsequently, this association was replicated in multiple large-scale studies in Caucasians\(^{13-16}\), African-Americans\(^{17}\) and Asians\(^{18}\).

IRF5 is a transcription factor constitutively expressed in lymphocytes and dendritic cells (DCs), but is induced in other cells by viral infection and IFN\(_\alpha\)\(^{19}\). Upon stimulation of TLR7/8/9 or viral infection, IRF5 becomes activated, dimerizes and translocates to nucleus, where it induces proinflammatory cytokines such as TNF\(_\alpha\) IL-12 and IL-6, as well as type I IFN\(^{12}\). IRF5 gene is encoded on chromosome 7q32.

During the course of association studies, it was revealed that IRF5 contains at least 3 polymorphisms of direct functional significance (Fig.1). IRF5 has more than 10 alternative isoforms, each contains one of exon 1A, 1B, 1C or 1D. The intron 1 SNP rs2004640 creates a splice donor site, and exon 1B is used only in the mRNA transcribed from the rs2004640T allele\(^{13}\). Secondly, there is an insertion/deletion (indel) polymorphism in exon 6 that results in indel of 10 amino acids in the PEST domain\(^{14-16}\).

The functional significance of this indel is not yet clear. Thirdly, a SNP in 3’-untranslated region (3’UTR), rs10954213A>G, abol-
ishes the poly A signal. The G allele leads to the usage of the second poly A signal located 648 bp downstream, resulting in longer mRNA and reduced mRNA level\textsuperscript{14-16}. The risk haplotype in Caucasians contains rs2004640T (exon 1B usage), exon 6 insertion and rs10954213A (elevated mRNA level)\textsuperscript{14-16}.

To examine the role of IRF5 polymorphisms in Japanese SLE, we carried out an association study on 277 patients and 201 controls\textsuperscript{9}. The intron 1 rs2004640T showed a tendency of association also in Japanese, but both the odds ratio (OR 1.24, 95% confidence interval [CI] 0.94-1.64, \( p=0.124 \)) and the allele frequency in the controls (0.301) were smaller in Japanese as compared with the Caucasians (OR 1.47, allele frequency 0.51)\textsuperscript{10}. On the other hand, OR and allele frequency in Japanese were very similar to those in a Korean population\textsuperscript{18}.

In contrast, 3 SNPs closely located to rs2004640 were identified, which had not previously been described in Caucasians (Fig.1). Interestingly, those SNPs showed stronger association with SLE than rs2004640 in Japanese. Specifically, rs41298401, located 6 bp downstream to rs2004640, showed the most significant protective association with Japanese SLE (G allele frequency, OR 0.65, 95%CI 0.46-0.93, \( p=0.017 \)). Another SNP, rs6953165, demonstrated positive association with SLE (allele frequency, OR 1.76, 95%CI 1.04-2.97, \( p=0.034 \)).

In contrast, exon 6 indel and poly A site SNP (rs10954213) did not show association with SLE in Japanese.

Haplotype analysis revealed considerable difference in the haplotype structure in Japanese and Caucasian populations. In Japanese, exon 6 insertion was in almost complete linkage disequilibrium (LD) with rs10954213G allele associated with low expression of IRF5 mRNA (\( r^2=0.98 \)); thus, the risk haplotype containing both the exon 6 insertion and rs10954213A allele was almost absent in Japanese. Such a difference in the haplotype structure is likely to be associated with the difference in the associated SNP between populations. Nevertheless, the association between IRF5 and SLE in both populations despite such a difference supports the crucial role of IRF5 in the development of SLE.

To elucidate the functional significance of intron 1 SNPs, we employed mRNA expression profile data of the lymphoblastoid B cell lines established from donors of the International HapMap project, deposited in GENEVAR database (Wellcome Trust Sanger Institute, http://www.sanger.ac.uk/humgen/genevar/). By utilizing these data in combination with the HapMap genotype data, one can statistically analyze association between any genotype and mRNA level of any gene as long as it is expressed in B cells. Actually, rs10954213 genotype was shown to be strongly correlated with mRNA level of IRF5\textsuperscript{19}, as was experimentally demonstrated\textsuperscript{14,15}. In addition, this genotype was also significantly correlated with a number of genes induced by IRF5\textsuperscript{19}, suggesting that this method can detect not only direct cis-acting effects, but also indirect trans-acting effects of SNPs on gene expression.

Using this approach, we examined association of intron 1 SNP genotypes with 31 genes induced by IRF5. The risk genotype of rs41298401 was positively associated with expression level of IFN\( \alpha \), STAT3, STAT5B and TMPO (thymopoietin), and negatively with IFNA10. Although the significance of these genes in relation to the molecular mechanisms of SLE is at this point unclear, these findings suggested that rs41298401 or other polymorphism which is in linkage disequilibrium with this SNP has an effect on the expression levels of some of the IRF5 inducible genes\textsuperscript{9}.

Recent studies in other populations have reported genetic effects of other IRF5 polymorphisms, such as rs729302 in the 5'-flanking region\textsuperscript{20}, CGGGG indel\textsuperscript{21} and rs3807306 in intron 1\textsuperscript{22}, and rs10488631 in the 3'-flanking region\textsuperscript{20,21}. Furthermore, association of IRF5 with other autoimmune or inflammatory diseases such as RA\textsuperscript{23}, Sjogren syndrome\textsuperscript{24} and inflammatory bowel diseases\textsuperscript{25} has been reported. Thus, IRF5 is undoubtedly an established susceptibility gene to SLE and other inflammatory diseases, but the causative SNP may vary among populations, and the molecular mechanism of association requires further study.

**STAT4**

STAT4, located at 2q32.2-q32.3, has recently been identified as a shared susceptibility gene to RA and SLE in Caucasians\textsuperscript{6}. Association with SLE was confirmed by two genome-wide association studies in Caucasians\textsuperscript{6,7}, and studies focused on the STAT4 in Caucasians\textsuperscript{26-29}, Colombians\textsuperscript{29} and Japanese\textsuperscript{30}.

STAT4 is a transcription factor expressed in lymphocytes, macrophages, and dendritic cells. STAT4 is essential for IL-12 signaling and induces IFN\( \gamma \) production and Th1 differentiation\textsuperscript{31}. STAT4 is also activated by type I IFNs\textsuperscript{32}. Furthermore, a recent study suggested the role of STAT4 in IL-23-induced IL-17 production\textsuperscript{33}. Based on these findings, STAT4 is considered an attractive candidate susceptibility gene to diseases rheumatic or autoimmune diseases.

**STAT1** gene is located adjacently to **STAT4** at 2q32.2. STAT1 is activated by type I IFNs and IFN\( \gamma \)\textsuperscript{34}. Moreover, STAT1 has been reported to be upregulated in peripheral blood mononuclear cells from SLE patients and in kidneys of lupus mice with nephritis\textsuperscript{35,36}. These observations suggest that **STAT1** is also a strong candidate susceptibility gene to SLE.
Table 1  Population attributable risk percent (PAR%) of SLE susceptibility alleles

<table>
<thead>
<tr>
<th>gene</th>
<th>allele</th>
<th>population</th>
<th>model</th>
<th>population frequency of the risk genotype</th>
<th>OR of the risk genotype</th>
<th>PAR%</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT4</td>
<td>rs7574865T</td>
<td>Japanese</td>
<td>dominant</td>
<td>0.565</td>
<td>2.19</td>
<td>40.2%</td>
<td>10</td>
</tr>
<tr>
<td>C8orf13-BLK</td>
<td>rs13277113A</td>
<td>Japanese</td>
<td>recessive</td>
<td>0.432</td>
<td>2.27</td>
<td>35.4%</td>
<td>11</td>
</tr>
<tr>
<td>IRF5</td>
<td>rs41268401C</td>
<td>Japanese</td>
<td>recessive</td>
<td>0.652</td>
<td>1.55</td>
<td>26.4%</td>
<td>9</td>
</tr>
<tr>
<td>HLA-DRB1</td>
<td>DRB1*1501</td>
<td>Japanese</td>
<td>dominant</td>
<td>0.124</td>
<td>2.97</td>
<td>19.6%</td>
<td>39</td>
</tr>
<tr>
<td>FCGR2B</td>
<td>rs1050501C</td>
<td>Japanese</td>
<td>recessive</td>
<td>0.053</td>
<td>2.19</td>
<td>5.9%</td>
<td>40</td>
</tr>
<tr>
<td>TNFRSF1B (TNFR2)</td>
<td>rs80195847G</td>
<td>Japanese</td>
<td>dominant</td>
<td>0.188</td>
<td>2.53</td>
<td>22.4%</td>
<td>41</td>
</tr>
<tr>
<td>TNFSF13 (APRIL)</td>
<td>rs11552708G</td>
<td>Japanese</td>
<td>dominant</td>
<td>0.803</td>
<td>2.01</td>
<td>44.7%</td>
<td>42</td>
</tr>
<tr>
<td>STAT4</td>
<td>rs7574865T</td>
<td>Caucasians</td>
<td>dominant</td>
<td>0.412</td>
<td>1.59</td>
<td>19.5%</td>
<td>6</td>
</tr>
<tr>
<td>C8orf13-BLK</td>
<td>rs13277113A</td>
<td>Caucasians</td>
<td>dominant</td>
<td>0.406</td>
<td>1.48</td>
<td>16.2%</td>
<td>10</td>
</tr>
</tbody>
</table>

PAR% were calculated based on the model (dominant or recessive) which gave a smaller $P$ value by $\chi^2$ test using 2 x 2 contingency tables for each allele.

In order to comprehensively examine the role of STAT1-STAT4 region for SLE, we selected 52 tag SNPs encompassing this region, and carried out an association study in Japanese\(^6\). Among the tag SNPs, rs10168266 in intron 5 as well as rs11889341 and rs7574865 in intron 3 were most significantly associated with SLE. In contrast, significant association was not detected for SNPs in the STAT1 region.

The rs7574865T allele, previously shown to be associated with SLE in Caucasians, was significantly increased in Japanese SLE (0.463) compared with controls (0.335, $p=4.9 \times 10^4$, OR 1.71, 95% CI 1.36-2.15). The association was compatible with the dominant model, under which the OR was 2.19 (T/T + G/T versus G/G). The SNPs rs11889341 and rs10168266 were in LD with rs7574865 and were also significantly associated with SLE. Logistic regression analysis failed to identify a single causative SNP of the three due to the strong LD.

Association of these SNPs was more strongly observed in SLE patients with nephritis or with anti-dsDNA antibodies both in the Caucasians\(^6,7\) and in the Japanese\(^6\).

The functional significance of these intronic SNPs remains unclear, but recent studies indicated that the risk genotype is associated with elevated mRNA level of STAT4\(^23,27\), but not with splicing isoform\(^26\).

C8orf13-BLK region

C8orf13-BLK region at 8p23.1 is a recently identified susceptibility region for SLE by two genome-wide association studies in Caucasian populations\(^6,7\). BLK encodes a B lymphoid specific tyrosine kinase of Src family\(^8\), whose function remains unclear. C8orf13 is a ubiquitously expressed gene, the function of which also remains unknown. The SLE-associated SNP, rs13277113, is located at the intergenic region between these genes, and the risk allele has been shown to be associated with low mRNA levels of BLK and high mRNA levels of C8orf13\(^6\).

To test whether this region is also associated with SLE in a Japanese population, we carried out an association study for 14 tag SNPs in this region. Eleven of the 14 SNPs exhibited evidence for association, among which rs13277113 showed the strongest association (allele frequency $p=4.75 \times 10^7$, OR 2.44, 95%CI 1.43-4.16). Most of the effects of other SNPs could be accounted for by LD with rs13277113. Thus, C8orf13-BLK region appears to contain shared susceptibility factor at least between Caucasians and Japanese\(^6\).

Contribution of each risk allele in different populations

Contribution of each susceptibility gene to diseases substantially varies among populations, at least partly because of differences in the allele frequencies, haplotype structure and possibly in the interacting genetic and environmental factors. To compare the impact of each risk allele in each population, we estimated population attributable risk percent (PAR%) of IRF5, STAT4, BLK as well as other previously established susceptibility genes to SLE in Japanese, such as HLA-DRB1*1501\(^9\), FCGR2B\(^14,15\), TNFRSF1B (TNFR2)\(^21\), and TNFSF13 (APRIL)\(^22\). PAR% is “used to estimate the excess rate of disease in the total study population of exposed and nonexposed that is attributable to the exposure”\(^23\), which, in this case, is the risk genotype. In a case-control study of a rare disease like SLE, PAR% can be approximated using the OR and the risk genotype frequency in the healthy
controls. The sum of PAR% for all risk alleles exceeds 100%, because each individual carries multiple risk alleles in a multifactorial disease like SLE.

As shown in Table 1, PAR% of STAT4 and BLK were considerably greater in the Japanese compared with the Caucasians. This is because both the OR and population frequency of the risk genotype are greater in Japanese.[6,11] Furthermore, PAR% of STAT4 and BLK are also greater than those of most of the other susceptibility genes in the Japanese.

Because PAR% may be affected by the difference in the method of ascertainment of each study, comparison of PAR% among different studies needs to be interpreted with caution. Nevertheless, these observations suggested substantial impact of these susceptibility genes on the development of SLE in Japanese.

Concluding remarks
Most of the susceptibility genes recently identified through large-scale studies have been successfully replicated in multiple populations. Furthermore, although pathways involving type I IFN, Th1/Th2 regulation and B cell receptor signaling had been heavily implicated, molecules such as IRF5, STAT4 and BLK had not been investigated in relation to the pathogenesis of SLE until the genetic association became evident. Thus, human genetics is finally beginning to disclose many new and solid paths to the understanding of the etiopathogenesis of lupus.

Our current findings emphasized that even for the susceptibility genes shared by multiple populations, difference in the associated risk allele or the degree of risk conferred by the allele should be taken into account for medical application.

The molecular mechanisms of association of these genes as well as the specific strategy how to utilize the genetics information for the diagnosis, treatment and prevention of lupus will require years of intensive research. At this point, however, it is fair to say that the lupus genetics research is on the right track.

Acknowledgement
This work was supported by Grant-in-Aid for Scientific Research (B) from Japan Society for the Promotion of Science (JSPS) and grants from the Ministry of Health, Labour and Welfare of Japan, Japan Rheumatism Foundation and The Naito Foundation.

References


40) Kyogoku C, Dijstelbloem HM, Tsuchiya N, Hatta Y, Kato

