Regulatory polymorphisms in EGR2 are associated with susceptibility to systemic lupus erythematosus

Keiko Myouzen¹, Yuta Kochi¹,*, Kenichi Shimane¹,2, Keishi Fujio², Tomohisa Okamura², Yukinori Okada³, Akari Suzuki¹, Tatsuya Atsumi⁵, Satoshi Ito⁶, Kazuki Takada⁷, Akio Mimori⁸, Shiro Ikegawa⁴, Ryo Yamada⁹, Yusuke Nakamura¹⁰ and Kazuhiko Yamamoto¹,²

¹Laboratory for Autoimmune Diseases, Center for Genomic Medicine (CGM), RIKEN, Tokyo 113-0033, Japan, ²Department of Allergy and Rheumatology, Graduation School of Medicine, The University of Tokyo, Tokyo 113-0033, Japan, ³Laboratory for Statistical Analysis and ⁴Laboratory for Bone and Joint Diseases, CGM, RIKEN, Tokyo 106-8639, Japan, ⁵Department of Medicine II, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan, ⁶Division of Clinical Immunology, Doctoral Program in Clinical Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba, Ibaraki 305-8575, Japan, ⁷Department of Medicine and Rheumatology, Graduate School, Tokyo Medical and Dental University, Tokyo 113-8519, Japan, ⁸Division of Rheumatic Diseases, International Medical Center of Japan, Tokyo 162-8655, Japan, ⁹Statistical Genetics, Center for Genomic Medicine, Graduate School of Medicine Kyoto University, Kyoto 606-8501, Japan and ¹⁰Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo 108-8639, Japan

Received October 31, 2009; Revised and Accepted February 24, 2010

Systemic lupus erythematosus (SLE) is an autoimmune disease induced by the combinations of environmental and genetic factors. Recently, mice in which the early growth response 2 (EGR2) gene, a zinc-finger transcription factor, is conditionally knocked out in CD2¹ T cells have been shown to develop a lupus-like autoimmune disease. Here, we evaluated if polymorphisms in the EGR2 gene influence SLE susceptibility in humans. We first analyzed the effect of SNPs in the EGR2 region on EGR2 expression, and a significant positive correlation with expression was identified in an SNP located at the 5′ flanking region of EGR2 (rs10761670, R = 0.23, P = 0.00072). We then performed a case–control association study using three sets of SLE cohorts by genotyping 14 tag SNPs in the EGR2 gene region. A peak of association with SLE susceptibility was observed for rs10761670 [Pooled: OR = 1.23 (95% CI 1.10–1.37), P = 0.00023]. This SNP was also associated with susceptibility to rheumatoid arthritis (RA) [OR = 1.15 (95% CI 1.05–1.26), P = 0.0019], suggesting that EGR2 is a common risk factor for SLE and RA. Among the SNPs in complete linkage disequilibrium with rs10761670 (r² = 1.0), two SNPs (rs1412554 and rs1509957) affected the binding of transcription factors and transcriptional activity in vitro, suggesting that they may be candidates of causal regulatory variants in this region. Therefore, EGR2 is a genetic risk factor for SLE, in which increased gene expression may contribute to SLE pathogenesis.

INTRODUCTION

Systemic lupus erythematosus (SLE, OMIM # 152700) is an autoimmune disease induced by the combinations of environmental and genetic factors. This disease is characterized by antinuclear autoantibodies, complement activation, hyperproduction of interferon and tissue destruction (1). Genetic studies using a candidate gene approach identified several SLE susceptibility genes, including HLA-DRB1, FCGR2B/3A/3B, PTPN22, STAT4 and IRF5 (2–7). In addition, recent genome-wide association studies (GWASs) have uncovered novel SLE susceptibility genes, including TNFAIP3, BANK1, ITGAM, PXK, KIAA1542 and C8orf13-BLK (8–11). Among these genes, genes such as PTPN22, STAT4, TNFAIP3 and

*To whom correspondence should be addressed at: Laboratory for Autoimmune Diseases, Center for Genomic Medicine, RIKEN, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Tel/Fax: +81 358008828; Email: ykochi@src.riken.jp

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
IRF5 are also associated with susceptibility to other autoimmune diseases, suggesting that some genetic dispositions are shared between SLE and other autoimmune diseases (7,12–14). These GWASs for SLE have enabled a comprehensive survey of the genome, and the statistical power of individual GWASs has been increasing as growing number of samples have been genotyped. However, a recent large-scale replication study that examined the suggested loci of GWAS demonstrated that the 26 confirmed risk loci only explain an estimated 8% of the total genetic susceptibility to SLE (15). This implicated many genes remained undiscovered probably due to their moderate risk. Therefore, the candidate gene approach is still a complementary tool for the identification of unknown genes with moderate risk that contribute to SLE susceptibility.

The early growth response 2 (Egr2) gene in mice is a member of the zinc-finger transcription factor Egr family (Egr1, 2, 3 and 4) that are expressed during thymic T-cell differentiation (16), and plays an essential role in hindbrain development and myelination of the peripheral nervous system (17). Whereas Egr4 is constitutively expressed in T cells, Egr1, 2, and 3 are up-regulated by T-cell receptor (TCR) engagement. Opposing functions of the Egrs have been described, where Egr1 enhances T-cell function by up-regulating IL-2 and other molecules that stimulate T cells (18). On the other hand, Egr2 and Egr3 are considered to negatively regulate T cells, which are demonstrated in a study using Egr2 and Egr3-deficient T cells (19). In that study, it was proposed that Egr2 and Egr3, which are located in the downstream of transcription factor NF-AT signaling, inhibited T-cell function by both up-regulating negative regulators such as Cbl-b and inhibiting the expression of T-cell activators such as Egr1 and NAB2. Interestingly, mice whose Egr2 gene was conditionally knocked out in CD25+ T cells were shown to develop a lupus-like autoimmune disease, characterized by accumulation of interferon-γ and interleukin-17-producing CD4+ T cells, loss of tolerance to nuclear antigens, massive infiltration of T cells into multiple organs and glomerulonephritis (20). More recently, we described a new subset of IL-10-secreting regulatory T cells, termed CD4+CD25+LAG3+ regulatory T cells, that characteristically express Egr2 (21). Because Egr2-transduced naïve CD4+ T cells differentiated into IL-10-secreting CD4+CD25-LAG3+ regulatory T cells, Egr2 was considered to be a key transcription factor for the differentiation of these cells. A significant role in intestinal immunity was suggested for these cells by the observation that the frequency of CD4+CD25-LAG3+ regulatory T cells in the CD4+CD25- T-cell population was comparatively higher in Peyer’s patch than in the spleen, and that experimentally induced colitis was effectively protected by these regulatory T cells.

In humans, chromosome 10q21, within which EGR2 is located, has been identified as a candidate locus for Crohn’s disease (CD) susceptibility in two independent GWASs (22,23). The landmark SNPs identified in these GWASs are located in the intergenic region between the ZNF365 and the EGR2 genes. T cells have been assumed to play a major role in the pathogenesis of CD, as an excess of T cells (both effector and regulatory T cells) were detected in the intestinal mucosa of patients (24,25). Indeed, genetic studies of CD identified several susceptibility genes that are involved in the differentiation of T cells, which further supports the significance of T cells for CD (25). Therefore, given the effects of Egr2 on murine T cells and the role of CD4+CD25-LAG3+Egr2+ regulatory T cells in the colitis model, EGR2 is considered to be a strong candidate gene for CD susceptibility at this locus.

The described evidence derived from the murine disease models and from human genetic studies implies the existence of disease causal variations in the EGR2 gene region, which might be shared between SLE and CD in humans. In the present study, we first searched for functional variants that may affect gene function or gene expression of EGR2, and then performed case–control association tests to examine the contribution of EGR2 to SLE susceptibility.

**RESULTS**

**Correlation between EGR2 expression and SNP genotypes**

To identify variants which potentially increase the risk of disease in the EGR2 gene region, we first searched for functional variants that might alter the amino acid sequence, or affect the expression level, of EGR2. The coding region of EGR2 was sequenced using DNA from 96 individuals affected with SLE (the power for detecting rare variants with frequency of 0.01 was estimated to be 0.86). In this sequencing analysis, no variation leading to amino acid substitution in EGR2 was discovered.

We then examined possible correlations between EGR2 expression and genotype. As EGR2 has been reported to be expressed in B cells as well as in T cells (26,27), we analyzed a data set of gene expression in Epstein-Barr virus (EBV)-transformed lymphoblastoid B cell lines from HapMap individuals (28). Analysis of the genotype data of Japanese HapMap samples identified three LD blocks in a Japanese population in an 80-kb region that includes the EGR2 gene on 10q21.3 (Fig. 1A). The correlation between SNP genotype and the expression of EGR2 in this region was assessed for data sets of three individual HapMap ethnic groups (JPT + CHB, CEU and YRI). The observation of a common peak of correlation in the 5′ flanking region of EGR2 among these three groups implied common regulatory factors for EGR2 expression in this chromosomal region (Fig. 1B). We then selected 69 SNPs from this 80-kb region that are common (minor allele frequency; MAF > 0.10) to all three HapMap ethnic groups, and assessed the correlation between their genotypes and the expression of EGR2 using the pooled data set of all populations (n = 210) (Fig. 1B). Rs10761670 (T/A) in the 5′ flanking region of EGR2 showed the highest correlation (R = 0.23). When the expression level of EGR2 was regressed with the number of alleles of rs10761670, the correlation was statistically significant (P = 0.00072, Fig. 1C). This correlation suggested that variations existed in the 5′ flanking region of EGR2 that affect gene expression, which may potentially increase the risk of disease.

**Case–control association study of the EGR2 gene region**

To evaluate the association between EGR2 polymorphisms and SLE susceptibility, we performed a case–control
association study in the EGR2 region. Tag SNPs were selected from the HapMap database for association tests. We selected a total of 14 tag SNPs with MAF > 0.10 from the LD blocks described above. Nine of these tag SNPs were from block 3, which spanned the 5′ region of EGR2 and contained the potential regulatory variants described above, with a threshold value of \( r^2 > 0.8 \), and the other five SNPs were from blocks 1 and 2 with a threshold value of \( r^2 > 0.5 \) (Table 1). We initially genotyped 376 SLE cases for the tag SNPs and compared their allele frequencies with those of 940 control individuals. We identified two SNPs (rs10761670 and rs955696) which showed significant association with SLE susceptibility [rs10761670: OR = 1.19 (95% CI 1.00–1.41), \( P = 0.049 \) and rs955696: OR = 1.23 (95% CI 1.02–1.48), \( P = 0.033 \), respectively, Table 1]. These SNPs are located in block 3 in the 5′ region of EGR2. To validate the case–control association test, we performed a replication study for these two SNPs using additional two sets of SLE cohorts (second set: 293 SLE cases and 881 controls; third set: 223 SLE cases and 658 controls). The association was replicated only for rs10761670 [second set: OR = 1.23 (95% CI 1.02–1.49), \( P = 0.029 \) and third set: OR = 1.30 (95% CI 1.05–1.62), \( P = 0.018 \), Table 2], although the possibility remained that the negative results in rs955696 was due to insufficient statistical power of the cohort sets (0.62 and 0.52 for second and third sets, respectively). No heterogeneity was detected among the associations of three cohorts (\( P = 0.83 \) in a Mantel–Haenszel analysis), and the combined analyses showed a statistically significant association of rs10761670 with SLE susceptibility [\( OR = 1.23 \) (95% CI 1.10–1.37), \( P = 0.00023 \), Table 2]. Both dominant and recessive models showed significant associations in rs10761670 in the pooled set [dominant model: \( OR = 1.33 \) (95% CI 1.09–1.62), \( P = 0.0045 \) and recessive model: \( OR = 1.32 \) (95% CI 1.11–1.57), \( P = 0.0016 \)]. After adjustment for gender by a logistic regression analysis, rs10761670 still showed a significant association with SLE susceptibility in the combined set [882 SLE cases and 2467 controls, \( OR = 1.23 \) (95% CI 1.09–1.38), \( P = 0.00070 \)]. As previous studies have demonstrated that multiple genes, including PTPN22, STAT4, TNFAIP3 and CD244 (29), increase the risk of both SLE and rheumatoid arthritis (RA), we examined the association of EGR2 with RA susceptibility using two sets of RA cohorts (first set, 1112 RA cases and 940 controls; second set, 830 RA cases and 881 controls). A significant association was observed in the first set [\( OR = 1.21 \) (95% CI 1.07–1.37), \( P = 0.0020 \)], but we did not identify a significant association in the second set [\( OR = 1.09 \) (95% CI 0.95–1.24), \( P = 0.22 \)]. The combined analyses showed a statistically significant association with RA susceptibility [\( OR = 1.15 \) (95% CI 1.05–1.26), \( P = 0.0019 \)] (Table 3). These results suggested that the EGR2 polymorphism increased the risk of RA as well as of SLE.

### Search for regulatory SNPs

To identify causal variants, we examined whether nuclear transcription factors bind to genomic sequences surrounding the SNPs and compared the binding ability of susceptible alleles with that of non-susceptible alleles. We selected 15 SNPs that are in complete linkage disequilibrium (LD) with rs10761670 (\( r^2 = 1.0 \)) using the genotype data of Japanese HapMap samples. These SNPs were examined by electrophoretic mobility shift assay (EMSA) using nuclear extracts from an EBV-transformed lymphoblastoid cell line (PSC cells). We found allelic differences in the binding ability of two
SNPs, rs1412554 and rs1509957. For rs1412554, the intensity of the shifted band was higher for the susceptible allele than for the non-susceptible allele. In contrast, a particular band was seen specifically to the non-susceptible allele for rs1509957 (Fig. 2A). Competition assays with unlabeled oligonucleotides showed that these complexes were specific for each oligonucleotide. Similar results were obtained when nuclear extracts of Jurkat cells were tested (Fig. 2A).

The transcriptional activity of rs1412554 and rs1509957 was then analyzed using luciferase assays. Allele specific constructs containing the SNPs and surrounding genomic sequences were constructed and transfected into Jurkat cells. Cells transfected with the susceptible allele of rs1412554 displayed a 1.2-fold greater enhancement of transcriptional activity than cells transfected with the non-susceptible allele (Fig. 2B). In contrast, only cells transfected with the

### Table 1. Association analysis of EGR2 with SLE susceptibility

<table>
<thead>
<tr>
<th>dbSNP ID</th>
<th>Allele (1/2)</th>
<th>LD blocks</th>
<th>Number of subjects</th>
<th>Frequency of allele 1</th>
<th>Odds ratio (95%CI)</th>
<th>P-value (allele)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10995312</td>
<td>C/T</td>
<td>Block 1</td>
<td>376</td>
<td>940</td>
<td>0.54</td>
<td>0.51</td>
</tr>
<tr>
<td>rs224278</td>
<td>A/G</td>
<td>Block 2</td>
<td>375</td>
<td>933</td>
<td>0.64</td>
<td>0.63</td>
</tr>
<tr>
<td>rs224285</td>
<td>A/C</td>
<td></td>
<td>374</td>
<td>934</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td>rs224292</td>
<td>G/A</td>
<td></td>
<td>376</td>
<td>934</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>rs11817939</td>
<td>A/G</td>
<td></td>
<td>374</td>
<td>934</td>
<td>0.89</td>
<td>0.87</td>
</tr>
<tr>
<td>rs2136613</td>
<td>A/G</td>
<td>Block 3</td>
<td>376</td>
<td>934</td>
<td>0.36</td>
<td>0.35</td>
</tr>
<tr>
<td>rs7078554</td>
<td>A/T</td>
<td></td>
<td>374</td>
<td>939</td>
<td>0.24</td>
<td>0.23</td>
</tr>
<tr>
<td>rs224307</td>
<td>G/A</td>
<td></td>
<td>375</td>
<td>934</td>
<td>0.66</td>
<td>0.63</td>
</tr>
<tr>
<td>rs10761670</td>
<td>T/A</td>
<td></td>
<td>375</td>
<td>939</td>
<td>0.55</td>
<td>0.51</td>
</tr>
<tr>
<td>rs10995335</td>
<td>G/A</td>
<td></td>
<td>375</td>
<td>940</td>
<td>0.90</td>
<td>0.88</td>
</tr>
<tr>
<td>rs10995337</td>
<td>G/A</td>
<td></td>
<td>374</td>
<td>939</td>
<td>0.91</td>
<td>0.89</td>
</tr>
<tr>
<td>rs3939306</td>
<td>T/C</td>
<td></td>
<td>376</td>
<td>940</td>
<td>0.27</td>
<td>0.26</td>
</tr>
<tr>
<td>rs949566</td>
<td>C/G</td>
<td></td>
<td>376</td>
<td>940</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>rs955696</td>
<td>C/T</td>
<td></td>
<td>376</td>
<td>938</td>
<td>0.30</td>
<td>0.26</td>
</tr>
</tbody>
</table>

aP-value (allele) was calculated using an allele frequency comparison test.
b95% CI = 95% confidence interval.

### Table 2. Replication study and combined analysis of EGR2 with SLE susceptibility

<table>
<thead>
<tr>
<th>dbSNP ID</th>
<th>Allele (1/2)</th>
<th>Case–control cohorts</th>
<th>Number of subjects</th>
<th>Frequency of allele 1</th>
<th>Odds ratio (95%CI)</th>
<th>P-value (allele)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10761670</td>
<td>T/A</td>
<td>2nd</td>
<td>287</td>
<td>876</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3rd</td>
<td>220</td>
<td>652</td>
<td>0.57</td>
<td>0.51</td>
</tr>
<tr>
<td>rs955696</td>
<td>C/T</td>
<td>Combined analysisa</td>
<td>882</td>
<td>2467</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>290</td>
<td>873</td>
<td>0.28</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3rd</td>
<td>223</td>
<td>654</td>
<td>0.31</td>
<td>0.27</td>
</tr>
<tr>
<td>rs10761670</td>
<td>T/A</td>
<td>Combined analysisc</td>
<td>889</td>
<td>2465</td>
<td>0.30</td>
<td>0.26</td>
</tr>
</tbody>
</table>

aP-value (allele) was calculated using an allele frequency comparison test.
b95% CI = 95% confidence interval.
cThe Mantel–Haenszel method was used for the combined analysis of first, second and third sets.

### Table 3. Association analysis of EGR2 with RA susceptibility

<table>
<thead>
<tr>
<th>dbSNP ID</th>
<th>Allele (1/2)</th>
<th>Case–control cohorts</th>
<th>Number of subjects</th>
<th>Frequency of allele 1</th>
<th>Odds ratio (95%CI)</th>
<th>P-value (allele)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs10761670</td>
<td>T/A</td>
<td>1st</td>
<td>1105</td>
<td>939</td>
<td>0.56</td>
<td>0.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2nd</td>
<td>827</td>
<td>877</td>
<td>0.53</td>
<td>0.51</td>
</tr>
<tr>
<td>rs955696</td>
<td>C/T</td>
<td>Combined analysisc</td>
<td>1932</td>
<td>1816</td>
<td>0.55</td>
<td>0.51</td>
</tr>
</tbody>
</table>

aP-value (allele) was calculated using an allele frequency comparison test.
b95% CI = 95% confidence interval.
cThe Mantel–Haenszel method was used for the combined analysis of first and second sets.
In the present study, we demonstrated that polymorphisms in the \textit{EGR2} gene are associated with SLE susceptibility, presumably through up-regulating the gene expression. An SNP in the 5' flanking region of \textit{EGR2}, termed rs10761670, showed the highest correlation between \textit{EGR2} expression and its genotype. Furthermore, in case–control association tests, the peak association with SLE susceptibility was observed for rs10761670. This correspondence between the two independent associations (gene expression and disease susceptibility) strongly indicates the presence of disease-causing regulatory variants in this region. Two SNPs, rs1412554 and rs1509957, which are in complete LD with rs10761670, and were shown to affect the binding of transcription factors and to modulate gene expression in \textit{in vitro} assays, are candidate SNPs for the causal variants, although other undiscovered variants may also contribute to the regulation of gene expression.

When the effective size of the \textit{EGR2} polymorphism that contributes to disease risk is measured as OR, this size is relatively small compared with that of other loci described in recent GWASs of SLE, suggesting that \textit{EGR2} polymorphism has a moderate risk for an individual. However, the risk allele of rs10761670 is common in the population (the allele frequency = 0.51). This indicates that the \textit{EGR2} polymorphism contributes to the onset of SLE in a wide population in contrast to rare but high penetrant variants. We calculated the population attributable risk proportion (PARP) to assess the reduction in susceptibility to SLE, if these risk alleles were removed from the population, and the PARP for rs10761670 in \textit{EGR2} was 0.088.

In GWASs of CD in Caucasian populations, two landmark SNPs (rs224136 and rs10761659) were located in the intergenic region between \textit{ZNF365} and \textit{EGR2} (at the 3' side of \textit{EGR2}). Because T cells are considered to play a significant role in the pathogenesis of CD, \textit{EGR2}, which is a potential regulator of TCR signaling, is a strong candidate for CD susceptibility at this locus. Therefore, genetic variations with disease-causing potential in this region could potentially be shared genetic risk factors for CD and SLE. However, the landmark SNP identified in the present study (rs10761670) is located in the 5' flanking region of \textit{EGR2} and is not in strong LD (r$^2 < 0.20$) with SNPs of CD (rs224136 and rs10761659) in either Japanese or Caucasian populations. This finding implies that causal variants may be different between SLE and CD or may differ between ethnic populations. Therefore, comparative analyses of both diseases between populations are needed to clarify this problem.

In mice, \textit{Egr2} conditional knockout in CD2$^+$ T cells results in the development of a lupus-like autoimmune disease, where complete loss of \textit{Egr2} function may lead to the loss of regulation of effector T cells and self-reactive T cells (20). In addition, transduction of \textit{Egr2} into naive CD4$^+$ T cells differentiates them to IL-10-secreting CD4$^+$CD25$^+$LAG3$^+$ regulatory T cells that would suppress self-reactive cells (21). These protective roles of \textit{Egr2} against autoimmunity in mice contradict our findings that enhanced expression of \textit{EGR2} MAY increase SLE susceptibility in humans. Similar observations were reported for the \textit{PTPN22} gene, another established susceptibility gene for both SLE and RA (4,30). \textit{PTPN22} encodes a tyrosine phosphatase, LYP that has an inhibitory effect on TCR signaling. For LYP, a non-synonymous SNP (rs2276601, R620W) was associated with disease susceptibility, and increased inhibition of TCR signaling was shown with the disease risk allele W620 (31). Furthermore, knockout of \textit{Pep}, the murine ortholog of \textit{PTPN22}, leads to various immune abnormalities, including expansion of effector T cells and increased antibody production (32). Taken together, enhanced negative regulation of T cells by inhibitory components such as \textit{EGR2} and \textit{PTPN22} may lead to tolerance.

**DISCUSSION**

Figure 2. Transcriptional activity is affected by rs1412554 and rs1509957. (A) EMSAs were performed with nuclear extracts from PSC and Jurkat cells. Biotin-labeled 25-bp oligonucleotides corresponding to each SNP (rs1412554 and rs1509957) served as probes for the susceptible (Sus) and the non-susceptible (Non-sus) allele, respectively, and the abilities of nuclear factors to bind to these oligonucleotides were evaluated (for rs1412554: Sus = T, Non-sus = C; for rs1509957: Sus = C, Non-sus = T). Unlabeled oligonucleotides were used for competition assays. A single asterisk (*) and double asterisks (**) indicate bands showing allelic differences in rs1412554 and rs1509957, respectively. PSC cells (Left) and Jurkat cells (Right). (B) Transcriptional activities of the 31-bp genomic sequences around SNPs (rs1412554 T/C) and rs1509957 C/T) were evaluated by luciferase assays. Each 31-bp oligonucleotide corresponding to these genomic sequences was inserted into the pGL3-promoter vector upstream of the SV 40 promoter and allele-specific constructs were transfected into Jurkat cells. Relative promoter activity was expressed as the ratio of the luciferase activity of each allele specific construct to the luciferase activity of the pGL3-promoter control vector. Data are presented as the mean ± SEM. Each experiment was independently repeated three times and each sample was studied in sextuplicate. *P < 0.05 by Student’s t-test. Rs1412554 (top) and rs1509957 (bottom).
up-regulation of the survival molecule Bcl-2 (33). As overexpression of mice, which concluded that Egr2 plays a central role in the up-regulation of the survival molecule Bcl-2 during the positive selection of thymocytes (34), up-regulated expression of Egr2 may also lead to increased selection of self-reactive thymocytes.

In conclusion, EGR2 is a genetic risk factor for SLE as well as for RA, for which increased gene expression may contribute to disease pathogenesis. Analyses of EGR2 in other autoimmune disorders, including CD, are needed to elucidate its precise role in autoimmunity.

MATERIALS AND METHODS

Subjects

Three independent sets of SLE (n = 376, 293 and 223) and two independent sets of RA (n = 1112 and 830) patients, and of control subjects (n = 940, 881 and 658), were enrolled in the study through medical institutes in Japan. The control subjects in the first set were the members of Midousuji Rotary Club and those in third set was recruited through Pharma SNP Consortium (35). All subjects were self-identified Japanese. SLE subjects met the revised criteria for SLE (1) and RA subjects met the revised ACR criteria for RA (36). All control subjects were confirmed to be not affected with autoimmune diseases including SLE and RA by reviewing their medical records. All subjects provided informed consent to their participation in the study, as approved by the Ethics Committee of the Center for Genomic Medicine, RIKEN. DNA was extracted from peripheral blood cells using a standard protocol.

SNP discovery

Unknown SNPs were revealed by direct sequencing of the DNA of 96 individuals affected with SLE. DNA fragments were amplified for sequencing with the appropriate primers and were purified using a MultiScreen PCR filter plate (Millipore, Billerica, MA, USA). The amplified DNAs were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) and signals were detected using an Applied Biosystems ABI 3700 DNA Analyzer.

Genotyping

We selected 14 tag SNPs from an 80-kb region on 10q21.3, which comprised three LD blocks and included the EGR2 gene, using the HapMap data of the Japanese population and Haploview software, v.4.0. These SNPs were genotyped using TaqMan SNP genotyping assays (Applied Biosystems) as indicated by the manufacturer. Fluorescence was detected using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). All SNPs were successfully genotyped with call rates >0.99 and were in Hardy–Weinberg equilibrium in control subjects (P > 0.01).

Electrophoretic mobility shift assays

PSC (an EBV-transformed lymphoblastoid cell line) and Jurkat cells were grown in RPMI 1640 medium supplemented with 10% fetal bovine serum and antibiotics. Nuclear extracts were prepared as previously described (37). In brief, following stimulation with 50 ng/ml phorbol myristate acetate (Sigma, St Louis, MO, USA) for 2 h, centrifuged cells were collected and resuspended in Buffer A (20 mM HEPES pH 7.6, 20% glycerol, 10 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 8.0, 1 mM DTT, 0.1% NP-40 and Protease inhibitor cocktail). The cells were then incubated on ice for 10 min, centrifuged and the pellets were resuspended in Buffer B (20 mM HEPES pH 7.6, 20% glycerol, 500 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA pH 8.0, 1 mM DTT, 0.1% NP-40 and Protease inhibitor cocktail). Following incubation on ice for 30 min and then centrifugation to remove cellular debris, the supernatant fraction containing nuclear proteins was collected.

Oligonucleotides (25 bp) were designed that corresponded to genomic sequences surrounding the SNPs. Single-stranded 25-bp oligonucleotide probes were labeled using a Biotin 3’ End DNA Labeling Kit (Pierce Biotechnology, Rockford, IL, USA) and sense and antisense oligonucleotides were then annealed. The LightShift Chemiluminescent EMSA kit (Pierce Biotechnology) was used for the detection of DNA–protein interactions. Biotin-labeled probes were incubated with nuclear extracts in a binding reaction mixture (10× binding buffer, 50 ng/µl poly dl-Dc, 50% glycerol, 5 mM MgCl₂ and 0.05% NP-40) for 20 min at 25°C. In competition studies, unlabeled oligonucleotides (100-fold excess) were preincubated with the nuclear extract and binding reaction mixture for 5 min before addition of the biotin-labeled probes. The DNA–protein complexes were separated on a non-denaturing 5% polyacrylamide gel in 1× TBE (Tris-borate-EDTA) running buffer for 70 min at 150 V. Gel protein–DNA complexes were transferred to a nitrocellulose membrane for 30 min at 380 mA and the transferred complexes were cross-linked to the membrane by exposure to UV light of 120 mJ/cm² for 1 min. DNA–protein complexes on the membrane were detected using the Chemiluminescent Nucleic Acid Detection Module (Pierce) and a LAS-3000 mini lumino-image analyzer (Fujiﬁlm, Tokyo, Japan). Allelic differences were analyzed using MultiGauge software (Fujiﬁlm) by measuring the intensity of the bands.

Luciferase assay

Oligonucleotides (31 bp) were designed that corresponded to genomic sequences that included susceptible or non-susceptible alleles. Complementary sense and antisense oligonucleotides were then annealed. To construct luciferase reporter plasmids, the pGL3-Promoter vector (Promega, Madison, WI, USA) was digested with MulI and BglII and a single copy of a 31-bp oligonucleotide was ligated into the vector upstream of the SV40 promoter using the TaKaRa Ligation kit ver. 2.1 (TaKaRa, Shiga, Japan). After confirmation of the sequence, these plasmids were purified using
a HiSpeed Plasmid Midi kit (Qiagen, Valencia, CA, USA). Jurkat cells (5 × 10⁵), grown as described above, were transfectioned with 2.5 μg of the constructs and 0.5 μg of the pRL-TK vector (an internal control for transfection efficiency) using the Lipofectamin2000 transfection reagent (Invitrogen, Carlsbad, CA, USA). After 24 h, the cells were collected and luciferase activity was measured using a Dual-Luciferase Reporter Assay system (Promega) and an Ultra Sensitive Tube Luminometer, Lumat LB 9507 (Berthold Technologies, Bad Wildbad, Germany). Each experiment was independently repeated three times and sextuplicate samples were assayed each time.

Statistical analysis

We used χ² contingency table tests to evaluate the significance of differences in allele frequency in the case–control subjects. We defined haplotype blocks using the Solid spine of LD definition of the Haplovie v4.0. We performed a Mantel–Haenszel analysis to calculate the pooled P-value and odds ratio of two independent association studies. We calculated the power of each cohort for testing association by Quanto Software (http://hydra.usc.edu/gxe/). To adjust for the confounding effects such as gender, we performed a logistic regression analysis using the STATISTICA software (StatSoft). We calculated PARP using the following formula; PARP = f (OR–1)/ (1 + f (OR–1)) × 100, where f is the allele frequency in the control subjects and OR is the odds ratio. PARP is defined as the reduction of disease incidence that would be achieved if the population had been entirely unexposed. Luciferase assay data were analyzed by Student’s t-test.

ACKNOWLEDGEMENTS

We thank K. Kobayashi, K. Shimada, M. Kitazato and all other members of the Laboratory for Autoimmune Disease, CGM, RIKEN for their advice and technical assistance. We also thank Dr A. Miyatake, the members of the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan, and the staffs of the BioBank Japan Project for clinical sample collection.

Conflicts of Interest statement. None declared.

FUNDING

This work was supported by grants from the CGM, RIKEN; the Ministry of Education, Culture, Sports, Science and Technology of Japan (Leading Project); and the Ministry of Health, Labor and Welfare of Japan (Research on Intractable Diseases).

REFERENCES
