HMG Advance Access published June 25, 2008



# A common *STAT4* risk haplotype for Systemic Lupus Erythematosus is over-expressed, correlates with anti-dsDNA production and shows additive effects with two *IRF5* risk alleles

Snaevar Sigurdsson[1], Gunnel Nordmark[2], Sophie Garnier[1], Elin Grundberg[3], Tony Kwan[3], Olof Nilsson[4], Maija-Leena Eloranta[2], Iva Gunnarsson[5], Elisabet Svenungsson[5], Gunnar Sturfelt[6], Anders A Bengtsson[6], Andreas Jönsen[6], Lennart Truedsson[7], Solbritt Rantapää-Dahlqvist[8], Catharina Eriksson[9], Gunnar Alm[10], Harald HH Göring[11], Tomi Pastinen[3,12], Ann-Christine Syvänen[1,*] and Lars Rönnblom[2]

[1)]Molecular Medicine, Department of Medical Sciences, Uppsala University, SE-75185 Uppsala, Sweden, [2)]Section of Rheumatology, Department of Medical Sciences, Uppsala University, SE-75185 Uppsala, Sweden [3)]McGill University and Genome Quebec Innovation Centre, Montreal, Quebec, H3A 1A4, Canada [4)]Department of Surgical Sciences, Uppsala University SE-7518 Uppsala, Sweden [5)]Department of Medicine, Rheumatology Unit, Karolinska Institutet/Karolinska University Hospital, SE-77176 Stockholm, Sweden [6)]Department of Rheumatology, Lund University, SE-22100, Lund, [7)]Institute of Laboratory Medicine Section of MIG, Lund University, SE-22100, Lund, Sweden [8)]Department of Rheumatology, University Hospital, SE-901 85 Umeå, Sweden [9)]Department of Clinical Immunology, University Hospital, SE-901 85 Umeå, Sweden [10)]Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden, [11)]Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas, USA, [12)]Department of Human Genetics, McGill University, Montreal, Quebec, Canada

*Corresponding author: Ann-Christine Syvänen, Department of Medical Sciences, Uppsala University, Molecular Medicine, Entrance 70, Research Department 2, 3rd floor, Uppsala University Hospital, 75185 Uppsala, Sweden. Phone: +46-18-6112959, Fax +46-18-55 36 01, E-mail: Ann-Christine.Syvanen@medsci.uu.se

**ABSTRACT**

Systemic Lupus Erythematosus (SLE) is the prototype autoimmune disease where genes regulated by type I interferon (IFN) are over-expressed and contribute to the disease pathogenesis. Because *STAT4* plays a key role in the type I IFN receptor signaling, we performed a candidate gene study of a comprehensive set of single nucleotide polymorphism (SNPs) in *STAT4* in Swedish patients with SLE. We found that ten out of 53 analyzed SNPs in *STAT4* were associated with SLE, with the strongest signal of association ($p = 7.1 \times 10^{-8}$) for two perfectly linked SNPs rs10181656 and rs7582694. The risk alleles of these ten SNPs form a common risk haplotype for SLE ($p = 1.7 \times 10^{-5}$; OR = 1.71). According to conditional logistic regression analysis the SNP rs10181656 or rs7582694 accounts for all of the observed association signal. By quantitative analysis of the allelic expression of *STAT4* we found that the risk allele of *STAT4* was over-expressed in primary human cells of mesenchymal origin, but not in B-cells, and that the risk allele of *STAT4* was over-expressed ($p = 8.4 \times 10^{-5}$; OR = 1.59) in cells carrying the risk haplotype for SLE compared to cells with a non-risk haplotype. The risk allele of the SNP rs7582694 in *STAT4* correlated to production of anti-dsDNA antibodies and displayed a multiplicatively increased, 1.82-fold risk of SLE with two independent risk alleles of the *IRF5* gene.

## INTRODUCTION

Systemic Lupus Erythematosus (SLE, OMIM # 15200) is an autoimmune disease characterized by production of antinuclear autoantibodies, immune complex formation and inflammation in multiple organs. Patients with SLE have raised serum levels of IFN-α, which correlates to both disease activity and severity (1), and with an increased expression level of type I IFN regulated genes peripheral blood cells (an IFN signature) (3-5). The reason for this IFN signature may be an ongoing production of type I IFN, which is caused by RNA- or DNA-containing immune complexes that trigger plasmacytoid dendritic cells (PDCs) to synthesize type I IFNs via the Toll like receptor (TLR) 7 or TLR9 (2). Because the type I IFNs cause activation and maturation of dendritic cells (DC), stimulation of Th1 and Tc cells as well as enhanced immunoglobulin production by B cells, this mechanism of a continuous activation of the type I IFN system could be directly involved in the etiopathogenesis of SLE (2, 3).

We have previously observed that the interferon regulatory factor 5 (*IRF5*) and tyrosine kinase 2 (*TYK2*) genes, which are involved in both the production and the effects of the type I IFNs are associated with susceptibility to develop SLE (4). These observations have been replicated in multiple subsequent studies (5, 6). The transcription factor IRF5 regulates the expression of type I IFN genes, while the Janus kinase TYK2 binds to the type I IFN receptor (IFNAR) and is required for signaling through this receptor. In addition, IRF5 is involved in cell adhesion, apoptosis, cell cycle regulation and early immune activation (7), while TYK2 interacts with the cellular receptors for IL-6, IL-10 and IL-12/23 cytokine families, in addition to IFNAR (8). Thus, both IRF5 and TYK2 have diverse functions in the immune system that can affect several pathways in autoimmune reactions.

In order to identify other molecules within the type I IFN system with a possible association with SLE, we studied the signal transducer and activator of transcription 4 (STAT4) transcription factor, which plays a key role in the type I IFN receptor signaling by being activated and translocated to the nucleus after IFNAR ligation (9). In addition, STAT4 transmits signals from the receptors for IL-12 and IL-23, and can therefore contribute to autoimmune responses by affecting the functions of several innate and adaptive immune cells (10). By analyzing 53 single nucleotide polymorphisms (SNPs) in *STAT4* for their association with SLE in Swedish patients and control individuals we identified a common risk haplotype type for SLE formed by ten SNPs in introns of *STAT4*. During the course of our study, an association of the SNP rs7574865 in *STAT4*, which tags this risk haplotype was found to be associated with SLE in three North American SLE- cohorts of European ancestry (11). Furthermore, two recently published genome-wide association studies of SLE confirmed this association (12, 13). In our study we show that the risk haplotype identified in *STAT4* is over-expressed in primary human cells of mesenchymal origin, but not in transformed B-cells, and that the risk haplotype of *STAT4* is over-expressed in mesenchymal cells (osteoblasts) carrying the risk haplotype, compared to cells with the non-risks haplotype. We also show that the risk allele of *STAT4* correlates with increased frequency of immunological manifestations in the SLE patients, including production of anti-dsDNA autoantibodies. Because IRF5 mediates type I IFN production and STAT4 affects the response to type I IFN, we also determined the joint effects of the risk-alleles of *STAT4* and two recently identified independent risk alleles of *IRF5* with strong effects on SLE (14). This analysis revealed a strong multiplicatively increased risk for SLE by these three polymorphisms located in *STAT4* and *IRF5*, which are two of the most strongly associated SLE-genes and belong to the same cell signaling pathway.

## RESULTS

### Identification of *STAT4* risk alleles and one major *STAT4* risk haplotype in SLE.

The STAT4 gene located on chromosome 2q32.3 consists of 24 exons spanning a 120 kb region (Figure 1).We genotyped 53 SNPs, distributed across *STAT4*, in DNA samples from 485 Swedish patients with SLE and 563 control individuals and observed 10 SNPs that yielded uncorrected p-values below 0.001, with p-values for the association signals calculated by Fischer's exact test ranging from $4.5 \times 10^{-4}$ to $7.1 \times 10^{-8}$ (Figure 1, Supplementary Table S2). The strongest signal of association (OR = 1.71 with 95% CI 1.40-2.08) was observed for two SNPs located in intron 3 of *STAT4*, namely the SNPs rs10181656 and rs7582694, which are in perfect LD ($r^2$=1) with each other in our genotype data. The 10 SNPs that display association signals with SLE are located in introns 3 to 24 of *STAT4*, and are all correlated with the most strongly associated SNPs rs10181656 and rs7582694 ($r^2 > 0.41$) (Figure 1 and Supplementary Table S3). We constructed haplotypes using the genotype data from these 10 SNPs and tested the association of the formed haplotypes with SLE (Table 1). This analysis defines a major risk haplotype for SLE that occurs with a frequency of 0.24 in the SLE-patients, compared to 0.17 in the controls and is strongly associated with SLE (unadjusted p-value = $1.7 \times 10^{-5}$ and OR 1.59 (95% CI 1.28 – 1.95). By conditional logistic regression analysis we found that the SNPs rs7582694 (or rs10181656) that displayed the strongest signals of association with SLE accounts for the association signals observed from all other SNPs in *STAT4* (Table 2). According to the HapMap phase II data there are five additional SNPs in *STAT4* that are in strong LD ($r^2 > 0.9$) with the SNPs rs10181656 and rs7582694. These are: rs11889341 ($r^2$=1.0), rs7574865 ($r^2$=1.0), rs7568275 ($r^2$=1.0), rs8179673 ($r^2$=0.95), rs10174238 ($r^2$=0.90). One of them, rs7574865, was recently shown to be associated with SLE with high

significance ($p = 1.87 \times 10^{-9}$, OR = 1.55) in SLE patients from the US with European ancestry (11), and two consequent genome-wide association studies confirm this finding (12, 13).

**Allelic and total expression levels of *STAT4***

Because there are no known common coding SNPs in *STAT4* that are in linkage disequilibrium with the SLE-associated *STAT4* SNPs, we investigated if a possible functional role for these intronic SNPs that accounted for the association signal with SLE could be to alter the expression of the STAT4 gene. *STAT4* has two well characterized isoforms, *STAT4* $\alpha$ and $\beta$, which could be relevant for SLE, because they have been shown to have distinct functions in mediating the response to IL12 in Th1-cells from mice (15). In the *STAT4* $\beta$ isoform, expression of an additional exon introduces a stop codon in the *STAT4* mRNA, which results in a truncated protein lacking a putative transcriptional activation domain in its C-terminus. We tested whether the disease associated *STAT4* SNPs would alter expression of the alternatively spliced *STAT4β* isoform by utilizing expression data from Affymetrix Exon 1.0 arrays in two population based panels of cells of both lymphoid (lymphoblast) and mesenchymal (osteoblast) origin (16). We did not find evidence for association with the expression level of the *STAT4β* isoform for any of the SLE-associated SNPs (data not shown).

Next we determined the allelic expression levels of *STAT4* (17) in pooled primary human osteoblast and pooled human lymphoblast cell lines. The SNP rs8179673, which is in almost perfect pair-wise LD ($r^2 = 0.95$) with our two strongest SLE-associated SNPs rs7582694 or rs10181656 was used as indicator of allelic imbalance in pre-mRNA extracted from the cells. This analysis showed that the SLE-risk allele of the SNP rs8179673 was significantly over-expressed in the pooled osteoblasts (p

= 0.038), but not in the lymphoblasts (Figure 2A-B). Consequently, we determined the allelic expression levels of the SLE-risk haplotype of *STAT4* in ten informative individual osteoblasts for which the SLE-risk haplotype could be unequivocally assigned and which were heterozygous for at least two of the SNPs rs1517352, rs3821236 and rs7574865 used as indicators of allelic imbalance. The SNP rs1517352 and rs3821236 were genotyped in our SLE-samples (Table 2, Supplementary Table S3) and the SNP rs7574865 is in perfect LD ($r^2 = 1.0$) with our two strongest SLE-associated SNPs. We observed significant over-expression of the SLE-risk haplotype in each of the osteoblasts carrying the SLE-risk haplotype, compared to its expression in osteoblasts with a non-risk haplotype ($p = 8.4 \times 10^{-5}$) (Figure 2C). We also determined the total expression level of *STAT4* in 40 individual osteoblasts with known *STAT4* haplotype status using data from Illumina Human Ref-8 Bead Arrays and found that *STAT4* mRNA was over-expressed in osteoblasts carrying the risk haplotype for SLE ($p = 0.0085$) compared to osteoblasts with non-risk haplotypes (Figure 2D).

**Correlation between *STAT4* risk allele and disease phenotype**

Information concerning disease manifestations and genotype data for the *STAT4* SNP rs7582694 were available for 475 SLE patients included in the association analysis described above, and from an additional cohort of 220 Swedish SLE patients. We correlated the presence of at least one risk allele with age at diagnosis of SLE, individual American College of Rheumatology (ACR) criteria, total number of ACR criteria and the presence of antibodies against double-stranded DNA (dsDNA). We found that patients with at least one risk allele of *STAT4* had a higher frequency of immunological disorder (unadjusted p-value = $1.6 \times 10^{-5}$) and presence of antibodies against dsDNA (unadjusted p-value = $5.3 \times 10^{-7}$) than patients with no risk alleles (Table 3).

**Additive effects of the risk alleles of *STAT4* and *IRF5* in SLE**

We have recently analyzed a comprehensive set of polymorphisms in *IRF5* located on chromosome 7q32 for their association with SLE in the same SLE patients and controls that were here analyzed for SNPs in *STAT4*(14). We demonstrated that two independently associated polymorphisms in *IRF5* account for the association signals with SLE, namely a previously unknown 5 bp (CGGGG) indel in the promoter region of *IRF5* (OR = 1.69, 95% CI 1.42-2.02) and the SNP rs10488631 or one of multiple perfectly correlated SNPs located 3' of *IRF5* (OR = 2.07, 95% CI 1.63 - 2.62). In the present study we investigated the joint effect on SLE of the risk allele of the SNP rs7582694 in *STAT4* and the two polymorphisms in *IRF5*. Figure 3A shows the distribution of risk alleles for SLE in *STAT4* and *IRF5* in patients and controls. The overall significance for the difference in risk allele counts between patients and controls in all groups is high, with a one-sided p-value of $1.7 \times 10^{-16}$, using an additive logistic regression model. No evidence for dominance or interactions was observed between the three risk alleles using multiple logistic regression analysis. Figure 3B shows the OR for SLE in the group of patients with two or more risk alleles, using the individuals with zero or one risk allele as reference. As can be seen, the risk for SLE increases as a function of the number of risk alleles in a multiplicative manner with a 1.82 –fold increase in OR for each additional risk allele. The OR for SLE is 2.08 (CI 1.53 – 2.83) for carriers of two risk alleles and 13.0 (3.78-44.3) for carriers of five or six risk alleles when compared to those with zero or one risk allele. No convincing correlation was found when the risk alleles in *STAT4* and *IRF5* were combined and correlated to disease phenotype in the patients with SLE (data not shown).

## DISCUSSION

In the present investigation we analyzed a panel of SNPs that capture the majority of the common variation of *STAT4* for their association with SLE, and identified ten SNPs with strong association signals. This result confirms a role for *STAT4* in SLE and corroborates previously reported observations of an association between individual SNPs in *STAT4* and risk to develop SLE (11-13). In our study the risk alleles of two highly correlated SNPs account for all association signals observed and define a single common risk haplotype with a strong association with SLE. However, we cannot exclude rare SNP alleles or other types of polymorphisms than SNPs as risk alleles for SLE. Identification of possible additional risk alleles must await complete re-sequencing of *STAT4* in SLE patients and controls.

When investigating the consequences of the risk allele on the expression of *STAT4*, we observed an increased expression of the risk allele of *STAT4* in primary cells of mesenchymal origin (human osteoblasts), but not in transformed B cells, which provides a functional hypothesis for the disease association in SLE. We recently demonstrated in comparative transcriptome analysis that human osteoblasts are closely related to human fibroblasts and chondrocytes (18). Interestingly STAT4 was one of the most significantly overexpressed genes in the osteoblasts as compared to the two other closely related mesenchymal cell types. The SNPs in *STAT4* that are associated with SLE are located in introns and could regulate the expression level or splicing of *STAT4*. Indeed, we show evidence supporting the former alternative, because the common risk haplotype alters expression of *STAT4* only in one of the cell types analyzed. Thus, our results indicate the potential presence of a tissue-specific intragenic enhancer, the function of which could be perturbed by common variants within the SLE risk haplotype. Obviously, different gene variants may have different effects on the

expression of *STAT4* between cell types, which could at least partially explain the diverse effects by *STAT4* in many different cells of the immune system (10).

In our study we also observed a correlation between the risk allele of *STAT4* and a higher frequency of antibodies against dsDNA among patients with the risk allele, which suggests that polymorphisms in *STAT4* contribute to the underlying autoimmune process in SLE. This effect could be due to the many different effects of STAT4 in the immune system. Besides its role in type I IFN signaling, STAT4 also transmits signals from e.g. IL-12 and IL-23, and is thus responsible for the IL-12 dependent activation of natural killer (NK) cells and production of IFN-γ, as well as for polarization of naïve CD4+ T-cells to IFN-γ producing Th1 effector cells (8). Recent studies also indicate that STAT4 regulates the IL-23 dependent expansion of TH17 cells, which are key actors in the pro-inflammatory immune response (19). As dendritic cells (DC) maturate functionally, they express more STAT4, which is required for efficient presentation of antigen (20). Activation of STAT4 is also responsible for the IL-12 induced production of IL-10 in murine NK cells (21) and Th1 cells (22). In the presence of type I IFN, IL-10 becomes mainly proinflammatory instead of anti-inflammatory (23).Thus, STAT4 has pleiotropic and mostly stimulatory effects in the immune system, and has been described as a critical regulator of inflammation *in vivo*, including the development of autoimmune diseases (10). A risk haplotype for SLE, which is associated with increased expression of STAT4, as suggested by our data from osteoblasts, is therefore of great interest. The complex role of STAT4 *in vivo* is illustrated by the observation that *STAT4* deficient lupus mice have aggravated nephritis and increased mortality, despite the fact that they have decreased autoantibody levels and complement depositions in the kidneys (24).

We have argued for a pivotal role for the type I IFN system in autoimmune diseases, including SLE (2), It has been proposed that type I IFNs stimulate an early autoimmune response, but may also later inhibit an on-going autoimmune response (25). Possible inhibition of the autoimmune response by type I IFNs could occur via increased concentration and activation of STAT1 mediated by the type I IFNs, which would decrease the function of STAT4, at least in part due to competition between STAT1 and STAT4 for binding to the IFNAR (26, 27). Therefore, it is possible that altered regulation of the expression of STAT4 in SLE may allow continued type I IFN stimulation of the STAT4-dependent autoimmune response.

Several transcription factors have been found to interact with the promoter of the STAT4 gene, increasing its expression, including AP-1 and the NF-κB (28).At least in DC, the proinflammatory cytokines TNF-α and IL-1 are potent stimulators of STAT4 expression via activation of NF-κB (28). Interestingly, IRF5 is major mediator of TLR-triggered expression of other proinflammatory cytokines than type I IFN, such as TNF-α (29), and in this way IRF5 could promote expression of the STAT4 gene. Thus the *STAT4* and *IRF5* genes, which both contain polymorphisms that confer increased risk for SLE, are also functionally linked.

When we determined the joint effect of the risk alleles of *STAT4* and two recently identified independent risk alleles in *IRF5*, we observed a multiplicative increase in OR by a factor of 1.82 for each additional risk allele and a considerably increased risk for SLE (OR=5.83) for those 18% of the SLE patients that carried four or more risk alleles. The importance of this finding is accentuated by the results of one of the recently published genome-wide SNP association studies, which defines *STAT4* and *IRF5,* along with HLA-DQA1 as the loci that display the strongest association with SLE

among all human genes (13). These observations , as well as the findings that polymorphisms in the *STAT4* gene also are associated with rheumatoid arthritis, (11, 30) and that polymorphisms in *IRF5* are associated with rheumatoid arthritis (31), inflammatory bowel diseases (IBD)(32) and multiple sclerosis (MS)(33), suggest that *STAT4* and *IRF5* can contribute to a general loss of tolerance, whereas other genes might contribute to the disease-specific phenotypes. In this respect *IRF5* may be unique because the CGGGG indel polymorphisms in *IRF5* is associated with SLE (14), IBDs (32) and MS (33), and thus seems to promote the autoimmune process. In contrast, the SNP rs10488631 and several proxies of this SNP downstream of *IRF5* are associated only with SLE and thus, appear to promote the disease-specific phenotype.

Our study demonstrates a strongly increased risk for SLE in individuals carrying multiple risk alleles of two genes from the type I IFN signaling pathway. When more susceptibility genes for SLE, or even a complete set of susceptibility genes have been identified, a signature of multiple risk alleles could be used to improve disease prediction (34). The observed connection between the risk allele of *STAT4* and production of a SLE-specific autoantibody also indicates that a genetic signature of risk alleles could be used in the future for improved classification of patients with SLE compared to the current ACR criteria. Combined genotype data from genes in the same or related molecular pathways may also be helpful when studying interactions between genes and the environment, for instance the susceptibility to disease flare during viral infections that is apparent in a subset of SLE patients (35).

## MATERIALS AND METHODS

### Patients and controls

Our study included 485 Swedish SLE patients fulfilling at least four of the classification criteria for SLE as defined by the American College of Rheumatology (ACR)(36) and 563 matched healthy control individuals from the same geographic areas as the patients. The patients were from the rheumatology clinics at the Lund (n = 144), Uppsala (n = 136) and Karolinska (Solna) (n = 205) University Hospitals. The SLE patients had a mean age at disease diagnosis of 33.8 ± 14.2 years (range 9-75). Disease manifestations were scored in accordance with the ACR criteria (36) and the mean number of criteria fulfilled was 5.7 ± 1.4 (range 4-10) out of the 11 ACR criteria. The control samples from Lund (n = 195) consisted of healthy blood donors from the same region as the patients. The control samples from Uppsala (n = 155) were from healthy blood donors that were matched with the patients for age, sex and region of living. The control samples from Stockholm (n = 213) were population-based control individuals matched for age and sex with the patients. DNA was extracted from blood samples of the patients and controls using standard procedures. For correlation of *STAT4* genotype with disease phenotype we also included 220 SLE patients from the rheumatology clinic of Umeå Univeristy Hospital. These patients had a mean age of diagnosis of 39.1 ± 15.7 years and a mean number of fulfilled ACR criteria of 5.4 ± 1.4. For studying the expression of the *STAT4* gene we used Epstein-Barr virus (EBV) transformed human B-cells (lymphoblastoid cells) from the Human Haplotype Map Caucasian panel, which is derived from individuals of northern European descent (37) and human primary cells of mesenchymal origin (osteoblasts) derived from unrelated Swedish donors undergoing elective hip replacement surgery at Uppsala University Hospital. The human osteoblasts were cultured as described earlier (38). The

study was approved by the regional ethical boards and all study subjects gave their informed consent to participate.

**Genotyping**

For genotyping we selected 64 SNPs from the *STAT4* gene, from 2 kb upstream of the first exon to 2 kb downstream of the last exon, thus including all exons and introns (Supplementary Table S1). The goal of the SNP selection was to cover the STAT4 gene region with approximately 1 SNP per 2 kb. The selection was guided by the Illumina Golden Gate Assay design scores and by the HapMap phase II data (20 Jan 2006 release, NCBI assembly B35, dbSNP b125). An Illumina score of 0.6 was used as the lower limit for inclusion in the panel. Tag-SNP information was obtained from the HapMap website, and for each group of SNPs in high LD ($r^2 > 0.95$, minor allele frequency > 0.05) at least one SNP was selected. The selected panel of 64 SNPs was genotyped in DNA samples from the Swedish patients with SLE and controls from Lund, Uppsala and Stockholm using the Golden Gate Assay (Illumina Inc, San Diego, CA, USA). Of the 64 SNPs genotyped, seven were non-polymorphic, two were failed due to poor genotype clustering and the genotypes of two SNPs deviated form Hardy-Weinberg equilibrium in the controls ($p < 0.01$). These SNPs were excluded from further analysis. For the 53 SNPs that fulfilled the quality control criteria, the sample success rate was 96.7% and the average genotype call rate in these samples was 99.7%. Four parent-offspring trios were included in the genotyping for inheritance check, and no Mendelian inheritance errors were observed. Thirty-two samples were genotyped in duplicates with the Golden Gate assay with 100% consistency between genotype calls. The 53 successfully genotyped SNPs capture 91% of the HapMap SNPs (MAF > 0.05) with $r^2 > 0.8$. The *IRF5* SNP rs10488631 was genotyped by the Illumina Golden Gate assay and the 5 bp (CGGGG) insertion-deletion polymorphism (indel) of

*IRF5* was genotyped by PCR amplification followed by size separation by electrophoresis using 4% agarose gels, or using an ABI 3770 capillary sequencer (Applied Biosystems, Foster City, CA, USA) as previously descibed (14). (14). The SNP rs7582694 in *STAT4* was genotyped in the SLE patients from Umeå using the SNPstream system (Beckman-Coulter Inc., Fullerton, CA, USA). The genotyping success rate was 90.6% and the reproducibility of genotyping was 100% according to 6% duplicated genotyping assays.

**Expression analysis of STAT4**

The *STAT4 β* isoform-specific expression was determined with Affymetrix Exon 1.0 arrays (Affymetrix, Santa Clara, CA, USA) using isoform-specific probe sets 2592363 at chr2:191604056-191604086 and 2592364 chr2:191604199-191604358 according to the procedures recommended by Affymetrix. The complete dataset for the EBV-transformed human B-cells (n = 57) has been described previously (16), and the complete data set for the human primary osteoblasts (n = 58) generated by the same procedures will be published separately. Allelic expression levels of *STAT4* were assessed in pooled lymphoblasts (n = 55) and osteoblasts (n = 55). The allele ratios were determined by quantitative sequencing of RT-PCR products from pre-mRNA and PCR product from genomic DNA in the pooled samples (39). Allelic expression ratios in the RNA samples were calculated using normalized heterozygote ratios from genomic DNA from the same cells as reference (39). The SNP rs8179673 was used as indicator of allelic imbalance in the expression of *STAT4* in the pooled samples. The sensitivity of pooled RNA analysis in the EBV-transformed human B-cells was assessed using known loci (n = 13) with allelic differences in expression in individual cell lines from previous studies by us (17) and others (40, 41). Of these 13 loci, 12 (92%) showed statistically significant over-expression of the expected allele in the pooled analysis (D. J.

Verlaan, E. Grundberg, B. Ge, R. Hoberman, K. Lam, V. Koka, J. Dias, S. Gurd, N. Martin, Ö. Ljunggren, O. Nilsson, A. Kindmark, H. Brandström, E. Harmsen, and T. Pastinen, manuscript in preparation). The individual osteoblasts have been genotyped for about half a million SNPs using Illumina 550K SNP Bead Arrays (Illumina Inc. San Diego, CA, USA). This data was used for assigning *STAT4* haplotypes to the individual osteoblasts using an algorithm in Phase 2.1 as described previously (42). Osteoblasts (n=10) for which the haplotype could be accurately assigned using this method, and which were heterozygous for at least two out of the three SNPs rs1517352, rs3821236 and rs7574865 used as indicators of the SLE-risk haplotype, were included in the allelic expression analysis of *STAT4* (39). The mean values of triplicate measurement of the allelic expression levels determined using two or three heterozygous SNPs were used. The total expression level of *STAT4* was determined in osteoblasts with known *STAT4* haplotype (N = 40) in two biological replicates with Human Ref-8 Bead Arrays from Illumina (Illumina Inc, San Diego, CA, USA) using the probe ILMN_8937. A variance-stabilizing transformation (VST), which takes advantage of a large number of technical replicates available from the Illumina BeadArrays was coupled with a robust spline normalization to obtain normalized mean expression values (43).

**Statistical analysis**

Fisher's exact test was used to test for Hardy-Weinberg equilibrium in the control samples and for comparing the allele counts in the cases and controls. Association tests, calculation of odds ratios (OR) and conditional logistic regression analysis were performed using the software PLINK. Linkage disequilibrium was calculated using Haploview v 4.0. The expression level of the *STAT4β* isoform was correlated with the genotypes of the SLE-associated SNPs in *STAT4* using a linear regression model. The expression level of the SLE-risk haplotype and the total expression level of

*STAT4* were correlated with the SLE-risk haplotype using a two-tailed t-test. A logistic regression model was used to estimate gene-gene interaction between the *STAT4* and *IRF5* SNPs and for the additive effects of the three SNPs. Fisher´s exact test was used to test for the difference in *STAT4* and *IRF5* risk allele counts between cases and controls. These analyses were performed using R-software and environment. Clinical criteria were analysed using StatView software. Fisher's exact test was used for comparing the frequencies of ACR criteria between cases and controls and continuous variables were compared using Mann-Whitney U-test.



**ACKNOWLEDGEMENTS**

We thank Rezvan Kiani Dehkordi at the Rheumatology Clinic, Uppsala University Hospital, for DNA collection and Anders Lundmark, Molecular Medicine, Department of Medical Sciences, Uppsala University, for assistance with the figures of the manuscript. Genotyping was performed using equipment available at the SNP Technology Platform in Uppsala, Sweden. Mr. Kevin Lam (M. Sc.) and Ms. Vonda Koka (M. Sc.) are acknowledged for performing the allelic expression experiments. Dr Pastinen holds a Canada Research Chair (CRC Tier II).

**CONFLICT OF INTEREST STATEMENT**

None of the authors declare any conflict of interest.

**FUNDING**

The study was supported by a Target Identification in Lupus (TIL) grant from the Alliance for Lupus Research, USA, by grants from the Swedish Research Council for Medicine, Agnes and Mac Rudberg's Foundation and from the Knut and Alice Wallenberg Foundation, the Swedish

Rheumatism Association, the Åke Wiberg Foundation, Center of Gender Medicine at Karolinska Institutet, the King Gustaf V 80-year Foundation and Ulla and Roland Gustafsson Foundation. Dr Pastinen is supported by funding from Genome Quebec and Genome Canada. The statistical analyses conducted in the AT&T Genomics Computing Center in San Antonio were supported in part by a gift from the AT&T Foundation and by the (US) National Center for Research Resources (RR013556).



## REFERENCES

1. Bengtsson, A., Sturfelt, G., Truedsson, L., Blomberg, J., Alm, G., Vallin, H. and Rönnblom, L. (2000) Activation of type I interferon system in systemic lupus erythematosus correlates with disease activity but not antiretroviral antibodies. *Lupus,* **9,** 664-671.
2. Rönnblom, L., Eloranta, M.L. and Alm, G.V. (2006) The type I interferon system in systemic lupus erythematosus. *Arthritis Rheum.,* **54,** 408-420.
3. Theofilopoulos, A.N., Baccala, R., Beutler, B. and Kono, D.H. (2005) Type I interferons (α/β) in immunity and autoimmunity. *Ann. Rev. Immunol.,* **23,** 307-336.
4. Sigurdsson, S., Nordmark, G., Göring, H.H., Lindroos, K., Wiman, A.C., Sturfelt, G., Jönsen, A., Rantapää-Dahlqvist, S., Möller, B., Kere, J. *et al.* (2005) Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am. J. Hum. Gen.,* **76,** 528-537.
5. Graham, R.R., Kozyrev, S.V., Baechler, E.C., Reddy, M.V., Plenge, R.M., Bauer, J.W., Ortmann, W.A., Koeuth, T., Gonzalez Escribano, M.F., Pons-Estel, B. *et al.* (2006) A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression

and is associated with increased risk of systemic lupus erythematosus. *Nat. Genet.,* **38,** 550-555.

6. Graham, D.S., Akil, M. and Vyse, T.J. (2007) Association of polymorphisms across the tyrosine kinase gene, TYK2 in UK SLE families. *Rheumatology (Oxford),* **46,** 927-930.

7. Hu, G., Mancl, M.E. and Barnes, B.J. (2005) Signaling through IFN regulatory factor-5 sensitizes p53-deficient tumors to DNA damage-induced apoptosis and cell death. *Cancer Res.,* **65,** 7403-7412.

8. Schindler, C., Levy, D.E. and Decker, T. (2007) JAK-STAT signaling: from interferons to cytokines. *J. Biol. Chem.,* **282,** 20059-20063.

9. Brierley, M.M. and Fish, E.N. (2002) Review: IFN-alpha/beta receptor interactions to biologic outcomes: understanding the circuitry. *J. Interferon Cytokine Res.,* **22,** 835-845.

10. Kaplan, M.H. (2005) STAT4: a critical regulator of inflammation in vivo. *Immunol. Res.,* **31,** 231-242.

11. Remmers, E.F., Plenge, R.M., Lee, A.T., Graham, R.R., Hom, G., Behrens, T.W., de Bakker, P.I., Le, J.M., Lee, H.S., Batliwalla, F. *et al.* (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N. Engl. J. Med.,* **357,** 977-986.

12. The International Consortium for Systemic Lupus Erythematosus Genetics (SLEGEN), Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J. and Langefeld, C.D. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat. Genet.,* **40,** 204-210.

13. Hom, G., Graham, R.R., Modrek, B., Taylor, K.E., Ortmann, W., Garnier, S., Lee, A.T., Chung, S.A., Ferreira, R.C., Pant, P.V. *et al.* (2008) Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX. *N. Engl. J. Med.,* **358,** 900-909.

14. Sigurdsson, S., Goring, H.H., Kristjansdottir, G., Milani, L., Nordmark, G., Sandling, J.K., Eloranta, M.L., Feng, D., Sangster-Guity, N., Gunnarsson, I. *et al.* (2008) Comprehensive evaluation of the genetic variants of interferon regulatory factor 5 (IRF5) reveals a novel 5 bp length polymorphism as strong risk factor for systemic lupus erythematosus. *Hum. Mol. Genet.,* **17,** 872-881.

15. Hoey, T., Zhang, S., Schmidt, N., Yu, Q., Ramchandani, S., Xu, X., Naeger, L.K., Sun, Y.L. and Kaplan, M.H. (2003) Distinct requirements for the naturally occurring splice forms Stat4alpha and Stat4beta in IL-12 responses. *Embo J.,* **22,** 4237-4248.

16. Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T.J., Sladek, R. and Majewski, J. (2008) Genome-wide analysis of transcript isoform variation in humans. *Nat. Genet.,* **40,** 225-231.

17. Pastinen, T., Ge, B., Gurd, S., Gaudin, T., Dore, C., Lemire, M., Lepage, P., Harmsen, E. and Hudson, T.J. (2005) Mapping common regulatory variants to human haplotypes. *Hum. Mol. Genet.,* **14,** 3963-3971.

18. Grundberg, E., Brandstrom, H., Lam, K.C., Gurd, S., Ge, B., Harmsen, E., Kindmark, A., Ljunggren, O., Mallmin, H., Nilsson, O. *et al.* (2008) Systematic Assessment of the Human Osteoblast Transcriptome in Resting and Induced Primary Cells. *Physiol. Genomics***,** in press.

19. Mathur, A.N., Chang, H.C., Zisoulis, D.G., Stritesky, G.L., Yu, Q., O'Malley, J.T., Kapur, R., Levy, D.E., Kansas, G.S. and Kaplan, M.H. (2007) Stat3 and Stat4 direct development of IL-17-secreting Th cells. *J. Immunol.,* **178,** 4901-4907.

20. Longman, R.S., Braun, D., Pellegrini, S., Rice, C.M., Darnell, R.B. and Albert, M.L. (2007) Dendritic-cell maturation alters intracellular signaling networks, enabling differential effects of IFN-alpha/beta on antigen cross-presentation. *Blood,* **109,** 1113-1122.

21. Grant, L.R., Yao, Z.J., Hedrich, C.M., Wang, F., Moorthy, A., Wilson, K., Ranatunga, D. and Bream, J.H. (2008) Stat4-dependent, T-bet-independent regulation of IL-10 in NK cells. *Genes Immun.***,** in press.

22. Rutz, S., Janke, M., Kassner, N., Hohnstein, T., Krueger, M. and Scheffold, A. (2008) Notch regulates IL-10 production by T helper 1 cells. *Proc. Natl. Acad. Sci. U. S. A.,* **105,** 3497-3502.

23. Sharif, M.N., Tassiulas, I., Hu, Y., Mecklenbrauker, I., Tarakhovsky, A. and Ivashkiv, L.B. (2004) IFN-alpha priming results in a gain of proinflammatory function by IL-10: implications for systemic lupus erythematosus pathogenesis. *J. Immunol.,* **172,** 6476-6481.

24. Jacob, C.O., Zang, S., Li, L., Ciobanu, V., Quismorio, F., Mizutani, A., Satoh, M. and Koss, M. (2003) Pivotal role of Stat4 and Stat6 in the pathogenesis of the lupus-like disease in the New Zealand mixed 2328 mice. *J. Immunol.,* **171,** 1564-1571.

25. Mangini, A.J., Lafyatis, R. and Van Seventer, J.M. (2007) Type I interferons inhibition of inflammatory T helper cell responses in systemic lupus erythematosus. *Ann. N. Y. Acad. Sci.,* **1108,** 11-23.

26. Nguyen, K.B., Watford, W.T., Salomon, R., Hofmann, S.R., Pien, G.C., Morinobu, A., Gadina, M., O'Shea, J.J. and Biron, C.A. (2002) Critical role for STAT4 activation by type 1 interferons in the interferon-gamma response to viral infection. *Science,* **297,** 2063-2066.

27. Miyagi, T., Gil, M.P., Wang, X., Louten, J., Chu, W.M. and Biron, C.A. (2007) High basal STAT4 balanced by STAT1 induction to control type 1 interferon effects in natural killer cells. *J. Exp. Med.,* **204,** 2383-2396.

28. Remoli, M.E., Ragimbeau, J., Giacomini, E., Gafa, V., Severa, M., Lande, R., Pellegrini, S. and Coccia, E.M. (2007) NF-{kappa}B is required for STAT-4 expression during dendritic cell maturation. *J. Leukoc. Biol.,* **81,** 355-363.

29. Kawai, T. and Akira, S. (2007) TLR signaling. *Semin. Immunol.,* **19,** 24-32.

30. Lee, H.S., Remmers, E.F., Le, J.M., Kastner, D.L., Bae, S.C. and Gregersen, P.K. (2007) Association of STAT4 with Rheumatoid Arthritis in the Korean Population. *Mol. Med.,* **13,** 455-460.

31. Sigurdsson, S., Padyukov, L., Kurreeman, F.A.S., Liljedahl, U., Wiman, A.-C., Alfredsson, L., Toes, R., Rönnelid, J., Klareskog, L., Huizinga, T.W.J. *et al.* (2007) Association of a haplotype in the promotor region of the interferon regulatory factor 5 gene with rheumatoid arthritis. *Arthritis Rheum.,* **56,** 2202-2210.

32. Dideberg, V., Kristjansdottir, G., Milani, L., Libioulle, C., Sigurdsson, S., Louis, E., Wiman, A.C., Vermeire, S., Rutgeerts, P., Belaiche, J. *et al.* (2007) An insertion-deletion polymorphism in the Interferon Regulatory Factor 5 (IRF5) gene confers risk of inflammatory bowel diseases. *Hum. Mol. Genet.,* **16,** 3008-3016.

33. Kristjansdottir, G., Sandling, J.K., Bonetti, A., Roos, I.M., Milani, L., Wang, C., Gustafsdottir, S.M., Sigurdsson, S., Lundmark, A., Tienari, P.J. *et al.* (2008) Interferon

Regulatory Factor 5 (IRF5) Gene Variants are Associated with Multiple Sclerosis in Three Distinct Populations. *J. Med. Genet.***,** in press.

34. Weedon, M.N., McCarthy, M.I., Hitman, G., Walker, M., Groves, C.J., Zeggini, E., Rayner, N.W., Shields, B., Owen, K.R., Hattersley, A.T. *et al.* (2006) Combining information from common type 2 diabetes risk polymorphisms improves disease prediction. *PLoS Med.,* **3,** e374.

35. Molina, V. and Shoenfeld, Y. (2005) Infection, vaccines and other environmental triggers of autoimmunity. *Autoimmunity,* **38**, 235-245.

36. Tan, E.M., Cohen, A.S., Fries, J.F., Masi, A.T., McShane, D.J., Rothfield, N.F., Schaller, J.G., Talal, N. and Winchester, R.J. (1982) The 1982 revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.,* **25,** 1271-1277.

37. Frazer, K.A. and Ballinger, D.G. and Cox, D.R. and Hinds, D.A. and Stuve, L.L. and Gibbs, R.A. and Belmont, J.W. and Boudreau, A. and Hardenbol, P. and Leal, S.M. *et al.* (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature,* **449,** 851-861.

38. Grundberg, E., Lau, E.M., Pastinen, T., Kindmark, A., Nilsson, O., Ljunggren, O., Mellstrom, D., Orwoll, E., Redlund-Johnell, I., Holmberg, A. *et al.* (2007) Vitamin D receptor 3' haplotypes are unequally expressed in primary human bone cells and associated with increased fracture risk: the MrOS Study in Sweden and Hong Kong. *J. Bone. Miner. Res.,* **22,** 832-840.

39. Ge, B., Gurd, S., Gaudin, T., Dore, C., Lepage, P., Harmsen, E., Hudson, T.J. and Pastinen, T. (2005) Survey of allelic expression using EST mining. *Genome Res.,* **15,** 1584-1591.

40. Cheung, V.G., Spielman, R.S., Ewens, K.G., Weber, T.M., Morley, M. and Burdick, J.T. (2005) Mapping determinants of human gene expression by regional and genome-wide association. *Nature,* **437,** 1365-1369.

41. Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., Bibikova, M., Chudin, E., Barker, D.L., Dickinson, T. *et al.* (2008) Differential allelic expression in the human genome: a robust approach to identify genetic and epigenetic cis-acting mechanisms regulating gene expression. *PLoS Genet.,* **4,** e1000006.

42. Marchini, J., Cutler, D., Patterson, N., Stephens, M., Eskin, E., Halperin, E., Lin, S., Qin, Z.S., Munro, H.M., Abecasis, G.R. *et al.* (2006) A comparison of phasing algorithms for trios and unrelated individuals. *Am. J. Hum. Genet.,* **78,** 437-450.

43. Lin, S.M., Du, P., Huber, W. and Kibbe, W.A. (2008) Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.,* **36,** e11.

**LEGENDS TO FIGURES 1 - 3**

**Figure 1.**

Schematic view of the genetic variation of *STAT4*. The positions of the exons are marked with vertical lines and the translation start site is indicated by an arrow. The negative $\log_{10}$ of the p-value for the association of each SNP with SLE determined by Fisher's exact test is shown by a dot in the diagram. The pair-wise linkage disequilibrium (LD)

$r^2$ –values between the SNPs in *STAT4* according to the combined data from the Swedish SLE patients and controls are shown in the lower part of the figure. The exact LD values (D' and $r^2$) are provided in Supplementary Table S3.

**Figure 2.**

Expression analysis of *STAT4*. Allelic expression analysis of STAT4 in (A) pooled primary human osteoblasts (n = 55) derived from unrelated Swedish individuals and (B) pooled human lymphoblastoid cell lines (n = 55) derived from unrelated individuals of European descent (the HapMap CEU individuals). The mean values of ratios (with SD in error bars) between the alleles (C/T) of the indicator SNP rs 8179673 in triplicate RNA samples compared to normalized heterozygote ratios in triplicate DNA samples from the corresponding cells are shown. (C) Allelic expression analysis of the SLE-risk haplotype of *STAT4* in informative individual primary osteoblasts (n = 10). for which the *STAT4* SLE-risk haplotype status is known and that are heterozygous for at least two of the indicator SNPs rs1517352 (A/C), rs 3821236 (A/G) and rs7574865 (G/T). The normalized mean (with SD in error bar) allelic expression levels of the SLE-risk haplotype of *STAT4* in each cell sample are shown for osteoblasts with the SLE-risk and non-risk haplotype, respectively. (D) Total expression analysis of individual primary osteoblasts with

known *STAT4* haplotype status using Illumina Human Ref-8 Bead Arrays. Mean normalized total expression levels ($log_2$ value on y-axis and SD of the mean in error bars) for homozygous (+/+, n=1) and heterozygous (+/-, n=16) carriers of the SLE-risk haplotype of *STAT4* (total n=17) and non-carriers (-/-) of the risk haplotype (n =23) are shown.

**Figure 3.**

Joint effects of risk alleles of *STAT4* and *IRF5*. A) Distribution of the number of patients and controls according to number of risk alleles for SLE. The risk alleles are the C-allele of the SNP rs7582694 in *STAT4* and the longer allele (4x CGGGG) of the 5 bp insertion-deletion polymorphisms and the C-allele of the SNP rs10488631 in *IRF5*. The black bars indicate patients and the light grey bars indicate controls. The p-values are for the difference in allele counts between patients and controls in the groups with two or more risk alleles, as compared to the groups with zero or one risk allele. B) Linear regression analysis showing a multiplicative effect of the risk alleles of the SNP rs7582694 in *STAT4* and of the longer allele of the 5 bp insertion-deletion polymorphisms and the C-allele of the SNP rs10488631 in *IRF5* on SLE. The odds ratios (OR) with 95% confidence intervals are shown as a function of number of risk alleles for SLE. The slope of the line corresponds to a 1.82-fold increase in OR for each additional risk allele.

**Table 1. Haplotype analysis of ten single nucleotide polymorphisms in *STAT4* associated with systemic lupus erythematosus.**

| | Haplotypes[b] | Haplotype frequencies[a] | | p values[c] | Odds Ratio (95% CI) |
|---|---|---|---|---|---|
| | | Cases | Controls | | |
| H1 | GGCCGCTCGC | 0.50 | 0.59 | 5.8E-05 | 0.70 (0.59-0.84) |
| H2 | AATAATCGCG | 0.24 | 0.17 | 1.7E-05 | 1.59 (1.28-1.98) |
| H3 | GGTAACCCGC | 0.06 | 0.07 | 0.59 | 0.91 (0.64-1.29) |
| H4 | GGCAACCCGG | 0.04 | 0.06 | 0.028 | 0.62 (0.40-0.95) |
| H5 | GGCAACCGCG | 0.05 | 0.04 | 0.058 | 1.51 (0.98-2.31) |
| H6 | GGCCGCCCGG | 0.02 | 0.02 | 0.90 | 1.04 (0.55-1.95) |
| H7 | GGTAACCGCG | 0.02 | 0.01 | 0.29 | 1.44 (0.73-2.85) |

[a] Haplotype frequencies estimated using Haploview v. 4.0.

[b] The ten SNPs included in the analysis are in order from left to right: rs3024886, rs3821236, rs3024877, rs1517352, rs13017460, rs10168266, rs16833249, rs10181656, rs7582694, rs16833260,

[c] Unadjusted P-values by chi square test using Haploview v 4.0.

[d] Odds ratio and 95 % confidence interval calculated using the formula (a x d)/(b x c) where a and b are the haplotype counts in cases and c and d are the haplotype counts in controls.

**Table 2. Conditional logistic regression analysis of single nucleotide polymorphisms in the *STAT4* gene in systemic lupus erythematosus.**

| SNP ID | Unadjusted P-values | |
|---|---|---|
| | Single Marker[a] | Conditional[b] on rs7582694 |
| rs3024908 | 0.16 | 0.22 |
| rs3024904 | 0.17 | 0.24 |
| rs3024936 | 0.81 | 0.91 |
| rs3024935 | 0.83 | 0.92 |
| rs3024896 | 0.04 | 0.94 |
| rs925847 | 0.53 | 0.23 |
| rs3024886 | 3.0E-05 | 0.83 |
| rs3821236 | 4.8E-05 | 0.93 |
| rs3024877 | 5.9E-05 | 0.46 |
| rs12999858 | 0.76 | 0.51 |
| rs1400654 | 0.04 | 0.14 |
| rs3771327 | 0.02 | 0.09 |
| rs3024866 | 1.7E-03 | 0.65 |
| rs3024851 | 0.03 | 0.12 |
| rs932169 | 0.53 | 0.77 |
| rs1517352 | 4.4E-04 | 0.63 |
| rs13017460 | 5.1E-04 | 0.68 |
| rs13389408 | 0.04 | 0.44 |
| rs1400656 | 0.04 | 0.15 |
| rs10168266 | 5.1E-05 | 0.80 |
| rs7594501 | 0.04 | 0.21 |
| rs2459611 | 0.81 | 0.58 |
| rs7601754 | 0.11 | 0.79 |
| rs3024921 | 0.69 | 0.72 |
| rs16833249 | 2.8E-04 | 0.80 |
| rs12998748 | 0.48 | 0.76 |
| rs6434435 | 0.85 | 0.26 |

| | | |
|---|---|---|
| rs10181656 | 8.8E-08 | NA |
| rs7582694 | 8.8E-08 | NA |
| rs16833260 | 3.5E-05 | 0.55 |
| rs6752770 | 0.05 | 0.56 |
| rs7599504 | 0.27 | 0.60 |
| rs4853543 | 0.61 | 0.37 |
| rs4073699 | 0.44 | 0.46 |
| rs17769077 | 0.85 | 0.92 |
| rs4341967 | 0.73 | 0.28 |
| rs10931483 | 0.70 | 0.29 |
| rs4583497 | 0.58 | 0.38 |
| rs4555370 | 0.69 | 0.32 |
| rs10804037 | 0.69 | 0.31 |
| rs17769459 | 0.65 | 0.91 |
| rs6712821 | 0.63 | 0.36 |
| rs1551443 | 0.62 | 0.37 |
| rs7566274 | 0.61 | 0.39 |
| rs2356350 | 0.98 | 0.23 |
| rs10189819 | 0.57 | 0.62 |
| rs7596818 | 0.44 | 0.81 |
| rs11685878 | 0.58 | 0.99 |
| rs4853546 | 0.43 | 0.79 |
| rs1031509 | 0.75 | 0.93 |
| rs10497711 | 0.74 | 0.73 |
| rs7572482 | 0.78 | 0.77 |
| rs897200 | 0.93 | 0.95 |

[a] Unadjusted p-value for the association with SLE based on additive logistic regression.

[b] P-value for the association using additive logistic regression conditional on the indicated marker. NA indicates that the marker can not be distinguished from the conditional model.

**Table 3. Correlation between the *STAT4* risk allele of the single nucleotide polymorphism rs7582694 and disease phenotype**

| Clinical criterion | Frequency for all patients (n = 695) | Frequency for patients with | | p value[a] |
|---|---|---|---|---|
| | | 0 risk allele (n = 300) | 1 or 2 risk alleles (n = 395) | |
| Malar rash (ACR1)[b] | 0.59 | 0.60 | 0.59 | 0.70 |
| Discoid rash (ACR2) | 0.25 | 0.28 | 0.24 | 0.19 |
| Photosensitivity (ACR3) | 0.69 | 0.71 | 0.67 | 0.36 |
| Oronasal ulcers (ACR4) | 0.25 | 0.23 | 0.25 | 0.59 |
| Arthritis (ACR5) | 0.77 | 0.79 | 0.76 | 0.32 |
| Serositis (ACR6) | 0.45 | 0.42 | 0.47 | 0.17 |
| Nephritis (ACR7) | 0.32 | 0.27 | 0.35 | 0.040 |
| CNS (ACR8) | 0.12 | 0.13 | 0.11 | 0.72 |
| Hematological (ACR9) | 0.60 | 0.56 | 0.63 | 0.073 |
| Immunological (ACR10)[c] | 0.66 | 0.57 | 0.73 | $1.6 \times 10^{-5}$ |
| ANA (ACR11) | 0.98 | 0.98 | 0.98 | 0.79 |
| Anti-dsDNA Ab ever[c] | 0.54 | 0.43 | 0.62 | $5.3 \times 10^{-7}$ |
| Number of ACR criteria[d] | 5.6 ± 1.4 | 5.5 ± 1.3 | 5.8 ± 1.4 | 0.022 |
| Age at diagnosis[d] | 35.5 ± 14.9 | 36.7 ± 15.4 | 34.5 ± 14.5 | 0.11 |

[a] Frequency of patients with 0 risk alleles compared to patients with 1 or 2 risk alleles fulfilling each clinical criterion, unadjusted p-value calculated by Fisher´s exact test

[b] American College of Rheumatology classification criteria (36).

[c] Antibodies to double-stranded DNA (dsDNA) is part of the immunological criterion

[d] Mean value with one standard deviation

[e] Unadjusted p-value calculated with the Mann-Whitney U-test

**FIGURES**

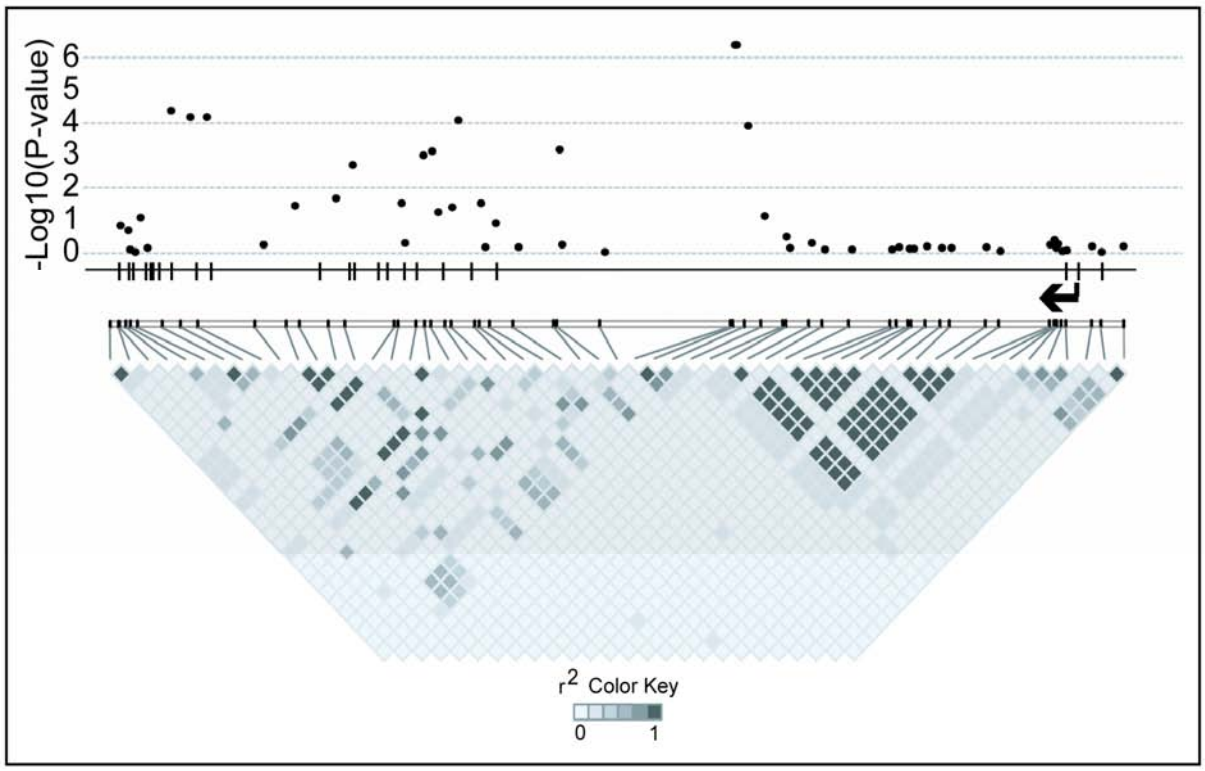



Figure 1

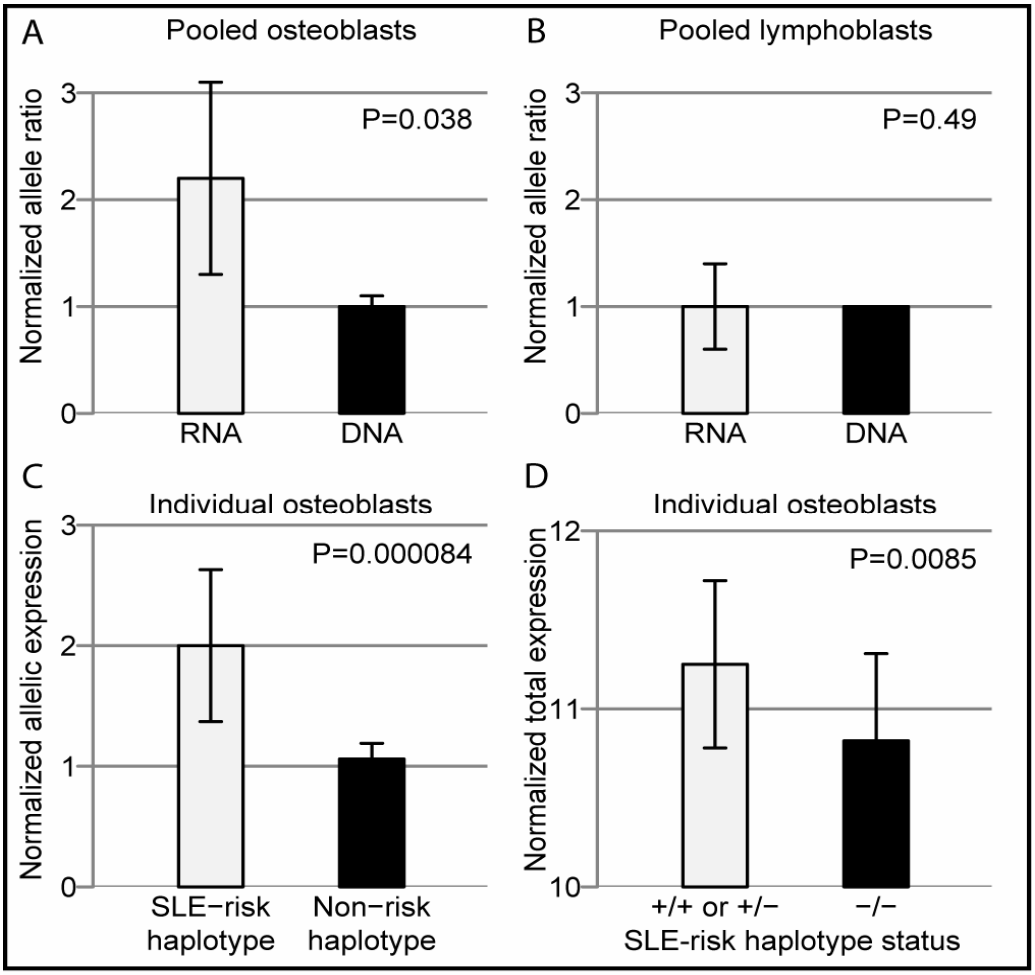
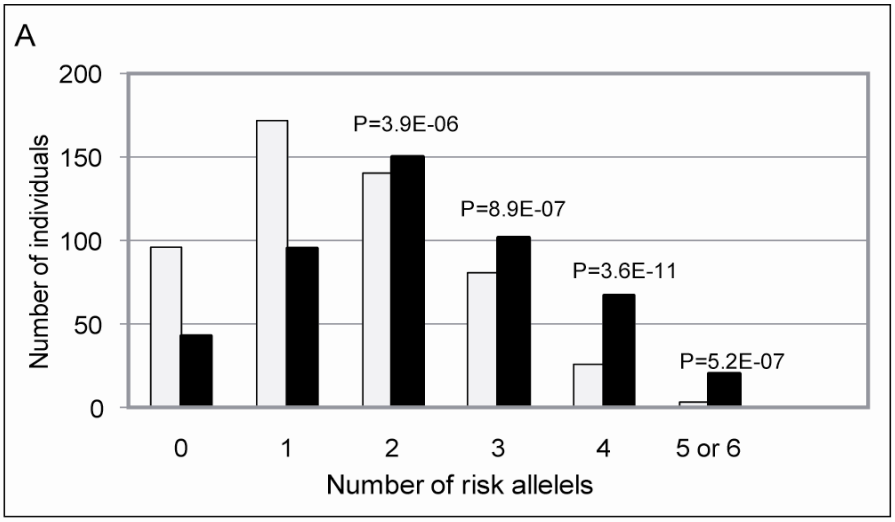



Figure 2



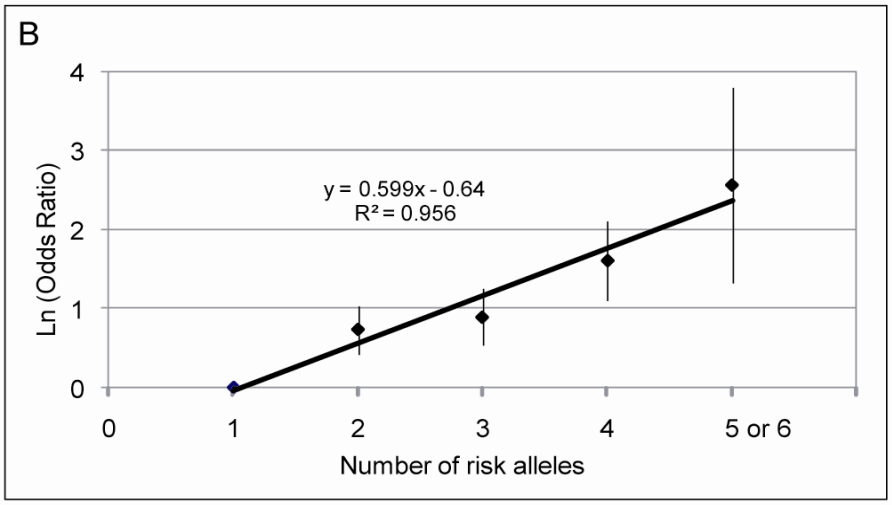




Figure 3

## ABBREVIATIONS

*Abbreviations*

ACR: American College of Rheumatology

CI: Confidence interval

CD: Crohn disease

DC: Dendritic cells

EBV: Epstein-Barr virus

HapMap: Haplotype mapping

IBD: Inflammatory bowel diseases

IFN: Interferon

IFNAR: Type I interferon receptor

IL: Interleukin

Indel: Insertion / deletion

IRF5: Interferon regulatory factor 5

MS: Multiple sclerosis

NC: Natural killer

OR: Odds ratio

PCR: Polymerase chain reaction

PDC: Plasmacytoid dendritic cells

RA: Rheumatoid arthritis

RT: Reverse transcriptase

SLE: Systemic lupus erythematosus

SNP: Single nucleotide polymorphism

STAT4: Signal transducer and activator of transcription 4

TLR: Toll-like receptor

TYK2: Tyrosine kinase 2