STAT4 and the Risk of Rheumatoid Arthritis and Systemic Lupus Erythematosus

Elaine F. Remmers, Ph.D., Robert M. Plenge, M.D., Ph.D., Annette T. Lee, Ph.D.,
Robert R. Graham, Ph.D., Geoffrey Hom, Ph.D., Timothy W. Behrens, M.D.,
Paul I.W. de Bakker, Ph.D., Julie M. Le, B.S., Hye-Soon Lee, M.D., Ph.D.,
Franak Batiwalla, Ph.D., Wentian Li, Ph.D., Seth L. Masters, Ph.D.,
Matthew G. Booty, B.S.; John P. Carulli, Ph.D., Leonid Padyukov, M.D., Ph.D.,
Lars Alfredsson, Ph.D., Lars Klareskog, M.D., Ph.D., Wei V. Chen, M.S.,
Christopher I. Amos, Ph.D., Lindsey A. Criswell, M.D., M.P.H.,
Matthew G. Booty, B.S., John P. Carulli, Ph.D., Leonid Padyukov, M.D., Ph.D.,
Lars Alfredsson, Ph.D., Lars Klareskog, M.D., Ph.D., Wei V. Chen, M.S.,
Christopher I. Amos, Ph.D., Lindsey A. Criswell, M.D., M.P.H.,
Michael F. Seldin, M.D., Ph.D., Daniel L. Kastner, M.D., Ph.D.,
and Peter K. Gregersen, M.D.

From the National Institute of Arthritis and Musculoskeletal and Skin Diseases, Bethesda, MD (E.F.R., J.M.L., S.L.M.,
M.G.B., D.L.K.); the Broad Institute, Cambridge, MA (R.M.P., R.R.G., P.I.W.B.);
Brigham and Women’s Hospital (R.M.P.);
the Feinstein Institute for Medical Research, Manhasset, NY (A.T.L., H.-S.L.,
F.B., W.L., P.K.G.); Genentech, South San Francisco, CA (G.H., T.W.B.);
Hanyang University College of Medicine, Seoul, South Korea (H.-S.L.); the Karolinska Institutet, Stockholm (L.P., L.A., L.K.);
the University of Texas M.D. Anderson Cancer Center, Houston (W.V.C., C.I.A.); the University
of California at San Francisco, San Francisco (W.V.C., C.I.A.); and the University of
California at Davis, Davis (M.F.S.). Address reprint requests to Dr. Gregersen at the Feinstein Institute for Medical Research, 350 Community Dr., Manhasset, NY 11030,
or at peterg@nshs.edu.

Drs. Remmers and Plenge contributed equally to this article.
RHEUMATOID ARTHRITIS IS THE MOST common cause of adult inflammatory arthritis and is associated with considerable disability and early mortality. Studies of twins clearly show a genetic contribution to disease susceptibility, and the siblings of patients with seropositive, erosive rheumatoid arthritis have an estimated risk of developing the disease of between 5 and 10 times that of the general population. The highly polymorphic HLA region is a major contributor to genetic risk of rheumatoid arthritis. Several other genes associated with more modest risks have recently been identified, including the Arg620→Trp variant of the intracellular phosphatase gene PTPN22. However, the definitive identification of additional risk genes outside the HLA region has been challenging.

We recently described a linkage peak with nearly genomewide significance on the long (q) arm of chromosome 2 in 642 families of European ancestry collected by the North American Rheumatoid Arthritis Consortium (NARAC). The region encompasses more than 50 million base pairs (Mb) of genomic DNA and has also been implicated in previous meta-analyses of linkage-study data. In the current study, we undertook a large case–control disease-association analysis of 13 selected candidate genes within the chromosome 2q linkage region.

METHODS

SUBJECTS

The NARAC case–control series included one affected member (the proband, if a DNA sample from the proband was available) of each family of European descent from the NARAC collection of affected sibling pairs and unrelated controls of self-identified European ancestry from the New York Cancer Project. The controls were additional subjects of self-reported European ancestry from the aforementioned collections. The controls were additional subjects of self-reported European ancestry from the New York Cancer Project.

The numbers of case patients and controls in the three rheumatoid arthritis and the three lupus case–control series are listed in Table 1. The institutional review boards of all investigative institutions approved these studies, and all participants provided written informed consent.

CANDIDATE GENES AND SELECTION OF SINGLE-NUCLEOTIDE POLYMORPHISMS

We selected candidate genes from within a linkage region of 52 Mb on chromosome 2q (the 2-LOD support interval) that we defined previously. For each selected gene, we initially used HapMap Phase I data to identify tag single-nucleotide polymorphisms (SNPs) that captured the majority of the then-known common SNP variation (i.e., that present on ≥5% of chromosomes) in the gene (defined as the sequence ranging from 10 kb upstream of the coding sequence to 10 kb downstream) using an r² threshold of 0.8 or greater. The r² correlation coefficient is a measure of linkage disequilibrium determined by the allelic correlation between SNPs; if r²=1, the markers are perfect predictors of one another.) Some SNPs were not genotyped directly; rather, their imputed genotypes were inferred from multimarker combinations (see Fig. S1 in the Supplementary Appen-
For fine mapping of the STAT1–STAT4 region, we used HapMap Phase II data to select additional SNPs that captured the majority of the known common variation in the region and that had a pairwise $r^2$ of more than 0.8. We also included all nonsynonymous coding SNPs reported in dbSNP, the SNP database of the National Center for Biotechnology Information; all SNPs within motifs conserved across species that were identified with the use of the University of California at Santa Cruz (UCSC) genome browser phastCons (conserved-elements track) (http://genome.ucsc.edu); and SNPs disrupting putative transcription-factor binding sites, identified with the use of the UCSC human–mouse–rat conserved transcription-factor binding sites track (http://genome.ucsc.edu).

**Genotyping**

DNA samples were obtained from all subjects. The samples were genotyped by means of one of two methods: a custom, highly multiplexed, bead-based array method, GoldenGate Genotyping (Illumina), and a multiplexed primer-extension method (SEQUENOM) (for details, see the Supplementary Appendix).

**Statistical Analysis**

All SNPs were tested for significant deviation from Hardy–Weinberg equilibrium in controls. Those with $P$ values of less than 0.005 were removed from the analysis. We also removed all SNPs with a minor allele frequency of less than 0.01, because of the reduced power to detect associations for rare SNPs. The remaining SNPs (including the imputed SNPs) were analyzed for an association with disease by means of comparison of the minor allele frequency in case patients and controls, with significance determined by means of a chi-square test. Odds ratios, and their 95% confidence intervals, for having the risk allele in chromosomes of case patients as compared with those of controls were also determined for selected SNPs.23 When combining data from different case–control series, we used a Mantel–Haenszel test (SPSS software, version 12.0.0; www.spss.com) to summarize the stratum-specific estimates.

Linkage-disequilibrium patterns in the STAT1–STAT4 region were determined with the use of Haploview software, version 3.32.24 The genotypes of 768 SNPs informative about European ancestry25 were used to adjust for the possibility of unmatched population structure in the case patients and the controls; STRAT software was used for structured association26 and EIGENSTRAT software for the correction of association-study results according to a method based on principal-components analysis.27

**Results**

**Candidate-Gene Screening in the Chromosome 2q Linkage Region**

We examined the 2-LOD support interval (Fig. 1A) of the previously identified7 linkage peak on chromosome 2 for the presence of genes that might influence rheumatoid arthritis. We evaluated 13 candidate genes (Fig. 1B; for further description, see Table S1 in the Supplementary Appendix). Association results for 82 tag or imputed SNPs within the selected candidate genes from an initial set of 525 independent case patients with rheumatoid arthritis and 1165 unrelated controls are shown in Figure 1B. In addition to a known association with a SNP in CTLA428 (rs3087243, $P=0.008$), we found an association with an unlinked SNP (located 15 Mb away; $r^2=0$) in STAT4 (rs7574865, $P=0.002$).

**Fine Mapping of Associations with Rheumatoid Arthritis in the STAT1–STAT4 Region**

The most significantly ($P=0.002$) associated SNP in the region, rs7574865, is in a linkage-disequilibrium block that extends from the middle of the STAT4 locus to the 3′ end of the gene (Fig. 2).
was, however, some evidence of longer-range disequilibrium that extended into STAT1 from the 3′ end of STAT4. We therefore included both genes in the fine mapping and in further analyses.

To map the location of the association with rheumatoid arthritis, we successfully genotyped the case patients and controls in the NARAC series for 63 SNPs located within the 209-kb STAT1–STAT4 region (average density, one SNP per 3.1 kb). These 63 SNPs captured 87% of the common variation (defined as a minor allele frequency of ≥0.05) in the HapMap Phase II data in the region.
with an $r^2$ value of more than 0.8. Four SNPs located within the large third intron of \textit{STAT4} had associations with rheumatoid arthritis with $P$ values of less than 0.001. The most significant $P$ value, $8.29 \times 10^{-5}$, was found for rs7574865 (Table 2). The four disease-associated SNPs were in strong linkage disequilibrium ($r^2 > 0.97$), and all had a minor allele frequency of 0.28 in the NARAC case patients with rheumatoid arthritis, as compared with 0.22 in the unrelated controls. Results for the complete set of 63 SNPs are given in Table S2 in the Supplementary Appendix.

**Replication of Associations of Variants in \textit{STAT4} with Rheumatoid Arthritis**

To confirm the associations found in the \textit{STAT1–STAT4} region, we genotyped subjects in the rheumatoid arthritis replication series for the same 63 SNPs. Among the case patients, we genotyped only those who were positive for anti–cyclic citrullinated peptide antibody, to minimize disease heterogeneity in this singleton case series. Four variants within intron 3 of \textit{STAT4} — the same four identified in our initial findings — were strongly associated with rheumatoid arthritis (e.g., rs7574865, $P = 6.26 \times 10^{-4}$) (Table 2). The complete results for the rheumatoid arthritis replication series are listed in Table S3 in the Supplementary Appendix.

We also performed analyses of the 63 SNPs in the combined NARAC and rheumatoid arthritis replication series (Fig. 3). In a combined Mantel–Haenszel analysis, the SNP most strongly associated with rheumatoid arthritis in the NARAC series, rs7574865, had a minor allele frequency of 0.27 in case patients and 0.22 in controls ($P = 2.81 \times 10^{-7}$; odds ratio for having the risk allele in chromosomes of patients vs. those of controls, 1.32; 95% confidence interval [CI], 1.19 to 1.46).

We genotyped the most significantly associated SNP from the NARAC case–control series, rs7574865, in 1529 case patients with recent-onset rheumatoid arthritis and in 881 controls from the Swedish Epidemiological Investigation of Rheumatoid Arthritis series. In this independent series, the minor allele frequency of rs7574865 was significantly greater in the case patients than in the controls ($P = 0.02$) (Table 2). The minor allele frequency for rs7574865 was lower in the Swedish patients with early rheumatoid arthritis (0.25) than in the North American patients with established rheumatoid arthritis (0.27), whereas the frequency in the controls was the same in both series (0.22).

A meta-analysis of the three independent case–control series for rheumatoid arthritis yielded strong evidence of an association of the minor allele of rs7574865 with disease susceptibility ($P = 4.64 \times 10^{-8}$). The odds ratio for having the risk allele in chromosomes of case patients as compared with those of controls was 1.27 (Table 2). Genotypic odds ratios for patients as compared with controls were 1.61 (95% CI, 1.28 to 2.03) for homozygotes and 1.27 (95% CI, 1.14 to 1.41) for heterozygotes.

In the NARAC patients with rheumatoid arthritis, of which 81% were positive for anti–cyclic citrullinated peptide antibody, the rs7574865 minor allele frequency did not differ significantly ($P > 0.05$) in the subgroup that was positive for the antibody (0.28) and the subgroup that was negative for the antibody (0.27). Logistic-regression analysis after accounting for the rs7574865 genotype in the combined NARAC and rheumatoid arthritis replication series showed that this one SNP could explain the signal across the \textit{STAT1–STAT4} region (data not shown). Furthermore, after accounting for the \textit{CTLA4} SNP associated with rheumatoid arthritis (rs3087243), the result for the
STAT4 rs7574865 remained significant. Thus, we concluded that the STAT4 SNP, or a variant in tight linkage disequilibrium with it, confers increased susceptibility to the development of rheumatoid arthritis.

**CORRECTION FOR DIFFERENCES BETWEEN CASE PATIENTS AND CONTROLS**

To address the possibility that case–control analyses may yield spurious associations due to undetected differences in population admixture or population substructure between case patients and controls, we genotyped the rheumatoid arthritis replication series for 768 SNPs informative about European ancestry, located throughout the genome. There was still strong evidence of association according to a structured association analysis (with STRAT software) (P=5×10⁻⁵) and an analysis using EIGENSTRAT software with correction for the four most significant principal components (P=2×10⁻⁴). Furthermore, when the genotypes of the SNPs informative about European ancestry were used to distinguish controls of predominantly northern European ancestry from those of predominantly southern European ancestry, we found an rs7574865 minor allele frequency of 0.22 in both groups, indicating that this allele frequency does not vary significantly between these subgroups (P>0.05).

**ASSOCIATION OF THE STAT4 VARIANT WITH SYSTEMIC LUPUS ERYTHEMATOSUS**

Since STAT4 lies within linkage peaks that have also been reported in patients with lupus, three lupus series of case and control subjects of European ancestry were also genotyped. We found that the minor allele frequency for rs7574865 was significantly increased in all three series among patients (0.29 to 0.31) as compared with controls (0.22 to 0.23) (P=9.56×10⁻⁶ to P=0.03) (Table 3). In a meta-analysis of the three series, we found strong evidence of association of the rs7574865 minor allele with lupus (P=1.87×10⁻⁹). The odds ratio for having the allele associated with lupus in chromosomes of patients as compared with those of controls was 1.55 (Table 3). Genotypic odds ratios were 2.41 (95% CI, 1.66 to 3.49) for homozygotes and 1.56 (95% CI, 1.30 to 1.88) for heterozygotes.

**DISCUSSION**

We have shown that a variant allele of STAT4 confers an increased risk for both rheumatoid arthritis and systemic lupus erythematosus. This finding provides support for the evolving concept that common risk genes underlie multiple autoimmune disorders and suggests the involvement of common pathways of pathogenesis among these different diseases.

STAT4 encodes a transcription factor that transmits signals induced by several key cytokines, including interleukin-12 and type 1 interferons, as well as interleukin-23. STAT4 is a latent cytosolic factor that, after activation by cytokines, is phosphorylated and accumulates in the nucleus. Activated STAT4 stimulates transcription of specific genes including interferon-γ, a key indicator of T-cell differentiation into type 1 helper T (Th1) cells. Therefore, STAT4-dependent signaling by interleukin-12 receptors plays a critical role in the development of a Th1-type T-cell response.
STAT4 has also been implicated in the optimal differentiation of a newly defined CD4+ T-cell lineage, designated Th17 cells. Dependent in part on the activity of interleukin-23, a cytokine related to interleukin-12, proinflammatory Th17 cells can play an important, if not predominant, role in chronic inflammatory disorders. Indeed, experiments that have targeted Th1 cells in models of autoimmune disease have often unwittingly targeted the Th17 lineage, because the key cytokines of the two lineages, interleukin-12 and interleukin-23, and their receptors share common subunits.

STAT4, a central player in both lineages, has proved to play a crucial role in experimental models of autoimmunity. STAT4-deficient mice are generally resistant to models of autoimmune disease, including arthritis. Furthermore, specific targeting of STAT4 by inhibitory oligodeoxynucleotides or antisense oligonucleotides can ameliorate disease in arthritis models, suggesting the utility of STAT4 as a therapeutic target.

Recent genetic data have shown that interleukin-23–receptor variants are associated with susceptibility to both Crohn’s disease and psoriasis; interleukin-12β polymorphisms have also been associated with a risk of psoriasis. Since both interleukin-12 and interleukin-23 act through STAT4, these data imply that a complex pattern of alterations in related pathways can lead to various forms of autoimmunity and chronic inflammation. STAT4 is also required for signaling in mature dendritic cells in response to type 1 interferons. Thus, there may be multiple mechanisms by which genetic variation in STAT4 can influence immune responses and predispose persons to autoimmunity. Indeed, in a murine model of lupus, STAT4 deficiency is associated with accelerated nephritis and increased mortality, in contrast to the protective effects in arthritis models.

Several family-based genome scans have revealed linkage of the chromosome 2q region with lupus, as well as with rheumatoid arthritis. We therefore extended our association studies to three independent lupus case–control series and found strong evidence that the STAT4 variant associated with rheumatoid arthritis was also associated with lupus. The identification of STAT4 as a common predisposition gene for both lupus and rheumatoid arthritis is similar to reported findings of broad associations of the intracellular phosphatase PTPN22 with these and other autoimmune diseases, such as type 1 diabetes mellitus, autoimmune thyroid disease, and myasthenia gravis. Clearly, the role of STAT4 in these other disorders should be examined. In addition, the influence of allelic variation on subgroups, manifestations, and outcomes of disease may shed further light on disease mechanisms. The majority of the North American patients with rheumatoid arthritis in our study had long-standing erosive disease, whereas the Swedish patients generally had disease of more recent onset. This may explain the somewhat weaker STAT4 association in the Swedish series. Given that lupus is a highly heterogeneous disorder, it will be important to study STAT4 polymorphisms in clinical subgroups, and in view of the knockout mouse data, this is particularly true with regard to the development of nephritis.

Genetic case–control studies such as ours must be carried out with careful attention to the possibility of false positive results, as has been emphasized elsewhere. First, multiple replication is essential for certainty about the basic findings. In addition, studies using unrelated case–control series run the risk of yielding spurious associations if there are unidentified differences in population structure between case patients and controls. To address this possibility, we genotyped the 1013 SNPs with Disease Susceptibility in Two North American Rheumatoid Arthritis Case–Control Series.

The significance of the associations (presented as 1 divided by the P value) is shown for 63 SNPs in the STAT1–STAT4 region, shown according to chromosomal position (from the March 2006 assembly of the University of California at Santa Cruz Genome Browser [hg18], National Center for Biotechnology Information build 36). The five SNPs with P values of less than 1×10⁻⁸ are labeled. The locations of the STAT1 and STAT4 genes are indicated by the blue arrows (pointing in the direction of transcription). Vertical black bars on the arrows represent the locations of exons. The association data are from the 1620 case patients and the 2635 controls in the North American Rheumatoid Arthritis Consortium series and the rheumatoid arthritis replication series.
case patients and the 1326 controls in the rheumatoid arthritis replication series for 768 ancestry-informative SNPs that were selected for reflecting differences in allele frequency among European subgroups and used two methods to control for such stratification. These methods did not reduce the significance of the association with disease. The association was probably robust to this correction because the allele frequency of the disease-associated STAT4 SNP does not vary among European subgroups.

Association studies cannot distinguish among multiple variants in strong linkage disequilibrium with one another, and a haplotype containing several variants could be required to confer a biologic effect. SNPs known to be in strong linkage disequilibrium with rs7574865, on the basis of HapMap CEU data, are listed in Table S4 in the Supplementary Appendix. All these variants are located in the third intron of the STAT4 gene, suggesting that splice variation or regulatory effects may explain the gene’s association with disease. Studies are under way to investigate these possibilities in various types of cells, including T cells, monocytes, macrophages, and dendritic cells. In addition, a complete resequencing of the STAT4 gene may yet reveal additional risk alleles. Nevertheless, even in the absence of precise molecular mechanisms, the discovery of these new disease associations with STAT4 should generate a variety of new hypotheses about the pathogenesis of autoimmunity.

Note added in proof: The association of STAT4 with rheumatoid arthritis has now been replicated in a Korean population.51

Supported by grants from the National Institutes of Health (N01-AR-1-2256, N01-AI95386, R01 AR44422, and N01-AR2-2263, to Dr. Gregersen; R01 AR050267, to Dr. Seldin; and R01 AR052300 and R01 AR44804, to Dr. Criswell), the Rosalind Russell Medical Research Center for Arthritis and the Kirkland Scholar Award (to Dr. Criswell), the Arthritis Foundation, the Boas Family, the Eileen Ludwig Greenland Center for Rheumatoid Arthritis, and the Intramural Research Program of the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The work was carried out in part at the General Clinical Research Center, Moffitt Hospital, University of California at San Francisco (UCSF), and at the General Clinical Research Center Feinstein Institute for Medical Research (FIMR), with grants from the National Center for Research Resources, Public Health Service (5-M01-RR-00079, to UCSF, and M01 RR018535). Dr. Plenge reports receiving consulting fees from Biogen Idec and lecture fees from Genentech. Drs. Hom and Behrens report being employees of Genentech; and Dr. Carulli, an employee of Biogen Idec. Dr. Criswell reports receiving consulting fees from Celera Diagnostics. Dr. Gregersen reports serving on the Abbott Scholar Award Advisory Committee and receiving honoraria from Biogen Idec, Genentech, and Roche Pharmaceuticals. No other potential conflict of interest relevant to this article was reported.

We thank the large number of investigators, practicing physicians, and research nurses who contributed data from their patients to the various collections used in our studies, including Dr. Michelle Petri for the Autoimmune Biomarkers Collaborative Network (ABCoN); Drs. Elena Massarotti, Claire Bom bardier, and Michael Weisman for the Study of New Onset Rheumatoid Arthritis (SONORA); the Multiple Autoimmune Diseases Genetics Consortium (MADGC); Marlena Kern, R.N., for the North American Rheumatoid Arthritis Consortium (NARAC); and Dr. Frederick Wolfe for the National Data Bank for Rheumatic Diseases, Wichita, Kansas, and the National Inception Cohort of Rheumatoid Arthritis patients. We also thank Dr. John O’Shea for thoughtful comments on an earlier version of this manuscript.

### Table 3. Associations of the SNP rs7574865 with Systemic Lupus Erythematosus.

<table>
<thead>
<tr>
<th>Case–Control Series</th>
<th>Allele</th>
<th>Minor Allele Frequency</th>
<th>Chi-Square Value</th>
<th>P Value</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UCSF</td>
<td>G</td>
<td>0.69</td>
<td>19.60</td>
<td>9.56x10^-6</td>
<td>1.61 (1.30–1.99)</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCoN</td>
<td>G</td>
<td>0.69</td>
<td>12.43</td>
<td>4.22x10^-4</td>
<td>1.52 (1.20–1.91)</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MADGC</td>
<td>G</td>
<td>0.71</td>
<td>4.66</td>
<td>0.03</td>
<td>1.46 (1.03–2.06)</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lupus meta-analysis</td>
<td>G</td>
<td>0.69</td>
<td>1.87x10^-9</td>
<td>1.55</td>
<td>1.34–1.79</td>
</tr>
<tr>
<td></td>
<td>T</td>
<td>0.31</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* UCSF denotes the University of California at San Francisco, ABCoN the Autoimmune Biomarkers Collaborative Network, and MADGC the Multiple Autoimmune Diseases Genetics Consortium. The lupus meta-analysis was a Mantel–Haenszel analysis; no significant evidence of heterogeneity was identified among the series (P>0.05).
40. Hildner KM, Schirmacher P, Aretya I, et al. Targeting of the transcription factor STAT4 by antisense phosphorothioate oligonucleotides suppresses collagen-