ORIGINAL ARTICLE

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 Batliwalla, 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.,
Michael F. Seldin, M.D., Ph.D., Daniel L. Kastner, M.D., Ph.D.,
and Peter K. Gregersen, M.D.

ABSTRACT

BACKGROUND

Rheumatoid arthritis is a chronic inflammatory disease with a substantial genetic component. Susceptibility to disease has been linked with a region on chromosome 2q.

METHODS

We tested single-nucleotide polymorphisms (SNPs) in and around 13 candidate genes — located within the previously linked chromosome 2q region — for association with rheumatoid arthritis. We then performed fine mapping of the *STAT1–STAT4* region in a total of 1620 case patients with established rheumatoid arthritis and 2635 controls, all from North America. Implicated SNPs were further tested in an independent case–control series of 1529 patients with early rheumatoid arthritis and 881 controls, all from Sweden, and in a total of 1039 case patients and 1248 controls from three systemic lupus erythematosus series.

RESULTS

A SNP haplotype in the third intron of *STAT4* were associated with susceptibility to both rheumatoid arthritis and systemic lupus erythematosus. Their minor alleles were present in 27% of chromosomes of patients with established rheumatoid arthritis, as compared with 22% of those of controls (for the SNP rs7574865, $P=2.81\times10^{-7}$; odds ratio for having the risk allele in chromosomes of patients vs. those of controls, 1.32). The association was replicated in Swedish patients with recent-onset rheumatoid arthritis ($P=0.02$) and matched controls. The haplotype marked by rs7574865 was strongly associated with lupus, being present on 31% of chromosomes of case patients and 22% of those of controls ($P=1.87\times10^{-9}$; odds ratio for having the risk allele in chromosomes of patients vs. those of controls, 1.55). Homozygosity of the risk allele, as compared with absence of the allele, was associated with a more than doubled risk for lupus and a 60% increased risk for rheumatoid arthritis.

CONCLUSIONS

A haplotype of *STAT4* is associated with increased risk for both rheumatoid arthritis and systemic lupus erythematosus, suggesting a shared pathway for these illnesses.

RHEUMATOID ARTHRITIS IS THE MOST common cause of adult inflammatory arthritis and is associated with considerable disability and early mortality.[1] Studies of twins clearly show a genetic contribution to disease susceptibility,[2] 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.[3] The highly polymorphic *HLA* region is a major contributor to genetic risk of rheumatoid arthritis.[4] Several other genes associated with more modest risks have recently been identified, including the Arg620→Trp variant of the intracellular phosphatase gene *PTPN22*.[5,6] 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[7] collected by the North American Rheumatoid Arthritis Consortium (NARAC).[8] 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.[9,10] 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[7,8] and unrelated controls of self-identified European ancestry from the New York Cancer Project[11] (www.amdec.org/amdec_initiatives/nycp.html). The rheumatoid-arthritis replication series consisted of singleton case patients with rheumatoid arthritis who were positive for anti–cyclic citrullinated peptide antibody, obtained through the Wichita Rheumatic Disease Data Bank,[12] the National Inception Cohort of Rheumatoid Arthritis Patients,[13] and the Study of New Onset Rheumatoid Arthritis,[14] and additional unrelated controls of European descent obtained from the New York Cancer Project. The Swedish case–control series included case patients and controls from the Epidemiological Investigation of Rheumatoid Arthritis Swedish inception cohort.[15,16]

Case patients with lupus were obtained from three sources: the University of California San Francisco (UCSF) patients were participants in the UCSF Lupus Genetics Project[17] and were recruited from UCSF Arthritis Clinics or from private rheumatology practices in northern California or by means of nationwide outreach. Medical records were reviewed to confirm that subjects met the criteria of the American College of Rheumatology for lupus.[18] The Autoimmune Biomarkers Collaborative Network (ABCoN) patients were recruited from the Hopkins Lupus Cohort[19] under the auspices of ABCoN[20] and also met the criteria of the American College of Rheumatology for lupus. Data from the Hopkins historical data base were used to determine fulfillment of ACR criteria. The Multiple Autoimmune Diseases Genetics Consortium (MADGC) patients were part of the MADGC collection.[21] The diagnosis of lupus based on ACR criteria was confirmed either by the treating physician or by the review of medical records. These three series included only case patients of self-described European descent 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.[7] 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 r^2 threshold of 0.8 or greater. (The r^2 correlation coefficient is a measure of linkage disequilibrium determined by the allelic correlation between SNPs; if $r^2 = 1$, the markers are perfect predictors of one another.) Some SNPs were not genotyped directly; rather, their imputed genotypes were inferred from multimarker combinations[22] (see Fig. S1 in the Supplementary Appen-

dix, available with the full text of this article at www.nejm.org).

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 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 ancestry[25] were used to adjust for the possibility of un-

Table 1. Case–Control Series for Rheumatoid Arthritis and Systemic Lupus Erythematosus.*

Series	No. of Case Patients	No. of Controls
Rheumatoid Arthritis		
NARAC	607	1309
Replication series	1013	1326
EIRA	1529	881
Lupus†		
UCSF	575	416
ABCoN	349	416
MADGC	115	416

* NARAC denotes the North American Rheumatoid Arthritis Consortium, EIRA the Epidemiological Investigation of Rheumatoid Arthritis, UCSF the University of California San Francisco, ABCoN the Autoimmune Biomarkers Collaborative Network, and MADGC the Multiple Autoimmune Diseases Genetics Consortium.
† The three sets of controls for the three lupus case–control series are 1248 independent samples that were randomly assigned to the three series.

matched population structure in the case patients and the controls; STRAT software was used for structured association[26] and EIGENSTRAT software for the correction of association-study results according to a method based on principal-components analysis.[27]
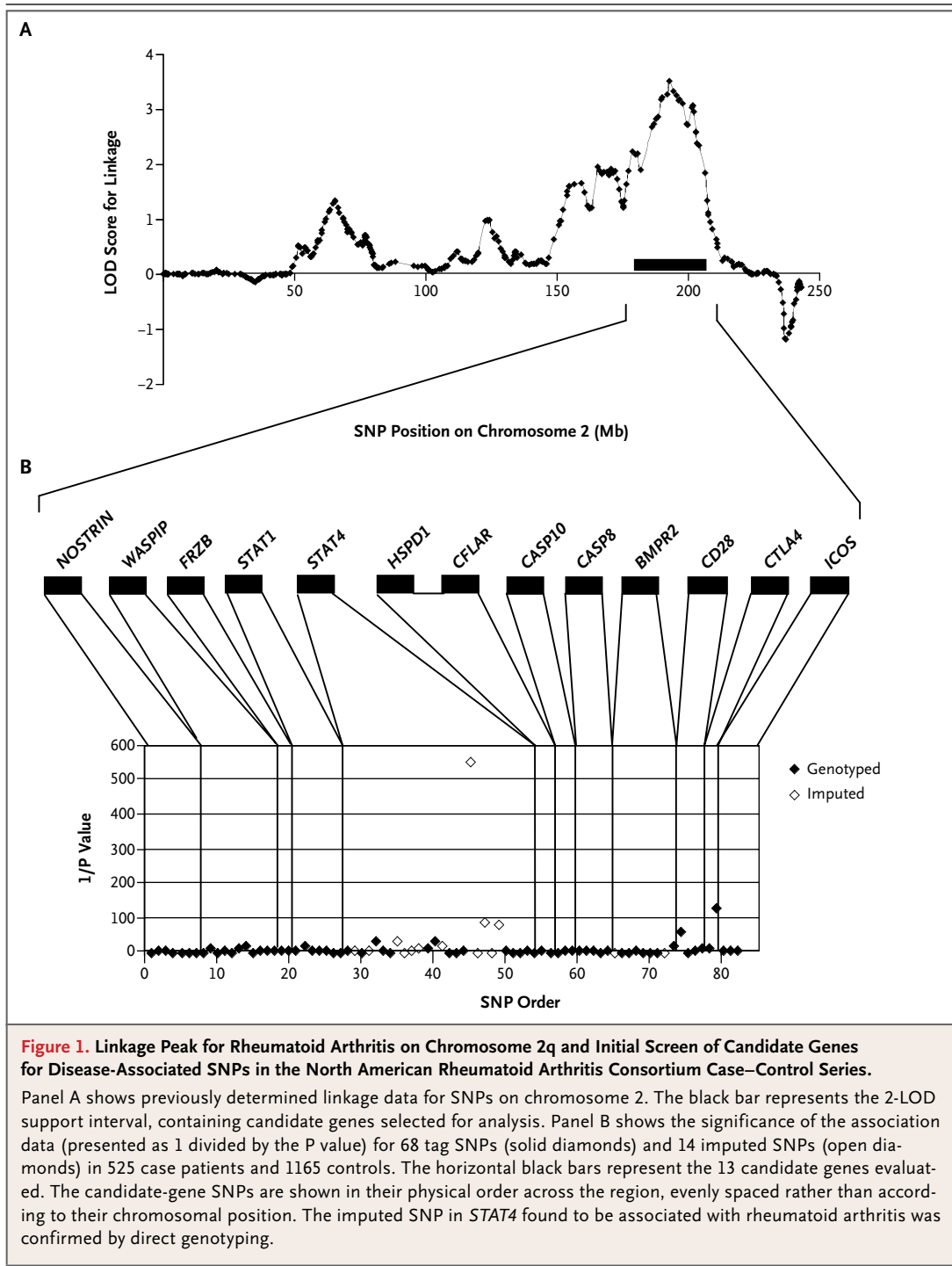
RESULTS

CANDIDATE-GENE SCREENING IN THE CHROMOSOME 2q LINKAGE REGION

We examined the 2-LOD support interval (Fig. 1A) of the previously identified[7] 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 CTLA4[28] (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<0.001) 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). There

Figure 1. Linkage Peak for Rheumatoid Arthritis on Chromosome 2q and Initial Screen of Candidate Genes for Disease-Associated SNPs in the North American Rheumatoid Arthritis Consortium Case–Control Series.

Panel A shows previously determined linkage data for SNPs on chromosome 2. The black bar represents the 2-LOD support interval, containing candidate genes selected for analysis. Panel B shows the significance of the association data (presented as 1 divided by the P value) for 68 tag SNPs (solid diamonds) and 14 imputed SNPs (open diamonds) in 525 case patients and 1165 controls. The horizontal black bars represent the 13 candidate genes evaluated. The candidate-gene SNPs are shown in their physical order across the region, evenly spaced rather than according to their chromosomal position. The imputed SNP in *STAT4* found to be associated with rheumatoid arthritis was confirmed by direct genotyping.

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 *STAT4* had associations with rheumatoid arthritis with P values of less than 0.001. The most significant P value, 8.29×10^{-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 *STAT4* WITH RHEUMATOID ARTHRITIS

To confirm the associations found in the *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 *STAT4* — the same four identified in our initial findings — were strongly associated with rheumatoid arthritis (e.g., rs7574865, P=6.26×10^{-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= 2.81×10^{-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–
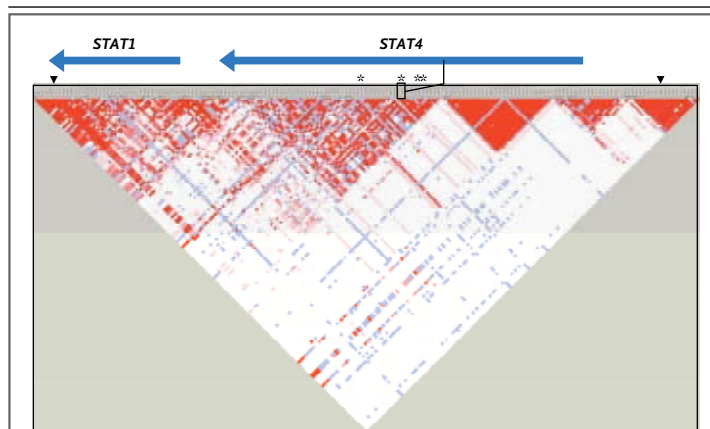
Figure 2. Linkage Disequilibrium Surrounding the *STAT4* SNP Found to be Associated with Rheumatoid Arthritis.

Linkage disequilibrium between pairs of SNPs across the *STAT1*–*STAT4* region is shown within the HapMap CEU population data (from persons of northern and western European ancestry). Blocks connecting pairs of SNPs are shaded according to the strength of the linkage disequilibrium between the SNPs, from 0.0 (white) to 1.0 (bright red), as measured by the disequilibrium coefficient D′. The lavender blocks indicate pairs of markers for which D′ is equal to 1.0 but the LOD score is less than 2.0. The imputed *STAT4* SNP, rs7574865, associated with rheumatoid arthritis is shown within the black rectangle. This SNP and three others (asterisks) were associated with disease susceptibility in both North American rheumatoid-arthritis case–control series. The locations of the *STAT1* and *STAT4* genes are indicated by the blue arrows (pointing in the direction of transcription). The region selected for fine mapping is located between the black arrowheads.

control series for rheumatoid arthritis yielded strong evidence of an association of the minor allele of rs7574865 with disease susceptibility (P=4.64×10^{-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 *STAT1*–*STAT4* region (data not shown). Furthermore, after accounting for the *CTLA4* SNP associated with rheumatoid arthritis (rs3087243), the result for the

Table 2. Associations of the SNP rs7574865 with Rheumatoid Arthritis.*

Case–Control Series	Allele	Minor Allele Frequency		Chi-Square Value	P Value	Odds Ratio (95% CI)
		Case Patients	Controls			
NARAC	G	0.72	0.78			
	T	0.28	0.22	15.49	8.29×10^{-5}	1.37 (1.17–1.60)
Replication series	G	0.74	0.78			
	T	0.26	0.22	11.70	6.26×10^{-4}	1.28 (1.11–1.47)
EIRA	G	0.75	0.78			
	T	0.25	0.22	5.08	0.02	1.18 (1.02–1.36)
Meta-analysis	G	0.74	0.78			
	T	0.26	0.22		4.64×10^{-8}	1.27 (1.16–1.36)

* NARAC denotes the North American Rheumatoid Arthritis Consortium (NARAC), and EIRA Epidemiological Investigation of Rheumatoid Arthritis. The RA meta-analysis was a Mantel–Haenszel analysis; no significant evidence (P>0.05) of heterogeneity was identified among the series.

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 IN POPULATION SUBSTRUCTURE 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) ($P=5\times10^{-5}$) and an analysis using EIGENSTRAT software with correction for the four most significant principal components ($P=2\times10^{-5}$). 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,[29-31] three lu-

pus 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) ($P=9.56\times10^{-6}$ to 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 ($P=1.87\times10^{-9}$). 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.[32]

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.[33] 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.[34,35]

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,[36] proinflammatory Th17 cells can play an important, if not predominant, role in chronic inflammatory disorders.[37] 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.[33]

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.[38] Furthermore, specific targeting of STAT4 by inhibitory oligodeoxynucleotides or antisense oligonucleotides can ameliorate disease in arthritis models,[39,40] 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[41] and psoriasis[42]; interleukin-12β polymorphisms have also been associated with a risk of psoriasis.[42,43] 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.[44,45] 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,[46] in contrast to the protective effects in arthritis models.[38]

Several family-based genome scans have revealed linkage of the chromosome 2q region with lupus, as well as with rheumatoid arthritis.[29-31] 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
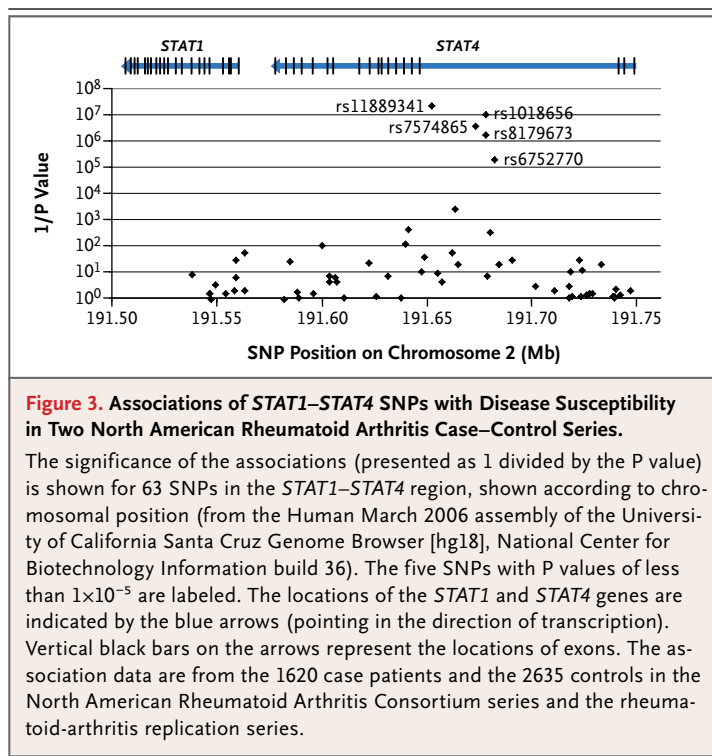
Figure 3. **Associations of *STAT1–STAT4* SNPs with Disease Susceptibility in Two North American Rheumatoid Arthritis Case–Control Series.**
The significance of the associations (presented as 1 divided by the P value) is shown for 63 SNPs in the *STAT1–STAT4* region, shown according to chromosomal position (from the Human March 2006 assembly of the University of California Santa Cruz Genome Browser [hg18], National Center for Biotechnology Information build 36). The five SNPs with P values of less than 1×10^{-5} 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.

PTPN22 with these and other autoimmune diseases,[6] such as type 1 diabetes mellitus,[47] autoimmune thyroid disease,[21] and myasthenia gravis.[48] 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.[49] 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.[25,50]

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

Case–Control Series	Allele	Minor Allele Frequency		Chi-Square Value	P Value	Odds Ratio (95% CI)
		Case Patients	Controls			
UCSF	G	0.69	0.78			
	T	0.31	0.22	19.60	9.56×10^{-6}	1.61 (1.30–1.99)
ABCoN	G	0.69	0.77			
	T	0.31	0.23	12.43	4.22×10^{-4}	1.52 (1.20–1.91)
MADGC	G	0.71	0.78			
	T	0.29	0.22	4.66	0.03	1.46 (1.03–2.06)
Lupus meta-analysis	G	0.69	0.78			
	T	0.31	0.22		1.87×10^{-9}	1.55 (1.34–1.79)

* UCSF denotes the University of California 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 (P>0.05) of heterogeneity was identified among the series.

To address this possibility, we genotyped the 1013 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[25] 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 au-

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

REFERENCES

1. Goronzy J, Weyand C. Rheumatoid arthritis. A. Epidemiology, pathology, and pathogenesis. In: Klippel J, Crofford L, Stone J, Weyand C, eds. Primer on the rheumatic diseases. 12th ed. Atlanta: Arthritis Foundation, 2001:209-17.

2. MacGregor AJ, Snieder H, Rigby AS, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis Rheum 2000;43:30-7.

3. Seldin MF, Amos CI, Ward R, Gregersen PK. The genetics revolution and the assault on rheumatoid arthritis. Arthritis Rheum 1999;42:1071-9.

4. Wordsworth BP, Bell JI. The immunogenetics of rheumatoid arthritis. Springer Semin Immunopathol 1992;14:59-78.

5. Begovich AB, Carlton VE, Honigberg LA, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330-7.

6. Gregersen PK, Lee HS, Batliwalla F, Begovich AB. PTPN22: setting thresholds for autoimmunity. Semin Immunol 2006; 18:214-23.

7. Amos CI, Chen WV, Lee A, et al. High-density SNP analysis of 642 Caucasian families with rheumatoid arthritis identifies two new linkage regions on 11p12 and 2q33. Genes Immun 2006;7:277-86.

8. Jawaheer D, Seldin MF, Amos CI, et al. Screening the genome for rheumatoid arthritis susceptibility genes: a replication study and combined analysis of 512 multicase families. Arthritis Rheum 2003;48: 906-16.

9. Choi SJ, Rho YH, Ji JD, Song GG, Lee YH. Genome scan meta-analysis of rheumatoid arthritis. Rheumatology (Oxford) 2006;45:166-70.

10. Etzel CJ, Chen WV, Shepard N, et al. Genome-wide meta-analysis for rheumatoid arthritis. Hum Genet 2006;119:634-41.

11. Mitchell MK, Gregersen PK, Johnson S, Parsons R, Vlahov D. The New York Cancer Project: rationale, organization, design, and baseline characteristics. J Urban Health 2004;81:301-10.

12. Wolfe F, Michaud K, Gefeller O, Choi HK. Predicting mortality in patients with rheumatoid arthritis. Arthritis Rheum 2003;48:1530-42.

13. Fries JF, Wolfe F, Apple R, et al. HLA-DRB1 genotype associations in 793 white patients from a rheumatoid arthritis inception cohort: frequency, severity, and treatment bias. Arthritis Rheum 2002;46: 2320-9.

14. Irigoyen P, Lee AT, Wener MH, et al. Regulation of anti-cyclic citrullinated peptide antibodies in rheumatoid arthritis: contrasting effects of HLA-DR3 and the shared epitope alleles. Arthritis Rheum 2005;52:3813-8.

15. Stolt P, Bengtsson C, Nordmark B, et al. Quantification of the influence of cigarette smoking on rheumatoid arthritis: results from a population based case-control study, using incident cases. Ann Rheum Dis 2003;62:835-41.

16. Padyukov L, Silva C, Stolt P, Alfredsson L, Klareskog L. A gene-environment interaction between smoking and shared epitope genes in HLA-DR provides a high risk of seropositive rheumatoid arthritis. Arthritis Rheum 2004;50:3085-92.

17. Thorburn CM, Prokunina-Olsson L, Sterba KA, et al. Association of PDCD1 genetic variation with risk and clinical manifestations of systemic lupus erythematosus in a multiethnic cohort. Genes Immun 2007;8:279-87.

18. Tan EM, Cohen AS, Fries JF, et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 1982;25:1271-7.

19. Petri M. Hopkins Lupus Cohort: 1999 update. Rheum Dis Clin North Am 2000; 26:199-213.

20. Bauer JW, Baechler EC, Petri M, et al. Elevated serum levels of interferon-regulated chemokines are biomarkers for active human systemic lupus erythematosus. PLoS Med 2006;3(12):e491.

21. Criswell LA, Pfeiffer KA, Lum RF, et al. Analysis of families in the Multiple Autoimmune Disease Genetics Consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. Am J Hum Genet 2005;76:561-71.

22. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet 2005;37:1217-23.

23. Bland JM, Altman DG. Statistics notes: the odds ratio. BMJ 2000;320:1468.

24. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

25. Seldin MF, Shigeta R, Villoslada P, et al. European population substructure: clustering of northern and southern populations. PLoS Genet 2006;2(9):e143.

26. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P. Association mapping in structured populations. Am J Hum Genet 2000;67:170-81.

27. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904-9.

28. Plenge RM, Padyukov L, Remmers EF, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet 2005;77:1044-60.

29. Moser KL, Neas BR, Salmon JE, et al. Genome scan of human systemic lupus erythematosus: evidence for linkage on chromosome 1q in African-American pedigrees. Proc Natl Acad Sci U S A 1998; 95:14869-74.

30. Gray-McGuire C, Moser KL, Gaffney PM, et al. Genome scan of human systemic lupus erythematosus by regression modeling: evidence of linkage and epistasis at 4p16-15.2. Am J Hum Genet 2000; 67:1460-9.

31. Cantor RM, Yuan J, Napier S, et al. Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23. Arthritis Rheum 2004;50:3203-10.

32. Gregersen PK, Behrens TW. Genetics of autoimmune disease — disorders of immune homeostasis. Nat Rev Genet 2006; 7:917-28.

33. Watford WT, Hissong BD, Bream JH, Kanno Y, Muul L, O'Shea JJ. Signaling by IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol Rev 2004; 202:139-56.

34. Morinobu A, Gadina M, Strober W, et al. STAT4 serine phosphorylation is critical for IL-12-induced IFN-gamma production but not for cell proliferation. Proc Natl Acad Sci U S A 2002;99:12281-6.

35. Nishikomori R, Usui T, Wu CY, Morinobu A, O'Shea JJ, Strober W. Activated STAT4 has an essential role in Th1 differentiation and proliferation that is independent of its role in the maintenance of IL-12R beta 2 chain expression and signaling. J Immunol 2002;169:4388-98.

36. Mathur AN, Chang HC, Zisoulis DG, et al. Stat3 and Stat4 direct development of IL-17-secreting Th cells. J Immunol 2007; 178:4901-7.

37. Bettelli E, Oukka M, Kuchroo VKT. T(H)-17 cells in the circle of immunity and autoimmunity. Nat Immunol 2007;8: 345-50.

38. Finnegan A, Grusby MJ, Kaplan CD, et al. IL-4 and IL-12 regulate proteoglycan-induced arthritis through Stat-dependent mechanisms. J Immunol 2002;169:3345-52.

39. Klinman DM, Gursel I, Klaschik S, Dong L, Currie D, Shirota H. Therapeutic potential of oligonucleotides expressing immunosuppressive TTAGGG motifs. Ann N Y Acad Sci 2005;1058:87-95.

40. Hildner KM, Schirmacher P, Atreya I, et al. Targeting of the transcription factor STAT4 by antisense phosphorothioate oligonucleotides suppresses collagen-in-

duced arthritis. J Immunol 2007;178:3427-36.

41. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. Science 2006;314:1461-3.

42. Cargill M, Schrodi SJ, Chang M, et al. A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. Am J Hum Genet 2007;80:273-90.

43. Tsunemi Y, Saeki H, Nakamura K, et al. Interleukin-12 p40 gene (IL12B) 3´-untranslated region polymorphism is associated with susceptibility to atopic dermatitis and psoriasis vulgaris. J Dermatol Sci 2002;30:161-6.

44. Fukao T, Frucht DM, Yap G, Gadina M, O'Shea JJ, Koyasu S. Inducible expression of Stat4 in dendritic cells and macrophages and its critical role in innate and adaptive immune responses. J Immunol 2001;166:4446-55.

45. Remoli ME, Ragimbeau J, Giacomini E, et al. NF-κB is required for STAT-4 expression during dendritic cell maturation. J Leukoc Biol 2007;81:355-63.

46. Jacob CO, Zang S, Li L, et al. Pivotal role of Stat4 and Stat6 in the pathogenesis of the lupus-like disease in the New Zealand mixed 2328 mice. J Immunol 2003; 171:1564-71.

47. Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004;36:337-8.

48. Vandiedonck C, Capdevielle C, Giraud M, et al. Association of the PTPN22*R620W polymorphism with autoimmune myasthenia gravis. Ann Neurol 2006;59:404-7.

49. Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn JN. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 2003;33:177-82.

50. Campbell CD, Ogburn EL, Lunetta KL, et al. Demonstrating stratification in a European American population. Nat Genet 2005;37:868-72.

51. Lee H-S, Remmers EF, Le JM, Kastner DL, Bae S-C, Gregersen PK. Association of *STAT4* with rheumatoid arthritis in the Korean population. Mol Med (in press).
Copyright © 2007 Massachusetts Medical Society.