

Current understanding of the genetic aetiology of rheumatoid arthritis and likely future developments

L. A. Criswell and P. K. Gregersen¹

Most of the work described herein was performed by the North American Rheumatoid Arthritis Consortium (NARAC). After a brief description of the NARAC and the multiplex family resource that has been developed by this consortium, we will summarize the current status of genome-wide screens using this valuable family collection. Next, we describe work that is under way to further delineate the genes on chromosome 18q that demonstrate linkage to rheumatoid arthritis (RA), including an analysis of candidate genes in the region and results of dense association mapping. We also describe an extensive analysis of functional single-nucleotide polymorphisms (SNPs) that is under way in collaboration with Celera Diagnostics, as well as studies designed to further dissect the phenotypic and genotypic heterogeneity of RA. We conclude by briefly summarizing our future plans to elucidate the genetic aetiology of RA.

The North American Rheumatoid Arthritis Consortium (NARAC) represents a collaboration among 14 US medical centres that has devoted substantial effort to identifying the genetic contribution to this important disorder [1]. In particular, the NARAC has created a unique registry and repository of multiplex rheumatoid arthritis (RA) families for use in gene mapping studies. This resource currently includes over 990 affected sibling pairs from 773 multiplex RA families. Entry criteria for families include the presence of erosive disease in at least one affected sibling and RA onset between the ages of 18 and 65 yr for at least one affected sibling. The presence of psoriasis, inflammatory bowel disease or systemic lupus erythematosus (SLE) in one or more affected siblings is an exclusionary criterion for families. All affected siblings undergo a standardized examination, radiography of hands and wrists and serological evaluation at the time of entry to the NARAC registry and repository. Clinical and genetic data from these families are available to the scientific community and can be accessed at the NARAC web site: www.naracdata.org.

Genome-wide screens

Genome-wide screens have been completed among 512 multiplex NARAC families (>600 affected sibling pairs, ASP) [2, 3]. In addition to confirming linkage to the major histocompatibility complex (MHC) on chromosome 6p ($P=4\times 10^{-12}$), seven additional chromosomal regions provide evidence of linkage to RA ($P<0.005$). Of interest, many of these regions overlap with genome screens in other autoimmune diseases [2]. An additional 202 ASPs are being genotyped in order to extend and confirm these findings.

Figure 1 displays linkage results for chromosomes 1, 6 and 18. The three curves shown correspond to the first and second cohorts (each approximately 205 families) and a combined analysis of all 512 multiplex families. As can be seen in the figure, several genomic regions reveal quite consistent evidence for linkage to RA. Additional work using these families has been performed to further delineate the contribution of the major histocompatibility complex (MHC) to RA susceptibility [4]. These results indicate that there are probably at least three independent genetic loci within the MHC that contribute to RA risk.

Studies of chromosome 18q

More recently, attention has focused on chromosome 18q in order to identify the genes responsible for the linkage signal in this region (Fig. 1). Importantly, in addition to the evidence of linkage observed in NARAC families, genome-wide screens among European and UK multiplex RA families also support linkage to RA in this region. Indeed, meta-analysis of these three international studies suggests that this is a high priority for subsequent work, as shown in Fig. 2.

Recent work on chromosome 18q has included both candidate gene and focused positional approaches. These studies have used a case-control association design, with approximately 400 NARAC cases and matched controls. Controls were identified through the New York Cancer Project (<http://www.amdec.org>) and were matched to cases on ethnicity (grandparental country of origin), sex and age. The candidate gene study has focused on the following genes in the region: transcription factor 4 (*TCF4*), mucosa-associated lymphoid tissue lymphoma translocation protein 1 (*MALTI*), PMA-induced protein 1 (*PMAIP1*), B-cell CLL/lymphoma 2 (*BCL2*) and tumour necrosis factor superfamily member 11A (*TNFRSF11A*), also known as receptor activator of NF- κ B or RANK. Although this work is ongoing, analyses to date examining single-nucleotide polymorphisms (SNPs) and haplotypes within each of these candidate genes have not revealed strong association with RA. For example, a single RANK haplotype is associated with RA at a significance level of 0.02 and one SNP in *MALTI* is associated with RA at a significance level of 0.03. However, given the number of SNPs examined, these associations are not convincing.

For this reason, NARAC investigators have initiated a dense mapping study of chromosome 18q. Specifically, 3072 SNPs across a 10 centimorgan (cM) region were selected in order to examine the approximately 10 million base pairs under the RA linkage peak on 18q. The SNP genotyping was performed in 460 NARAC cases and matched controls by Illumina Inc. (San Diego, CA, USA), a commercial SNP typing company participating in the International HapMap project (<http://www.hapmap.org/>). Two thousand seven hundred and nineteen of the 3072 SNPs were successfully typed, resulting in 2.5 million genotypes.

Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco, CA and ¹Robert S. Boas Center for Genomics and Human Genetics, North Shore Long Island Jewish Research Institute, Manhasset, NY, USA.

Correspondence to: L. A. Criswell, Division of Rheumatology, 374 Parnassus Avenue, Box 0500, San Francisco, CA, 94143-0500, USA.
E-mail: lac@itsa.ucsf.edu

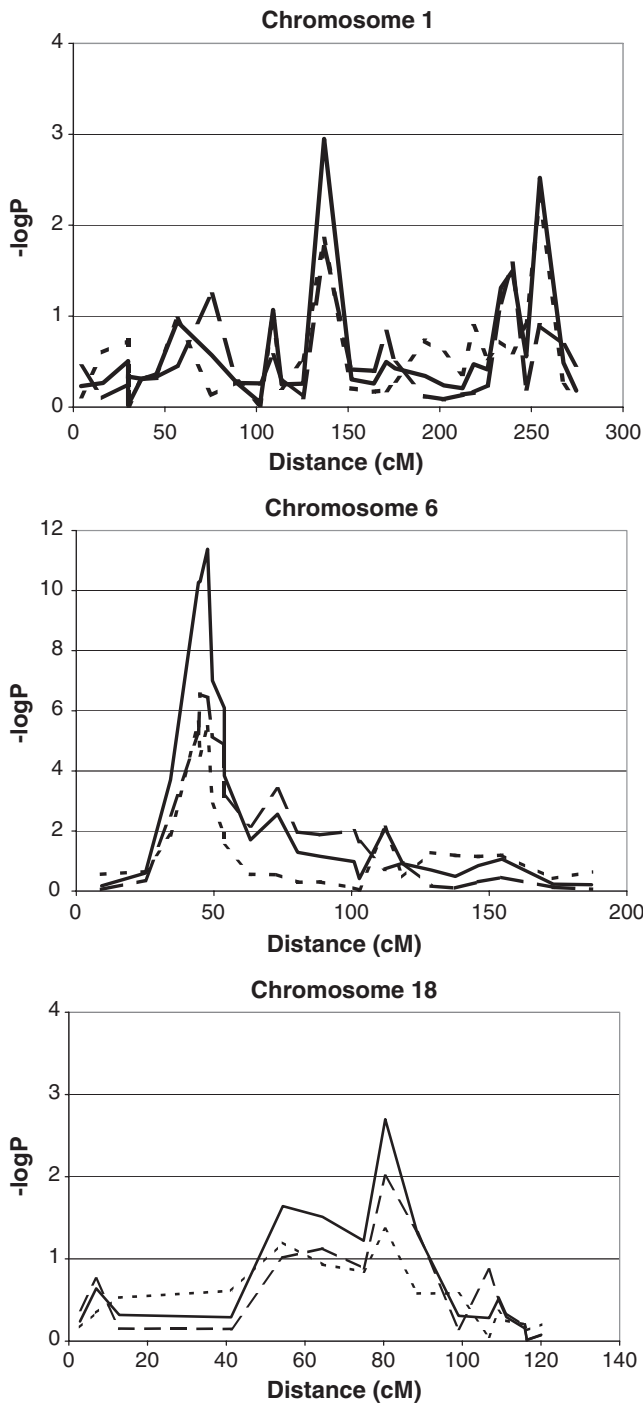


FIG. 1. Linkage results for three chromosomes analysed in a genome-wide screen of 512 multiplex NARAC families. Graphs show linkage results for the first and second cohorts (broken lines) and a combined analysis of all families (solid line). The x axes correspond to distance (in cM) along each of the three chromosomes and the y axes correspond to $-\log P$ value for each genetic marker [3] © 2003 American College of Rheumatology. Reproduced with permission by John Wiley & Sons, Inc.

Two thousand two hundred and ninety-seven of the 2719 successfully typed SNPs had minor allele frequencies $\geq 5\%$, producing a total of 2.1 million informative genotypes for analysis.

Figure 3 displays the results of this dense mapping association study, with the strength of association ($-\log_{10}$ of P value) shown along the ordinate and the location of each SNP along the abscissa.

As shown in the figure, a number of SNPs demonstrate strong evidence of association with RA. Of particular interest, however, are two regions in which clusters of associated SNPs are observed. Twelve SNPs were significantly associated with RA ($P \leq 0.001$) at 54.9 Mb (region 1) and three SNPs were associated with RA ($P \leq 0.005$) at 53.2 Mb (region 2). Several SNPs in these two regions were subsequently genotyped among 613 additional RA cases and 518 additional controls, yielding an odds ratio (OR) of 1.5 ($P = 0.000036$) for the most significantly associated SNP in region 2 and an OR of 1.2 ($P = 0.0085$) for the most significantly associated SNP in region 1. Although these results support the relevance of this genomic region to RA susceptibility, neither of the associated SNPs are within a gene or apparent regulatory region. Thus, additional work is needed to replicate these findings before embarking on a comprehensive analysis of the genes within this region.

Genome-wide functional SNP screen in RA

NARAC investigators have also pursued a functional genome-wide scan in collaboration with Dr Ann Begovich and her colleagues at Celera Diagnostics (Alameda, CA, USA). Celera Diagnostics has identified and developed genotyping assays for a large number of SNPs that are likely to be functionally relevant due to their location in coding or regulatory regions. Approximately 12 000 functional SNPs have been screened for association with RA using NARAC families and other collections of RA cases and controls. The most exciting results to date, however, emerged from the initial study that included a discovery set of 475 seropositive RA cases and matched controls and a replication set of 840 NARAC cases and matched controls. This initial study focused on 96 functional SNPs or SNPs in candidate genes or linkage peaks. Very strong association was observed for a missense SNP (R620W) in a protein tyrosine phosphatase (*PTPN22*), with an OR of 1.7 ($P = 0.0007$) in the discovery set and an OR of 2.1 ($P = 3.4 \times 10^{-9}$) in the replication set. Furthermore, a stratified analysis of the NARAC (replication) cases suggested that the SNP was primarily or more strongly associated with seropositive RA (i.e. OR = 2.4, $P < 0.0001$ for seropositive compared with OR = 1.1, $P = 0.64$ for seronegative RA) [5].

Interest in *PTPN22*, an intracellular protein tyrosine phosphatase, relates to its negative regulatory effects on T-cell activation through binding to intracellular kinases, such as Csk. Recent work in *PTPN22* knockout animals [6] suggests that this gene might be relevant to the development of autoimmunity. Indeed, in addition to the RA association described above, this same missense SNP has now been reported to be associated with several other autoimmune diseases, including type 1 diabetes [7], SLE [8] and Hashimoto's thyroiditis [9]. Additional work confirms the association of this missense SNP with RA and SLE [10, 11]. Of interest, however, *PTPN22* does not appear to be associated with multiple sclerosis [9, 12]. This differential association, in conjunction with other considerations, suggests that *PTPN22* may play a primary role in autoantibody production.

The role of genetic factors in phenotypic heterogeneity in RA

The NARAC multiplex family collection has also proved to be a valuable resource for studies addressing the phenotypic heterogeneity of RA. Like most complex human diseases, RA is characterized by variation in the specific phenotype expressed by individual patients. It is likely that at least some of this heterogeneity reflects the action of different genetic factors, or interactions between genes or between genes and environmental

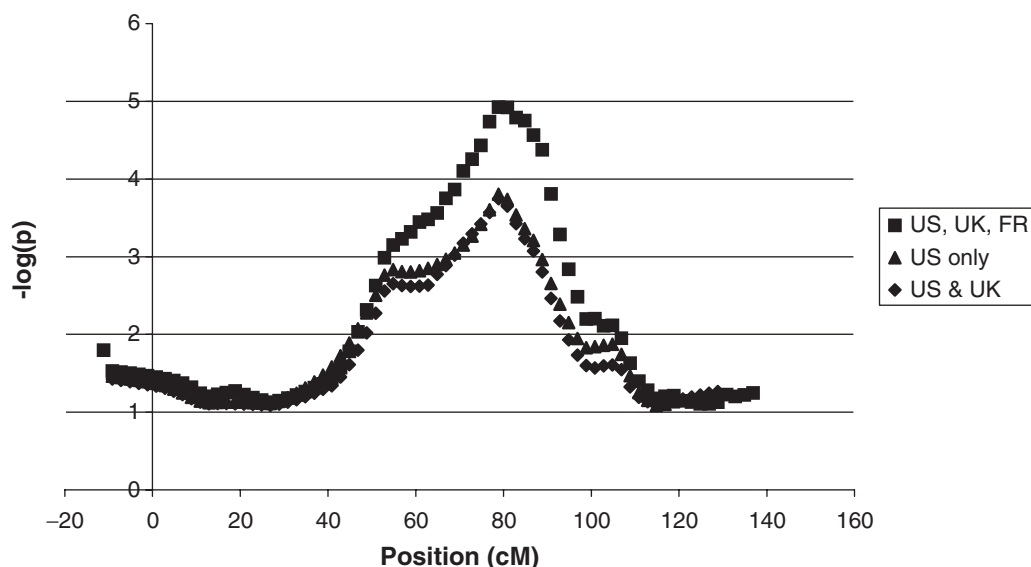


FIG. 2. Meta-analysis of three international studies supports RA susceptibility locus on chromosome 18q.

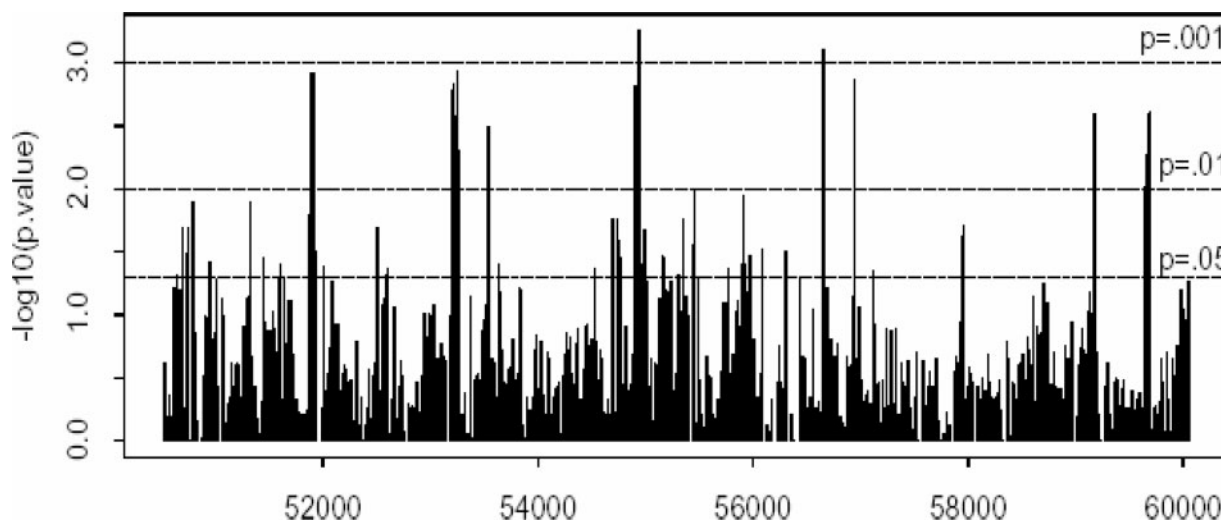


FIG. 3. Dense association mapping across 10cM of chromosome 18q [18]. Reproduced with permission from Blackwell Publishing.

factors. Here we describe two studies using the NARAC collection that address this issue.

Familial clustering of RA disease features

As described previously, RA patients enrolled in the NARAC family collection undergo careful phenotypic characterization in a standardized fashion. Thus, patients are classified according to the presence and titre of specific autoantibodies (IgM RF and anti-CCP), joint involvement according to the Joint Alignment and Motion (JAM) score [13], the presence and extent of erosive disease on hand/wrist radiographs, functional status according to the Health Assessment Questionnaire (HAQ) score [14], age and calendar year of RA onset, the presence of nodules or other extra-articular manifestations, and the presence of other autoimmune diseases.

We examined the presence and extent of familial clustering for each of these disease features among 1097 affected individuals (i.e. RA patients) from 512 multiplex families. Multivariate methods were used in order to account for the impact of other

factors, such as disease duration and sex, on the development of specific disease features. Results demonstrated strong evidence of familial clustering for several disease features [15], including seropositivity (multivariate OR = 4.3, $P < 0.0001$), nodules (OR = 2.3, $P < 0.0001$) and age at RA diagnosis (multivariate regression coefficient, $\beta = 0.44$, $P < 0.0001$). Results for each RA disease feature examined are shown in Table 1. The presence of familial clustering of specific disease features supports the hypothesis that genetic factors influence the specific RA phenotype expressed in individual patients. However, it is important to bear in mind that shared environmental factors or exposures may also contribute to familial clustering. These results also highlight the importance of considering the specific RA phenotype in genetic and epidemiologic studies of this complex disorder.

Linkage analysis of quantitative RA traits

As previously mentioned, genome-wide screens using the NARAC multiplex family collection has contributed substantially to our

understanding of the number and location of potentially relevant genetic risk factors for RA [2, 3]. These analyses have employed non-parametric methods of linkage analysis and focused on the standard RA phenotype as defined by American College of Rheumatology criteria [16]. In view of the complexity of the disease, we hypothesized that analysis of more homogeneous RA phenotypes might facilitate the identification of genes and elucidate underlying disease mechanisms. Further, we hypothesized that the investigation of quantitative RA phenotypes will be a powerful way of dissecting the phenotypic and genotypic heterogeneity of the disease.

For these reasons we reanalysed the genome screen data focusing on two quantitative RA phenotypes. Specifically, we analysed titres of two autoantibodies that were systematically characterized in NARAC patients at the time of study entry. These included IgM RF and anti-CCP titres, both of which were performed at the University of Washington, Department of Laboratory Medicine, Immunology Division. RF testing was performed using a latex-enhanced nephelometric assay (Behring Diagnostics, San Jose, CA, USA) and anti-CCP titres were determined based on a second-generation enzyme-linked immunosorbent assay (Inova Diagnostics). The genetic markers

TABLE 1. Familial clustering of disease features in 1097 affected individuals from 512 multiplex NARAC families

Disease characteristic	Odds ratio	P
Rheumatoid factor positive	4.3	<0.0001
Presence of nodules	2.3	<0.0001
Other extra-articular manifestations ^a	1.7	0.5
Autoimmune thyroid disease	1.9	0.3
Other autoimmune diseases ^b	1.5	0.4
	Regression coefficient	
Age at RA diagnosis ^c	0.44	<0.0001
Year of RA diagnosis	0.010	0.001
JAM score ^d	0.12	0.001
HAQ score ^e	0.16	0.0001

Results shown are from multivariate logistic and linear regression analyses with disease duration, sex and age at diagnosis as covariates.

^aThese include vasculitis, rheumatic lung disease, Felty's syndrome, scleritis and scleromalacia.

^bOther autoimmune diseases include Sjögren's syndrome, dermatomyositis/polymyositis, polyarteritis nodosa, idiopathic thrombocytopenic purpura, myasthenia gravis, multiple sclerosis, scleroderma and undifferentiated connective tissue disease

^cAge at RA diagnosis not included as covariate for analysis of this disease feature.

^dJoint Alignment and Motion score.

^eHAQ score = Health Assessment Questionnaire score.

analysed in this quantitative linkage analysis consisted of the same 380 microsatellite markers typed for the original genome screen, that is for RA defined qualitatively (i.e. presence vs. absence of RA). Subjects included 1002 affected individuals from 491 multiplex NARAC families. The Multipoint Engine for Rapid Likelihood Inference (MERLIN) statistical package [17] (<http://www.sph.umich.edu/csg/abecasis/Merlin/>) was used to perform the genome-wide linkage analysis for each quantitative trait (i.e. autoantibody titres).

Figure 4 displays linkage results for chromosome 6 for both quantitative traits and for RA defined qualitatively. These results are notable for the strength and consistency of linkage evidence to anti-CCP titre and for RA defined qualitatively, in contrast to RF titre, which demonstrated much weaker evidence of linkage. As expected, the linkage peaks on this chromosome correspond to the HLA region on the short arm of chromosome 6, though the peak is quite broad, covering a substantial portion of this chromosome. Other chromosomes displayed different patterns of linkage results, with some genomic regions of overlap between the two quantitative traits and other regions exhibiting notable differences in the size or location of linkage peaks, as was the case for chromosome 6. Although we view these quantitative linkage results as preliminary, overall the findings support the genetic heterogeneity of RA related to specific phenotypes. Further work in this area is likely to be fruitful due to the potential power of quantitative trait methods and the likelihood that application of these methods to complex disorders like RA will assist with the identification of genes related to distinct pathways. To the extent that these pathways are associated with variation in disease outcome and response to treatment, these findings should ultimately translate into better diagnostic and therapeutic tools.

Future plans

There are a number of other projects under way now or planned for the near future that are likely to increase our understanding of the genetic aetiology of RA. For example, we have recently initiated a SNP-based genome screen on a larger set of multiplex RA families recruited by the NARAC with the hope that this analysis will further delineate the multiple genomic regions that contribute to the disease and provide more precise locations of the causative genes. We will continue to carefully consider specific phenotypes based on the expectation that certain genes or genomic regions will be relevant to broad autoimmunity phenotypes, whereas other genes or regions may be most relevant to very narrowly defined phenotypes (e.g. production of specific autoantibodies). Collaboration with multiple groups nationally and internationally, in addition to for-profit entities, will become even more critical to our ability to continue to make progress in this area

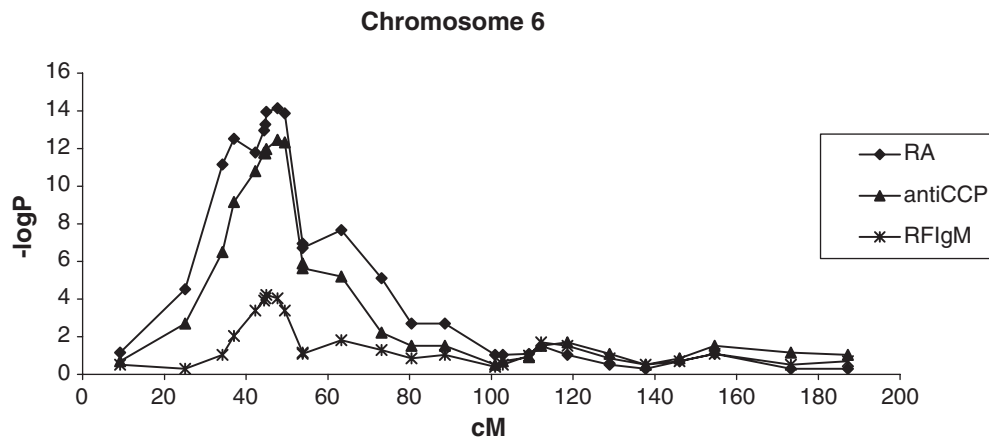


FIG. 4. Results of linkage analysis for two quantitative traits (RF and CCP titre) and for RA defined qualitatively.

due to the high costs of applying state-of-the-art genomics technologies and the large numbers of subjects required to ensure adequate statistical power, particularly for more specific phenotype definitions. Further, the direction of future research efforts is likely to be driven to a great extent by the specific new technologies that are developed in the coming years. Finally, functional studies and other complementary approaches are going to be required to fully elucidate the relevance of genetics findings in terms of underlying disease biology and to translate these basic findings into better diagnostic and therapeutic tools.

Acknowledgements

Much of the work described herein was performed by the NARAC and we are indebted to the many investigators who have made this effort possible, including Drs Michael Seldin, Daniel Kastner, Christopher Amos, Damini Jawaheer and Mark Wener. We are also grateful for the important contributions of Marlena Kern, Elaine Remmers, Annette Lee, Wei Chen and Raymond Lum, as well as the participating NARAC families. We would also like to acknowledge the efforts of Drs Ann Begovich and Victoria Carlton of Celera Diagnostics, Alameda, CA, USA. Financial support for this work was provided by the National Institutes of Health (NIAMS and NIAID) and the Arthritis Foundation. These studies were performed in part in the General Clinical Research Center, Moffitt Hospital, University of California, San Francisco, with funds provided by the National Center for Research Resources (5 M01 RR00079, C.S. Public Health Service).

The authors have declared no conflicts of interest.

References

- Gregersen PK. The North American Rheumatoid Arthritis Consortium—bringing genetic analysis to bear on disease susceptibility, severity, and outcome. *Arthritis Care Res* 1998;11:1–2.
- Jawaheer D, Seldin MF, Amos CI *et al*. A genome-wide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet* 2001;68:927–36.
- 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.
- Jawaheer D, Li W, Graham RR *et al*. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. *Am J Hum Genet* 2002;71:585–94.
- Begovich AB, Carlton VEH, Honinberg LA *et al*. A missense SNP in the protein tyrosine phosphatase *PTPN22* is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;75:330–7.
- Hasegawa K, Martin F, Huang G, Tumas D, Diehl L, Chan AC. PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. *Science* 2004;303:685–9.
- 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.
- Kyogoku C, Langefeld CD, Ortmann WA *et al*. Genetic association of the R620W polymorphism of protein tyrosine phosphatase *PTPN22* with human SLE. *Am J Hum Genet* 2004;75:504–7.
- Criswell LA, Pfeiffer KA, Lum RF *et al*. Analysis of families in the Multiple Autoimmune Disease Genetics Consortium (MADGC): the *PTPN22* 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 2005;76:561–71.
- Lee AT, Li W, Liew A *et al*. The *PTPN22* R620W polymorphism associates with RF positive rheumatoid arthritis in a dose-dependent manner but not with HLA-SE status. *Genes Immun* 2004;6:129–33.
- Orozco G, Sanchez E, Gonzalez-Gay MA *et al*. Association of a functional single-nucleotide polymorphism of *PTPN22*, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum* 2005;52:219–24.
- Begovich AB, Caillier SJ, Alexander HC *et al*. The R620W polymorphism of the protein tyrosine phosphatase *PTPN22* is not associated with multiple sclerosis. *Am J Hum Genet* 2005;76:184–7.
- Spiegel TM, Spiegel JS, Paulus HE. The joint alignment and motion scale: a simple measure of joint deformity in patients with rheumatoid arthritis. *J Rheumatol* 1987;14:887–92.
- Fries JF, Spitz PW, Young DY. The dimensions of health outcomes: The Health Assessment Questionnaire, Disability and Pain Scales. *J Rheumatol* 1982;9:789–93.
- Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. Clustering of disease features within 512 multicase rheumatoid arthritis families. *Arthritis Rheum* 2004;50:736–41.
- Arnett FC, Edworthy SM, Bloch DA *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
- Abecasis GR, Cherny SS, Cookson WO, Cardon LR. Merlin—rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002;30:97–101.
- Gregersen PK. Pathways to gene identification in rheumatoid arthritis: *PTPN22* and beyond. *Immunol Rev* 2005;204:74–86.