A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis

Ann B Begovich
Victoria E. H. Carlton
Lee A Honigberg
Steven J Schrodi
Anand P Chokkalingam, et al.

Available at: http://works.bepress.com/steve_schrodi/4/
A Missense Single-Nucleotide Polymorphism in a Gene Encoding a Protein Tyrosine Phosphatase (PTPN22) Is Associated with Rheumatoid Arthritis

Ann B. Begovich, 1, * Victoria E. H. Carlton, 1 Lee A. Honigberg, 2 Steven J. Schrodi, 1 Anand P. Chokkalingam, 1 Heather C. Alexander, 1 Kristin G. Ardlie, 1 Qiqing Huang, 2 Ashley M. Smith, 2 Jill M. Spoerke, 2 Marion T. Conn, 2 Monica Chang, 1 Sheng-Yung P. Chang, 1 Randall K. Saiki, 1 Joseph J. Catanese, 1 Diane U. Leong, 1 Veronica E. Garcia, 1 Linda B. McAllister, 1 Douglas A. Jeffery, 2 Annette T. Lee, 5 Franak Batliwalla, 5 Elaine Remmers, 6 Lindsey A. Criswell, 7 Michael F. Seldin, 8 Daniel L. Kastner, 6 Christopher I. Amos, 4 John J. Sninsky, 1 and Peter K. Gregersen 5

1Celera Diagnostics, Alameda, CA; 2Celera Genomics, South San Francisco, CA; 3Genomics Collaborative, Inc., Cambridge, MA; 4Department of Epidemiology, University of Texas MD Anderson Cancer Center, University of Texas, Houston; 5Center for Genomics and Human Genetics, North Shore- LIJ Research Institute, Manhasset, NY; 6Genetics and Genomics Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda; 7Rosalind Russell Medical Research Center for Arthritis, Department of Medicine, University of California, San Francisco; and 8Rowe Program of Human Genetics, Department of Medicine, University of California, Davis

Rheumatoid arthritis (RA) is the most common systemic autoimmune disease, affecting ~1% of the adult population worldwide, with an estimated heritability of 60%. To identify genes involved in RA susceptibility, we investigated the association between putative functional single-nucleotide polymorphisms (SNPs) and RA among white individuals by use of a case-control study design; a second sample was tested for replication. Here we report the association of RA susceptibility with the minor allele of a missense SNP in PTPN22 (discovery-study allelic Pp; replication-study allelic P), which encodes a hematopoietic-specific protein tyrosine phosphatase also known as “Lyp.” We show that the risk allele, which is present in ~17% of white individuals from the general population and in ~28% of white individuals with RA, disrupts the P1 proline-rich motif that is important for interaction with Csk, potentially altering these proteins’ normal function as negative regulators of T-cell activation. The minor allele of this SNP recently was implicated in type 1 diabetes, suggesting that the variant phosphatase may increase overall reactivity of the immune system and may heighten an individual carrier’s risk for autoimmune disease.

Rheumatoid arthritis (RA [MIM 180300]) is characterized by immune cell–mediated destruction of the joint architecture and is two to three times more common in women than in men (Firestein 2003). A strong genetic component is indicated (Seldin et al. 1999; MacGregor et al. 2000), and genome scans have identified multiple regions linked to disease (Cornelis et al. 1998; Shiozawa et al. 1998; Jawaheer et al. 2001, 2003; MacKay et al. 2002; Fisher et al. 2003). Although interesting associations have been reported (Suzuki et al. 2003; Tokuhiro et al. 2003), only alleles at the HLA-DRB1 locus have consistently demonstrated both linkage and association (Seldin et al. 1999).

To identify genes involved in genetic predisposition to RA, we performed a case-control association study (called “discovery study”) with assays for 87 putative functional SNPs (Botstein and Risch 2003) in RA candidate genes and/or linkage regions. The discovery study, consisting of 475 individuals with RA and 475 individually matched controls, was obtained by Genomics Collaborative, Inc. (GCI). Case samples were collected from throughout the United States, and they met the 1987 American College of Rheumatology (ACR) diagnostic criteria for RA. All case samples were from white in-
individuals with an age at onset of RA of between 18 and 68 years and a positive rheumatoid factor of \( \geq 20 \) IU. Individuals with psoriasis, systemic lupus erythematosus (SLE), ankylosing spondylitis, or Reiter syndrome were excluded. Control samples were taken from a pool of healthy white individuals from the United States with no medical history of RA or of any of the autoimmune disorders listed above. A single control was matched to each case on the basis of sex, age (\( \pm 5 \) years), and ethnicity (grandparental country/region of origin). All protocols and recruitment sites have been approved by national and/or local institutional review boards, and all subjects were enrolled with informed written consent.

We found association with the minor allele (T) of a missense SNP (R620W [rs2476601, 1858C\( \rightarrow \)T]) in the protein tyrosine phosphatase non-receptor type 22 gene (PTPN22) (allelic frequency 13.8% in cases, 8.8% in controls) \( (P = 6.6 \times 10^{-4}, \) allelic odds ratio \( [OR] 1.65, 95\% CI 1.23–2.20).\)

Replication in a second study (called the “replication study”) confirmed association. Cases in the replication study were obtained by the North American Rheumatoid Arthritis Consortium (NARAC) and consisted of patients with RA or of any of the autoimmune disorders listed above. A single control was matched to each case on the basis of sex, age, site). Informed written consent was obtained from every participating recruitment centers of NARAC (NARAC Web site). Controls were selected from a pool of healthy white individuals from the United States with the local institutional review board’s approval was secured at each recruitment site. The enrollment criteria for family participation are described in detail elsewhere (Jawaheer et al. 2001). In brief, at least two siblings must satisfy the 1987 ACR criteria for RA, at least one sibling must have documented erosions on hand radiographs, and at least one sibling must have disease onset between the ages of 18 and 60 years. The presence of psoriasis, inflammatory bowel disease, or SLE was an exclusionary criterion for the sib pair. Controls were selected from 20,000 individuals who are part of the New York Cancer

### Table 1

<table>
<thead>
<tr>
<th>GENOTYPE</th>
<th>DISCOVERY STUDY</th>
<th>REPLICATION STUDY</th>
<th>ALL SIBS*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CONTROLS (n = 475)</td>
<td>CASES (n = 475)</td>
<td>SIBS* (n = 926)</td>
</tr>
<tr>
<td>TT</td>
<td>.006 (3)</td>
<td>.013 (6)</td>
<td>.010 (9)</td>
</tr>
<tr>
<td>TC</td>
<td>.164 (78)</td>
<td>.251 (119)</td>
<td>.154 (143)</td>
</tr>
<tr>
<td>CC</td>
<td>.830 (394)</td>
<td>.737 (350)</td>
<td>.836 (774)</td>
</tr>
</tbody>
</table>

NOTE.—Genotypes were generated by use of kinetic, allele-specific PCR (Germer et al. 2000): 0.3 ng of DNA in a 15-\( \mu \)l reaction containing the allele-specific primers 5'-CTCCACTTCCTGTAT/C and a common primer, 5'-CCCA-TCCCACACTTTATTTTATAC. Genotyping accuracy was 100%, as determined by internal comparisons by use of a different assay (Iannone et al. 2000) for the same marker on \( \sim 6\% \) of the samples. All case and control populations were in Hardy-Weinberg equilibrium.

* Independent cases comprising one randomly selected sib per NARAC family.

† All affected individuals in NARAC.

### Table 2

**Association of PTPN22 with RA**

<table>
<thead>
<tr>
<th>DATA SET</th>
<th>GENOTYPE</th>
<th>NO. OF CASES</th>
<th>NO. OF CONTROLS</th>
<th>RESULTS FOR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CRUDE ANALYSIS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR</td>
</tr>
<tr>
<td>Discovery:</td>
<td>TT</td>
<td>6</td>
<td>3</td>
<td>2.35 (5.6–9.07) (P = .32)</td>
</tr>
<tr>
<td></td>
<td>TC</td>
<td>119</td>
<td>78</td>
<td>1.72 (1.25–2.36) (P = .001)</td>
</tr>
<tr>
<td></td>
<td>TT+TC</td>
<td>125</td>
<td>81</td>
<td>1.74 (1.27–2.38) (P = .0005)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>350</td>
<td>394</td>
<td>... (\ldots) (\ldots) (\ldots) (\ldots) (\ldots) (\ldots)</td>
</tr>
<tr>
<td>Replication:</td>
<td>TT</td>
<td>10</td>
<td>9</td>
<td>2.63 (1.06–6.53) (P = .03)</td>
</tr>
<tr>
<td></td>
<td>TC</td>
<td>126</td>
<td>143</td>
<td>2.09 (1.59–2.74) (P &lt; .0001)</td>
</tr>
<tr>
<td></td>
<td>TT+TC</td>
<td>136</td>
<td>152</td>
<td>2.12 (1.62–2.76) (P &lt; .0001)</td>
</tr>
<tr>
<td></td>
<td>CC</td>
<td>327</td>
<td>774</td>
<td>... (\ldots) (\ldots) (\ldots) (\ldots) (\ldots) (\ldots)</td>
</tr>
</tbody>
</table>

* Values were calculated by use of standard contingency tables (or two-tailed Fisher’s Exact Test).

† CLR was performed as described elsewhere (Breslow and Day 1980) to account for individual matching of controls to cases.

‡ To assess whether the observed associations were independent of the HLA-DRB1 genotype, we further adjusted for this known risk factor. HLA-DRB1 genotyping was performed as described elsewhere (Jawaheer et al. 2002), and, for this analysis, samples were binned according to their HLA-DRB1 genotype: high risk, two shared epitopes (HR-2SE); high risk, one shared epitope (HR-1SE; 4K,X; or 4R,X); low risk, one shared epitope (LR-1SE; 1R,X; or 10,X); and low risk, zero shared epitopes (LR-0SE; X,X) (Fries et al. 2002).

§ ORs were calculated relative to the major allele homozygote (CC).

† P value calculated by use of Fisher’s Exact Test.

∗ Because of the infrequency of the TT genotype, we combined it with the TC genotype and repeated the analysis.
PTPN22 knockdown by RNAi increases antigen-receptor signaling in Jurkat T-cell line. Knockdown and NF-κB transcriptional response after T-cell receptor stimulation for cells transfected with two control siRNAs (Scramble) and two PTPN22 siRNAs. First, 4.5 μg siRNA (Dharmacon) and 1 μg pNF-κB-luc plasmid (Stratagene) were electroporated into 2 × 10⁶ Jurkat cells (ATCC, clone E6-1, homozygous for 620R) in Cell Line Nucleofector Solution V, according to manufacturer’s protocol #S18 (amaxa). After 24 h, 5.5 × 10⁶ cells were stimulated with anti-CD3 (2 μg/ml [BD Pharmingen, clone UCHT1]), anti-CD28 (2 μg/ml [BD Pharmingen, clone CD28.2]), and anti-mouse IgG1 (2 μg/ml [BD Pharmingen, clone A85-1]). Cells were lysed and assayed for luminescence 6 h after stimulation by use of Bright-Glo (Promega) and a 96-well MicroBeta plate reader (Wallac). PTPN22 siRNAs that were effective for knockdown were found by screening seven candidate siRNA sequences that were designed by use of Web-based tools from Ambion and Dharmacon. siRNA sequences used in this study were PTPN22.1 (5′-AAAGCAGACAAACCTATCCT), PTPN22.2 (5′-GAGGATTCAGCTACATCAAT), Scramble.1 (5′-AAGACGGCATCAAGGTGAAC), and Scramble.2 (5′-AATCTCCGAACGTGTCACGT). RNA was harvested according to the manufacturer’s protocol (RNeasy 96 [Qiagen]), and TaqMan real-time quantitative RT-PCR (SDS-7900 [Applied Biosystems]) was used to measure PTPN22 expression levels. PTPN22 mRNA levels in each sample were normalized to total RNA amounts (RiboGreen [Molecular Probes]) and then expressed relative to PTPN22 levels in cells that were untransfected with siRNA. The PTPN22 TaqMan sequences were 5′-GGCCCAAAGCAAGAAAATTACTAAA (forward), 5′-TGTCTGCCTTGTACTTGGTAGATTG (reverse), and 5′-TTCAGCTTCAGAAATT (MGB probe).

Project (NYCP), a population-based prospective study of the genetic and environmental factors that cause disease (New York Cancer Project Web site). Two control individuals were matched to a single randomly chosen affected sib on the basis of sex, age (birth decade), and ethnicity (grandparental country/region of origin). These 463 independent cases (referred to as “single sibs”) and their 926 matched controls were used for all analyses reported in this study, except where noted.

The association for single sibs was as follows: allele frequency 15.8% in cases and 8.7% in controls (P = 5.6 × 10⁻⁸, allelic OR 1.97, 95% CI 1.55–2.50). An increase of the risk allele frequency was apparent when all affected siblings (n = 840) were analyzed (allele frequency 16.8%). Genotypic analyses produced similar results, showing increased frequencies of both TT and TC genotypes in the cases, compared with the controls (table 1).

We used contingency tables and conditional-logistic regression (CLR) (Breslow and Day 1980) to assess the association of the PTPN22 R620W genotype with RA (table 2). In the discovery study, both the TC (ORCLR 1.69, P = .0012) and TT (ORCLR 2.26, P value not significant) genotypes conferred an increased risk for dis-
Table 3

Association of PTPN22 with RF+ and RF− Disease

<table>
<thead>
<tr>
<th>Data Set and Genotype</th>
<th>RF Status</th>
<th>No. of Cases</th>
<th>No. of Controls</th>
<th>Results for Crude Analysis*</th>
<th>CLR*</th>
<th>CLR, Adjusted for HLA*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>OR* 95% CI</td>
<td>P</td>
<td>OR* 95% CI</td>
</tr>
<tr>
<td>Discovery:*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT + TC</td>
<td>RF†</td>
<td>125</td>
<td>81</td>
<td>1.74</td>
<td>1.27–2.38</td>
<td>.0005</td>
</tr>
<tr>
<td>CC</td>
<td>RF†</td>
<td>350</td>
<td>394</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Replication:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TT + TC</td>
<td>RF†</td>
<td>119</td>
<td>122</td>
<td>2.38</td>
<td>1.78–3.18</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>CC</td>
<td>RF†</td>
<td>266</td>
<td>648</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>TT + TC</td>
<td>RF†</td>
<td>17</td>
<td>30</td>
<td>1.17</td>
<td>.60–2.29</td>
<td>.64</td>
</tr>
<tr>
<td>CC</td>
<td>RF†</td>
<td>61</td>
<td>126</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

* Values were calculated by use of standard contingency tables.

† CLR was performed as described elsewhere (Breslow and Day 1980) to account for individual matching of controls to cases.

To assess whether the observed associations were independent of the HLA-DRB1 genotype, we further adjusted for this known risk factor. For this analysis, samples were binned according to their HLA-DRB1 genotype: high risk, two shared epitopes (HR-2SE); high risk, one shared epitope (HR-1SE; 4K,X; or 4R,X); low risk, one shared epitope (LR-1SE; 1R,X; or 10,X); and low risk, zero shared epitopes (LR-0SE; X,X) (Fries et al. 2002).

d ORs were calculated relative to the major allele homozygote (CC).

e All patients in the discovery cohort are RF/H11001.

f Because of the infrequency of the TT genotype, we combined it with the TC genotype for these analyses.

ease compared with the CC genotype. Further adjustment for the HLA-DRB1 genotype, the strongest known genetic risk factor (Seldin et al. 1999), had little impact on risk estimates. Similar results were observed in the replication study. The susceptible TT and TC genotypes were strongly associated with rheumatoid factor–positive (RF+) disease, even after adjustment for the HLA genotype (P < .0001 in discovery study, P ≈ .0005 in replication study) (table 3); however, there was no evidence for association with rheumatoid factor–negative (RF−) disease. This interesting observation has been replicated in an additional cohort of patients with recent-onset RA (A. T. Lee and P. K. Gregersen, unpublished data). There was no consistent sex difference in association of the risk allele with RA (data not shown).

PTPN22 is located on chromosome 1p13, ∼9 Mb centromeric to a microsatellite marker, D1S1631, that
shows linkage to RA in the NARAC sib pairs (SIBPAL \( P = .0011 \)) (Jawaheer et al. 2003) from which our replication study was drawn. Linkage analysis using Allegro (Gudbjartsson et al. 2000) yielded an insignificant LOD score of 0.15 for PTPN22 R620W in these sib pairs, indicating that R620W is not solely responsible for the linkage seen at D1S1631. We then conducted stratified analyses based on PTPN22 R620W genotypes of the probands. Among 7, 76, and 193 affected sib pairs for which the proband had the TT, TC, or CC genotype, respectively, the respective mean identity-by-descent–sharing values were 0.598, 0.505, and 0.486, and the respective nonparametric linkage (NPL) scores were 0.94, 0.35, and −0.46 (Allegro) (Gudbjartsson et al. 2000). These results suggested that R620W has a weak effect on the 1p13 linkage signal. Li et al. (2004) have recently shown that there is large variability in NPL scores when families are stratified by the genotype of a single, randomly selected sib, and they developed a program (Genotype-IBD Sharing Test [GIST]) in which families are weighted on the basis of the genotypes of all family members. GIST analysis that was conditional on the T allele indicated significant evidence for linkage \( (P < .0001) \). The apparent difference between the significance of the stratified analysis and the GIST results may reflect the increased power of the GIST algorithm, or, alternatively, the rarity of the TT genotype may lead to poor convergence of the GIST test statistic to the asymptotic distribution. Taken together, all these data suggest that, whereas the SNP may account for a small part of the linkage signal seen on 1p, it is clearly not responsible for the entire signal.

We also genotyped this SNP in several additional control populations (table 4). The observed frequency of the risk allele in 560 additional white subjects (8.4%) was consistent with results in our two RA control populations. When all three control data sets of whites \( (n = 1,961) \) were combined, the T allele was present in 16.7% of individuals at an allele frequency of 8.7%. The T allele frequency was lower in 99 Mexican Americans (3.5%) and 409 African Americans (2.4%). This allele was not detected in 100 Han Chinese or 21 Africans. A Fisher’s Exact Test showed that these differences in allele frequencies were highly significant \( (P < 1 \times 10^{-10}) \). It will be important to expand these studies to determine whether presence of the T allele in the African American and Mexican American populations is due to admixture with whites.

PTPN22 encodes a 110-kD cytoplasmic protein tyrosine phosphatase that consists of an N-terminal phosphatase domain and a long C-terminal region containing several proline-rich motifs (Cohen et al. 1999). The mouse ortholog, PEP (encoded by the murine gene \( Ptpn8) \), has been shown to be a potent down-regulator of T-cell receptor–dependent responses through its association with the SH3 domain of Csk (Cloutier and Veillette 1999). Although PEP and PTPN22 are only 70% identical at the amino acid level (Cohen et al. 1999), overexpression data suggest PTPN22 may play a role similar to that of PEP (Hill et al. 2002). To confirm that PTPN22 functions as a negative regulator of T-cell activation, we used RNA interference (RNAi) to decrease expression in Jurkat cells and then measured the response to antigen-receptor stimulation. mRNA knockdown by two independent siRNAs increased T-cell receptor–dependent activation by approximately two-fold, as measured by an NF-κB-reporter response (fig. 1). This
Figure 3  RNA expression profile of PTPN22. Expression of the PTPN22 major splice variant (GenBank accession number NM_015967) was determined by kinetic RT-PCR analysis by use of total RNA from sorted hematopoietic cells (AllCells) and selected tissues (Clontech), as described elsewhere (Rogge et al. 2000). Data for additional tissues are provided online (table A1 [online only]). To eliminate genomic DNA amplification, amplification primers (5′-GGTGAGGAAGCTGGAGAAT and 5′-GGGAGAAGAACGATCTTGATGTA) were selected from different exons. Each RNA sample was also amplified with seven housekeeping genes, as described elsewhere (Rogge et al. 2000). The level of expression of these housekeeping genes was used to normalize the amount of message in all samples. The normalized expression levels of PTPN22 in all normal tissues (excluding tumor cell lines and purified hematopoietic cells) were averaged, and the results from each individual sample were expressed as a fold change relative to this average. A positive number indicates more PTPN22 message in the indicated sample relative to the average of all samples, whereas a negative number means there is less PTPN22 message relative to the overall average.

is consistent with recent results in PEP-deficient mice (Hasegawa et al. 2004). These animals have rather subtle phenotypic alterations in T-cell function, with enhanced activation of Lck and expansion of memory T cells, but apparently normal naive T-cell function.

The first proline-rich domain (P1) of PEP binds the SH3 domain of the negative regulatory kinase Csk (Cloutier and Veillette 1996; Gregorieff et al. 1998). Bottini and colleagues (2004) have recently shown that the R620W SNP (which is located in P1) affects the binding of PTPN22 to Csk. We confirmed this finding by cotransfection of full-length cDNA clones for PTPN22 (R620 or W620) and Csk into 293T cells, followed by immunoprecipitation of PTPN22. Although both PTPN22 constructs were expressed at similar levels, 2.5–3-fold less Csk was coimmunoprecipitated by the W620 protein, relative to the R620 protein (fig. 2). These data suggest that the association of this missense SNP with RA may be due to the inability of the variant phosphatase to bind Csk and down-regulate T-cell activation.

Prior studies have suggested that PTPN22 is expressed primarily in hematopoietic tissues (Cohen et al. 1999; Hill et al. 2002; Chien et al. 2003), such as thymus, spleen, bone marrow, and peripheral blood mononuclear cells (PBMCs). We confirmed and extended these findings by examining numerous tissues and specific PBMC subsets by use of semiquantitative kinetic RT-PCR (fig. 3 and table A1 [online only]). Expression appeared to be largely confined to hematopoietic tissues and was present in all subtypes of normal human PBMCs tested, including resting CD3+ T cells, CD4+ T cells, CD8+ T cells, B cells, monocytes, neutrophils, dendritic cells, and natural killer (NK) cells (fig. 3). However, there appears to be a hierarchy, with NK cells and neutrophils expressing the most PTPN22 message and B cells and CD4+ T cells expressing the least. These data raise the possibility that the association of PTPN22 with RA could also be a result of functional changes in these other cell types. In particular, monocytosis and monocyte activation are characteristic features of RA, and NK and NK-T cells may also be involved in the pathogenesis of RA (Kojo et al. 2001; Dalbeth and Callan 2002). As a consequence, a complete analysis of PTPN22 function in these cell populations will be required to fully elucidate the role of this protein in autoimmunity.

Familial clustering of multiple autoimmune disorders in the same family is well documented, and first-degree relatives of RA probands have a significantly higher prevalence of type 1 diabetes (T1D [MIM 222100]) than the general population (Lin et al. 1998). In addition,
genomewide scans for autoimmune diseases, in both mouse and human, show colocalization of susceptibility loci (Becker et al. 1998; Griffiths et al. 1999), suggesting a possible shared genetic basis for autoimmunity. The finding that the minor allele of the PTPN22 SNP reported here is also associated with T1D (Bottini et al. 2004) supports the hypothesis that there are common genetic variants that contribute to general immune dysregulation and susceptibility to autoimmunity (Marrack et al. 2001; Wandstrat and Wakeland 2001). It will be important to examine PTPN22 associations in a wide variety of autoimmune diseases and to determine its functional role in each disease. Although a common underlying mechanism for autoimmunity is tempting to postulate, given the expression of this molecule in many immunologically relevant cell types, the possibility remains that PTPN22 may act in different ways in different autoimmune diseases.

**Acknowledgments**

We are grateful to the patients with RA, the controls, and the collaborating clinicians, for participation in this study; members of the Celera Diagnostics High Throughput, Biomarker, and Computational Biology groups, for invaluable help; T. L. Bugawan and E. Trachtenberg, for HLA typing; R. Lundsten, M. Kern, H. Khalili, and A. Rodrigues-Brown, for database and sample management of the NARAC collection; J. Lemaire and S. Mahan, for database and sample management at GCI; and T. White, S. Broder, R. Zamoyska, X. Hu, S. Dalrymple, J. Buggy, K. Van Orden, M. Kale, and P. Young, for discussions and insightful comments on this manuscript. Collection of the NARAC cohort has been funded by a National Arthritis Foundation grant and the National Institutes of Health, acting through the National Institute of Arthritis and Musculoskeletal and Skin Diseases and the National Institute of Allergy and Infectious Diseases (contracts N01-AR-7-2232 and R01-AR44222). 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 RR-00079, U.S. Public Health Service). Support was also provided by the National Institutes of Health through grant HG02275.

**Electronic-Database Information**

The accession number and URLs for data presented herein are as follows:

- GenBank, http://www.ncbi.nlm.nih.gov/Genbank/ (for PTPN22 major splice variant [accession number NM_015967])
- Genotype-IBD Sharing Test (GIST), http://phg.mc.vanderbilt.edu/GIST.shtml
- Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for RA and T1D)

**References**


