

# Genetic Association Between the *PRKCH* Gene Encoding Protein Kinase C $\eta$ Isozyme and Rheumatoid Arthritis in the Japanese Population

Yoichiro Takata,<sup>1</sup> Daisuke Hamada,<sup>1</sup> Katsutoshi Miyatake,<sup>1</sup> Shunji Nakano,<sup>1</sup> Fumio Shinomiya,<sup>2</sup> Charles R. Scafe,<sup>3</sup> Vincent M. Reeve,<sup>3</sup> Dai Osabe,<sup>4</sup> Maki Moritani,<sup>1</sup> Kiyoshi Kunika,<sup>1</sup> Naoyuki Kamatani,<sup>5</sup> Hiroshi Inoue,<sup>1</sup> Natsuo Yasui,<sup>1</sup> and Mitsuo Itakura<sup>6</sup>

**Objective.** Analyses of families with rheumatoid arthritis (RA) have suggested the presence of a putative susceptibility locus on chromosome 14q21–23. This large population-based genetic association study was undertaken to examine this region.

**Methods.** A 2-stage case–control association study of 950 unrelated Japanese patients with RA and 950 healthy controls was performed using >400 gene-based common single-nucleotide polymorphisms (SNPs).

**Results.** Multiple SNPs in the *PRKCH* gene encoding the  $\eta$  isozyme of protein kinase C (PKC $\eta$ ) showed significant single-locus disease associations, the most significant being SNP c.427+8134C>T (odds ratio 0.72, 95% confidence interval 0.62–0.83,  $P = 5.9 \times 10^{-5}$ ). Each RA-associated SNP was consistently mapped to 3 distinct regions of strong linkage disequilibrium (i.e., linkage disequilibrium or haplotype blocks)

in the *PRKCH* gene locus, suggesting that multiple causal variants influence disease susceptibility. Significant SNPs included a novel common missense polymorphism of the *PRKCH* gene, V374I (rs2230500), which lies within the ATP-binding site that is highly conserved among PKC superfamily members. In circulating lymphocytes, *PRKCH* messenger RNA was expressed at higher levels in resting T cells (CD4+ or CD8+) than in B cells (CD19+) or monocytes (CD14+) and was significantly down-regulated through immune responses.

**Conclusion.** Our results provide evidence of the involvement of *PRKCH* as a susceptibility gene for RA in the Japanese population. Dysregulation of PKC $\eta$  signal transduction pathway(s) may confer increased risk of RA through aberrant T cell-mediated autoimmune responses.

Rheumatoid arthritis (RA), a common systemic inflammatory disease with autoimmune features, affects ~1% of the adult population worldwide. The pathogenesis of RA involves an aberrant immune response to self antigens, which induces persistent synovial inflammation and subsequent cartilage destruction and bone erosions. Although the etiology of RA remains largely unknown, numerous epidemiologic studies have provided strong evidence of a major genetic component of RA susceptibility (1,2). In twin studies, a higher concordance rate for RA in monozygotic than in dizygotic twins has been shown, with heritability approaching 60% (3). Family-based analyses have shown an increased risk of RA in siblings of affected probands compared with the general population (relative risk in siblings [ $\lambda_s$ ] 2–17) (2).

An association between disease risk and the HLA locus has been well established in different ethnic groups. In particular, several HLA–DRB1 alleles (\*0401,

Supported in part by grants from the Japan Biological Information Consortium, in affiliation with the New Energy and Industrial Technology Development Organization, and the Ministry of Education, Science, and Technology (Knowledge Cluster Initiative).

<sup>1</sup>Yoichiro Takata, MD, Daisuke Hamada, MD, PhD, Katsutoshi Miyatake, MD, Shunji Nakano, MD, PhD, Maki Moritani, PhD, Kiyoshi Kunika, PhD, Hiroshi Inoue, MD, PhD, Natsuo Yasui, MD, PhD: The University of Tokushima, Tokushima, Japan; <sup>2</sup>Fumio Shinomiya, MD, PhD: Mima Hospital, Tokushima, Japan; <sup>3</sup>Charles R. Scafe, BSc, Vincent M. Reeve, BSc: Applied Biosystems, Foster City, California; <sup>4</sup>Dai Osabe, MSc: Japan Biological Information Consortium, and Fujitsu, Tokyo, Japan; <sup>5</sup>Naoyuki Kamatani, MD, PhD: Tokyo Women's Medical University, Tokyo, Japan; <sup>6</sup>Mitsuo Itakura, MD, PhD: The University of Tokushima, Tokushima, and Japan Biological Information Consortium, Tokyo, Japan.

Address correspondence and reprint requests to Mitsuo Itakura, MD, PhD, Division of Genetic Information, Institute for Genome Research, The University of Tokushima, 3-18-15 Kuramotocho, Tokushima-city, Tokushima 770-8503, Japan. E-mail: itakura@genome.tokushima-u.ac.jp.

Submitted for publication March 29, 2006; accepted in revised form August 30, 2006.

\*0404, \*0405, \*0408, \*0101, \*0102, \*1001, and \*1402) are strongly associated with susceptibility to RA (4). These DRB1 alleles encode a conserved amino acid sequence (QKRAA, QRRAA, or RRRRAA) in the third hyper-variable region of the molecule, which is commonly called the shared epitope (5). The development of severe disease has also been associated with the HLA-DRB1 shared epitope (6). However, it is estimated that this association accounts for only one-third of the genetic component of RA (1), implying that several other non-HLA genes contribute to RA susceptibility.

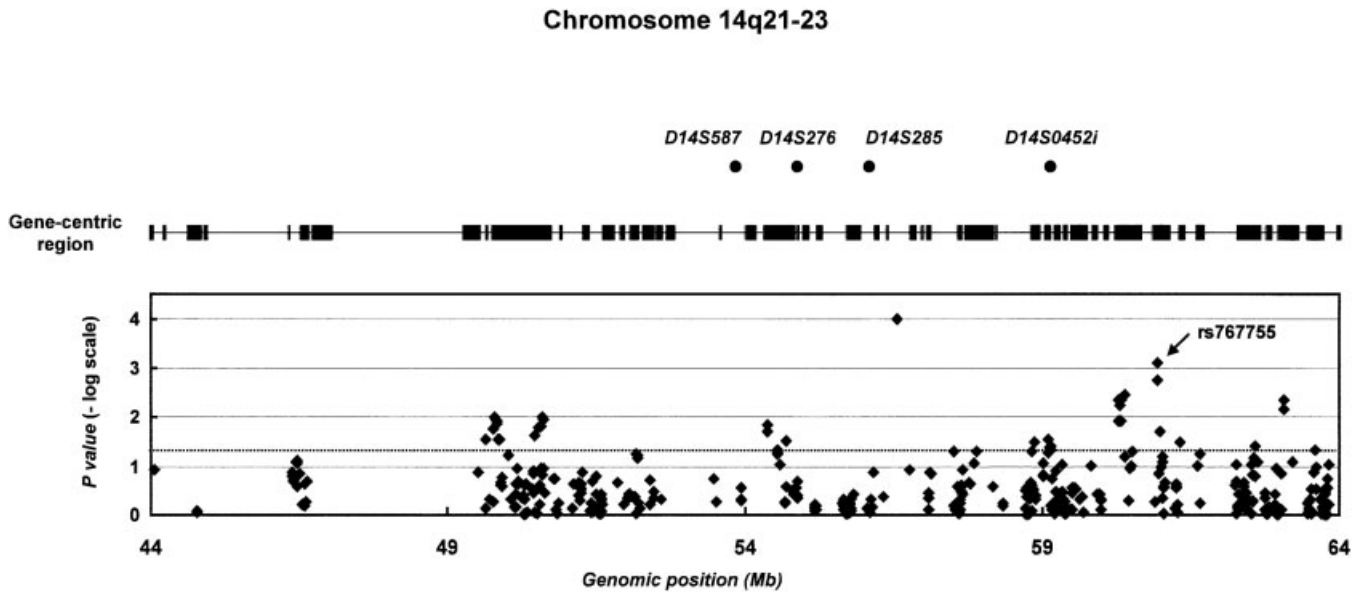
To search for gene(s) underlying susceptibility to RA, genome-wide linkage screening in families with RA has been conducted by at least 4 independent research groups focusing on different ethnicities. The European Consortium on Rheumatoid Arthritis Families focused on Europeans (7,8), Shiozawa et al focused on Japanese (9), the North American Rheumatoid Arthritis Consortium focused on Caucasians living in the US (10,11), and the Arthritis Research Campaign: UK National Repository of Multicase RA Families focused on Caucasians living in the UK (12,13). In these studies, linkage to the HLA locus on chromosome 6p was confirmed, and a number of potentially important non-HLA loci were identified. Several loci (e.g., 1p, 1q, 2q, 5q, 13q, 14q, 16p, 18p, and Xq) overlapped  $\geq 2$  genome screens (8); however, other than the HLA region, no obvious consensus regarding which chromosomal regions would be most likely to contain RA susceptibility genes was obtained. This is consistent with other complex genetic disorders such as type 1 diabetes mellitus (DM), and could be due to a number of factors, including the limited statistical power of individual studies (false-positive or false-negative findings), intra- and interpopulation differences in the effect size of each gene, and genetic heterogeneity.

One way to circumvent and minimize these problems is to use a meta-analysis, in which data across independent genome scans are combined. Indeed, Fisher et al (14) completed a meta-analysis using the rank-based genome scan meta-analysis method on the early genome scan data obtained in France, the US, the UK, and Japan, and reported linkage signals on 6p (HLA), 6q, 12p, and 16cen ( $P < 0.01$ ), as well as 8 novel regions with suggestive evidence ( $P < 0.05$ ). Choi et al (15) also applied the genome scan meta-analysis method to data obtained from European and Caucasian populations, and reported evidence of the involvement of 7 chromosomal regions, including the HLA region. More recently, Etzel et al (16) applied a novel meta-analytical

method developed by Loesgen et al (17) to Caucasian-only populations, and reported overwhelming evidence of linkage in the HLA region, strong evidence of 8p and 16p ( $P < 0.01$ ), and marginal evidence of 1q, 2q, 5q, and 18q ( $P < 0.05$ ). Taken together, results of these meta-analyses highlight several regions of genetic linkage for RA, especially on chromosomes 1q, 6p, 12p, 16p, and 18q, that can be further studied in fine mapping and candidate gene analyses.

Chromosome 14q is one of the regions for which evidence of linkage to RA was found in the meta-analysis by Fisher et al (14), although the reported significance level was not overwhelming when compared with other possible linkage regions. We noticed that in this region, microsatellite markers showing at least some evidence of significance in each study were tightly clustered in a narrow genomic region of  $\sim 3$  Mb. These microsatellite markers were D14S587 ( $P = 0.037$  for Caucasians living in the US by sibpair analysis with the SibPal program) (10), D14S276 ( $P = 0.03$  for Caucasians living in the UK; logarithm of odds [LOD] 0.9 for Japanese) (9,12), and D14S285 ( $P = 0.049$  for Europeans by single-point analysis) (7). In addition, a recent whole-genome case-control association study in a Japanese population identified another closely located marker, D14S0452i, to be significantly associated with RA ( $P = 0.001$  by Fisher's exact test) (18). These adjacent markers on chromosome 14q23 may be indicative of a common RA susceptibility gene across racial/ethnic groups.

Interestingly, the chromosome 14q21-q23 region has also been implicated repeatedly in susceptibility to systemic lupus erythematosus (SLE). Suggestive evidence of linkage to this region was reported in Caucasian populations in the US and Canada (D14S276; LOD 2.81,  $P = 0.00016$ ) (19), in Mexican Americans and Caucasians in the US (D14S63; nonparametric linkage [NPL] 1.70,  $P = 0.04$  and D14S258; NPL 2.02,  $P = 0.02$ ) (20), in the Swedish population (D14S592; LOD 1.15) (21), and in the Finnish population (D14S587; NPL 2.20,  $P = 0.02$ ) (22), although no single study has obtained significant linkage results. It is well known that different autoimmune diseases share susceptibility loci (23), and there are numerous reports describing concurrence of RA and SLE in individuals (24) or in families (25,26). In this study, therefore, we hypothesized that a gene predisposing to RA and/or SLE may exist in chromosome 14q21-q23, and we investigated this region in a large-scale association study with a set of  $>400$  single-nucleotide polymorphism (SNP) markers in 950 unre-



**Figure 1.** Results of a single-nucleotide polymorphism (SNP) association study on chromosome 14q21–23. Allelic  $P$  values (negative log scale) of 378 gene-based SNPs (also referred to as gene-centric evenly spaced common SNPs) from the exploratory sample are plotted against genomic positions. Dashed line indicates a  $P$  value of 0.05. SNPs were selected mainly in gene-coding regions rather than in intergenic regions. Locations of 4 putative rheumatoid arthritis-linked microsatellite markers (D14S587, D14S276, D14S285, and D14S0452i) are also shown. A landmark SNP located in the *PRKCH* gene showing evidence of association in the 2-stage case-control study, rs767755, is indicated.

lated Japanese patients with RA and 950 controls. A similar approach was recently used to identify the *SEC8L1* gene, located on chromosome 7q31, as a putative RA susceptibility gene in Japanese individuals (27). Our approach was useful in demonstrating an association between RA and multiple variants of the *PRKCH* gene, which encodes the  $\eta$  isozyme of protein kinase C (PKC $\eta$ ).

## PATIENTS AND METHODS

**Subjects.** The study was conducted in accordance with the tenets of the Declaration of Helsinki and approved by the Ethics Committee for Human Genome and Gene Research at The University of Tokushima. Written informed consent was obtained from all participants prior to enrollment in the study. A total of 1,900 Japanese subjects, consisting of 950 RA patients (200 men and 750 women; mean  $\pm$  SD age at enrollment  $61.7 \pm 12.4$  years, mean  $\pm$  SD age at RA diagnosis  $48.7 \pm 13.4$  years) and 950 healthy controls (205 men and 745 women; mean  $\pm$  SD age at enrollment  $39.1 \pm 15.2$  years), were recruited through the orthopedic clinic at Tokushima University Hospital and its affiliates, and the rheumatology clinic at Tokushima Kensei Hospital (Tokushima, Japan). Detailed information on patients and control subjects was previously reported (27). All RA patients satisfied the revised criteria of the American College of Rheumatology (formerly, the American Rheumatism Association) (28), with a minimum disease duration of 3 years. Blood specimens were collected from all

study subjects. An additional set of 46 DNA samples was obtained from unrelated healthy Japanese volunteers (24 men and 22 women) at The University of Tokushima, and was used for optimization of the SNP genotyping assay and estimating allele frequencies, as previously described (29). Genomic DNA samples were extracted from peripheral blood leukocytes according to standard methods.

**SNP selection and genotyping.** For SNP-based fine mapping of the chromosome 14q susceptibility locus, we reviewed 4 RA whole-genome linkage scans published in March 2002 or earlier (7,9,10,12), and searched for the primary target region likely to contain gene(s) predisposing to RA. We chose an  $\sim$ 20-Mb genomic segment between markers D14S288 and D14S63 on chromosome 14q21–23, centering around the 3 putative RA-associated markers (D14S587, D14S276, and D14S285) (Figure 1). This region also included D14S0452i, which recently was reported to be associated with RA in Japanese subjects (18). According to the human genome data (National Center for Biotechnology Information [NCBI] Build 35.1), the target segment contained a total of 170 known or predicted genes.

SNPs used for genotyping were identified from either the NCBI Database of Single-Nucleotide Polymorphisms or the Celera Discovery System database. For this study, we selected gene-based SNPs, also referred to as gene-centric evenly spaced common SNPs (27,30), using the following criteria: suitability for designing high-throughput TaqMan SNP genotyping assays; mapping to the gene-centric region, which we assigned to an interval between 10 kb upstream from the transcription start site and 10 kb downstream from the final exon; an average distance of 10 kb between adjacent SNPs whenever possible; and common SNPs with a minor allele

frequency of  $>0.15$  based on genotyping data from 46 Japanese control individuals, supplied by Applied Biosystems (Foster City, CA) (Scafe CR, Reeve VM: unpublished observations).

A total of 422 high-quality SNPs, with a median inter-SNP distance of 10.3 kb, were selected and used for further association testing. The mean  $\pm$  SD minor allele frequency was  $0.33 \pm 0.10$ . Of 422 SNPs, 406 (96.2%) were mapped within 67 genes (39.4% of the total target genes), and the remaining 16 were mapped in intergenic regions (for further information on the 422 SNPs, see supplementary Table 1, available online at <http://www.mrw.interscience.wiley.com/suppmat/0004-3591/suppmat/>). There were significant numbers of allele associations among the SNPs, mostly between 2 or 3 adjacent SNPs, i.e., SNPs in linkage disequilibrium (data not shown). When the SNP data set was analyzed using the method described by Gabriel et al (31), there were 77 genomic regions showing significant linkage disequilibrium (commonly known as a linkage disequilibrium block or a haplotype block).

All but 2 SNPs were genotyped by TaqMan assays in a 384-well plate format, using commercially available reagents (TaqMan Universal polymerase chain reaction [PCR] Master Mix, no AmpErase UNG; Applied Biosystems). Fluorescence was detected using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Genotyping accuracy was previously demonstrated by showing 100% concordance with results obtained by direct sequencing and by randomly retyping (30). Two SNPs in the *PRKCH* gene, rs2230500 and rs2230501, were genotyped by PCR direct sequencing with BigDye Terminator Cycle Sequencing kit version 1.1 and an ABI 3730xl DNA analyzer (Applied Biosystems).

**Two-stage association study.** For the SNP association study, we applied a 2-stage case-control analysis strategy in 1,900 subjects by randomly assigning them to 2 independent subsets. The exploratory study consisted of 380 patients and 380 controls, while the validation study consisted of an independent sample set of 570 patients and 570 controls. There were no significant differences in any clinical parameter in either the control or the patient group between these 2 subsets (data not shown). All 422 SNPs were first tested in the exploratory sample, and SNPs exhibiting significant allele associations ( $P < 0.05$ ) were further tested in the independent validation sample. Significant SNPs ( $P < 0.05$ ) in the validation sample were tentatively regarded as putative RA-associated SNPs and evaluated further. The standard Bonferroni correction was used for statistical adjustment for multiple comparisons. The overall power of our 2-stage case-control design, which was calculated with CaTS version 0.0.1 (32), was  $>0.8$ , assuming a multiplicative genetic model with a genotype relative risk (GRR) of 1.4.

**Resequencing and polymorphism screening of the PKC $\eta$  gene (*PRKCH*).** To identify novel DNA polymorphism(s)/mutation(s) in the *PRKCH* gene locus, all coding exons and a 26-kb region in intron 2 were resequenced in 24 randomly chosen individuals (12 RA patients and 12 controls). For direct sequencing, PCR amplicons were purified using ExoSAP-IT (Amersham Biosciences, Piscataway, NJ) and, subsequently, a cycle sequencing reaction was performed as described above. Details of the primers and PCR conditions are available from the authors upon request.

**RNA extraction, complementary DNA preparation, and reverse transcription (RT)-PCR analysis.** Synovial specimens were obtained from the knee joints of patients with RA or patients with osteoarthritis (OA) who underwent arthroplasty. Fibroblast-like synoviocytes (FLS) derived from synovial tissue were cultured in Dulbecco's modified Eagle's medium containing 10% fetal calf serum. Total cellular RNA from FLS was extracted using an RNeasy kit (Qiagen, Valencia, CA). Other RNAs from various human tissue samples (Human Total RNA Master Panel II, Human Blood Fractions MTC Panel) were purchased from BD Biosciences (Palo Alto, CA). RNAs were reverse-transcribed with random hexamer primers and Superscript II (Invitrogen, Carlsbad, CA).

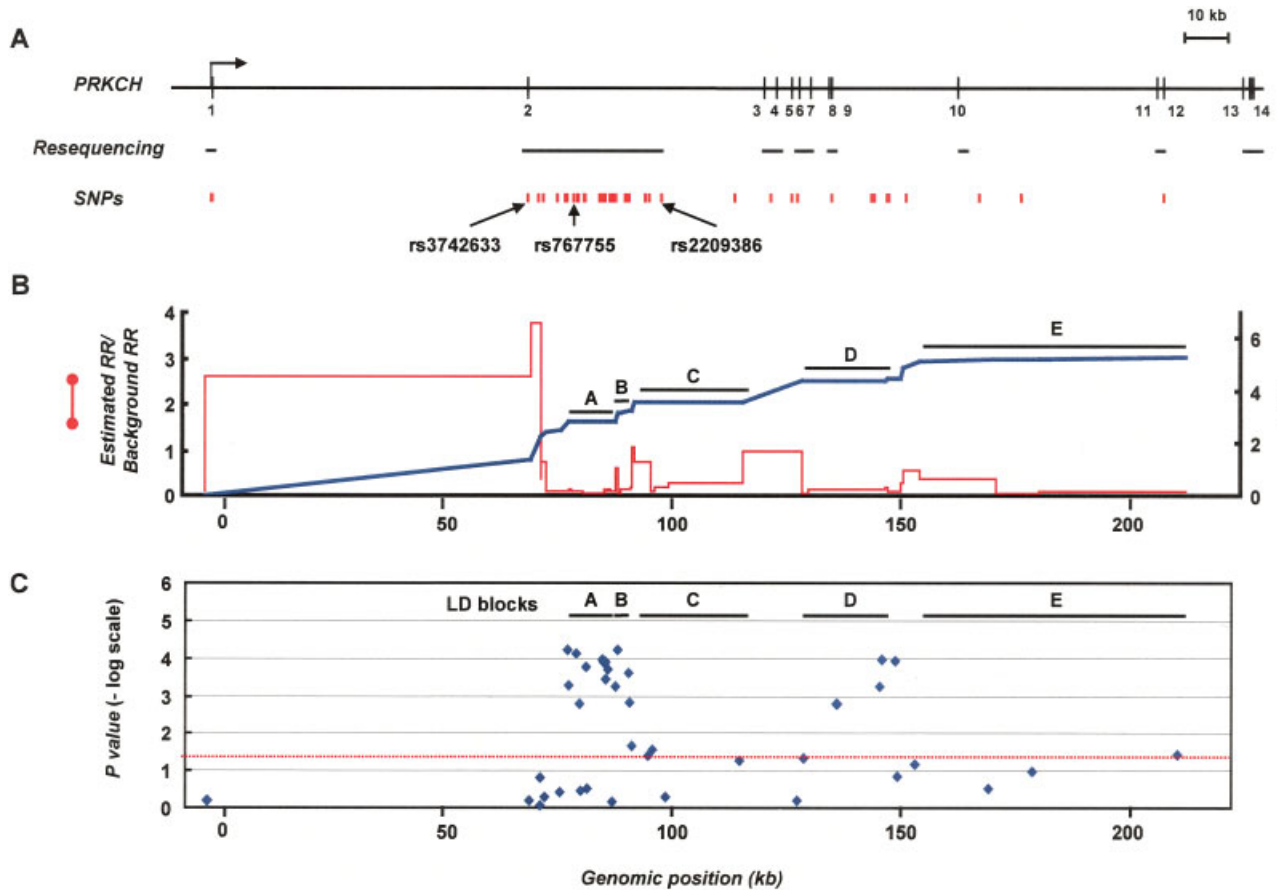
For conventional RT-PCR analysis, the following specific primers for human *PRKCH* and *ACTB* ( $\beta$ -actin) genes were designed based on established nucleotide sequences (GenBank accession nos. NM\_006255 and NM\_001101, respectively): for *PRKCH*, 5'-GGG-AGT-TTC-ACT-GAA-GCT-AC-3' (forward) and 5'-CAG-CGT-TTA-TGG-ACG-ACA-CA-3' (reverse); for *ACTB*, 5'-TGA-CCC-AGA-TCA-TGT-TTG-AGA-3' (forward) and *ACTB* 5'-GTG-GTG-GTG-AAG-CTG-TAG-CC-3' (reverse). RT-PCR was performed using AmpliTaq Gold DNA polymerase (PerkinElmer, Yokohama, Japan), under the following conditions: 95°C for 10 minutes, followed by 30 cycles at 95°C for 30 seconds, 56°C for 10 seconds, and 72°C for 30 seconds, and then a final extension for 5 minutes at 72°C. PCR products were subjected to electrophoresis on a 2% agarose gel, followed by staining with ethidium bromide. Real-time quantitative RT-PCR analysis was performed on an ABI Prism 7900HT sequence detection system, using TaqMan Gene Expression Assay probes (for *PRKCH*, Hs00178933\_m1; for *ACTB*, Hs99999903\_m1). The expression level of the *PRKCH* transcript was quantified using the threshold cycle ( $C_t$ ) method, according to the recommendations of the manufacturer (Applied Biosystems), and normalized to the amount of *ACTB*.

**Statistical analysis.** Tests for deviation from Hardy-Weinberg equilibrium were conducted for each SNP locus, using a chi-square goodness-of-fit test. Association tests on individual SNPs were performed with  $2 \times 2$  contingency tables of SNP allele and case-control status with the chi-square test, using SNPalyze Pro version 5.0 (Dynacom, Chiba, Japan). Odds ratios (ORs) were calculated with 95% confidence intervals (95% CIs). Pairwise linkage disequilibrium between the SNPs was assessed by the 2 most common measures, the absolute Lewontin's coefficient ( $|D'|$ ) and Pearson's correlation coefficient ( $r^2$ ) (33). Graphical Overview of Linkage Disequilibrium (GOLD) software was used to graphically depict the linkage disequilibrium results (34). The LDMAP program was used to construct a metric linkage disequilibrium map (35), discriminating regions of conserved linkage disequilibrium with additive distances in linkage disequilibrium units. Haplotypes were inferred from SNP genotyping data using Phase version 2.1 (36), and haplotype-tagging SNPs were determined with SNPalyze Pro version 5.0.

## RESULTS

**Chromosome 14q-specific SNP association study.** To map the RA susceptibility genes on chromosome 14q21-23, we first selected 422 SNPs scattered





**Figure 2.** Human *PRKCH* gene structure, single-nucleotide polymorphisms (SNPs), and linkage disequilibrium block pattern. **A**, Genomic organization of the *PRKCH* gene, consisting of 14 exons (vertical lines), is shown at the top. Horizontal lines represent introns. Genomic regions screened by resequencing included all the coding exons and an ~26-kb sequence of intron 2 between rs3742633 and rs2209386. Red vertical lines show the locations of SNPs used for linkage disequilibrium analysis or an association study. SNPs rs3742633, rs2209386, and rs767755 (landmark SNP) are indicated. **B**, Graph of the linkage disequilibrium unit (LDU) map and estimate of the local recombination rate (RR) (generated by the Phase program) between each pair of consecutive SNPs, plotted against their genomic positions. The LDU map identifies regions of strong linkage disequilibrium as plateaus (i.e., linkage disequilibrium or haplotype blocks) and characterizes steps that reflect recombination events. There are 5 distinct linkage disequilibrium blocks, blocks A–E, in the *PRKCH* gene region. **C**, Allelic  $P$  values (negative log scale) of 43 SNPs in the *PRKCH* gene region, plotted against genomic positions. Dashed line indicates a  $P$  value of 0.05.

throughout the region, and genotyped all in an exploratory sample (380 RA patients and 380 controls). Among 422 SNPs, 44 were dropped from the analysis either because of observed deviations from Hardy-Weinberg equilibrium ( $P < 0.05$ ) in control subjects or because of their low allele frequencies (minor allele frequency  $< 0.10$ ), which resulted in insufficient power to detect statistical significance. Consequently, 378 SNPs were tested for a single-locus association between SNP allele frequencies and case-control status, and significant associations ( $P < 0.05$ ) were detected in 35 SNPs (Figure 1).

These associations were further tested in an

independent validation sample (570 RA patients and 570 controls). Three SNPs (rs767755, rs1255784, and rs1255717) located within 2 genes (*PRKCH* and *PPP2R5E*) showed significant differences ( $P < 0.05$ ) in allele frequencies between patients and controls in this stage; however, for rs1255784 and rs1255717, the directions of association in this second cohort were the opposite of the effects seen in the exploratory sample, and results of association tests were not significant when the exploratory and validation sample sets were pooled (data not shown), indicating that their associations were false-positive findings. A consistently reproducible association was observed only for rs767755. When all of the

genotyping data were pooled (a total of 950 patients and 950 controls), the overall OR for this SNP was 0.73 (95% CI 0.62–0.86,  $P = 1.3 \times 10^{-4}$ ). The association remained significant after stringent Bonferroni correction for the 378 tests performed in the pooled sample ( $P_{\text{corr}} = 0.049$ ). Overall, our association test identified 1 landmark SNP, rs767755, to be associated with RA. This SNP was located in intron 2 of the *PRKCH* gene, encoding the PKC $\eta$  isozyme (for full association results, see supplementary Table 1).

**RA-associated SNPs in the *PRKCH* (PKC $\eta$ ) gene region.** The human *PRKCH* gene consists of 14 exons spanning ~229 kb of genomic DNA (Figure 2A). To examine the linkage disequilibrium and haplotype structure around rs767755 (landmark SNP), and those across the entire *PRKCH* gene, we searched for additional SNPs using the public database and, concurrently, by a resequencing strategy. In the public database, 2 ungenotyped SNPs, rs3742633 and rs2209386, were found to be located 16.5 kb upstream and 10.0 kb downstream from rs767755, respectively (Figure 2A). We first resequenced the entire interval between these 2 SNPs, covering an ~26-kb sequence of intron 2 of the *PRKCH* gene, in 24 individuals (12 RA patients and 12 controls). Within that region, a total of 63 DNA polymorphisms, including 61 single-base substitutions and 2 DNA insertion/deletion polymorphisms, were identified (see supplementary Table 2, available online at <http://www.mrw.interscience.wiley.com/suppmat/0004-3591/suppmat/>). We next searched for *PRKCH* exonic polymorphisms by sequencing all coding exons and exon–intron boundaries, and identified 5 SNPs: 2 SNPs in the 5'-untranslated region (rs11622327 and rs1951964), 2 synonymous SNPs (V374V in exon 9 [rs2230501] and N558N in exon 12 [rs1088680]), and 1 nonsynonymous SNP (V374I in exon 9 [rs2230500]).

We then selected a parsimonious set of 43 SNPs in the *PRKCH* gene region based on their locations, allele frequencies, and the preliminary results of a linkage disequilibrium analysis using the resequencing data (see Table 1 for a list of the 43 SNPs). We genotyped these SNPs for our entire sample (950 RA patients and 950 controls), and the extent and distribution of linkage disequilibrium within the *PRKCH* gene were assessed using the 2 most common measures of pairwise linkage disequilibrium,  $|D'|$  and  $r^2$  (33). We observed  $\geq 5$  linkage disequilibrium blocks, based on  $|D'|$  values of  $>0.8$ . These findings were also supported by the results of  $r^2$  analysis, since there were no apparent correlations among SNPs from any of the individual blocks (see supplementary Figure 1, available online at

<http://www.mrw.interscience.wiley.com/suppmat/0004-3591/suppmat/>).

To define the linkage disequilibrium block structure in a more integrated manner, we constructed a metric linkage disequilibrium units map using the LD-MAP program (35), which assigns a location in linkage disequilibrium units to each SNP, identifies regions of strong linkage disequilibrium (i.e., linkage disequilibrium blocks) as plateaus, and characterizes steps reflecting recombination hot spots or recombination events defining the relationships between blocks. In the *PRKCH* gene locus, we observed 5 distinct linkage disequilibrium blocks (blocks A–E) (Figure 2B), a finding similar to the  $|D'|$  results. The linkage disequilibrium block pattern in the linkage disequilibrium units map was consistent with local recombination rates, as estimated by a Phase algorithm (36) (Figure 2B).

When allele frequencies for the 43 *PRKCH* SNPs were compared between 950 patients and 950 controls, 21 SNPs (11 located within block A, 4 in block B, and 6 in block D) showed a statistically significant association with RA ( $P < 0.01$ ) (Table 1). As shown in Figure 2C, the distribution of RA-associated SNPs was directly dependent on linkage disequilibrium structure. It is noteworthy that the associations of each SNP in block A (significant  $P$  values ranged from  $5.9 \times 10^{-5}$  to  $1.6 \times 10^{-3}$ ) (Table 1) were attributable to increased frequencies of the major alleles in patients, whereas those for SNPs in blocks B and D were attributable to increased frequencies of the minor alleles in patients (significant  $P$  values ranged from  $6.2 \times 10^{-5}$  to  $1.5 \times 10^{-3}$  and from  $1.1 \times 10^{-4}$  to  $1.6 \times 10^{-3}$ , respectively), suggesting that multiple causal variants/alleles may be involved.

Logistic regression analysis revealed that the associations between these SNPs and RA remained statistically significant after adjustment for age and sex (data not shown). In addition, allele frequencies for the significant SNPs were compared in subgroups stratified for sex, duration of disease, rheumatoid factor positivity, and the presence of the HLA–DRB1 shared epitope, and no obvious differences were detected (data not shown). Overall, these data imply that there are 3 independent, linkage disequilibrium block–dependent (blocks A, B, and D) associations between *PRKCH* and RA. Among the significant SNPs, all but 1 were intronic, being located in intron 2, 5, or 9. A significant association result was also obtained for a nonsynonymous SNP found in exon 9 (V374I [rs2230500] [block D]).

Haplotypes generally contain more information than individual SNPs (31,37), because of their higher levels of genetic heterozygosity (or allele diversity). We

**Table 1.** Association analysis of *PRKCH* SNPs\*

SNP	Location	Position 1†	Position 2‡	Allele		MAF§		$\chi^2$	P	OR (95% CI)¶	LD block
				Major	Minor	Patients	Controls				
rs11622327	5'-UTR	-224	60,858,349	A	C	0.128	0.122	0.25	0.62	-	-
rs1951964	5'-UTR	-83	60,858,490	C	G	0.128	0.122	0.22	0.64	-	-
rs3742633	Exon 2	+69,149	60,927,722	A	G	0.191	0.185	0.22	0.64	-	-
rs1957907	Intron 2	+71,411	60,929,984	G	T	0.365	0.368	0.03	0.87	-	-
c.427+2247T>G (this study)#	Intron 2	+71,432	60,930,005	T	G	0.148	0.132	2.09	0.15	-	-
rs1957910	Intron 2	+72,484	60,931,057	A	G	0.394	0.404	0.43	0.51	-	-
c.427+6489G>C (this study)#	Intron 2	+75,674	60,934,247	G	C	0.166	0.156	0.81	0.37	-	-
c.427+8134C>T (this study)#	Intron 2	+77,319	60,935,892	C	T	0.170	0.222	16.1	0.000059	0.72 (0.62–0.83)	A
rs12887002**	Intron 2	+77,690	60,936,263	C	T	0.169	0.213	12.0	0.00054	0.75 (0.63–0.88)	A
rs12147491	Intron 2	+79,174	60,937,747	G	A	0.387	0.450	15.6	0.000077	0.77 (0.68–0.88)	A
rs7142290	Intron 2	+79,962	60,938,535	C	T	0.448	0.499	9.97	0.0016	0.81 (0.71–0.92)	A
rs7141535	Intron 2	+80,132	60,938,705	G	A	0.220	0.232	0.82	0.36	-	A
rs1547148	Intron 2	+81,377	60,939,950	A	G	0.170	0.218	14.1	0.00017	0.73 (0.62–0.87)	A
rs7141628	Intron 2	+81,629	60,940,202	A	T	0.221	0.236	1.09	0.30	-	A
rs7160675	Intron 2	+84,874	60,943,447	T	C	0.171	0.221	14.9	0.00011	0.73 (0.62–0.85)	A
rs7140526	Intron 2	+85,004	60,943,577	A	T	0.170	0.219	14.7	0.00013	0.73 (0.62–0.86)	A
rs2352084	Intron 2	+85,334	60,943,907	T	C	0.170	0.219	14.4	0.00014	0.73 (0.62–0.87)	A
rs767757	Intron 2	+85,554	60,944,127	G	T	0.401	0.458	12.7	0.00036	0.79 (0.69–0.90)	A
rs767755 (landmark SNP)**	Intron 2	+85,628	60,944,201	A	G	0.170	0.219	14.7	0.00013	0.73 (0.62–0.86)	A
rs7145570	Intron 2	+86,001	60,944,574	T	C	0.395	0.455	13.9	0.0002	0.78 (0.70–0.90)	A
c.427+17895A>G (this study)#	Intron 2	+87,080	60,945,653	A	G	0.168	0.163	0.14	0.71	-	A
rs10149839	Intron 2	+87,737	60,946,310	A	G	0.488	0.433	11.9	0.00056	1.25 (1.10–1.43)	B
rs912620	Intron 2	+88,244	60,946,817	G	T	0.477	0.412	16.1	0.000062	1.30 (1.13–1.48)	B
rs2882830	Intron 2	+90,385	60,948,958	G	A	0.476	0.417	13.4	0.00026	1.27 (1.12–1.44)	B
rs4902055	Intron 2	+90,678	60,949,251	A	G	0.418	0.368	10.1	0.0015	1.24 (1.09–1.39)	B
c.427+21986T>C (this study)#	Intron 2	+91,171	60,949,744	T	C	0.227	0.197	5.21	0.022	1.20 (1.01–1.39)	C
rs912621	Intron 2	+94,769	60,953,342	A	G	0.336	0.368	4.22	0.040	0.87 (0.76–0.99)	C
rs2209386††	Intron 2	+95,664	60,954,237	G	A	0.336	0.370	4.83	0.028	0.86 (0.75–0.99)	C
rs11848272**	Intron 2	+98,382	60,956,955	T	A	0.465	0.475	0.41	0.52	-	C
rs12435952††	Intron 2	+114,403	60,972,976	C	T	0.081	0.099	3.74	0.053	-	C
rs2296273††	Intron 4	+126,931	60,985,504	A	G	0.277	0.283	0.19	0.66	-	D
rs2296274††	Intron 5	+128,358	60,986,931	A	G	0.502	0.469	3.92	0.048	1.14 (0.99–1.28)	D
rs2230500	Exon 9 (V374I)	+135,419	60,993,992	G	A	0.247	0.204	10.0	0.0016	1.28 (1.11–1.50)	D
rs2230501	Exon 9 (V374V)	+135,421	60,993,994	A	C	0.247	0.204	10.0	0.0016	1.28 (1.11–1.50)	D
rs912616††	Intron 9	+144,657	61,003,230	G	A	0.218	0.174	11.9	0.00056	1.33 (1.12–1.56)	D
rs912618††	Intron 9	+144,771	61,003,344	A	G	0.218	0.174	11.9	0.00057	1.33 (1.13–1.57)	D
rs959728††	Intron 9	+145,198	61,003,771	C	T	0.289	0.234	15.0	0.00011	1.33 (1.15–1.54)	D
rs3783789††	Intron 9	+147,980	61,006,553	A	G	0.229	0.179	14.7	0.00012	1.37 (1.17–1.60)	D
rs3783786††	Intron 9	+148,404	61,006,977	A	G	0.510	0.486	2.14	0.14	-	-
rs2181985**	Intron 9	+152,090	61,010,663	G	A	0.355	0.383	3.31	0.069	-	E
rs4902064**	Intron 10	+168,004	61,026,577	C	T	0.336	0.352	1.08	0.30	-	E
rs12147132**	Intron 10	+177,290	61,035,863	A	G	0.146	0.165	2.49	0.11	-	E
rs1088680	Exon 12	+208,406	61,066,979	C	T	0.272	0.243	4.32	0.038	1.17 (1.01–1.35)	E

\* LD = linkage disequilibrium; 5'-UTR = 5'-untranslated region.

† Numbers refer to the corresponding position starting with +1 at the adenine of initiation codon ATG.

‡ Physical position on chromosome 14 (National Center for Biotechnology Information build 35.1).

§ Minor allele frequencies (MAFs) obtained from the pooled (first- and second-stage) population (950 patients and 950 controls) are shown.

¶ Odds ratios (ORs) reflect the effects of minor alleles on susceptibility to rheumatoid arthritis. 95% CI = 95% confidence interval.

# Single-nucleotide polymorphism (SNP) names are given using the human *PRKCH* cDNA sequence (GenBank accession no. NM\_006255) as a reference sequence, according to the MDI/Human Genome Variation Society Mutation Nomenclature Recommendations (<http://www.hgvs.org/mutnomen/>).

\*\* Gene-centric evenly spaced common SNPs.

†† SNPs from the public database.

therefore investigated the haplotypes in linkage disequilibrium blocks A, B, and D. In each of these 3 linkage disequilibrium blocks, 5 haplotypes that accounted for >98% of haplotypes comprising the respective sites were inferred using the Phase algorithm (36) (desig-

nated A1–A5, B1–B5, and D1–D5, respectively) (Table 2). As indicated in Table 2, the minimum numbers of SNPs required to tag these haplotypes (haplotype-tagging SNPs) were 5 for haplotypes A1–A5 (c.427+8134C>T, rs7142290, rs7141535, rs767757, and

**Table 2.** Haplotype association analysis

Linkage disequilibrium block, haplotype	Haplotype-tagging SNP alleles*	Frequency		$\chi^2$	<i>P</i>	Permutation <i>P</i>
		Patients	Controls			
<b>Block A</b>						
A1	CCGGA	0.430	0.375	11.6	0.00067	0.00082
A2	CTATA	0.219	0.230	0.68	0.41	0.41
A3	TTGTA	0.169	0.219	14.9	0.00011	0.00008
A4	CCGGG	0.112	0.117	0.23	0.63	0.62
A5	CTGGG	0.056	0.047	1.78	0.18	0.19
<b>Block B</b>						
B1	AGGA	0.499	0.557	12.5	0.0004	0.00047
B2	GTAG	0.399	0.335	16.6	0.000046	0.00006
B3	GTAA	0.054	0.047	1.16	0.28	0.29
B4	GTGA	0.024	0.028	0.41	0.52	0.52
B5	GGAG	0.011	0.022	7.51	0.0061	0.0063
<b>Block D</b>						
D1	AGC	0.488	0.526	5.22	0.022	0.022
D2	GGC	0.194	0.208	1.05	0.31	0.32
D3	GAT	0.217	0.170	13.1	0.0003	0.0003
D4	GGT	0.061	0.058	0.16	0.69	0.71
D5	GAC	0.028	0.034	1.11	0.29	0.3

\* The order of haplotype-tagging single-nucleotide polymorphisms (SNPs), shown in 5' to 3', is (left to right), for linkage disequilibrium block A, c.427+8134C>T, rs7142290, rs7141535, rs767757, c.427+17895A>G; for linkage disequilibrium block B, rs10149839, rs912620, rs2882830, rs4902055; for linkage disequilibrium block D, rs2296274, V374I, rs959728.

c.427+17895A>G), 4 for haplotypes B1–B5 (rs10149839, rs912620, rs2882830, and rs4902055), and 3 for haplotypes D1–D5 (rs2296274, rs2230500 [V374I], and rs959728).

Comparisons of haplotype frequencies between patient and control groups revealed 3 common haplotypes, A1, B2, and D3, to be susceptibility haplotypes with increased frequencies in RA patients (*P* values ranged from  $4.6 \times 10^{-5}$  to  $6.7 \times 10^{-4}$ ; permutation test-based *P* values ranged from  $6.0 \times 10^{-5}$  to  $8.2 \times 10^{-4}$ ), while 3 other common haplotypes, A3, B1, and D1, and a less common haplotype, B5, which were protective haplotypes, had lower frequencies in patients (*P* values ranged from  $1.1 \times 10^{-4}$  to  $2.2 \times 10^{-2}$ ; permutation test-based *P* values ranged from  $8.0 \times 10^{-5}$  to  $2.2 \times 10^{-2}$ ). However, these differences were no more significant than those observed when the individual SNPs that constituted each linkage disequilibrium block were tested independently.

**Genotype combination effects among SNPs residing in different linkage disequilibrium blocks.** Identification of linkage disequilibrium block-dependent associations (blocks A, B, and D) between the *PRKCH* gene and RA prompted us to examine the genotype combination effects among SNPs residing in individual linkage disequilibrium blocks. First, we selected 1 SNP from each of the 3 linkage disequilibrium blocks for which allele associations with RA were most significant. From block A we selected c.427+8134C>T; from block

B, rs912620; and from block D, rs959728. Among the 3 SNPs, all pairwise  $r^2$  measures of linkage disequilibrium were confirmed to be  $<0.1$  (supplementary Figure 1). For simplicity, we then designated the putative susceptibility alleles (c.427+8134C, rs912620-T, and rs959728-T) and protective/neutral alleles (c.427+8134T, rs912620-G, and rs959728-C) as allele 1 and allele 0, respectively. At each single SNP site, the presence of the susceptibility allele in both single (0/1) and double (1/1) doses conferred an increased OR relative to the 0/0 genotype (absence of susceptibility allele) (Table 3). The differences were significant for all SNPs in individuals homozygous for the susceptibility allele (1/1 genotype; *P* values ranged from  $2.7 \times 10^{-5}$  to  $2.8 \times 10^{-3}$ ); however, in heterozygotes (0/1 genotype), only rs959728 reached statistical significance (*P* = 0.0026).

Using genotyping data for these 3 SNP loci, we calculated the ORs for all observed 2-SNP or 3-SNP combinations. In the 2-SNP-combination analysis, the ORs were generally increased with an increased number of susceptibility alleles (Table 3), and the increase was most significant among individuals homozygous for  $\geq 1$  susceptibility SNP allele, depending on the number(s) of second susceptibility SNP alleles. The same tendency was observed in the 3-SNP-combination analysis (Table 3); however, statistical significance was obtained mainly for the effect of the susceptibility allele at rs959728 in individuals who were double-homozygous for suscepti-



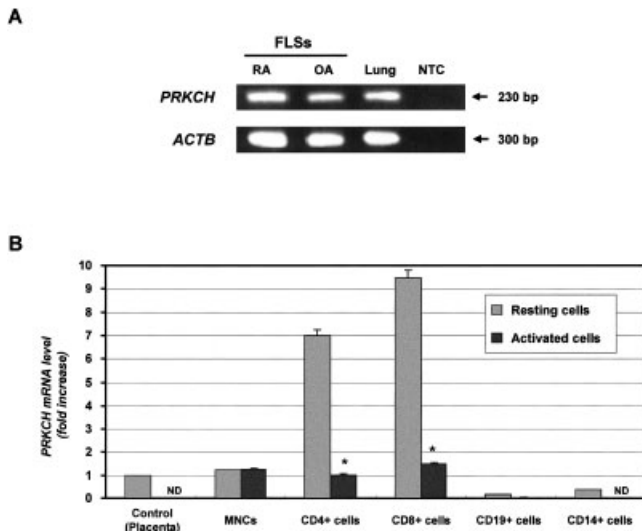
**Table 3.** SNP genotype combination effects on association\*

	Frequency		$\chi^2$	<i>P</i>	OR (95% CI)
	RA Patients	Controls			
<b>Single SNP</b>					
c.427+8134C>T					
0/0	0.025	0.048	–	–	1.00 (referent)
0/1	0.289	0.348	2.86	0.09	1.56 (0.93–2.60)
1/1	0.685	0.604	8.91	0.0028	2.13 (1.28–3.55)
rs912620					
0/0	0.282	0.336	–	–	1.00 (referent)
0/1	0.483	0.504	1.58	0.21	1.14 (0.93–1.41)
1/1	0.235	0.160	17.6	0.000027	1.75 (1.35–2.28)
rs959728					
0/0	0.502	0.584	–	–	1.00 (referent)
0/1	0.417	0.363	9.05	0.0026	1.34 (1.11–1.62)
1/1	0.081	0.053	9.35	0.0022	1.79 (1.23–2.61)
<b>2-SNP</b>					
c.427+8134C>T and rs912620					
0/0 and 0/0	0.025	0.043	–	–	1.00 (referent)
0/1 and 0/0	0.131	0.160	1.22	0.27	1.37 (0.78–2.39)
0/1 and 0/1	0.152	0.185	1.31	0.62	1.38 (0.79–2.40)
1/1 and 0/0	0.126	0.133	2.59	0.11	1.59 (0.90–2.79)
1/1 and 0/1	0.331	0.315	4.53	0.033	1.77 (1.04–3.01)
1/1 and 1/1	0.228	0.156	10.7	0.0011	2.45 (1.42–4.24)
c.427+8134C>T and rs959728					
0/0 and 0/0	0.019	0.035	–	–	1.00 (referent)
0/1 and 0/0	0.180	0.225	1.56	0.21	1.47 (0.80–2.71)
0/1 and 0/1	0.099	0.107	2.72	0.099	1.7 (0.90–3.23)
0/1 and 1/1	0.011	0.016	0.16	0.69	1.22 (0.46–3.27)
1/1 and 0/0	0.303	0.325	3.24	0.072	1.72 (0.95–3.12)
1/1 and 0/1	0.313	0.242	8.41	0.0037	2.38 (1.31–4.33)
1/1 and 1/1	0.069	0.037	12.4	0.00043	3.46 (1.71–7.00)
rs912620 and rs959728					
0/0 and 0/0	0.207	0.245	–	–	1.00 (referent)
0/0 and 0/1	0.065	0.090	0.61	0.43	0.86 (0.59–1.26)
0/1 and 0/0	0.212	0.271	0.30	0.59	0.93 (0.71–1.21)
0/1 and 0/1	0.243	0.199	7.23	0.0072	1.45 (1.11–1.90)
0/1 and 1/1	0.028	0.034	0.00	0.98	0.99 (0.58–1.72)
1/1 and 0/0	0.083	0.068	3.75	0.053	1.46 (0.99–2.13)
1/1 and 0/1	0.109	0.074	9.16	0.0025	1.73 (1.21–2.48)
1/1 and 1/1	0.043	0.018	12.6	0.00039	2.84 (1.57–5.16)
<b>3-SNP</b>					
c.427+8134C>T and rs912620 and rs959728					
0/0 and 0/0 and 0/0	0.019	0.034	–	–	1.00 (referent)
0/1 and 0/0 and 0/0	0.105	0.124	1.74	0.19	1.53 (0.81–2.90)
0/1 and 0/0 and 0/1	0.022	0.037	0.03	0.87	1.07 (0.48–2.35)
0/1 and 0/1 and 0/0	0.075	0.101	0.73	0.39	1.33 (0.69–2.56)
0/1 and 0/1 and 0/1	0.073	0.068	3.67	0.055	1.92 (0.98–3.75)
1/1 and 0/0 and 0/0	0.082	0.087	2.50	0.11	1.69 (0.88–3.26)
1/1 and 0/0 and 0/1	0.038	0.045	1.29	0.26	1.52 (0.74–3.16)
1/1 and 0/1 and 0/0	0.137	0.169	1.40	0.24	1.45 (0.78–2.71)
1/1 and 0/1 and 0/1	0.170	0.126	8.05	0.0045	2.43 (1.30–4.53)
1/1 and 0/1 and 1/1	0.024	0.019	3.68	0.055	2.27 (0.98–5.29)
1/1 and 1/1 and 0/0	0.083	0.067	5.70	0.017	2.23 (1.15–4.34)
1/1 and 1/1 and 0/1	0.104	0.072	8.39	0.0038	2.59 (1.35–4.98)
1/1 and 1/1 and 1/1	0.040	0.017	12.3	0.00044	4.22 (1.86–9.60)

\* For all analyses, n = 948–950 rheumatoid arthritis (RA) patients and 939–948 controls. SNP = single-nucleotide polymorphism; OR = odds ratio; 95% CI = 95% confidence interval.

bility alleles at c.427+8134C>T and rs912620. The highest OR was observed in individuals homozygous for all susceptibility alleles (1/1-1/1-1/1; OR 4.22, 95% CI

1.86–9.60,  $P = 4.4 \times 10^{-4}$ ). Essentially the same results were obtained with several different 3-SNP sets chosen from blocks A, B, and D (data not shown).



**Figure 3.** *PRKCH* mRNA expression in fibroblast-like synoviocytes (FLS) and peripheral blood fractions. **A**, *PRKCH* and *ACTB* gene expressions were analyzed by conventional reverse transcriptase–polymerase chain reaction (RT-PCR), using FLS RNA isolated from representative patients with rheumatoid arthritis (RA) or osteoarthritis (OA). Lung total RNA was used as a positive control for *PRKCH* gene expression. NTC = no template control (without cDNA). **B**, Levels of *PRKCH* mRNA in peripheral blood fractions. *PRKCH* mRNA levels in mononuclear cells (MNCs) (B cells, T cells, and monocytes), CD4+ cells (T helper/inducer cells), CD8+ cells (T suppressor/cytotoxic cells), CD19+ cells (B cells), and CD14+ cells (monocytes) were determined by real-time quantitative RT-PCR and normalized to *ACTB* levels. Activation was induced by incubation with 2  $\mu$ l/ml pokeweed mitogen (PWM) and 5  $\mu$ g/ml concanavalin A (Con A) for 3 days (MNCs), 5  $\mu$ g/ml Con A for 3–4 days (CD4+ cells), 5  $\mu$ g/ml phytohemagglutinin for 3 days (CD8+ cells), or 2  $\mu$ l/ml PWM for 4 days (CD19+ cells). Data are presented as the mean and SEM ( $n = 3$ ) fold change compared with the *PRKCH* expression in placenta (set at 1.0). \* =  $P < 0.001$  versus resting cells, by Student's *t*-test. ND = experiments not done.

**Expression of *PRKCH* in synovial tissue fibroblasts and peripheral lymphocytes.** As a first step in evaluating the role of PKC $\eta$  in inflammatory arthritis, we investigated the expression of *PRKCH* messenger RNA (mRNA) in FLS from patients with RA or OA. As shown in Figure 3A, *PRKCH* expression was demonstrated in FLS from both RA and OA patients, at levels similar to that in the lung. We next investigated *PRKCH* expression in peripheral blood fractions (Figure 3B). *PRKCH* was expressed in resting CD4+ cells (T helper/inducer cells) or CD8+ cells (T suppressor/cytotoxic cells) at higher levels than in resting CD19+ cells (B cells) or CD14+ cells (monocytes). The *PRKCH* levels in these cells were significantly down-regulated by activation (–85% in CD4+ cells and –84% in CD8+ cells;

$P < 0.001$ ). Levels of *PRKCH* mRNA in the mononuclear cell fraction, however, were not significantly affected. These findings indicated that expression of the *PRKCH* gene in certain cell types is regulated through immune responses.

## DISCUSSION

The current study was designed to determine whether gene(s) located on chromosome 14q are associated with susceptibility to RA in the Japanese population. Our SNP association scan covers an ~20-Mb genomic region on chromosome 14q21–23. We focused on the gene-coding region of this area, and chose SNPs yielding priorities in their allele frequencies, rather than relying on linkage disequilibrium information and resources that recently became publicly available (e.g., the International HapMap Project), since a high-density SNP map of the region was not available when we started this study. In addition, to reduce the time and cost of genotyping, we adopted a 2-stage strategy, in which we proceeded to the validation stage of genotyping only if the results from the exploratory stage offered some possibility of overall significance. To test associations, we used the so-called “joint analysis” strategy (32), in which we analyzed the pooled data from both exploratory and validation stages, using a  $P$  value of 0.05 with Bonferroni correction for 378 comparisons (the total number of SNP loci used for analysis). A recent report indicated that joint analysis is more efficient and powerful than the standard 2-stage strategy, which considers the second cohort to be a completely independent replication panel and tests statistics for the second stage data alone (32).

Our association study strategy was successful in identifying a significant association for 1 landmark SNP (rs767755), located in intron 2 of the *PRKCH* gene encoding PKC $\eta$ . Subsequent analysis of additional SNPs within this gene revealed multiple SNPs scattered in the 3 distinct linkage disequilibrium blocks (blocks A, B, and D) to be significantly associated with RA. Associations were observed in a linkage disequilibrium block-dependent manner. Moreover, the results of genotype combination analysis using SNPs residing in different linkage disequilibrium blocks suggest that 2- or 3-way combinations of “at risk” genotypes increase the risk of RA additively or synergistically. Significant SNPs included a novel missense SNP substituting V for I at residue 374 of the PKC $\eta$  (V374I). Of note, the V374 residue is located within the conserved ATP-binding site with the consensus sequence GXGXXGX<sub>16</sub>K, and mu-

tations in this motif reportedly result in abolished functional activities for other PKC isozymes (38,39). An investigation is now under way in our laboratory to determine whether this variant has functional consequences.

Other significant SNPs were noncoding SNPs, located in intron 2, 5, or 9; these observations are not surprising given the recent reports of similar findings in RA (40) and also in other common diseases with genetic associations, such as type 2 DM (41) and asthma (42). Although the exact causal connection between these intronic SNPs and RA risk is not known, they may alter PKC $\eta$  expression or function by affecting splicing. In this regard, we searched for potential DNA regulatory sites and protein-binding sites in SNPs but identified no such motifs (TRANSFAC database [43]; data not shown).

One limitation of this genetic association study is that our controls were not matched for age. With respect to age at the time of the interview and subsequent blood sampling, the control subjects were significantly younger than the patients (mean  $\pm$  SD age 39.1  $\pm$  15.2 years versus 61.7  $\pm$  12.4 years, respectively;  $P < 0.001$ ). Owing to the late onset of RA, it is possible that at least some of the controls who were included in this study will develop RA in the future. This potential misclassification in the control group due to the inclusion of undiagnosed cases, as well as inappropriate case definition, would bias and weaken any association between *PRKCH* and RA.

There are, however, several arguments against such bias. First, the mean  $\pm$  SD age at RA diagnosis in the patients was 48.7  $\pm$  13.4 years, rather similar to the mean age of controls, although the difference was still statistically significant ( $P < 0.0001$ ). Second, our controls were interviewed extensively about their health and reported no family history of any autoimmune diseases. Therefore, they would have only a low risk of developing RA, and the overall population incidence is estimated to be <0.5–1%. Thus, the sample size of our association study should still provide adequate power to detect an association. Finally, there was no significant difference in *PRKCH* genotype distribution between our controls and a group of Japanese individuals from the same geographic area who had type 2 DM ( $n = 711$ ), from whom samples were collected for a different genetic association study (ref. 30 and Takata Y: unpublished observations). The mean  $\pm$  SD age of the type 2 DM group was 62.6  $\pm$  11.4 years, which was not significantly different from that of RA patients in this study. We thus believe our comparison to be acceptable. However, due to the cross-sectional, population-based nature of this

study, our findings need to be confirmed in other independent populations and/or by family-based tests of association. In the first instance, the use of populations collected for large prospective cohort studies would be one obvious approach.

The *PRKCH* gene is an excellent functional candidate for susceptibility to RA, since the disease is believed to be the result of misdirected immune responses by autoreactive T cells, and PKC plays a critical role in signal transduction controlling T cell activation. The PKC gene family consists of  $\geq 11$  members, including *PRKCH*/PKC $\eta$ , and their gene products (isozymes) are usually categorized into the following 3 groups based on structure and cofactor/activator requirements: conventional (PKC $\alpha$ , PKC $\beta 1$ , PKC $\beta 2$ , and PKC $\gamma$ ), novel (PKC $\delta$ , PKC $\epsilon$ , PKC $\eta$ , PKC $\theta$ , and PKC $\mu$ ), and atypical (PKC $\zeta$  and PKC $\lambda/\iota$ ). Conventional PKCs are dependent on calcium and diacylglycerol for their functional activity, whereas novel PKCs are calcium independent, and atypical PKCs are calcium and diacylglycerol independent.

Individual PKC isozymes have been shown to exhibit biologically distinct and even opposing cellular functions in different cell types (44), and recent data suggest that at least some isozymes of PKC are involved in critical function(s) of T cells. For example, PKC $\theta$ , which is highly expressed in T cells, has been extensively documented to play essential roles in T cell activation, signaling, and gene transcription regulation (45). In addition, PKC $\beta$  has been suggested to be a key element in proper T cell migration (46), while PKC $\zeta$  has been described as maintaining the integrity of the actin cytoskeleton and mediating interleukin-induced T cell proliferation (47). It is also noteworthy that gold sodium thiomalate, which has been used as a therapeutic agent for RA for many years, is now recognized as a PKC inhibitor, suppressing mitogen-induced T cell proliferation (48).

Unfortunately, the physiologic function of PKC $\eta$  in T cells has not yet been documented, and the pathophysiologic mechanism (or pathway) by which *PRKCH* gene polymorphisms may influence RA risk remains unknown. However, this study is the first to show that the *PRKCH* gene is expressed at high levels in helper/inducer (CD4+) or suppressor/cytotoxic (CD8+) T cells, in a manner analogous to that of PKC $\theta$ , and that its expression is down-regulated through immune responses. These observations suggest that PKC $\eta$  functions, at least in part, in signaling pathways unique to T cells, and, therefore, that dysregulation of its gene expression and/or function might be associated with an increased risk of RA.

Another possibility is that PKC $\eta$  may function as a regulator in both the pathogenic pathway and certain aspects of the cytokine signaling cascade in monocytes and macrophages, although, in our experiments, the *PRKCH* gene expression level in monocytes (CD14+ cells) was low compared with that in T cells. Interestingly, plasma levels of nitric oxide, a mediator of inflammation, were shown to be elevated in patients with severe RA, and a positive correlation between PKC $\eta$  and inducible nitric oxide synthase expression was observed in peripheral blood monocyte-derived macrophages from RA patients (49). Although these findings provide evidence supporting the possible involvement of PKC $\eta$  in immunologic activities of monocytes and macrophages, further investigation is obviously required before firm conclusions can be drawn.

In conclusion, we have identified the *PRKCH* gene encoding PKC $\eta$  as a putative candidate gene conferring genetic susceptibility to RA in a Japanese population. Although PKC $\eta$  certainly has biologic plausibility as an RA gene, replication studies in other independent populations and/or by family-based tests of association are essential for determining whether our observations are consistent. Further investigations into the molecular mechanisms by which PKC $\eta$  alters RA susceptibility are also required. These studies may ultimately lead to the development of novel therapies that modulate the PKC $\eta$  pathway in patients with RA and other autoimmune disorders.

#### ACKNOWLEDGMENTS

We are grateful to the many patients and volunteers who participated in this study. We greatly appreciate the excellent technical assistance of Dr. Rumi Katashima, Ms Kazue Tsugawa, and Ms Yukiko Sakamoto (The University of Tokushima). We wish to thank Dr. Shunichi Shiozawa (Kobe University) for his efforts in the revision of this manuscript, and the many past and present members of our research group for helpful discussions and assistance.

#### AUTHOR CONTRIBUTIONS

Dr. Itakura had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Study design.** Drs. Takata, Hamada, and Itakura.

**Acquisition of data.** Drs. Takata, Hamada, Miyatake, Nakano, and Shinomiya, Mr. Scafe, Mr. Reeve, and Mr. Osabe, and Drs. Moritani, Kunika, and Yasui.

**Analysis and interpretation of data.** Drs. Takata and Hamada.

**Manuscript preparation.** Drs. Takata, Inoue, and Itakura.

**Statistical analysis.** Drs. Takata, Kamatani, Inoue, and Itakura.

#### REFERENCES

1. Deighton CM, Walker DJ, Griffiths ID, Roberts DF. The contribution of HLA to rheumatoid arthritis. *Clin Genet* 1989;36:178–82.
2. Seldin MF, Amos CI, Ward R, Gregersen PK. The genetics revolution and the assault on rheumatoid arthritis [review]. *Arthritis Rheum* 1999;42:1071–9.
3. MacGregor AJ, Snieder H, Rigby AS, Koskenvuo M, Kaprio J, Aho K, et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. *Arthritis Rheum* 2000;43:30–7.
4. Wordsworth BP, Lanchbury JS, Sakkas LI, Welsh KI, Panayi GS, Bell JI. HLA-DR4 subtype frequencies in rheumatoid arthritis indicate that DRB1 is the major susceptibility locus within the HLA class II region. *Proc Natl Acad Sci U S A* 1989;86:10049–53.
5. Gregersen PK, Silver J, Winchester RJ. The shared epitope hypothesis: an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. *Arthritis Rheum* 1987;30:1205–13.
6. MacGregor A, Ollier W, Thomson W, Jawaheer D, Silman A. HLA-DRB1\*0401/0404 genotype and rheumatoid arthritis: increased association in men, young age at onset, and disease severity. *J Rheumatol* 1995;22:1032–6.
7. Cornelis F, Faure S, Martinez M, Prud'homme JF, Fritz P, Dib C, et al. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc Natl Acad Sci U S A* 1998;95:10746–50.
8. Osorio y Fortea J, Bukulmez H, Petit-Teixeira E, Michou L, Pierlot C, Cailleau-Moindrault S, et al. Dense genome-wide linkage analysis of rheumatoid arthritis, including covariates. *Arthritis Rheum* 2004;50:2757–65.
9. Shiozawa S, Hayashi S, Tsukamoto Y, Goko H, Kawasaki H, Wada T, et al. Identification of the gene loci that predispose to rheumatoid arthritis. *Int Immunol* 1998;10:1891–5.
10. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Monteiro J, et al. A genomewide screen in multiplex rheumatoid arthritis families suggests genetic overlap with other autoimmune diseases. *Am J Hum Genet* 2001;68:927–36.
11. Jawaheer D, Seldin MF, Amos CI, Chen WV, Shigeta R, Etzel C, 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.
12. MacKay K, Eyre S, Myerscough A, Milicic A, Barton A, Laval S, et al. Whole-genome linkage analysis of rheumatoid arthritis susceptibility loci in 252 affected sibling pairs in the United Kingdom. *Arthritis Rheum* 2002;46:632–9.
13. Eyre S, Barton A, Shephard N, Hinks A, Brintnell W, MacKay K, et al. Investigation of susceptibility loci identified in the UK rheumatoid arthritis whole-genome scan in a further series of 217 UK affected sibling pairs. *Arthritis Rheum* 2004;50:729–35.
14. Fisher SA, Lanchbury JS, Lewis CM. Meta-analysis of four rheumatoid arthritis genome-wide linkage studies: confirmation of a susceptibility locus on chromosome 16. *Arthritis Rheum* 2003;48:1200–6.
15. Choi SJ, Rho YH, Ji JD, Song GG, Lee YH. Genome scan meta-analysis of rheumatoid arthritis. *Rheumatology (Oxford)* 2006;45:166–70.
16. Etzel CJ, Chen WV, Shepard N, Jawaheer D, Cornelis F, Seldin MF, et al. Genome-wide meta-analysis for rheumatoid arthritis. *Hum Genet* 2006;119:634–41.
17. Loesgen S, Dempfle A, Golla A, Bickeboller H. Weighting schemes in pooled linkage analysis. *Genet Epidemiol* 2001;21 Suppl 1:S142–7.
18. Tamiya G, Shinya M, Imanishi T, Ikuta T, Makino S, Okamoto K, et al. Whole genome association study of rheumatoid arthritis using 27,039 microsatellites. *Hum Mol Genet* 2005;14:2305–21.



19. Gaffney PM, Kearns GM, Shark KB, Ortmann WA, Selby SA, Malmgren ML, et al. A genome-wide search for susceptibility genes in human systemic lupus erythematosus sib-pair families. *Proc Natl Acad Sci U S A* 1998;95:14875-9.
20. Shai R, Quismorio FP, Li L, Kwon OJ, Morrison J, Wallace DJ, et al. Genome-wide screen for systemic lupus erythematosus susceptibility genes in multiplex families. *Hum Mol Genet* 1999;8:639-44.
21. Lindqvist AK, Steinsson K, Johanneson B, Kristjansdottir H, Arnasson A, Grondal G, et al. A susceptibility locus for human systemic lupus erythematosus (hSLE1) on chromosome 2q. *J Autoimmun* 2000;14:169-78.
22. Koskenmies S, Lahermo P, Julkunen H, Ollikainen V, Kere J, Widen E. Linkage mapping of systemic lupus erythematosus (SLE) in Finnish families multiply affected by SLE. *J Med Genet* 2004;41:e2-5.
23. Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs. *Nat Med* 2001;7:899-905.
24. Panush RS, Edwards NL, Longley S, Webster E. 'Rheupus' syndrome. *Arch Intern Med* 1988;148:1633-6.
25. Taneja V, Singh RR, Malaviya AN, Anand C, Mehra NK. Occurrence of autoimmune diseases and relationship of autoantibody expression with HLA phenotypes in multicase rheumatoid arthritis families. *Scand J Rheumatol* 1993;22:152-7.
26. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al, on behalf of the Grupo Latinoamericano de Estudio del Lupus Eritematoso (GLADEL). Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. *Arthritis Rheum* 2005;52:1138-47.
27. Hamada D, Takata Y, Osabe D, Nomura K, Shinohara S, Egawa H, et al. Association between single-nucleotide polymorphisms in the SEC8L1 gene, which encodes a subunit of the exocyst complex, and rheumatoid arthritis in a Japanese population. *Arthritis Rheum* 2005;52:1371-80.
28. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.
29. De La Vega FM, Isaac H, Collins A, Scafe CR, Halldorsson BV, Su X, et al. The linkage disequilibrium maps of three human chromosomes across four populations reflect their demographic history and a common underlying recombination pattern. *Genome Res* 2005;15:454-62.
30. Kato H, Nomura K, Osabe D, Shinohara S, Mizumori O, Katashima R, et al. Association of single-nucleotide polymorphisms in the suppressor of cytokine signaling 2 (SOCS2) gene with type 2 diabetes in the Japanese. *Genomics* 2006;87:446-58.
31. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. *Science* 2002;296:2225-9.
32. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* 2006;38:209-13.
33. Ardlie KG, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nat Rev Genet* 2002;3:299-309.
34. Abecasis GR, Cookson WO. GOLD: graphical overview of linkage disequilibrium. *Bioinformatics* 2000;16:182-3.
35. Maniatis N, Collins A, Xu CF, McCarthy LC, Hewett DR, Tapper W, et al. The first linkage disequilibrium (LD) maps: delineation of hot and cold blocks by diplotype analysis. *Proc Natl Acad Sci U S A* 2002;99:2228-33.
36. Stephens M, Donnelly P. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet* 2003;73:1162-9.
37. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, et al. Haplotype variation and linkage disequilibrium in 313 human genes. *Science* 2001;293:489-93.
38. Ohno S, Konno Y, Akita Y, Yano A, Suzuki K. A point mutation at the putative ATP-binding site of protein kinase C  $\alpha$  abolishes the kinase activity and renders it down-regulation-insensitive: a molecular link between autophosphorylation and down-regulation. *J Biol Chem* 1990;265:6296-300.
39. Li W, Yu JC, Shin DY, Pierce JH. Characterization of a protein kinase C- $\delta$  (PKC- $\delta$ ) ATP binding mutant: an inactive enzyme that competitively inhibits wild type PKC- $\delta$  enzymatic activity. *J Biol Chem* 1995;270:8311-8.
40. Suzuki A, Yamada R, Chang X, Tokuhira S, Sawada T, Suzuki M, et al. Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 2003;34:395-402.
41. Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, Hara M, et al. Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 2000;26:163-75.
42. Van Eerdewegh P, Little RD, Dupuis J, Del Mastro RG, Falls K, Simon J, et al. Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 2002;418:426-30.
43. Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, et al. TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res* 2000;28:316-9.
44. Dempsey EC, Newton AC, Mochly-Rosen D, Fields AP, Reyland ME, Insel PA, et al. Protein kinase C isozymes and the regulation of diverse cell responses. *Am J Physiol Lung Cell Mol Physiol* 2000;279:L429-38.
45. Baier G. The PKC gene module: molecular biosystematics to resolve its T cell functions. *Immunol Rev* 2003;192:64-79.
46. Volkov Y, Long A, McGrath S, Ni Eidhin D, Kelleher D. Crucial importance of PKC- $\beta$ (I) in LFA-1-mediated locomotion of activated T cells. *Nat Immunol* 2001;2:508-14.
47. Gomez J, Martinez de Aragon A, Bonay P, Pitton C, Garcia A, Silva A, et al. Physical association and functional relationship between protein kinase C  $\zeta$  and the actin cytoskeleton. *Eur J Immunol* 1995;25:2673-8.
48. Hashimoto K, Whitehurst CE, Matsubara T, Hirohata K, Lipsky PE. Immunomodulatory effects of therapeutic gold compounds: gold sodium thiomalate inhibits the activity of T cell protein kinase C. *J Clin Invest* 1992;89:1839-48.
49. Pham TN, Rahman P, Tobin YM, Khraishi MM, Hamilton SF, Alderdice C, et al. Elevated serum nitric oxide levels in patients with inflammatory arthritis associated with co-expression of inducible nitric oxide synthase and protein kinase C- $\eta$  in peripheral blood monocyte-derived macrophages. *J Rheumatol* 2003;30:2529-34.