ARTHRITIS & RHEUMATISM
Vol. 58, No. 8, August 2008, pp 2275–2286
DOI 10.1002/art.23623


# Genome-Wide Association Study of Rheumatoid Arthritis in the Spanish Population

## *KLF12* as a Risk Locus for Rheumatoid Arthritis Susceptibility

Antonio Julià,[1] Javier Ballina,[2] Juan D. Cañete,[3] Alejandro Balsa,[4] Jesus Tornero-Molina,[5] Antonio Naranjo,[6] Mercedes Alperi-López,[2] Alba Erra,[2] Dora Pascual-Salcedo,[4] Pere Barceló,[7] Jordi Camps,[8] and Sara Marsal[1]

***Objective.*** **To identify new genes associated with susceptibility to rheumatoid arthritis (RA), using a 2-stage genome-wide association study.**

***Methods.*** **Following a liability-based study design, we analyzed 317,503 single-nucleotide polymorphisms (SNPs) in 400 patients with RA and 400 control subjects. We selected a group of candidate SNPs for replication in an independent group of 410 patients with RA and 394 control subjects. Using data from the 3 previous genome-wide association studies in RA, we also looked for genomic regions showing evidence of common association signals. Finally, we analyzed the presence of genome-wide epistasis using the binary test implemented in the PLINK program.**

***Results.*** **We identified several genomic regions showing evidence of genome-wide association ($P < 1 \times 10^{-5}$). In the replication analysis, we identified *KLF12* SNP rs1324913 as the most strongly associated SNP ($P = 0.01$). In our study, we observed that this SNP showed higher significance than *PTPN22* SNP rs2476601, in both the genome-wide association studies and the replication analyses. Furthermore, the integration of our data with those from previous genome-wide association studies showed that *KLF12* and *PTPRT* are the unique loci that are commonly associated in 3 different studies ($P = 0.004$ and $P = 0.002$ for *KLF12* in the Wellcome Trust Case Control Consortium study and the Brigham and Women's Rheumatoid Arthritis Sequential Study genome-wide association study, respectively). The genome-wide epistasis analysis identified several SNP pairs close to significance after multiple test correction.**

***Conclusion.*** **The present genome-wide association study identified *KLF12* as a new susceptibility gene for RA. The joint analysis of our results and those from previous genome-wide association studies showed genomic regions with a higher probability of being genuine susceptibility loci for RA.**

Rheumatoid arthritis (RA) is one of the most prevalent autoimmune diseases in the world (1). In RA, chronic inflammation of the synovial joints leads to progressive articular damage, which can result in major functional disability (2). The etiology of RA is unknown, but several family aggregation and twin studies (3,4) clearly demonstrate a heritable component of the disease. Part of this genetic component of susceptibility has been consistently associated with the HLA class II locus variation. The remaining 50–75% of the genetic component includes several other genomic regions that are more difficult to identify due to their lower penetrance or more complex models of action (5,6).

Linkage scans and candidate gene studies have

Supported by the Spanish Ministry of Education and Science (Proyectos Singulares y Estratégicos, PSE-010000-2006-6) and by Schering-Plough SA.

[1]Antonio Julià, Alba Erra, MD, Sara Marsal, MD, PhD: Institut de Recerca, Hospital Universitari Vall d'Hebron, Barcelona, Spain; [2]Javier Ballina, MD, PhD, Mercedes Alperi-López, MD: Hospital Universitario Central de Asturias, Oviedo, Asturias, Spain; [3]Juan D. Cañete, MD, PhD: Hospital Clinic i Provincial de Barcelona, Barcelona, Spain; [4]Alejandro Balsa, MD, PhD, Dora Pascual-Salcedo, PhD: Hospital Universitario La Paz, Madrid, Spain; [5]Jesus Tornero-Molina, MD, PhD: Hospital Universitario de Guadalajara, Castilla-La Mancha, Spain; [6]Antonio Naranjo, MD, PhD: Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain; [7]Pere Barceló, MD: Hospital Universitari Vall d'Hebron, Barcelona, Spain; [8]Jordi Camps: Barcelona Supercomputing Centre, Barcelona, Spain.

Dr. Tornero-Molina has received consulting fees, speaking fees, and/or honoraria from Wyeth and Bristol-Myers Squibb (more than $10,000 each).

Address correspondence and reprint requests to Sara Marsal, MD, PhD, Unitat de Recerca de Reumatologia, Hospital Universitari Vall d'Hebron, Passeig Vall d'Hebron 119-129, 08035 Barcelona, Spain. E-mail: smarsal@ir.vhebron.net.

Submitted for publication December 5, 2007; accepted in revised form April 14, 2008.

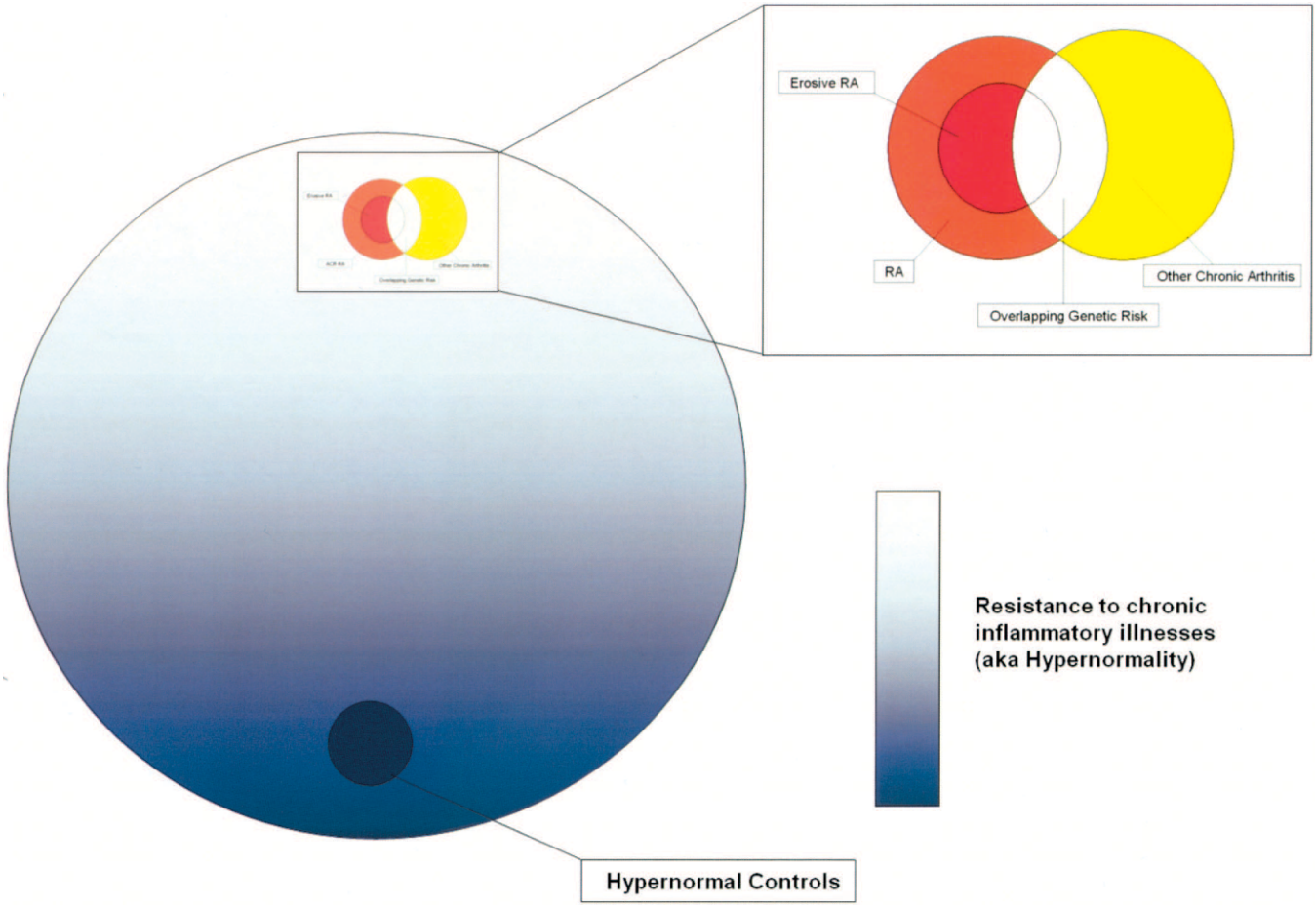



**Figure 1.** Liability-based model used in the present genome-wide association study. In this model, a continuous latent risk variable is truncated to specify affection, which in this case is a chronic inflammatory disease. Those individuals with the highest liability have a specific group of genes that condition to a particular outcome (i.e., disease diagnostic). Those individuals with the lowest liability are resistant to inflammatory illnesses; hence, they are "hypernormal" for these conditions. RA = rheumatoid arthritis.

successfully identified a small number of candidate genes for RA susceptibility, ranging from the robust association of *PTPN22* (7) to the more modest association of *PADI4* (8) and *CTLA4* (9). Together with these important steps in the characterization of RA genetic architecture, the candidate strategy has also produced a large number of genes that have failed to show convincing association (10). Linkage scans, an extremely powerful methodology for identifying genes with simple genetic models of inheritance, have several limitations for common diseases such as RA (11,12). Recently, genome-wide association studies have enabled the combination of 2 fundamental advantages of the previous approaches: the unbiased analysis of whole genome linkage scans and the power and resolution of case–control studies (13).

To date, 3 genome-wide association studies in RA have been performed, providing important advances in the characterization of genetic susceptibility in RA. The Wellcome Trust Case Control Consortium (WTCCC) performed an unprecedented genome-wide analysis of 7 common diseases in the UK population (14). This approach enabled not only the identification of strong candidate regions for each disease but also the identification of common susceptibility regions between different diseases. More recently, 2 genome-wide studies using North American and Swedish cohorts identified and replicated *TRAF-C5* (15) and *TNFAIP3* (16) as new genetic loci strongly associated with positive anti–cyclic citrullinated peptide antibodies in RA subtype susceptibility. These important findings demonstrate the effectiveness of the genome-wide association study approach and represent important steps toward the identification of RA genetic architecture.

Here, we report the results of a 2-stage genome-

wide association study performed in the Spanish population. In contrast to the 3 previous genome-wide studies, we used a design based on disease liability to both RA and chronic inflammatory diseases. We also performed a replication analysis of a selected group of new candidate single-nucleotide polymorphisms (SNPs) in an independent sample. In order to look for common associated genomic regions, we contrasted our results with those of the 3 previous genome-wide association studies. Finally, we also analyzed more complex genetic models through a genome-wide analysis of gene–gene interactions (i.e., epistasis) associated with RA susceptibility.

## PATIENTS AND METHODS

**Study design.** We performed a 2-stage genome-wide association study in RA. In the first stage, 400 patients with RA and 400 control subjects were analyzed for 317,503 genomic SNPs. From these results, a selection of new candidate SNPs was further genotyped in an independent group of 410 patients with RA and 394 control subjects. In the genome-wide analysis, both case and control groups were formed by 2 subgroups (n = 200 each) based on the liability model shown in Figure 1. This model assumes that there is a continuous latent risk of chronic inflammatory diseases. Those individuals with the lowest risk of developing any type of chronic inflammatory diseases are defined as "hypernormal" (17). In the high-risk zone, the continuous variable is truncated to specify a chronic inflammatory disease. Therefore, this model integrates the increasing evidence of shared genetic risk for common inflammatory diseases (14,18,19) and the specific genetic variants that determine each particular condition.

Informed consent was obtained from all individuals, according to the Declaration of Helsinki. The study was approved by the Institut de Recerca de l'Hospital Universitari Vall d'Hebron ethics committee.

**Whole-genome association study subjects.** Patients with RA were recruited from 5 Spanish hospitals: Hospital Universitario Central de Asturias, Hospital Universitario de Guadalajara, Hospital Clínic i Provicial de Barcelona, Hospital Universitario de La Paz (Madrid), and Hospital Universitari Vall d'Hebron (Barcelona). All patients fulfilled the revised American College of Rheumatology (ACR; formerly, the American Rheumatism Association) 1987 criteria for the classification of RA (20). Two hundred patients were selected for having a longstanding disease with severe radiologic and functional disability (longstanding RA). The remaining cases were selected from among a group of patients with RA who were attending early arthritis clinics and had been followed up for a minimum of 2 years (early RA).

Control subjects were selected according to the liability model described previously. In order to capture the genetic component that is specific for RA and different from other chronic arthritides, we selected a group of 200 patients with non-RA inflammatory arthritis (non-RA). This group of patients and those with early RA were selected from the same early arthritis clinics, and the non-RA group comprised spondylarthritis (34%), undetermined arthritis (26%), psoriatic arthritis (20%), connective tissue disorders (15%), and other less common inflammatory arthropathies (5%). To increase the efficiency of our study, we selected a group of 200 individuals with the lowest liability for RA or any other chronic inflammatory disease, whom we here describe as hypernormal control subjects. Using the randomized control collection of IRCIS BioBank (Hospital San Joan de Reus, Tarragona, Spain), we selected only those individuals whose age placed them at risk of RA (>40 years old), were Caucasian, and had a 3-generation Spanish origin. We reduced the genetic liability in this group by excluding those individuals with ≥1 first-degree relative with a chronic inflammatory disease (including autoimmune diseases). All 4 subgroups had the female-to-male sex distribution (3:1 ratio) that is characteristic of RA (21).

**Replication study subjects.** We collected an independent group of 410 patients with RA (347 women and 63 men) from the same 5 hospitals. All patients fulfilled the ACR 1987 revised criteria for the classification of RA and were Caucasian and of Spanish origin. A control group of similar size (n = 394 [284 women and 110 men]) was obtained from the Spanish National DNA Bank repository (Banco Nacional de ADN, Salamanca, Spain). All control subjects were Caucasian and of Spanish origin, were older than age 30 years, and did not have an autoimmune disease.

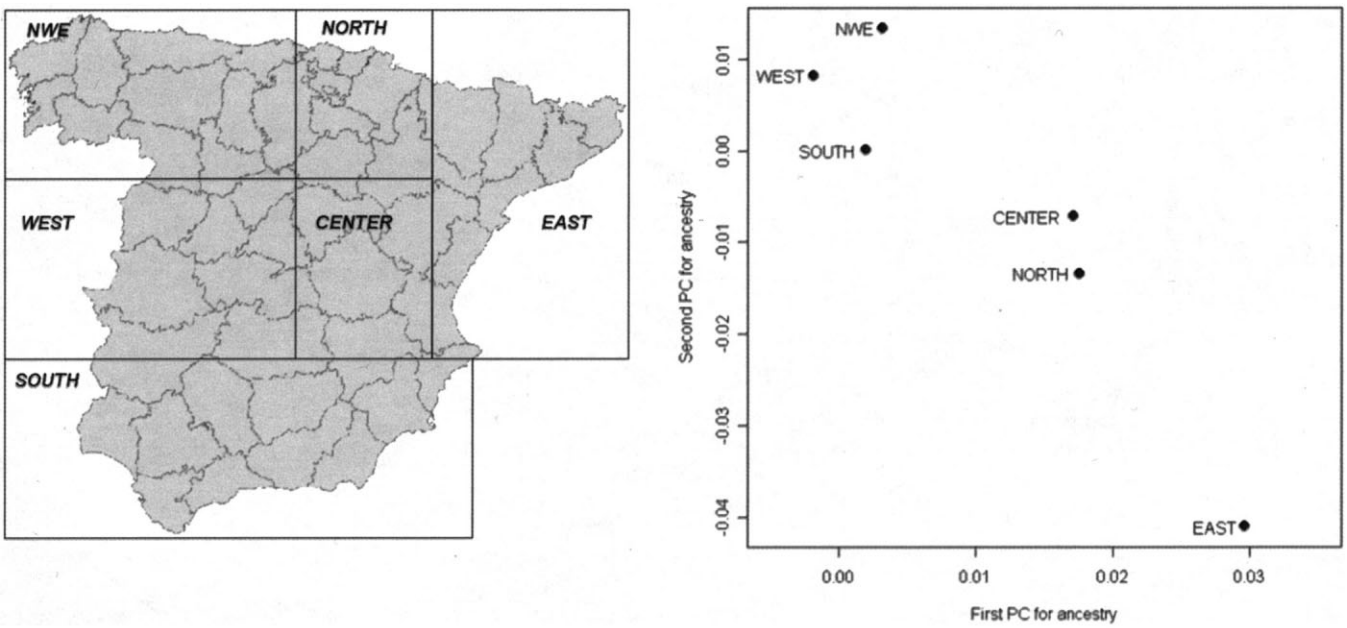
**Sample preparation for whole-genome and replication genotyping.** Although most samples analyzed were from local DNA collections, ~20% of them were extracted from whole blood using the Flexigene purification system (Qiagen, Chatsworth, CA).

More than 317,000 SNPs were genotyped in each of the 800 individuals in the genome-wide association study, using the HumanHap300 BeadArray system (Illumina, San Diego, CA). The selection of highly informative markers (tagSNPs) included in this system provides strong coverage of the whole genome (22). Samples were amplified, labeled, and hybridized according to the Illumina Infinium II assay. After scanning in an Illumina BeadArray reader, fluorescence intensities were automatically converted to genotypes using Illumina BeadStudio software version 2.0.

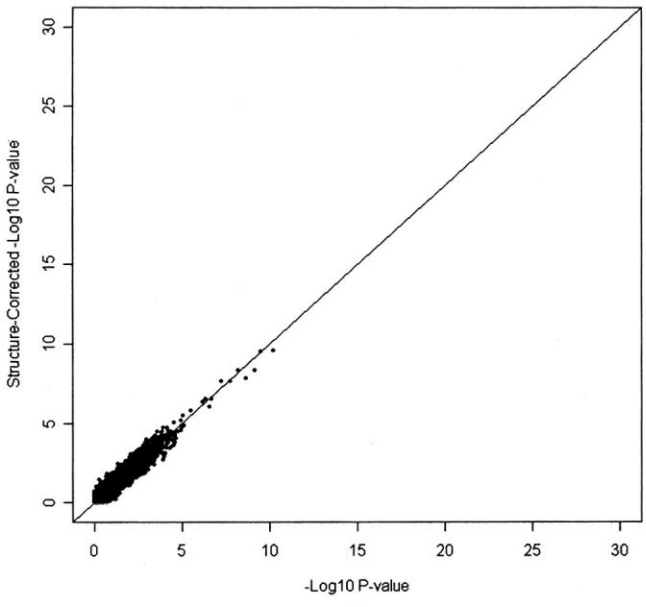
Replication genotyping and cross-platform quality control were performed using the MassARRAY SNP genotyping system (Sequenom, San Diego, CA) (23). Genotype calling was performed using the automatic call system implemented in Sequenom Typer software. All genotyping assays were performed at the Centro Nacional de Genotipado (Barcelona, Spain).

**Whole-genome association study quality control.** For the 800 individuals analyzed, the average genotyping rate was 98.8%. To this data we subsequently applied several quality control filters, as follows: 1) exclusion of those SNPs with more than 10% of values missing (2.2%), 2) exclusion of individuals with more than 10% of values missing (none), 3) uninformative SNPs (minor allele frequency < 0.01) (0.1%), and 4) SNPs under Hardy-Weinberg disequilibrium ($P < 0.0001$) (0.3%). Although 6,590 SNPs (2%) were in very high pairwise linkage disequilibrium ($r^2 > 0.99$), we did not exclude them from further analyses.

**Population structure analysis.** The presence of structure in a population can be an important confounder in genetic association studies. In order to detect strong variability components in our genome-wide association study samples, we performed the principal components analysis (PCA) implemented in Eigenstrat software (24). The PCA technique




A




B

**Figure 2. A,** Principal components (PCs) informative for ancestry. For 246 individuals in the genome-wide association study, complete information regarding the province of birth of all 4 grandparents was available. Based on this information, individuals were divided into 6 geographic regions in Spain. The graph shows the 2 principal components informative for ancestry, demonstrating a west-to-east trend. **B,** Scatter plot of *P* values before (x-axis) and after (y-axis) correction for structure. Correcting for geographic structure using the ancestry-informative PCs as covariates does not show a trend with signals above and below the diagonal line.

effectively decomposes the variability present in high-dimensional data sets into lower dimensions. The top axes of variation (i.e., the principal components) should reflect the geographic trends (if such trends exist) in our sample. As a more indirect measure of population structure, we also calculated the genomic inflation factor ($\lambda_{observed}$) (25), a measure of

the "overdispersion" of the association statistic (i.e., allelic chi-square). The closer this value is to the null value ($\lambda_{null}$ = 1), the lower the probability of the presence of population structure in the sample.

**Genome-wide association analysis, SNP selection, and replication analysis.** We performed chi-square allelic tests for the 299,918 SNPs that remained after quality control filtering using PLINK software (26). Following the liability model described previously, we performed 3 different analyses: a global comparison (all patients with RA versus all control subjects), an extreme liability comparison (patients with long-standing RA versus hypernormal control subjects), and a chronic arthritis liability comparison (all patients with RA and those with non-RA inflammatory arthritis versus hypernormal control subjects). Using Benjamini and Hochberg correction for multiple testing, only HLA class II SNPs and a single marker in chromosome 3 (rs11129989) were significant. Tables showing the complete results are available online at http://www.urr.cat.

Several criteria have been proposed for selecting SNPs for replication that do not withstand conservative multiple test correction methods. Some studies have used the significance rank to select a relative arbitrary number of SNPs (27), while others use biologic information to favor a group of candidate SNPs (28). In our study, we observed an increased number of non-HLA SNPs showing strong signals ($P < 1 \times 10^{-5}$) in the extreme group comparison (7 SNPs) compared with the global comparison (1 SNP). Therefore, we decided to use the information from this comparison in 2 different selection strategies.

In one strategy, we began by genotyping a group of highly significant HLA SNPs (n = 13) (data not shown) in the replication group. All of these SNPs were positively replicated. Next, we calculated a bootstrapped *P* value for the genome-wide association study extreme comparison (n = 1,000 resamplings). From this resample-based rank, we selected all non-HLA SNPs that had higher significance values than any of the positively replicated HLA class II SNPs (7 SNPs). In the second strategy, we selected those SNPs with significance of $P < 1 \times 10^{-3}$ in the extreme group comparison (326 SNPs) and that also had significance of $P < 1 \times 10^{-3}$ when tested in the early RA versus hypernormal control subject data sets (27 SNPs). Both methods yielded a total number of 34 SNPs that were genotyped and analyzed in the replication group. In order to provide a measure of contrast of our results in the replication group, we included SNPs from known candidates for RA. This included *PTPN22* (rs2476601), *CTLA4* (rs231804), and *PADI4* (rs2240340). Like the whole-genome association study analysis, the replication association analysis was performed using the allelic chi-square test ($P < 0.05$).

**Genome-wide scan for epistasis.** We performed the binary test of epistasis (SNP × SNP method) implemented in PLINK. Performing the $\sim 45 \times 10^{-9}$ pairwise analyses would take several weeks in a typical workstation. In order to make it a feasible analysis, we modified the PLINK software so that it could be run in MareNostrum, a supercomputer with 10.240 64-bit Myrinet-connected processors with a final calculation capacity of 94.21 Teraflops (Barcelona Supercomputing Centre, Barcelona, Spain). Chromosome X SNPs were excluded from the analyses. Tables showing the extended results are available online at http://www.urr.cat.

**Comparison with previous genome-wide association studies.** Using the available data from each of the 3 previous genome-wide association studies, we looked for genomic regions that share indicative association signals with our study. For the WTCCC study, results for all 500,000 SNPs are available (14), while in the Brigham and Women's Rheumatoid Arthritis Sequential Study (BRASS) (16) and North American Rheumatoid Arthritis Consortium (NARAC) and the Swedish Epidemiological Investigation of Rheumatoid Arthritis (EIRA) (15) studies, only those SNPs with *P* values of $<1 \times 10^{-3}$ and $<1 \times 10^{-4}$, respectively, are directly accessible. In order to perform a more informative analysis, we selected the most significant SNPs in one study ($P < 1 \times 10^{-4}$) (study 1) and searched for those neighboring SNPs in the other study showing an indicative significance ($P < 0.005$) (study 2). We considered 2 SNPs from different studies to be suggestive of a common association if their genomic distance was <200 kb. The results for all analyses are available online at http://www.urr.cat.

## RESULTS

Using a liability-based design, we genotyped 317,503 SNPs in 400 patients with RA and 400 control subjects. After applying several filtering criteria, 299,918 high-quality SNPs were finally selected for subsequent analyses.

**Population structure.** A dense set of SNPs covering the genome enables the robust identification of population outliers, using multidimensional analysis techniques (24,26). In our study, using the PCA technique, we identified and removed 41 outlier individuals (17 patients with RA and 24 control subjects). Analyzing the top principal components, we found that 2 of them captured a west-to-east trend, although they were less efficient in reflecting the north-to-south geographic variation (Figure 2A). Adding these 2 principal components as covariates in the genome-wide association analysis did not show a strong trend in the data (Figure 2B). This is in agreement with the low genomic inflation factor detected ($\lambda_{observed}$ = 1.01, $\lambda_{null}$ = 1). Therefore, the results reported do not correct for structure.

**Genome-wide association study findings.** We performed allelic association analyses to identify those loci associated with RA susceptibility and with general susceptibility to chronic inflammatory arthritis. Results for the strongest signals ($P < 1 \times 10^{-5}$) outside the HLA region are shown in Table 1. Except for rs2225966, rs2002842, and rs1328132, the other 19 SNPs are intronic or located within 100 kb from the closest gene. SNP rs11129989 in the extreme liability

**Table 1.** SNPs showing the strongest evidence of association in the GWAS analysis*

| GWAS analysis | SNP | Chr | Gene | MA | MAF | OR | *P* |
|---|---|---|---|---|---|---|---|
| Global | rs2002842 | 18 | *SALL3* | A | 0.49 | 1.61 | $5.52 \times 10^{-6}$ |
| Extreme liabilities | rs11129989 | 3 | *ZNF662* | G | 0.08 | 0.32 | $2.47 \times 10^{-7}$ |
| Extreme liabilities | rs1328132 | 6 | *OFCC1* | T | 0.20 | 0.46 | $3.52 \times 10^{-6}$ |
| Extreme liabilities | rs11086843 | 20 | *PTPRT* | C | 0.54 | 1.96 | $3.79 \times 10^{-6}$ |
| Extreme liabilities | rs9878975 | 3 | *AGO61* | C | 0.11 | 0.41 | $6.71 \times 10^{-6}$ |
| Extreme liabilities | rs2060396 | 2 | *CTNNA2* | A | 0.20 | 0.47 | $7.24 \times 10^{-6}$ |
| Extreme liabilities | rs2225966 | 1 | *LPHN2* | C | 0.26 | 0.50 | $8.99 \times 10^{-6}$ |
| Extreme liabilities | rs7968375 | 12 | *MANSC1* | A | 0.36 | 0.52 | $9.06 \times 10^{-6}$ |
| Chronic arthritis | rs6739713 | 2 | *R3HDM1* | G | 0.36 | 0.56 | $1.47 \times 10^{-6}$ |
| Chronic arthritis | rs946908 | 14 | *DAAM1* | C | 0.16 | 0.51 | $1.60 \times 10^{-6}$ |
| Chronic arthritis | rs2822383 | 21 | *C21orf81* | T | 0.27 | 2.11 | $3.21 \times 10^{-6}$ |
| Chronic arthritis | rs309137 | 2 | *DARS* | C | 0.43 | 0.58 | $3.22 \times 10^{-6}$ |
| Chronic arthritis | rs309160 | 2 | *DARS* | A | 0.43 | 0.58 | $3.29 \times 10^{-6}$ |
| Chronic arthritis | rs1108929 | 1 | *LOC127540* | A | 0.20 | 0.55 | $3.66 \times 10^{-6}$ |
| Chronic arthritis | rs11129989 | 3 | *ZNF662* | G | 0.11 | 0.49 | $3.82 \times 10^{-6}$ |
| Chronic arthritis | rs4314247 | 4 | *KIAA0992* | G | 0.45 | 0.58 | $4.44 \times 10^{-6}$ |
| Chronic arthritis | rs10915577 | 1 | *AJAP1* | A | 0.50 | 1.73 | $5.38 \times 10^{-6}$ |
| Chronic arthritis | rs4624474 | 21 | *BRWD1* | T | 0.48 | 1.74 | $6.28 \times 10^{-6}$ |
| Chronic arthritis | rs309143 | 2 | *DARS* | G | 0.23 | 0.56 | $6.47 \times 10^{-6}$ |
| Chronic arthritis | rs1324913 | 13 | *KLF12* | A | 0.28 | 0.58 | $6.53 \times 10^{-6}$ |
| Chronic arthritis | rs6986405 | 8 | *SGCZ* | A | 0.35 | 1.83 | $8.86 \times 10^{-6}$ |
| Chronic arthritis | rs2823580 | 21 | *C21orf34* | C | 0.25 | 2.04 | $9.51 \times 10^{-6}$ |
| Chronic arthritis | rs2807873 | 1 | *HLX1* | T | 0.24 | 0.57 | $9.73 \times 10^{-6}$ |

* Single-nucleotide polymorphisms (SNPs) from the whole-genome association study (GWAS) analyses showing the strongest significance values ($P < 1 \times 10^{-5}$). Only SNPs outside the HLA region (25–35 Mb from chromosome 6) are shown. Chr = chromosome; MA = minor allele; MAF = minor allele frequency; OR = odds ratio.

analysis is the unique non-HLA SNP that was still significant after correction for multiple testing (corrected $P = 0.013$). This SNP was selected for replication in the independent sample.

As expected, several HLA class II–region SNPs showed a strong association in the global and extreme liability analyses ($P < 1 \times 10^{-9}$) but were also the strongest markers in the chronic arthritis analysis (data not shown). In particular, SNPs rs6457617 and rs9275390 were statistically significant in all 3 analyses after correction for multiple testing. Both of these SNPs are between *HLA-DQA1* and *HLA-DQA2*, 5 kb apart from each other.

**Replication study findings.** Using 2 different approaches, we selected a total group of 34 candidate SNPs for replication in an independent cohort of 410 patients with RA and 394 control subjects. The estimated genotyping error rate was extremely low (0.3%), indicating strong reproducibility of the results.

The results for the final 38 SNPs are shown in Table 2. Among all markers tested, only 5 SNPs showed a nominal association ($P < 0.05$). Two of them (rs10864382 and rs7006821) showed an effect opposite to that detected in the genome-wide association study analysis. The other 3 SNPs were the coding SNP from *PTPN22* (rs2476601; $P = 0.022$) and 2 intronic SNPs, one from the third intron of *SVOP* (rs7313861; $P = 0.043$) and the other from the first intron of *KLF12* (rs1324913; $P = 0.013$). All 3 SNPs showed a good correlation with the size of the genetic effect detected in the genome-wide association study analysis (for the genome-wide association studies and the replication studies, respectively, the odds ratios [ORs] were 1.47 and 1.49 for rs2476601, 1.33 and 1.23 for rs731861, and 0.73 and 0.77 for rs1324913).

**Genome-wide epistasis.** We performed a genome-wide analysis of all SNP × SNP combinations and their association with susceptibility to RA and chronic arthritis. Although correction for the $>45 \times 10^{-9}$ tests performed determined a very high significance threshold ($P = 1.2 \times 10^{-12}$), we observed several SNP pairs that were very close to this value (Table 3).

**Genomic regions common with those in previous genome-wide association studies.** The integration of our results with those from the previous genome-wide association studies identified several genomic regions showing common association signals. The closest SNPs (<50 kb) are shown in Table 4. Within this group, 2 SNPs from *CSMD2* showed the most significant common association ($P = 2.99 \times 10^{-5}$ for rs10914783 in WTCCC,

**Table 2.** Results for selected SNPs in the replication study*

| SNP | Chr | Gene | Selection criteria | MA | MAF | *P*, global | OR, global | OR | *P* |
|---|---|---|---|---|---|---|---|---|---|
| rs10889271 | 1 | *INADL* | Bootstrap | T | 0.4 | 0.0016 | 0.72 | 1.00 | 0.977 |
| rs10864382 | 1 | *SLC2A5* | SC | C | 0.37 | 0.0058 | 1.36 | 0.79 | 0.028 |
| rs2807873 | 1 | *HLX1* | Bootstrap | T | 0.23 | 0.00028 | 0.65 | 1.17 | 0.162 |
| rs524331 | 1 | *TRIM67* | SC | T | 0.42 | $5.50 \times 10^{-5}$ | 1.54 | 0.83 | 0.069 |
| rs2240340 | 1 | *PADI4* | Known | A | NA | NA | NA | 1.09 | 0.411 |
| rs2476601 | 1 | *PTPN22* | Known | A | 0.12 | 0.029 | 1.47 | 1.49 | 0.022 |
| rs10490105 | 2 | *FANCL* | SC | A | 0.21 | $2.99 \times 10^{-5}$ | 0.61 | 1.22 | 0.084 |
| rs2060396 | 2 | *CTNNA2* | Bootstrap | A | 0.23 | 0.0016 | 0.69 | 1.01 | 0.958 |
| rs6739713 | 2 | *R3HDM1* | SC | G | 0.38 | 0.1188 | 0.85 | 1.11 | 0.311 |
| rs231804 | 2 | *CTLA4* | Known | C | 0.45 | 0.02059 | 0.79 | 0.83 | 0.060 |
| rs7609518 | 2 | *GPC1* | Bootstrap | C | 0.31 | 0.0051 | 1.39 | 0.95 | 0.625 |
| rs11129989 | 3 | *ZNF662* | MTS | G | 0.1 | $7.62 \times 10^{-5}$ | 0.54 | 1.30 | 0.092 |
| rs4677179 | 3 | *RYBP* | SC | A | 0.22 | 0.0032 | 1.48 | 1.11 | 0.436 |
| rs6802500 | 3 | *PDZRN3* | SC | T | 0.4 | $2.18 \times 10^{-5}$ | 1.59 | 1.03 | 0.777 |
| rs1022079 | 4 | *LOC132321* | SC | A | 0.29 | 0.00097 | 1.48 | 1.12 | 0.366 |
| rs306364 | 4 | *LOC132321* | Bootstrap | A | 0.49 | 0.086 | 1.2 | 0.95 | 0.585 |
| rs4314247 | 4 | *KIAA0992* | SC | G | 0.45 | 0.0023 | 0.73 | 0.87 | 0.154 |
| rs289079 | 4 | *PCDH7* | SC | T | 0.49 | 0.00085 | 1.43 | 0.93 | 0.477 |
| rs7725585 | 5 | *DAB2* | SC | A | 0.45 | 0.0001 | 1.53 | 1.07 | 0.494 |
| rs713584 | 5 | *SPOCK* | SC | A | 0.35 | 0.0021 | 0.72 | 1.01 | 0.937 |
| rs3130299 | 6 | *NOTCH4* | SC | G | 0.3 | 0.00048 | 0.68 | 0.82 | 0.069 |
| rs682946 | 6 | *COL9A1* | SC | C | 0.33 | $4.00 \times 10^{-5}$ | 1.61 | 1.06 | 0.624 |
| rs1565441 | 6 | *FRMD1* | SC | T | 0.49 | $8.86 \times 10^{-5}$ | 1.51 | 0.91 | 0.358 |
| rs3823833 | 7 | *ICA1* | SC | C | 0.41 | 0.00035 | 0.69 | 0.93 | 0.503 |
| rs9656200 | 7 | *GPR85* | SC | A | 0.12 | 0.0014 | 0.62 | 0.93 | 0.596 |
| rs7793728 | 7 | *Sep-07* | SC | G | 0.22 | $4.05 \times 10^{-5}$ | 0.62 | 0.95 | 0.672 |
| rs7006821 | 8 | *EYA1* | SC | C | 0.09 | 0.00018 | 2.29 | 0.68 | 0.030 |
| rs1241799 | 11 | *B3GAT1* | SC | A | 0.1 | 0.011 | 1.62 | 0.93 | 0.691 |
| rs1468796 | 12 | *TMPO* | SC | T | 0.42 | $3.68 \times 10^{-5}$ | 1.56 | 0.84 | 0.091 |
| rs7313861 | 12 | *SVOP* | SC | T | 0.42 | 0.0079 | 1.33 | 1.23 | 0.043 |
| rs1324913 | 13 | *KLF12* | SC | A | 0.28 | 0.0047 | 0.73 | 0.77 | 0.013 |
| rs1886925 | 13 | *SLC10A2* | SC | A | 0.42 | 0.0022 | 0.73 | 1.03 | 0.768 |
| rs769426 | 17 | *OR1A1* | SC | G | 0.3 | 0.597 | 0.94 | 1.08 | 0.489 |
| rs2002842 | 18 | *SALL3* | SC | A | 0.49 | $5.52 \times 10^{-6}$ | 1.61 | 0.90 | 0.287 |
| rs1329820 | 20 | *C20orf23* | Bootstrap | A | 0.23 | 0.0012 | 1.52 | 1.04 | 0.735 |
| rs6030267 | 20 | *PTPRT* | Bootstrap | A | 0.21 | 0.0021 | 1.52 | 0.94 | 0.646 |
| rs2823580 | 21 | *C21orf34* | SC | C | 0.24 | 0.028 | 1.32 | 0.92 | 0.490 |
| rs2836982 | 21 | *BRWD1* | SC | C | 0.5 | 0.0022 | 1.37 | 1.12 | 0.272 |

* Among all 38 single-nucleotide polymorphisms (SNPs) tested, 5 showed nominal significance. The 2 most associated SNPs are *KLF12* rs1324913 followed by *PTPN22* rs2476601. Two SNPs (*SLC2A5* rs10864382 and *EYA1* rs7006821) show an effect opposite to the estimated effect in the genome-wide association study analysis. Chr = chromosome; MA = minor allele; MAF = minor allele frequency; OR = odds ratio; SC = subgroup comparison; NA = not applicable; MTS = multiple testing significance.

**Table 3.** Top pairwise SNP × SNP interactions identified in all 3 genome-wide comparisons*

| GWAS | SNP 1 | Chr | Gene | SNP 2 | Chr | Gene | *P* |
|---|---|---|---|---|---|---|---|
| Global | rs9752494 | 2 | *PPM1B* | rs1569020 | 12 | *GPR133* | $1.22 \times 10^{-12}$ |
| Global | rs10465885 | 1 | *GJA5* | rs2302502 | 18 | *PTPRM* | $3.62 \times 10^{-11}$ |
| Global | rs950675 | 2 | *TPO* | rs1569020 | 12 | *GPR133* | $4.84 \times 10^{-11}$ |
| Global | rs12755965 | 1 | *GJA5* | rs6776932 | 3 | *ACPP* | $5.41 \times 10^{-11}$ |
| Extreme liability | rs259401 | 6 | *RAB32* | rs2322140 | 17 | *DNAH9* | $7.73 \times 10^{-11}$ |
| Extreme liability | rs2244817 | 8 | *SULF1* | rs3826296 | 17 | *AKAP1* | $8.56 \times 10^{-11}$ |
| Extreme liability | rs2244817 | 8 | *SULF1* | rs998113 | 17 | *AKAP1* | $9.52 \times 10^{-11}$ |
| Chronic arthritis | rs10171653 | 2 | *RTN4* | rs7033413 | 9 | *GLIS3* | $5.69 \times 10^{-12}$ |
| Chronic arthritis | rs2580768 | 2 | *RTN4* | rs7033413 | 9 | *GLIS3* | $2.63 \times 10^{-11}$ |
| Chronic arthritis | rs4849025 | 2 | *CNTNAP5* | rs2392829 | 8 | *PXDNL* | $9.07 \times 10^{-11}$ |

* The binary test implemented in PLINK was used to identify several single-nucleotide polymorphism (SNP) pairs close to the threshold for correction for multiple testing ($P = 1.11 \times 10^{-12}$). Interaction association can be detected by neighboring SNPs, as can be seen for the *SULF1–AKAP1* interaction in the extreme liability analysis and the *RTN4–GLIS3* interaction in the chronic arthritis analysis. GWAS = genome-wide association study; Chr = chromosome.

**Table 4.** Genomic loci showing common signals between the present genome-wide association study and the 3 previous genome-wide association studies*

| Region | Gene in region | Study 1 | Top SNP | *P* | Study 2 | Top SNP | *P* |
|---|---|---|---|---|---|---|---|
| 13q22 | *KLF12* | URR | rs1887346 | $6.03 \times 10^{-5}$ | BRASS | rs9318225 | $2.00 \times 10^{-3}$ |
| | | URR | rs9318228 | $3.66 \times 10^{-5}$ | | | |
| 1p35.1–p34.3 | *CSMD2* | URR | rs1108929 | $2.56 \times 10^{-5}$ | WTCCC | rs10914783 | $2.99 \times 10^{-5}$ |
| 5p15 | *TAS2R1* | URR | rs13159275 | $8.92 \times 10^{-5}$ | WTCCC | rs10513046 | 0.0021 |
| 11p15.1 | *NAV2* | URR | rs10833197 | $4.86 \times 10^{-5}$ | WTCCC | rs2568127 | 0.00051 |
| 4p14-p12 | *ATP8A1* | BRASS | rs10517039 | $2.00 \times 10^{-6}$ | URR | rs4370169 | 0.0021 |
| | | | | | URR | rs6447164 | 0.0044 |
| | | | | | URR | rs10517035 | 0.0009 |
| | | | | | URR | rs10517038 | 0.0048 |
| | | | | | URR | rs3811768 | 0.00042 |
| 10q22–q23 | *NRG3* | BRASS | rs10509440 | $6.00 \times 10^{-5}$ | URR | rs12358407 | 0.0043 |
| 10q23.1 | *KIAA1128* | BRASS | rs10491033 | $1.00 \times 10^{-7}$ | URR | rs1572430 | 0.00057 |
| 12q24.1 | *TBX5* | BRASS | rs10507251 | $4.00 \times 10^{-5}$ | URR | rs11830449 | 0.00029 |
| 1p35.1–p34.3 | *CSMD2* | WTCCC | rs10914783 | $2.99 \times 10^{-5}$ | URR | rs1108929 | $2.56 \times 10^{-5}$ |
| | | | | | URR | rs10799004 | 0.00047 |
| | | | | | URR | rs10799006 | 0.0022 |
| 1p31.1 | *IFI44* | WTCCC | rs11162922 | $1.80 \times 10^{-6}$ | URR | rs7416587 | 0.0038 |
| | | | | | URR | rs4384179 | 0.0035 |
| 5q14.1 | *CMYA5* | WTCCC | rs7343 | $8.28 \times 10^{-5}$ | URR | rs1129770 | 0.0049 |
| 6q23 | *EYA4* | WTCCC | rs2677821 | $2.48 \times 10^{-13}$ | URR | rs2327358 | 0.00073 |
| 8q13.3 | *EYA1* | WTCCC | rs4133002 | $6.17 \times 10^{-5}$ | URR | rs13274769 | 0.0028 |
| 8q23 | *OXR1* | WTCCC | rs16874205 | $9.43 \times 10^{-10}$ | URR | rs13364828 | 0.0025 |
| 10p12 | *PTER* | WTCCC | rs12269329 | $5.80 \times 10^{-6}$ | URR | rs11253931 | 0.0012 |
| 15q21.3 | *WDR72* | WTCCC | rs1711029 | $3.61 \times 10^{-12}$ | URR | rs1021744 | 0.0047 |
| 18q23 | *SALL3* | WTCCC | rs2941794 | $1.27 \times 10^{-10}$ | URR | rs2002842 | 0.00092 |
| | | | | | URR | rs2941811 | 0.0047 |
| 22q13.1 | *C1QTNF6* | WTCCC | rs743777 | $7.92 \times 10^{-6}$ | URR | rs229527 | 0.0038 |

* The top single-nucleotide polymorphisms (SNPs) from the first study (study 1; $P < 1 \times 10^{-3}$) were examined in the second study (study 2; $P < 0.005$) for signals within a 50-kb flanking region. No matches with the North American Rheumatoid Arthritis Consortium and the Swedish Epidemiological Investigation of Rheumatoid Arthritis studies were found at this genomic distance. URR = Unitat de Recerca de Rheumatologia (present study); BRASS = Brigham and Women's Rheumatoid Arthritis Sequential Study; WTCCC = Wellcome Trust Case Control Consortium.

and $P = 2.56 \times 10^{-5}$ for rs1108929 in our study). This association is the strongest detected, even after extending the analysis to a distance of 200 kb.

We found 2 genomic regions to be common in both the WTCCC study and BRASS. *KLF12* SNPs rs1887346 and rs9318228 in our study were associated with BRASS SNP rs9318225 ($P = 0.002$) (Table 4) and with WTCCC SNP rs1887346 ($P = 0.0049$). *PTPRT* intronic SNPs rs6030267 ($P = 4.08 \times 10^{-5}$) and rs11086843 ($P = 2.07 \times 10^{-6}$) were commonly associated with BRASS SNP rs10485690 ($P = 5 \times 10^{-4}$) and WTCCC SNP rs2223542 ($P = 0.0015$).

## DISCUSSION

We performed a 2-stage genome-wide association study for RA in the Spanish population, using 400 patients with RA and 400 control subjects. From these results, we selected a group of candidate SNPs and performed a replication study in an independent group of 410 patients with RA and 394 control subjects. In our study, we found *KLF12* to have stronger significance than previously associated non-HLA SNPs. We also integrated our association results with those of the 3 previous genome-wide association studies in RA. *KLF12* and *PTPRT* are the 2 unique genes that are in common regions in our study and both the WTCCC study and BRASS. Finally, we performed a genome-wide analysis for epistasis and found several SNP pairs with statistical values close to significance even after correction for multiple testing.

In the present study, we followed a liability model that could underlie susceptibility to chronic inflammatory diseases and, thus, susceptibility to RA. Results of several recent studies support this model. *NALP1* has been recently associated with vitiligo and several other autoimmune diseases, including RA (18). Fc receptor–

like protein (29) and STAT-4 (19) have been associated with susceptibility to both RA and systemic lupus erythematosus. *PTPN22* itself was studied and associated with RA after demonstrating its association with susceptibility to type 1 diabetes mellitus (7). The WTCCC genome-wide scan also provides several genomic regions linking chronic inflammatory diseases such as RA, type 1 diabetes mellitus, and Crohn's disease (14). Following this model, we selected individuals in whom the risk of developing chronic inflammatory disease was lowest (hypernormal controls) and also individuals in whom a different chronic inflammatory arthritis (non-RA) was diagnosed. In order to increase the contrast, we also included individuals in the RA group who had a highly erosive phenotype. This strategy always adds substantial power to the traditional case–control design (17), although the difficulties associated with obtaining such individuals generally limit its extended use.

Several lines of evidence support the association of *KLF12* with RA susceptibility. First, in our population, SNP rs1324913 showed a stronger association with RA compared with *PTPN22* SNP rs2476601. This was observed in the genome-wide and the replication analyses. Although this does not imply a stronger genetic effect (for *PTPN22* and *KLF12*, the estimated ORs were 1.49 and 1.3, respectively), it provides further evidence of association. The allelic association analyses performed in the present study assume a multiplicative genetic model (30) (i.e., the risk of developing the disease multiplied by a factor for each susceptibility allele carried). However, exploration of alternative genetic models (i.e., dominant, recessive, and genotypic) can give additional information on relevant genetic associations. In our replication analysis, the dominant model of rs1324913 had a significance of $P = 0.005$; no other replicated SNP showed such an increase in significance (data not shown). In order to check the consistency of this observation, we performed the same model analysis in our genome-wide association study data. We observed that the dominant model also had the strongest association in the extreme liability analysis ($P_{Dominant} = 1 \times 10^{-5}$ versus $P_{Multiplicative} = 1 \times 10^{-4}$) and the global analysis ($P_{Dominant} = 6 \times 10^{-4}$ versus $P_{Multiplicative} = 5 \times 10^{-3}$).

Other important evidence supporting *KLF12* association is that, together with *PTPRT*, they are the only 2 genomic regions commonly found when comparing our study with the BRASS and WTCCC studies. In the extreme liability analysis, 2 SNPs in the *KLF12* transcribed region had a significance of $P < 1 \times 10^{-4}$ (rs1887346 and rs9318228 with $P = 6.02 \times 10^{-5}$ and $P = 3.66 \times 10^{-5}$, respectively). Both SNPs are only 8.2 kb apart from each other, and are only 1.7 kb and 9.9 kb, respectively, from the SNP rs9318225 in BRASS ($P = 2 \times 10^{-3}$) (Table 1). When comparing our study with the WTCCC, we found rs1184596 (same intron, 170 kb upstream) to have an indicative association ($P = 4.9 \times 10^{-3}$). Although it is more distant to SNPs rs1887346 and rs9318228, it is closer to the replicated *KLF12* SNP rs1324913 (38 kb). This finding supports the replicability of this association in different populations.

*KLF12* (activator protein 2$\alpha$ [AP-2$\alpha$] repressor) is a member of the family of Kruppel-like transcriptional regulatory factors (31), which play fundamental roles in differentiation and development. *KLF12* is known to repress the transcription of *AP-2*$\alpha$ transcription factor after binding to the general correpressor protein C-terminal binding protein 1 (32). The expression patterns of AP-2$\alpha$–regulated genes, including the gene for tumor necrosis factor $\alpha$ (TNF$\alpha$), have been implicated in malignant transformation and stress responses (33,34). Thus, genetic variations could increase susceptibility to RA through various mechanisms: either by facilitating the transformation of local connective cells (32) or by promoting lymphocyte survival (35). More intriguingly, the recent characterization of AP-2–mediated TNF$\alpha$ gene expression in B19 parvoviral infection could add an alternative mechanism. This type of infection can produce a chronic inflammatory arthritis that can fulfill the diagnostic criteria for RA. For many years, B19 has been studied as a possible trigger for RA, although with controversial results (36,37). The observed genetic association of *KLF12* with RA and its direct implication in TNF$\alpha$ regulation could represent a new perspective on genetic and environmental interactions associated with RA susceptibility.

The integration of our results with those of previous genome-wide association studies identified several relevant genomic regions. *PTPRT* and *KLF12* are the only loci associated in both the WTCCC study and BRASS. Protein tyrosine phosphatase (PTP) receptor T, the most frequently mutated PTP in human cancers, has been recently characterized as a key inhibitor of STAT-3 (38). STAT-3, in turn, mediates transcriptional activation in response to several cytokines, including RA-associated interleukin-6. The common signal at *CSMD2* in our study and the WTCCC study is the strongest association detected. *CSMD2* is a recently cloned gene (39) whose functionality still needs to be identified. Recently, *CSMD1*, a gene with high structural similarity to *CSMD2*, has been implicated in the inhibition of complement activation (40), which could suggest a com-

mon functionality. In addition, the results from our analysis along with previous genome-wide association studies also identified several SNPs from the newly associated *TNFAIP3* locus (rs643571 at $P = 6 \times 10^{-3}$ and rs6915853 at $P = 2 \times 10^{-3}$ in our study). This finding confirms the power of our analysis method to identify genomic regions that are relevant to RA susceptibility.

Although the same technologic platform was used, comparison of our study with the NARAC and EIRA studies did not reveal evidence of association for any of the *TRAF1-C5*–region SNPs. This lack of association, which was also observed in the WTCCC study (15), could be largely attributable to the restricted analysis of anti-CCP antibody–positive RA patients in the NARAC and EIRA cohorts. Therefore, an exact comparison would have required specifically addressing the association in anti-CCP antibody–positive patients only. However, a more detailed analysis of the association in this region showed additional results. We found several indicative signals (4 SNPs at a significance level of $P < 0.01$) in the 5′– and 3′–untranslated regions of *FBXW2*. This gene is 109 kb from *TRAF1* and encodes for a protein that participates in the ubiquitin/proteasome degradation system (41). Therefore, our analysis suggests that, in this genomic region, *FBXW2* seems to be associated with RA susceptibility.

There is increasing evidence that gene–gene interactions (epistasis) could be of major relevance in susceptibility to complex diseases (42,43). A recent study demonstrated that complete genome-wide analysis has more power to detect relevant SNP pairs than methods that use filtering strategies, even after correction for multiple testing (43). In the present study, we performed an exhaustive analysis of all autosomal SNP pairs and found several of them with significance close to the threshold of significance for multiple testing. An important observation is that no known main-effect SNP was observed in this group of top SNPs. In fact, HLA class II SNPs appear only at the level of $P = 1 \times 10^{-8}$ to $P = 1 \times 10^{-7}$. This could probably indicate that, although the possibility of epistasis with this region cannot be discarded, other regions with marginal main effects seem to show stronger interactions associated with disease risk. This also confirms the need to perform exhaustive analyses in the search for epistasis. To our knowledge, none of the top SNP pairs ($P < 1 \times 10^{-10}$) belong to genes from a known common biologic pathway. Protein phosphatase 1B, a regulator of NF-κB transcription factor, has a strong interaction in the global analysis with protein G–coupled receptor 133, which, to date, has no associated biologic function. In the extreme liability analysis, the human sulfatase 1 gene (*SULF1*), a heparan sulfatase involved in tumor progression and inflammation (44), is also interacting with 2 SNPs from the A kinase anchor protein 1 gene (*AKAP1*). The latter has been associated with cAMP-mediated signal transduction and messenger RNA trafficking (45).

The present study is one of the first genome-wide association analyses in RA. Using a liability-based design, we found several new candidate SNPs for RA and chronic inflammatory arthritis. We performed a replication analysis in an independent subset of SNPs, from which *KLF12* emerged as a new candidate susceptibility gene for RA. A comparison of our results with those from the 3 previous genome-wide association studies confirmed the relevance of the *KLF12* locus and also identified several other regions of interest for subsequent studies. In order to search for more complex genetic models involved in RA susceptibility, we performed a genome-wide analysis for epistasis. The results presented here add important aspects to the continuing definition of RA genetic architecture. Consistent replication of these results in different populations will confirm the association of the genomic regions to RA susceptibility.

## ACKNOWLEDGMENTS

We thank Professor David Clayton for his critical review of the manuscript. We are indebted to the Barcelona Supercomputing Centre support team. We also are grateful to Genoma España for their support to this project, and especially to Professor J. Bertranpetit CeGen (National Genotyping Center), Professor J. Benitez (CeGen-Nodo de Madrid), Professor A. Carracedo (CeGen-Nodo de Santiago de Compostela), Dr. A. Garcia, and Professor A. Orfao (Banco Nacional de ADN).

## AUTHOR CONTRIBUTIONS

Dr. Marsal 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.** Julià, Ballina, Tornero-Molina, Alperi-López, Erra, Barceló, Marsal.

**Acquisition of data.** Julià, Ballina, Cañete, Balsa, Tornero-Molina, Naranjo, Alperi-López, Erra, Pascual-Salcedo, Marsal.

**Analysis and interpretation of data.** Julià, Ballina, Cañete, Tornero-Molina, Alperi-López, Camps, Marsal.

**Manuscript preparation.** Julià, Marsal.

**Statistical analysis.** Julià, Marsal.

## REFERENCES

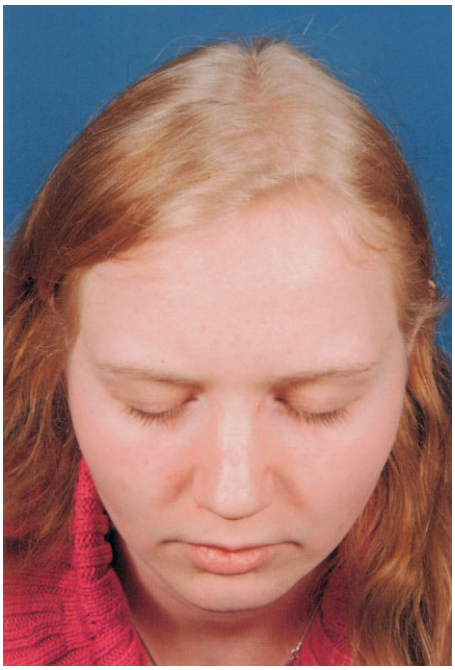
1. Marrack P, Kappler J, Kotzin BL. Autoimmune disease: why and where it occurs [review]. Nat Med 2001;7:899–905.

2. Firestein GS. Evolving concepts of rheumatoid arthritis [review]. Nature 2003;423:356–61.
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. Cornelis F, Faure S, Martinez M, Prud'homme JF, Fritz P, Dib C, et al, for the European Consortium on Rheumatoid Arthritis Families. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. Proc Natl Acad Sci U S A 1998;95:10746–50.
5. Jawaheer D, Li W, Graham RR, Chen W, Damle A, Xiao X, et al. Dissecting the genetic complexity of the association between human leukocyte antigens and rheumatoid arthritis. Am J Hum Genet 2002;71:585–94.
6. Chapman J, Clayton D. Detecting association using epistatic information. Genet Epidemiol 2007;31:894–909.
7. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–7.
8. Suzuki A, Yamada R, Chang X, Tokuhiro 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.
9. Rodriguez MR, Nunez-Roldan A, Aguilar F, Valenzuela A, Garcia A, Gonzalez-Escribano MF. Association of the CTLA4 3′ untranslated region polymorphism with the susceptibility to rheumatoid arthritis. Hum Immunol 2002;63:76–81.
10. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies [review]. Genet Med 2002;4:45–61.
11. Risch N, Merikangas K. The future of genetic studies of complex human diseases. Science 1996;273:1516–7.
12. Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet 2001;17:502–10.
13. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits [review]. Nat Rev Genet 2005;6:95–108.
14. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–78.
15. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, Ding B, et al. TRAF1-C5 as a risk locus for rheumatoid arthritis: a genomewide study. N Engl J Med 2007;357:1199–209.
16. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, Maller J, et al. Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 2007;39:1477–82.
17. Morton NE, Collins A. Tests and estimates of allelic association in complex inheritance. Proc Natl Acad Sci U S A 1998;95:11389–93.
18. Jin Y, Mailloux CM, Gowan K, Riccardi SL, LaBerge G, Bennett DC, et al. NALP1 in vitiligo-associated multiple autoimmune disease. N Engl J Med 2007;356:1216–25.
19. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977–86.
20. 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.
21. Doran MF, Pond GR, Crowson CS, O'Fallon WM, Gabriel SE. Trends in incidence and mortality in rheumatoid arthritis in Rochester, Minnesota, over a forty-year period. Arthritis Rheum 2002;46:625–31.
22. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al, for the International HapMap Consortium. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449:851–61.
23. Buetow KH, Edmonson M, MacDonald R, Clifford R, Yip P, Kelley J, et al. High-throughput development and characterization of a genomewide collection of gene-based single nucleotide polymorphism markers by chip-based matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Proc Natl Acad Sci U S A 2001;98:581–4.
24. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–9.
25. Devlin B, Roeder K, Wasserman L. Genomic control, a new approach to genetic-based association studies [review]. Theor Popul Biol 2001;60:155–66.
26. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
27. Winkelmann J, Schormair B, Lichtner P, Ripke S, Xiong L, Jalilzadeh S, et al. Genome-wide association study of restless legs syndrome identifies common variants in three genomic regions. Nat Genet 2007;39:1000–6.
28. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 2007. E-pub ahead of print.
29. Kochi Y, Yamada R, Suzuki A, Harley JB, Shirasawa S, Sawada T, et al. A functional variant in FCRL3, encoding Fc receptor-like 3, is associated with rheumatoid arthritis and several autoimmunities. Nat Genet 2005;37:478–85.
30. Balding DJ. Handbook of statistical genetics. West Sussex (UK): John Wiley & Sons; 2001.
31. Bieker JJ. Kruppel-like factors: three fingers in many pies. J Biol Chem 2001;276:34355–8.
32. Schuierer M, Hilger-Eversheim K, Dobner T, Bosserhoff AK, Moser M, Turner J, et al. Induction of AP-2$\alpha$ expression by adenoviral infection involves inactivation of the AP-2rep transcriptional corepressor CtBP1. J Biol Chem 2001;276:27944–9.
33. Kannan P, Buettner R, Chiao PJ, Yim SO, Sarkiss M, Tainsky MA. N-ras oncogene causes AP-2 transcriptional self-interference, which leads to transformation. Genes Dev 1994;8:1258–69.
34. Wajapeyee N, Somasundaram K. Cell cycle arrest and apoptosis induction by activator protein 2$\alpha$ (AP-2$\alpha$) and the role of p53 and p21WAF1/CIP1 in AP-2$\alpha$-mediated growth inhibition. J Biol Chem 2003;278:52093–101.
35. Zhou M, McPherson L, Feng D, Song A, Dong C, Lyu SC, et al. Kruppel-like transcription factor 13 regulates T lymphocyte survival in vivo. J Immunol 2007;178:5496–504.
36. Simpson RW, McGinty L, Simon L, Smith CA, Godzeski CW, Boyd RJ. Association of parvoviruses with rheumatoid arthritis of humans. Science 1984;223:1425–8.
37. Takahashi Y, Murai C, Shibata S, Munakata Y, Ishii T, Ishii K, et al. Human parvovirus B19 as a causative agent for rheumatoid arthritis. Proc Natl Acad Sci U S A 1998;95:8227–32.
38. Zhang X, Guo A, Yu J, Possemato A, Chen Y, Zheng W, et al. Identification of STAT3 as a substrate of receptor protein tyrosine phosphatase T. Proc Natl Acad Sci U S A 2007;104:4060–4.
39. Lau WL, Scholnick SB. Identification of two new members of the CSMD gene family small star, filled. Genomics 2003;82:412–5.
40. Kraus DM, Elliott GS, Chute H, Horan T, Pfenninger KH, Sanford SD, et al. CSMD1 is a novel multiple domain complement-regulatory protein highly expressed in the central nervous system and epithelial tissues. J Immunol 2006;176:4419–30.
41. Yang CS, Yu C, Chuang HC, Chang CW, Chang GD, Yao TP, et al. FBW2 targets GCMa to the ubiquitin-proteasome degradation system. J Biol Chem 2005;280:10083–90.

42. Julia A, Moore J, Miquel L, Alegre C, Barcelo P, Ritchie M, et al. Identification of a two-loci epistatic interaction associated with susceptibility to rheumatoid arthritis through reverse engineering and multifactor dimensionality reduction. Genomics 2007;90:6–13.
43. Evans DM, Marchini J, Morris AP, Cardon LR. Two-stage two-locus models in genome-wide association. PLoS Genet 2006;2:e157.
44. Kodaira Y, Nair SK, Wrenshall LE, Gilboa E, Platt JL. Phenotypic and functional maturation of dendritic cells mediated by heparan sulfate. J Immunol 2000;165:1599–604.
45. Furusawa M, Taira T, Iguchi-Ariga SM, Ariga H. AMY-1 interacts with S-AKAP84 and AKAP95 in the cytoplasm and the nucleus, respectively, and inhibits cAMP-dependent protein kinase activity by preventing binding of its catalytic subunit to A-kinase-anchoring protein (AKAP) complex. J Biol Chem 2002;277:50885–92.

DOI 10.1002/art.23652

## *Clinical Image: Blonde by prescription*



The patient, a 25-year-old woman who worked as a farmer, was diagnosed as having discoid lupus erythematosus. Her inflammatory skin lesions were successfully treated with hydroxychloroquine (HCQ; 400 mg/day). Four months after the initiation of therapy, generalized and symmetric depigmentation of the patient's scalp and body hair (so-called "acquired poliosis") was noted. HCQ is a Toll-like receptor 9 pathway–inhibiting antimalarial agent that is also used in the therapeutic management of autoimmune diseases such as systemic lupus erythematosus and rheumatoid arthritis. Generalized drug-induced poliosis in patients with strawberry-blonde hair, as well as vitiligo in African patients with dark skin, have been reported to occur during therapy with antimalarial agents. The underlying mechanisms remain elusive. It has been suggested that binding of antimalarial agents to eumelanin and pheomelanin disturbs melanogenesis. Discontinuation of the medication usually leads to repigmentation of the hair and skin. In our patient, newly grown hair 3 months after cessation of the HCQ therapy exhibited her natural hair color. In those cases in which the therapy must be continued, the use of hair dye is the preferred way to achieve cosmetic improvement.

Stephan Meller, MD
Peter Arne Gerber, MD
Bernhard Homey, MD
*Heinrich-Heine University*
*Dusseldorf, Germany*