Rheumatoid Arthritis in Latin Americans Enriched for Amerindian Ancestry Is Associated With Loci in Chromosomes 1, 12, and 13, and the HLA Class II Region

David López Herráez,[1] Manuel Martínez-Bueno,[1] Laura Riba,[2] Ignacio García de la Torre,[3] Mónica Sacnún,[4] Mario Goñi,[5] Guillermo A. Berbotto,[6] Sergio Paira,[7] Jorge Luis Musuruana,[8] César E. Graf,[9] Alejandro J. Alvarellos,[10] Osvaldo D. Messina,[11] Alejandra M. Babini,[12] Ingrid Strusberg,[13] Juan Carlos Marcos,[†] Hugo Scherbarth,[14] Alberto J. Spindler,[15] Ana Quinteros,[16] Sergio M. A. Toloza,[17] José Luis C. Moreno,[18] Luis J. Catoggio,[19] Guillermo Tate,[20] Alicia Eimon,[21] Gustavo Citera,[22] Antonio Catalán Pellet,[23] Gustavo G. Nasswetter,[24] Mario H. Cardiel,[25] Pedro Miranda,[26] Francisco Ballesteros,[26] Jorge A. Esquivel-Valerio,[27] Marco A. Maradiaga-Ceceña,[28] Eduardo M. Acevedo-Vásquez,[29] Conrado García García,[30] the GENAR Study Collaborators, Teresa Tusié-Luna,[2] Bernardo A. Pons-Estel,[31] and Marta E. Alarcón-Riquelme[32]

Objective. To identify susceptibility loci for rheumatoid arthritis (RA) in Latin American individuals with admixed European and Amerindian genetic ancestry.

Methods. Genotyping was performed in 1,475 patients with RA and 1,213 control subjects, using a

customized BeadArray containing 196,524 markers covering loci previously associated with various autoimmune diseases. Principal components analysis (EigenSoft package) and Structure software were used to identify outliers and define the population substructure. REAP software was used to define cryptic relatedness and duplicates, and genetic association analyses were conducted using Plink statistical software.

Results. A strong genetic association between RA and the major histocompatibility complex region was observed, localized within *BTNL2/DRA–DQB1–DQA2* ($P = 7.6 \times 10^{-10}$), with 3 independent effects. We identified an association in the *PLCH2-HES5-TNFRSF14-MMEL1* region of chromosome 1 ($P = 9.77 \times 10^{-6}$), which was previously reported in Europeans, Asians, and Native Canadians. We identified one novel putative association in *ENOX1* on chromosome 13 ($P = 3.24 \times 10^{-7}$). Previously reported associations were observed in the current study, including *PTPN22*, *SPRED2*, *STAT4*, *IRF5*, *CCL21*, and *IL2RA*, although the significance was relatively moderate. Adjustment for Amerindian ancestry improved the association of a novel locus in chromosome 12 at *C12orf30* (*NAA25*) ($P = 3.9 \times 10^{-6}$). Associations with the *HLA* region, *SPRED2*, and *PTPN22* improved in individuals positive for anti–cyclic citrullinated peptide antibodies.

Conclusion. Our data define, for the first time, the contribution of Amerindian ancestry to the genetic architecture of RA in an admixed Latin American population by confirming the role of the *HLA* region and supporting the association with a locus in chromosome 1. In addition, we provide data for novel putative loci in chromosomes 12 and 13.

Rheumatoid arthritis (RA) is a painful chronic inflammatory condition that typically leads to destruction of the joints and disability. Understanding the genetic basis of RA is necessary for the development of a more personalized approach to treating the disease. Therefore, achieving an understanding of the mechanisms of RA and its genetic architecture is required groundwork for improving our knowledge of this disabling condition.

The prevalence of RA is particularly high among the native populations of the Americas, hereafter called Amerindians. For example, the prevalence of RA is 10-fold higher among Native Canadians/Native Americans than among Europeans (3% and 0.3%, respectively) (1,2), and recent studies have demonstrated that unaffected relatives of RA patients with Amerindian ancestry have an increased prevalence of anti–cyclic citrullinated peptide (anti-CCP) antibodies and higher serum cytokine levels compared with their European counterparts within the same geographic areas (3). Furthermore, our group recently reported that in the population of the Yucatán region of Mexico, in which Amerindian ancestry is common, the prevalence of RA is 5-fold higher than that in other regional populations in Mexico (4,5). In general, individuals of Amerindian ancestry (6,7) have more severe disease, with more joint destruction.

Despite the clear need to understand the risk of RA in admixed populations with enriched Amerindian ancestry, association studies to define risk loci derived from individuals with Amerindian ancestry have proven difficult. Populations in Latin America are genetically highly heterogeneous, admixed between Europeans, Amerindians, and Sub-Saharan Africans, as a result of the relatively recent history of encounter. Thus, although several susceptibility loci in individuals of European or Asian origin have been identified through genome-wide association scans and fine mapping in recent years (8–18), little is known about genetic susceptibility to RA in admixed populations with enriched Amerindian ancestry. In particular, we do not know how results obtained in European or Asian populations apply to admixed populations.

The goal of the present study was to identify novel loci that are associated with RA in populations that may be derived from Amerindian ancestry. To do this, we assembled the largest cohort to date of Latin American patients with RA and performed genotyping using a customized array containing loci that were previously shown to be genetically associated with various autoimmune diseases (19). We observed a clear genetic association with the HLA region, with the strongest peaks in the regions of *BTNL2/DRA* and *DQB1/DQA2*. In addition, we observed a strong association of the previously identified *PLCH2-HES5-TNFRSF14-MMEL1* region in chromosome 1p36.32. Finally, a novel suggestive locus on chromosome 13q14.11 (*ENOX1*) was identified. Associations with other known loci were detected, although the significance was relatively moderate. Adjustment for Amerindian ancestry improved the genetic association with several single-nucleotide polymorphisms (SNPs) in the *C12orf30/NNA25* gene on chromosome 12, suggesting that Amerindian ancestry in the admixed population probably contributes this RA risk locus.

PATIENTS AND METHODS

Blood sample collection. Peripheral blood samples were obtained through a multicenter collaborative project known as "Genómica de artritis reumatoide" (GENAR) involving centers in Argentina, Mexico, Chile, and Peru (see Appendix A for GENAR members and their locations). To be included in the study, patients were required to be older than age 18 years, to fulfill at least 4 of the 1987 American College of Rheumatology criteria for the classification of RA (20), to have had clinical disease beginning at least 2 years prior to inclusion in our study, and to sign an informed consent form approved by the local institutional review board. Individuals with a past or present diagnosis of juvenile RA, concomitant or underlying inflammatory autoimmune disease, or spondyloarthropathies were excluded. Patients and control subjects resided in the same geographic area. The treating rheumatologists subjectively assessed the ethnicity of the individuals. For genotyping, 2,688 samples were obtained from subjects from Argentina (816 patients and 644 controls), Mexico (281 patients and 380 controls), Peru (243 patients and 111 controls), and Chile (135 patients and 78 controls). The male:female ratios among patients and controls were 1:8 and 1:9, respectively. A total of 768 patients were positive for anti-CCP antibodies. The remainder were either negative or had no available data regarding anti-CCP antibody status.

Microarray genotyping. Genotyping was performed using an Illumina Infinium High-Density BeadArray designed by the Immunochip Consortium (19). Genomic DNA extracted from the peripheral blood samples was genotyped (as a service of the Robert S. Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, Manhasset, New York) according to Illumina protocols. NCBI build 36 (hg18) mapping was used. Images and associated data files were generated using an Illumina iScan System. Clustering (i.e., SNP calling) was performed for the entire collection of samples together, using the Illumina GenomeStudio GenTrain 2.0 algorithm. Subsequently, mandatory concordance between self-reported and X chromosome heterogeneity–estimated sex for each sample was examined.

Data set filtering. Prior to performance of the statistical analyses, management and filtering of the working data set were conducted using Plink version 1.07 (21). The starting data set included 2,667 subjects and 196,524 markers (718 small insertion/deletions and 195,806 SNPs). We applied the following filters: minimum total call rate per sample of 90% (this rate excluded 5 subjects), minimum call rate per marker of 98%, minor allele frequency threshold of 0.5%, significant deviation from Hardy-Weinberg equilibrium in patients and controls ($P < 0.0001$), significant deviation from Hardy-Weinberg equilibrium in controls ($P < 0.01$), and a cutoff P value of 0.00001 for differential missingness in no-call genotypes between patients and controls.

Stratification analyses. HapMap data were scanned to detect the presence of each marker identifier contained in the Immunochip. Markers showing different absolute allele frequencies between HapMap populations were used to study the genetic substructure of the working data set. A total of 10,847 such markers were identified. In order to have independent markers (i.e., noncorrelated [$r^2 > 0.2$]), Plink linkage disequilibrium (LD) analysis was run using a 50-marker moving window, sliding every 5 markers. Additionally, markers present in regions known to be associated with RA risk, including the HLA region, were excluded from this analysis. We obtained a subset of 6,337 independent markers common to our samples and the HapMap population sets: Yoruba (YRI; n = 60 individuals), European (CEU; n = 60), Chinese (CHB; n = 45), and Japanese (JPT; n = 45). Furthermore, 2 additional sets of Immunochip-genotyped subjects from Spain (369 Europeans and 34 Nahuas [indigenous Meso-Americans serving as a proxy for Amerindian ancestry]) were used as in-house reference populations.

To estimate the admixture proportion in our samples, we performed analyses with Structure version 2.2 software (22), using the same set of independent markers for all of the described subjects (for best-fit clustering with $K = 4$ for 5 iterations). Analyses using Structure software revealed 10 individuals for whom the estimated Sub-Saharan African or East Asian genetic ancestry was >10%, and these individuals were also removed from further analyses (see Supplementary Figure 1, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract). Using the same set of markers and the admixture proportions estimated with Structure, we investigated the cryptic relatedness in our samples using a REAP software program for pairwise kinship coefficient estimation specifically designed for admixed samples (23). We removed one individual from each pair of samples with kinship coefficient estimates of >0.044 (n = 118 subjects). This value excluded subjects having up to a potential one-third degree of kinship, as recommended by Thornton et al (23). Because we observed that this value excluded subjects with >70% Amerindian ancestry, we used a coefficient of 0.15 for the analysis adjusting for Amerindian ancestry, which is somewhat more liberal. This approach excluded 36 individuals who were among those extracted when using the estimate of 0.044 but who were clearly duplicates.

Principal components analysis (PCA) was performed using the smartpca program in the EigenSoft 4.0 beta package (24) and identified 141 outliers in our Immunochip set (σ threshold = 6 with 5 iterations along 15 eigenvectors). Finally, after removal of outlier and "close-related" subjects, an additional PCA was performed with the same settings as described above using only the case–control samples with the subset of 6,337 markers to compute the values of the most significant axes of variation for inclusion as covariates in the subsequent association testing. Two more individuals were detected as outliers. Our final filtered working data set had 2,396 individuals (1,317 patients and 1,079 controls) and 114,788 markers. Of the patients, 750 were anti-CCP antibody positive. The final inflation factor to be used for genomic control (GC) correction was 1.039 (see Supplementary Figure 2, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract), which was obtained when using the first 4 principal components.

Analysis of correlation between eigenvectors and ancestry proportions obtained from Structure. We used linear regression analysis to test the correlation between the stratification results obtained from the PCA and the Structure analysis. We made a pairwise comparison of the genetic components obtained from Structure for the best-fit clustering result at $K = 4$, with the following average proportions: for Amerindians, 45.40%; for Europeans, 50.18%; for Sub-

Table 1. Main identified genetic associations with rheumatoid arthritis in Latin Americans enriched for Amerindian ancestry and P values after stratification for the presence of anti-CCP antibodies*

SNP	Location	Locus	Alleles	MAF, cases	MAF, controls	OR (95% CI)	P†	FDR	P‡
rs2395178	6p21.32	*BTNL2/DRA*	G/C	0.2419	0.3372	0.64 (0.65–0.63)	7.60×10^{-10}	3.41×10^{-6}	6.59×10^{-10}
rs9275224	6p21.32	*DQB1/DQA2*	A/G	0.2061	0.3066	0.61 (0.53–0.71)	5.05×10^{-10}	3.41×10^{-6}	3.75×10^{-11}
rs3135335	6p21.32	*BTNL2/DRA*	G/C	0.2421	0.3356	0.65 (0.57–0.74)	1.71×10^{-9}	4.86×10^{-6}	1.94×10^{-9}
rs3129867	6p21.32	*BTNL2/DRA*	G/C	0.2347	0.3229	0.66 (0.58–0.74)	6.10×10^{-9}	2.77×10^{-5}	6.89×10^{-9}
rs3998158	6p21.32	*DQB1/DQA2*	G/A	0.3814	0.2865	1.48 (1.31–1.69)	5.66×10^{-8}	3.57×10^{-5}	3.85×10^{-9}
rs9501626	6p21.32	*BTNL2/DRA*	A/C	0.1649	0.2319	0.64 (0.55–0.74)	5.11×10^{-8}	1.09×10^{-4}	6.59×10^{-10}
rs34102154	6p21.32	*DRB1/DQA1*	A/G	0.1685	0.2392	0.67 (0.58–0.77)	2.83×10^{-7}	6.39×10^{-4}	1.94×10^{-9}
rs1493522	13q14.11	*ENOX1*	C/G	0.0335	0.0712	0.47 (0.35–0.62)	3.24×10^{-7}	1.05×10^{-3}	7.89×10^{-5}
rs11585747	1p36.32	*HES5/TNFRSF14*	A/G	0.477	0.4173	1.37 (1.22–1.55)	9.34×10^{-6}	2.27×10^{-3}	6.25×10^{-5}
rs2296442	1p36.32	*PLCH2*	A/G	0.4878	0.4311	1.37 (1.20–1.53)	9.77×10^{-6}	3.09×10^{-3}	2.60×10^{-5}

* SNP = single-nucleotide polymorphism; MAF = minor allele frequency; OR = odds ratio; 95% CI = 95% confidence interval; FDR = false discovery rate.
† Corrected for genomic controls.
‡ Obtained when stratifying for individuals positive for anti–cyclic citrullinated peptide (anti-CCP) antibodies, after correction for genomic control.

Saharan Africans, 2.23%; for East Asians, 2.19%. Linear regression analyses were performed using in-house R scripts (http://www.r-project.org/).

Statistical analysis. Allelic associations for each marker were tested using the filtered data set with Plink additive logistic regression, testing for the minor allele as reference (21). The Tracy-Widom test was used to evaluate the significance of the principal components. Association tests included the axes of the 10 most significant principal components as covariates. In a separate analysis, and in order to study the contribution of the Amerindian genetic component in RA, we performed an additional association analysis using exclusively principal component 1 as covariate. For each association analysis, we generated files of adjusted significance values, defining statistical significance as $P < 4 \times 10^{-6}$ for GC-corrected tests or $P < 0.05$ for Bonferroni-adjusted single-step tests. The dependency of association for genetic variants with the lowest P values was systematically examined by stepwise conditional testing. Plots were generated using in-house R scripts (http://www.r-project.org/) and a modified version of LocusZoom (25). For the latter, user-supplied LD tables were generated from our Immunochip-genotyped data. LD was computed using Haploview version 4.2 (26).

RESULTS

PCA and Structure analysis. PCA was used to investigate the underlying genetic structure of the genotyped samples. This methodology allows for analysis without assigning individuals to any putative population of origin. The results were largely concordant with the 4 clusters obtained in the Structure analysis. The plot of principal component 2 versus principal component 1 in the first PCA (see Supplementary Figure 1) resembled the pattern previously observed for this type of analysis (27,28). The first 2 principal components explained the greatest proportion (68.7%) of the genetic variation

observed in the data. Population outliers are also shown in the plot. Principal component 2 differentiated the YRI population from all others. Principal component 1 exhibited a gradient of genetic variation extending from the South American continent across Eurasia, with East Asians (Chinese and Japanese) grouping together, and a European cluster composed of HapMap CEU– and in-house–typed Spaniards. Rather than forming clusters, individuals from Mexico and South America were spread along the principal component 1 axis, supporting the conclusion that these populations consist of admixed individuals with varying degrees of European and/or Amerindian genetic ancestry. This is even clearer in the plot of principal component 2 versus principal component 1 in the last PCA following exclusion of the outliers and cryptic relatedness. In this case, principal component 1 explained 67% of the genetic variation. This variation runs continuously from its minimal value, corresponding to a Peruvian individual with 99.89% of Amerindian ancestry (<0.1% of European ancestry), to its maximum value, an Argentinian individual with <0.1% of Amerindian ancestry (99.28% of European ancestry).

Three putative risk loci revealed by association analyses using the first 4 principal components. Next, we tested the genetic association of individual SNPs with RA in our cohort. Filters were applied as described in Patients and Methods, followed by logistic regression analysis using the additive model. Importantly, to account for the number of SNPs contained in the genotyping array, we considered only associations significant at P values less than the nominal alpha level of 4×10^{-6} for GC-corrected tests, corresponding to P values less
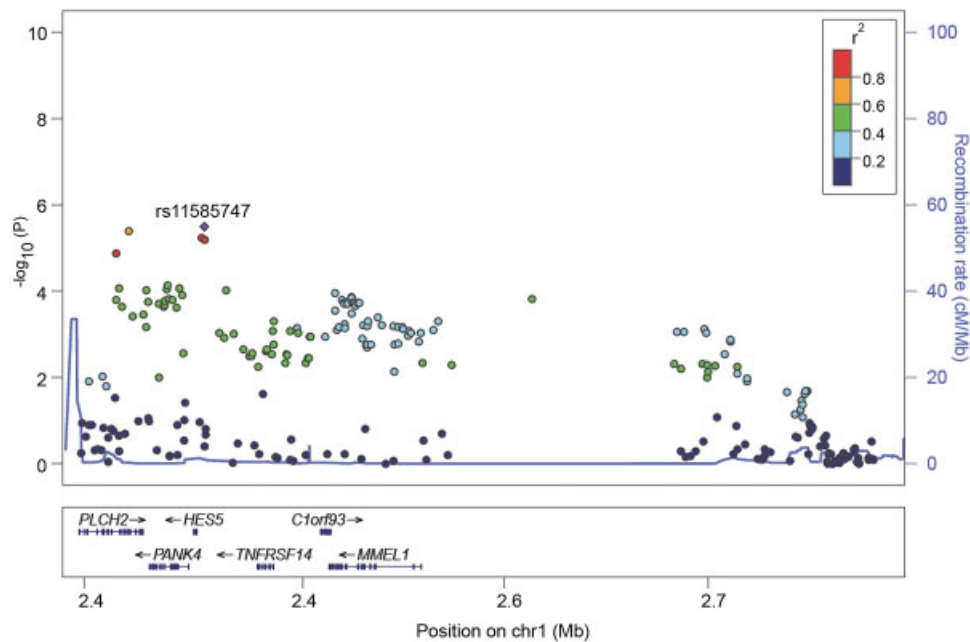
Figure 1. Detailed LocusZoom plot of the genetic association in the region on chromosome 1p36. The colors represent the r^2 between single-nucleotide polymorphisms. The diamond symbol represents the best hit; for the sake of clarity, only the best hit is shown.

than 0.05 after Bonferroni correction for multiple testing. The Manhattan plot shown in Supplementary Figure 3 (available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract) shows the results of the individual SNP allelic associations tested using the additive model, with the minor allele as reference and corrected for the top 10 primary components. Several loci reached our threshold of statistical significance and were located at 6p21.32 (rs2395178 and rs9275224), 1p36.32 (rs11585747 and rs2296442), and 13q14.11 (rs1493522) (Table 1). Supplementary Table 1 (available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract) shows the best results with GC-corrected P values less than 6.0×10^{-3}.

Locus 1p36.32 includes a 150-kb region containing the genes *PLCH2, PANK4, HES5, TNFRSF14, C1orf93*, and *MMEL1*. The latter 3 genes were previously shown to be associated with RA in Europeans, East Asians, and, recently, in a Native American population (2,11,17). Although the top SNPs were located 5′ of those genes (Figure 1), conditional analysis suggested that the associations identified in 1p36.32 were not independent of those described previously. This is somewhat expected, because genetic admixture is known to result in large LD blocks. Indeed, inspection of the LD pattern confirmed that the markers were in high LD,

with an average r^2 value of 0.88 extending across 40.6 kb (data not shown). Therefore, the associations we identified in our admixed cohort supported previous results that implicated 1p36.32 in RA susceptibility. In addition, rs2296442 ($P = 9.77 \times 10^{-6}$, odds ratio 1.37) (see Table 1) is a coding nonsynonymous substitution in *PLCH2*.

We then evaluated locus 13q14.11, which contained only a single RA-associated marker (rs1493522) (see Table 1). Tests performed to condition on this SNP repeatedly suggested associations for up to 36 markers in the same genomic region of locus 13q14.11, indicating that the association signal was not completely dependent on rs1493522. Therefore, we tested each of these 36 associated markers in a stepwise conditional manner to verify possible new independent effects across the region. All tests performed gave virtually the same results as those of the conditional analysis on rs1493522 using an additive model. The analysis using Haploview showed high LD in the 13q14.11 locus, with an average r^2 value >0.86 for a genomic region spanning 134 kb when conditioned on rs1493522 (Figure 2). This SNP showed an average r^2 value of 0.67 with the other SNPs throughout the region (see Supplementary Figure 4, available on the *Arthritis & Rheumatism* web site at http://online library.wiley.com/doi/10.1002/art.37923/abstract). Thus, locus 13q14.11, including *ENOX1*, contains 2 independent effects, but the precise SNP(s) responsible for the
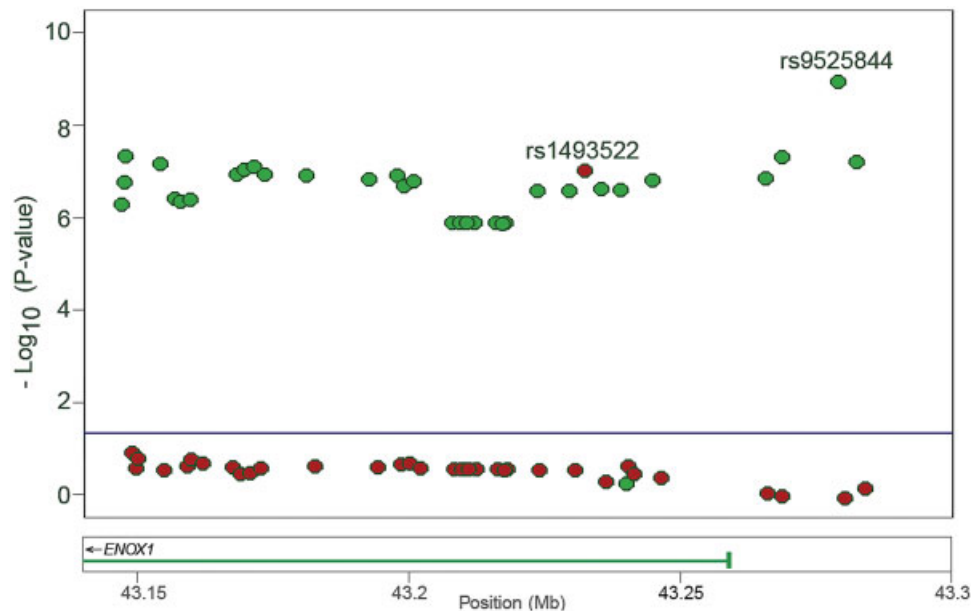
Figure 2. Detailed LocusZoom plot of the chromosome 13q14.11 region. For single-nucleotide polymorphisms (SNPs) above the purple line, the red data point represents the genetic association prior to conditional analysis, and the green data points represent the genetic association after conditioning on SNP rs1493522.

second independent effect could not be identified due to high LD. The conditional association is represented by rs9525844 ($P = 5.48 \times 10^{-9}$).

Finally, the most robust association signals in our study were at 6p21.32, the HLA class II region (Table 1). A detailed visual inspection of this region revealed 2 main subregions of associated markers driven by SNPs rs2395178 (between *BTNL2* and *DRA)*, rs9275224 (between *DQB1* and *DQA2*), and rs3998158 (between *DQB1* and *DQA2*) (Figure 3; see also Supplementary Figure 5A, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract). We selected a region that covered approximately the whole span of the HLA class II region (roughly from position 32.5 Mb to position 32.8 Mb of the human reference genome assembly hg18).

To determine the independence of the associations detected in the HLA, we performed conditional analysis, conditioning on each of the most significant SNPs. Conditioning on rs2395178, located on the first subregion, showed that rs9275224 was the only SNP in the HLA region that remained significantly associated; conditioning on the other SNPs in the subregion showed similar results. Conditioning on rs9275224 showed that SNPs in the first subregion remained significantly associated (see Supplementary Table 4, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract). Stepwise conditional analysis for the other markers covering the HLA class II region produced similar results, indicating that the associations at these 2 regions are independent. However, conditioning on rs3998158 (see Supplementary Figure 5B and Supplementary Table 5, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract) revealed RA associations at 4 markers that did not previously show association. These associated markers (rs9394113, rs9357152, rs5000632, and rs9378125) were all intergenic between *DQB1* and *DQA2*. All other associations lost significance. Thus, our results suggested that there are several independent effects within the HLA class II region.

When we investigated only RA patients who were positive for anti-CCP antibodies, the genetic association at the HLA region became more significant by one order of magnitude (Table 1). In contrast, significance for the chromosome 1 and chromosome 13 regions was reduced similarly. Among the previously identified genes that we also observed to be associated, the association at *PTPN22* and *SPRED2* was improved when anti-CCP antibody–positive status was analyzed (Table 2).

Secondary association analysis using principal component 1 as covariate. Because of the increased prevalence of RA in Amerindian populations, we assessed
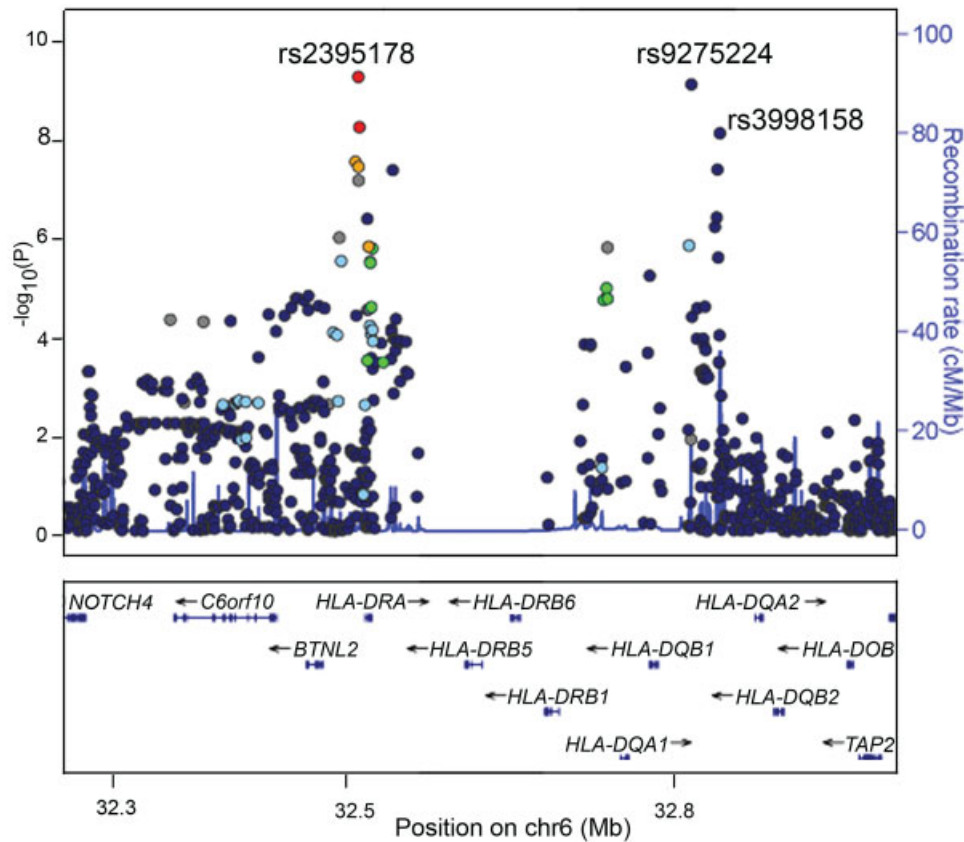
Figure 3. LocusZoom plot showing the region in chromosome 6p from 32.0 Mb to 33.4 Mb and the major hits (rs2395178 and rs9275224) within the HLA region prior to conditional analysis.

the contributions of Amerindian ancestry to the genetic associations in our cohort. First, we determined whether Structure and each of the top PCAs assigned similar ancestries to individuals. The 2 main ancestries (Amerindian and European) were almost completely dependent when inversely correlated (b = −1.007, R^2 = 0.9752, F test $P < 2.2 \times 10^{-16}$). We performed linear regression analysis to compare the Amerindian ancestry estimated by Structure and eigenvectors obtained from the PCAs. Amerindian ancestry correlated with principal component 1 (b = −1.11, R^2 = 0.978, F test $P < 2.2 \times 10^{-16}$). These results confirmed that the PCA and Structure analysis provided equivalent results in detecting the main axis of genetic variation in the studied individuals. Therefore, in the secondary association analysis, we used principal component 1 as a covariate in the additive analysis and applied a GC correction afterward. Identical association results were obtained by adjusting for Amerindian ancestry using Structure (data not shown). This analysis maintained the genetic association

Table 2. Previously known rheumatoid arthritis–associated loci that show association in Hispanic populations*

SNP	Chr	Gene	P†	P‡
rs10752747	1	MMEL1	6.61×10^{-5}	0.0019
rs2985859	1	MMEL1	7.92×10^{-5}	0.0015
rs1217421	1	PTPN22	0.003	0.0001
rs12039973	1	ATP6V1G3/PTPRC	0.003	NS
rs871974	2	SPRED2	0.0005	2.52×10^{-5}
rs3024921	2	STAT4	0.0008	NS
rs932169	2	STAT4	0.001	0.002
rs7579215	2	RAPH1/CD28	0.005	NS
rs57448334	4	IL2/IL21	0.0008	NS
rs6887783	5	PTTG1-ATP10B	4.74×10^{-5}	0.003
rs9373821	6	PREP/PRDM1	0.0008	NS
rs35000415	7	IRF5	4.56×10^{-5}	0.0002
rs2070197	7	IRF5	5.76×10^{-5}	0.0002
rs10814136	9	CCL21	0.001	NS
rs3118471	10	IL2RA	0.004	NS
rs17208015	18	CD226	0.003	NS

* SNP = single-nucleotide polymorphism; Chr = chromosome; NS = not significant.
† Corrected for genomic controls.
‡ Obtained when stratifying for individuals positive for anti–cyclic citrullinated peptide antibodies, after correction for genomic control.
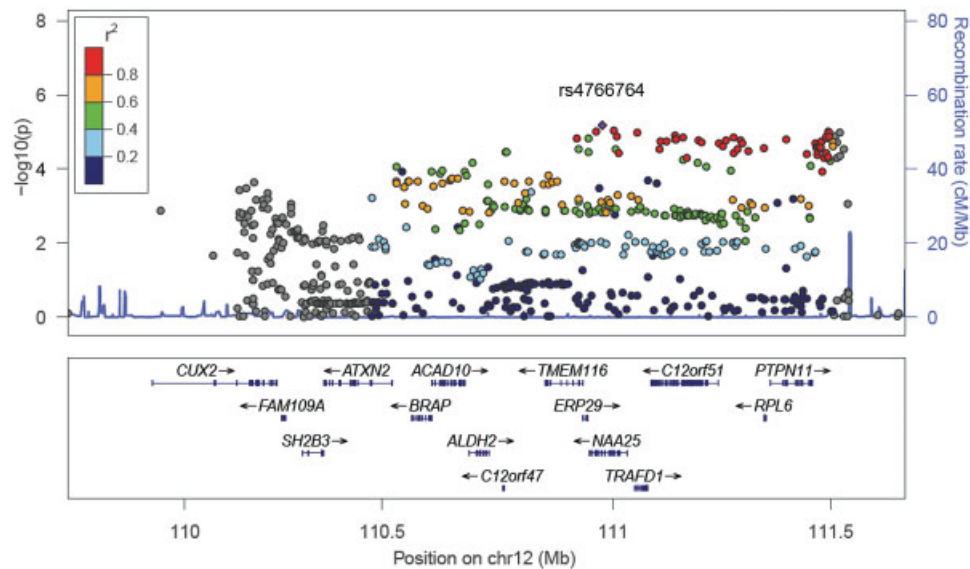
Figure 4. Detailed LocusZoom plot of associations detected at 12q24.3 when using primary component 1 as a covariate in the logistic regression analyses. The colors represent the r^2 between single-nucleotide polymorphisms. The diamond symbol represents the best hit.

within the HLA region, i.e., SNP rs2395178 remained as the best hit ($P = 3.6 \times 10^{-9}$) (see Supplementary Table 3, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/ abstract), and for several other SNPs within *BTNL2/ HLA−DRA*, but the P values were one order of magnitude lower.

Outside the major histocompatibility complex, our principal component 1–adjusted analysis revealed a new genetic association with a SNP located in an intron of *C12orf30* on chromosome 12q24.13 (rs4766764; $P = 3.9 \times 10^{-6}$). This gene, also called *NAA25*, encodes the *N*-acetyltransferase 25, NatB auxiliary unit, a protein that is thought to be important in cell cycle progression and is part of the *N*-acetyltransferase complex that performs cotranslational N-terminal acetylation of methionine residues that are followed by an asparagine. Several additional SNPs in this region, probably representing a tight LD block, were also newly associated with RA in the secondary association analyses, expanding the associated region at 12q24.13 to include *NAA25-PTPN11-RPH3A-TRAFD1* (see Supplementary Table 3 and Figure 4). The region lies close to a *SH2B3*, a gene that was previously associated with RA in Europeans (29). Although detailed fine mapping will be required to determine the strongest genetic associations within this locus, these results suggest that Amerindian ancestry contributes to the genetic risk of RA in Latin Americans primarily through the *HLA* and *NAA25* regions.

DISCUSSION

Amerindian groups in the US have an increased prevalence of RA, along with more severe disease and joint destruction (3,6,7,30–32). Although progress has been made in identifying the genes that influence RA susceptibility, little is known about genetic susceptibility in Hispanic populations in the Americas with a large admixture of Amerindian ancestry. This is largely attributable to the difficulty of performing association studies in admixed populations and the fact that the power required to identify genetic associations in admixed populations is larger than that required for more homogeneous populations. Here, we report the first genetic analysis of RA in Latin American populations, using the largest Latin American RA cohort described to date and carefully selected healthy control subjects. Our analysis of >114,000 SNPs, covering all genes associated with autoimmune diseases as of December 2010 (the so-called Immunochip), identified 2 regions that have previously been associated with RA and revealed 2 novel genetic associations.

The strongest association observed was in the HLA class II region; this association has been previously reported in RA association studies in Amerindian and admixed Hispanic populations. A genetic association of RA with the shared epitope (SE) alleles of *HLA–DRB1* was previously identified in individuals of Native American as well as Mexican American ancestry, and

DRB1*0901 was independently associated with the presence of anti-CCP antibodies (31,33). In addition, in a relatively small study of Peruvian patients with RA, classic genotyping for HLA showed that among the DRB1* alleles with the SE sequence, only *DRB1*1402* (located within a haplotype including *DQB1*0301* and *DQA1*0401*) was significantly more common in patients versus control subjects, and that the TNFα6 allele of the tumor necrosis factor α microsatellite was more common in patients, regardless of whether they had this haplotype (34). In a second, larger study in the Mexican Mestizo population, in which the proportion of individuals with European ancestry is larger than that in the Peruvian population (35), associations with several *DRB1* SE alleles were identified (36). In the primary genetic association analysis in our study, the strongest associations were located outside of the *DRB1* gene but in close proximity (15–20 Kb); the closest SNP was rs34102154 (GC-corrected $P = 1.15 \times 10^{-7}$).

Interestingly, we did identify a strong and independent association within the *DQB1–DQA2* region when performing conditional analysis with SNP rs3998158. Our results suggest that the genetic architecture of the HLA region in these populations is very complex and includes several independent genetic effects within the HLA class II locus, but the ancestry adjustment did not reveal new associations. It is possible that the observed complexity of RA associations within the HLA region reflects the diverse origin of our sample population, and more detailed RA association studies of this unique genetic region, including classic HLA typing, are warranted.

Our results also demonstrated an association at the 1p36.32 locus covering *PLCH2, PANK4, HES5, TNFRSF14, C1orf93*, and *MMEL1*. RA association at this locus was also described in a recent study in a Native Canadian cohort (2) and in earlier studies in Asian and European cohorts (11,37). Although these studies referred to the region as *TNFRSF14-MMEL1* because *TNFRSF14* appeared to be the best candidate in the region, the precise gene responsible for the association at this locus is unknown. In our study, the second strongest association at this locus was a nonsynonymous substitution (rs2296442) in *PLCH2*. This gene encodes phospholipase Cη2, a member of the PLCη family of the phosphoinositide-specific PLC superfamily of enzymes that cleave phosphatidylinositol 4,5-diphosphate to generate the second messengers inositol 1,4,5-trisphosphate and diacylglycerol (38). During cell signaling, these molecules play a key role in promoting calcium mobilization and the activation of effector functions within

neuronal cells, but very little is known about the function of *PLCH2* in immune cells.

A single SNP (rs1493522) in chromosomal region 13q14.11 showed a suggestive association with RA in the initial analysis of our cohort. However, after conditioning for rs1493522, several SNPs within a very tight LD block showed strong association. The first SNP is located within the first intron of *ENOX1*, which encodes ecto-NOX disulfide-thiol exchanger 1. Members of the ecto-NOX family exhibit both hydroquinone (NADH) oxidase and protein disulfide-thiol exchange activities and are involved in plasma membrane electron transport pathways that can influence cellular functions as diverse as cellular defense, intracellular redox homeostasis, and control of cell growth and survival (39). The other SNPs that were significant after conditioning are located in the distant promoter region of *ENOX1*. Thus, our results suggest 2 genetic effects within this gene.

Although our study is the first to implicate *ENOX1* in RA susceptibility, a recent genome-wide association study in patients with RA showed an association between *ENOX1* and patient responses to tocilizumab, a humanized monoclonal antibody targeting the interleukin-6 (IL-6) receptor (40). Therefore, *ENOX1* may participate in an IL-6–related pathway. Furthermore, a region of shared homozygosity including a single recessive variant in *ENOX1* was identified in 4 siblings with adult-onset autoimmune myasthenia gravis (41).

When the genetic association was tested in a group of individuals positive for anti-CCP antibodies, the genetic association improved for the HLA region, particularly for SNP rs9275224 within *DQB1/DQA2*, and at the non-HLA loci *SPRED2* and *PTPN22*, in agreement with previous studies (42,43). Other loci, such as *TAGAP* and *BACH2*, were observed to be associated in the anti-CCP antibody–positive set but not in the overall set (see Supplementary Table 2, available on the *Arthritis & Rheumatism* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.37923/abstract).

It is well established that the prevalence of RA is increased in Amerindian populations, but the genetic basis for this observation has not been revealed. We addressed this issue by performing association analysis with adjustment for Amerindian ancestry. This analysis revealed that several SNPs in a region showing tight LD in chromosome 12q24.3 were significant at a threshold of 4×10^{-6}. These SNPs surround *C12orf30/NAA25* and are close to other genes, including *SH2B3*, which was previously associated with RA. This same region has been associated with type 1 diabetes mellitus in Estonians (44) and most interestingly, with juvenile arthritis

(45), which was an exclusion criterion for our study. Therefore, it will be interesting to test whether, in our sample, this locus could be associated with early age at onset of RA (45,46). It should be noted that several of the results require replication.

The results of this large-scale association study provide new insights into the genetic basis of RA in Latin American individuals with enriched Amerindian ancestry and identified risk loci associated with Amerindian ancestry, a poorly studied but important risk factor in RA. In addition, our analysis in a large cohort with heterogeneous, admixed ancestry provides new insight into the complex genetic architecture of RA. Therefore, these studies address longstanding questions regarding the genetics of RA and provide an impetus for future studies to define how these loci impact the clinical manifestations and severity of the disease.

ACKNOWLEDGMENTS

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Alarcón-Riquelme 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 conception and design. Alarcón-Riquelme.
Acquisition of data. Herráez, Riba, García de la Torre, Sacnún, Goñi, Berbotto, Paira, Musuruana, Graf, Alvarellos, Messina, Babini, Strusberg, Marcos, Scherbarth, Spindler, Quinteros, Toloza, Moreno, Catoggio, Tate, Eimon, Citera, Pellet, Nasswetter, Cardiel, Miranda, Ballesteros, Esquivel-Valerio, Maradiaga-Ceceña, Acevedo-Vásquez, García, Tusié-Luna, Pons-Estel, Alarcón-Riquelme.
Analysis and interpretation of data. Herráez, Martínez-Bueno, Alarcón-Riquelme.

REFERENCES

1. Molokhia M, McKeigue P. Risk for rheumatic disease in relation to ethnicity and admixture. Arthritis Res 2000;2:115–25.
2. El-Gabalawy HS, Robinson DB, Daha NA, Oen KG, Smolik I, Elias B, et al. Non-HLA genes modulate the risk of rheumatoid arthritis associated with HLA-DRB1 in a susceptible North American Native population. Genes Immun 2011;12:568–74.
3. El-Gabalawy HS, Robinson DB, Smolik I, Hart D, Elias B, Wong K, et al. Familial clustering of the serum cytokine profile in the relatives of rheumatoid arthritis patients. Arthritis Rheum 2012; 64:1720–9.
4. Pelaez-Ballestas I, Sanin LH, Moreno-Montoya J, Alvarez-Nemegyei J, Burgos-Vargas R, Garza-Elizondo M, et al. Epidemiology of the rheumatic diseases in Mexico: a study of 5 regions based on the COPCORD methodology. J Rheumatol Suppl 2011;86:3–8.
5. Wang S, Ray N, Rojas W, Parra MV, Bedoya G, Gallo C, et al. Geographic patterns of genome admixture in Latin American Mestizos. PLoS Genet 2008;4:e1000037.
6. Del Puente A, Knowler WC, Pettitt DJ, Bennett PH. High incidence and prevalence of rheumatoid arthritis in Pima Indians. Am J Epidemiol 1989;129:1170–8.
7. Jacobsson LT, Knowler WC, Pillemer S, Hanson RL, Pettitt DJ, Nelson RG, et al. Rheumatoid arthritis and mortality: a longitudinal study in Pima Indians. Arthritis Rheum 1993;36:1045–53.
8. Thomson W, Barton A, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis association at 6q23. Nat Genet 2007;39: 1431–3.
9. Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J, et al. Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. Nat Genet 2008;40:1156–9.
10. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet 2009;41:1313–8.
11. Kurreeman FA, Stahl EA, Okada Y, Liao K, Diogo D, Raychaudhuri S, et al. Use of a multiethnic approach to identify rheumatoid-arthritis-susceptibility loci, 1p36 and 17q12. Am J Hum Genet 2012;90:524–32.
12. 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.
13. Plenge RM, Padyukov L, Remmers EF, Purcell S, Lee AT, Karlson EW, et al. Replication of putative candidate-gene associations with rheumatoid arthritis in >4,000 samples from North America and Sweden: association of susceptibility with PTPN22, CTLA4, and PADI4. Am J Hum Genet 2005;77:1044–60.
14. 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.
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. 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.
17. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, Burtt NP, et al. Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 2008;40:1216–23.
18. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, Kastner DL, et al. REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 2009;41:820–3.
19. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther 2011;13:101.
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. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a toolset for whole-genome association and population-based linkage analysis. Am J Hum Genet 2007;81: 559–75.
22. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155: 945–59.
23. Thornton T, Tang H, Hoffmann TJ, Ochs-Balcom HM, Caan BJ, Risch N. Estimating kinship in admixed populations. Am J Hum Genet 2012;91:122–38.
24. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. PLoS Genet 2006;2:e190.

25. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. Bioinformatics 2010;26:2336–7.

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

27. Lopez Herraez D, Bauchet M, Tang K, Theunert C, Pugach I, Li J, et al. Genetic variation and recent positive selection in world-wide human populations: evidence from nearly 1 million SNPs. PLoS One 2009;4:e7888.

28. Li Q, Yu K. Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. Genet Epidemiol 2008;32:215–26.

29. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet 2011;7:e1002004.

30. Hernandez-Cruz B, Cardiel MH, Villa AR, Alcocer-Varela J. Development, recurrence, and severity of infections in Mexican patients with rheumatoid arthritis: a nested case-control study. J Rheumatol 1998;25:1900–7.

31. Del Rincon I, Escalante A. HLA–DRB1 alleles associated with susceptibility or resistance to rheumatoid arthritis, articular deformities, and disability in Mexican Americans. Arthritis Rheum 1999;42:1329–38.

32. Hirsch R, Lin JP, Scott WW Jr, Ma LD, Pillemer SR, Kastner DL, et al. Rheumatoid arthritis in the Pima Indians: the intersection of epidemiologic, demographic, and genealogic data. Arthritis Rheum 1998;41:1464–9.

33. El-Gabalawy HS, Robinson DB, Hart D, Elias B, Markland J, Peschken CA, et al. Immunogenetic risks of anti-cyclical citrullinated peptide antibodies in a North American Native population with rheumatoid arthritis and their first-degree relatives. J Rheumatol 2009;36:1130–5.

34. Castro F, Acevedo E, Wollheim FA, Sandberg-Wollheim M. Tumor necrosis factor microsatellites and HLA-DRB1*, HLA-DQA1*, and HLA-DQB1* alleles in Peruvian patients with rheumatoid arthritis. Ann Rheum Dis 2001;60:791–5.

35. Sanchez E, Webb RD, Rasmussen A, Kelly JA, Riba L, Kaufman KM, et al. Genetically determined Amerindian ancestry correlates with increased frequency of risk alleles for systemic lupus erythematosus. Arthritis Rheum 2010;62:3722–9.

36. Ruiz-Morales JA, Vargas-Alarcon G, Flores-Villanueva PO, Villarreal-Garza C, Hernandez-Pacheco G, Yamamoto-Furusho JK, et al. HLA-DRB1 alleles encoding the "shared epitope" are associated with susceptibility to developing rheumatoid arthritis whereas HLA-DRB1 alleles encoding an aspartic acid at position 70 of the beta-chain are protective in Mexican Mestizos. Hum Immunol 2004;65:262–9.

37. Danoy P, Wei M, Johanna H, Jiang L, He D, Sun L, et al. Association of variants in MMEL1 and CTLA4 with rheumatoid arthritis in the Han Chinese population. Ann Rheum Dis 2011;70: 1793–7.

38. Zhou Y, Wing MR, Sondek J, Harden TK. Molecular cloning and characterization of PLC-eta2. Biochem J 2005;391(Pt 3):667–76.

39. Scarlett DJ, Herst PM, Berridge MV. Multiple proteins with single activities or a single protein with multiple activities: the conundrum of cell surface NADH oxidoreductases. Biochim Biophys Acta 2005;1708:108–19.

40. Wang J, Bansal AT, Martin M, Germer S, Benayed R, Essioux L, et al. Genome-wide association analysis implicates the involvement of eight loci with response to tocilizumab for the treatment of rheumatoid arthritis. Pharmacogenomics J 2012. E-pub ahead of print.

41. Landoure G, Knight MA, Stanescu H, Taye AA, Shi Y, Diallo O, et al. A candidate gene for autoimmune myasthenia gravis. Neurology 2012;79:342–7.

42. Eyre S, Bowes J, Diogo D, Lee A, Barton A, Martin P, et al. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat Genet 2012;44:1336–40.

43. Morgan AW, Thomson W, Martin SG, Carter AM, Erlich HA, Barton A, et al. Reevaluation of the interaction between HLA–DRB1 shared epitope alleles, PTPN22, and smoking in determining susceptibility to autoantibody-positive and autoantibody-negative rheumatoid arthritis in a large UK Caucasian population. Arthritis Rheum 2009;60:2565–76.

44. Douroudis K, Kisand K, Nemvalts V, Rajasalu T, Uibo R. Allelic variants in the PHTF1-PTPN22, C12orf30 and CD226 regions as candidate susceptibility factors for the type 1 diabetes in the Estonian population. BMC Med Genet 2010;11:11.

45. Prahalad S, Hansen S, Whiting A, Guthery SL, Clifford B, McNally B, et al. Variants in TNFAIP3, STAT4, and C12orf30 loci associated with multiple autoimmune diseases are also associated with juvenile idiopathic arthritis. Arthritis Rheum 2009;60: 2124–30.

46. Thompson SD, Sudman M, Ramos PS, Marion MC, Ryan M, Tsoras M, et al. The susceptibility loci juvenile idiopathic arthritis shares with other autoimmune diseases extend to PTPN2, COG6, and ANGPT1. Arthritis Rheum 2010;62:3265–76.

APPENDIX A: MEMBERS OF THE GENÓMICA DE ARTRITIS REUMATOIDE COLLABORATION

Members of the Genómica de Artritis Reumatoide Collaboration who contributed samples and clinical data are as follows: Carlos Meineri (Instituto Lucha Antipoliomelítica de Rosario, Santa Fé, Argentina); Susana Roverano (Hospital José María Cullen, Santa Fé, Argentina); Donna L. González (Hospital J. B. Iturraspe, Santa Fé, Argentina); Jorge Chemez (Hospital San Martín, Paraná, Argentina); Veronica Saurit, Francisco Caeiro (Hospital Privado de Córdoba, Córdoba, Argentina); Norma G. Vila, Maria C. De la Vega (Hospital Cosme Argerich, Buenos Aires, Argentina); Paula B. Alba (Hospital Italiano de Córdoba, Córdoba, Argentina); Ana Bertoli, Mariela Ramos (Instituto Reumatológico Strusberg, Córdoba, Argentina); Cecilia Musumeci, Roberto Haag, Amelia Granel (Hospital Interzonal General de Agudos San Martín, La Plata, Argentina); Jorge A. López, Alicia Testa (Hospital Interzonal General de Agudos Oscar E. Alende, Mar del Plata, Argentina); Alberto Berman (Centro Medico Privado de Reumatología, Tucumán, Argentina); Dora Lía Vázquez (Fundación Instituto para la Promoción de la Salud y Educación, San Miguel de Tucumán, Argentina); Santiago E. Agüero, Sergio Orellana (Hospital Interzonal San Juan Bautista, Catamarca, Argentina); Rodolfo Pardo Hidalgo (Hospital G Rawson, San Juan, Argentina); Enrique Soriano, Cecilia Castel, Lorena Narváez (Hospital Italiano de Buenos Aires, Buenos Aires, Argentina); Eduardo Mysler (Organización Médica Integral, Buenos Aires, Argentina); Roberto M. Arana (Centro de Educación Médica e Investigaciones Clínicas, Buenos Aires, Argentina); José Maldonado Cocco (Instituto de Rehabilitación Psicofísica, Ciudad Autónoma de Buenos Aires, Argentina); Antonio Mónaco, Mariel Cutri (Hospital Bernardino Rivadavia, Ciudad Autónoma de Buenos Aires, Argentina); María del Rosario Maliandi (Hospital de Clínicas de Buenos Aires, Buenos Aires, Argentina); Jorge Mariano Cucho (Hospital Nacional Guillermo Almenara Irigoyen, Lima, Peru); Cecilia Trejo, Jorge Saavedra (Hospital San Juan de Dios, Santiago de Chile, Chile); Carmen Medina (Reumatología Hospital San Borja-Arriarán, Santiago de Chile, Chile); Marco González (Unidad de Reumatología Hospital General de México Eduardo Liceaga, Mexico City, Mexico); Jacqueline Rodriguez Amado, Mario A. Garza-Elizondo (Hospital Universitario de Nuevo León, Nuevo León, Monterrey, Mexico); Josefina Munguía-Rojas (Centro de Estudios de Investigación Básica y Clínica, S.C., Guadalajara, Mexico).