Association analysis in a Latin American population revealed ethnic differences in rheumatoid arthritis-associated SNPs in Caucasian and Asian populations

P. Castro-Santos1,2, R. A. Verdugo3,4, R. Alonso-Arias3,5, M. A. Gutiérrez6,7, J. Suazo8, J. C. Aguilón9, J. Olloquequi9, C. Pinochet10, A. Lucía11, L. A. Quiñones12,13,14✉ & R. Díaz-Peña1,14✉

Large genome-wide association studies (GWAS) have increased our knowledge of the genetic risk factors of rheumatoid arthritis (RA). However, little is known about genetic susceptibility in populations with a large admixture of Amerindian ancestry. The aim of the present study was to test the generalizability of previously reported RA loci in a Latin American (LA) population with admixed ancestry. We selected 128 single nucleotide polymorphisms (SNPs) in linkage equilibrium, with high association to RA in multiple populations of non-Amerindian origin. Genotyping of 118 SNPs was performed in 313 RA patients/487 healthy control subjects by mid-density arrays of polymerase chain reaction (PCR). Some of the identified associations were validated in an additional cohort (250 cases/290 controls). One marker, the SNP rs2451258, located upstream of T Cell Activation RhoGTPase Activating Protein (TAGAP) gene, showed significant association with RA ($p = 5 \times 10^{-3}$), whereas 18 markers exhibited suggestive associations ($p < 0.05$). Haplotype testing showed association of some groups of adjacent SNPs around the signal transducer and activator of transcription 4 (STAT4) gene ($p = 9.82 \times 10^{-3}$ to $2.04 \times 10^{-3}$) with RA. Our major finding was little replication of previously reported genetic associations with RA. These results suggest that performing GWAS and admixture mapping in LA populations has the potential to reveal novel loci associated with RA. This in turn might help to gain insight into the ‘pathogenomics’ of this disease and to explore trans-population differences for RA in general.

Rheumatoid arthritis (RA) is an autoimmune inflammatory rheumatic disease that affects mainly synovial joints among many tissues and organs. It affects approximately 1% of the population worldwide1 and, although this

1Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Talca, Chile. 2Inmunología, Centro de Investigaciones Biomédicas (CINBIO), Universidad de Vigo, Vigo, Spain. 3Programa de Genética Humana, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile. 4Departamento de Oncología Básico Clínica, Facultad de Medicina, Universidad de Chile, Santiago, Chile. 5Immunology Department, Hospital Universitario Central de Asturias, Oviedo, Spain. 6Rheumatology, Aimértef Naval Hospital, Viña del Mar, Valparaíso, Chile. 7Valparaíso University, Viña del Mar, Valparaíso, Chile. 8Instituto de Investigación en Ciencias Odontológicas, Facultad de Odontología, Universidad de Chile, Santiago, Chile. 9Immune Regulation and Tolerance Research Group, Programa de Inmunología, ICBM, Facultad de Medicina, Universidad de Chile, Santiago, Chile. 10Hospital Regional de Talca, Talca, Chile. 11Universidad Europea de Madrid (Faculty of Sports Sciences) and Research Institute Hospital 12 de Octubre (‘i + 12’), Madrid, Spain. 12Laboratory of Chemical Carcinogenesis and Pharmacogenetics, Department of Basic-Clinical Oncology, Faculty of Medicine, University of Chile, Santiago, Chile. 13Latin American Network for Implementation and Validation of Clinical Pharmacogenomics Guidelines (RELIVAF-CYTED), Madrid, Spain. 14These authors contributed equally: L. A. Quiñones and R. Díaz-Peña. ✉E-mail: lquinone@med.uchile.cl; roberto.diaz@uautonoma.cl
condition can develop at any age, RA affects women more frequently than men and is mainly diagnosed between the ages of 40–60 years. In Latin America (LA), differences towards women seem to be higher, whereas prevalence has been estimated between 0.2–0.5%²⁻³. In Chile, there are data showing that the overall prevalence of RA based on clinical examination is 0.46%⁶.

The etiology of RA is multifactorial and partially unknown because of the complex interactions between genetic and environmental factors. Approximately 50% of RA risk is thought to be genetic, and one-third of this risk is associated with the human leukocyte antigen (HLA) locus⁷, specifically HLA-DRB1 shared alleles (SE), which encode a common amino acid sequence⁸. Since 2007 about 101 RA risk loci have emerged from genome-wide association studies (GWAS) and subsequent GWAS meta-analyses⁹, mostly in individuals from European and/or Asian populations (Supplementary Table 1). In fact, none of the GWAS pertaining to RA has been performed in LA populations (Supplementary Table 1).

It is generally accepted that many common risk variants are shared between multiethnic populations, but allele frequencies of disease-associated single nucleotide polymorphisms (SNPs) vary significantly among ethnic groups due to genetic drift or selection⁹. Linkage between causal variants and tag SNPs included in genotyping microarrays might vary depending on population-specific pattern of recombination which in turn, is largely affected by population size, founder effects and admixture processes. In addition, populations with different histories may carry distinct causal mutations even in similar loci. All of these factors can preclude generalization of genetic associations from one population to another, and suggest testing for locus- or haplotype-wise rather than SNP-wise generalization¹⁰.

López Herráez et al.¹¹ examined susceptibility loci for RA in LA populations. In this study, a strong association with HLA region was observed with three independent effects, probably due to the diverse origin of the patients (Argentina, Mexico, Chile, and Peru). Some of the RA associations previously reported in GWAS were also replicated in the study by López Herráez and coworkers, but with moderate significant values (including protein tyrosine phosphatase, non-receptor type 22 (PTPN22) and signal transducer and activator of transcription 4 (STAT4) genes). However, in general, genetic association studies on RA have not been robustly replicated in LA populations. Therefore, the aim of the present study was to carry out a high-density SNP genotyping in candidate genes to test their association with susceptibility to RA in the Chilean population, in order to provide insight on the cross-ethnic generalizability of known European and Asian RA risk loci to LA populations.

**Results**

In the present study, five hundred and sixty-three (42.0%) of the included individuals suffered RA. Supplementary Table 2 shows the characteristics of the RA patients that were used for the analysis. The mean age was 48 and 58 years for cohort 1 and 2, respectively, and 84.7% and 81.0% of the patients were women. The mean duration of the disease was 8 years. Anti-cyclic citrullinated peptide (CCP) antibodies were detected in a total of 218 patients being positive in 164 of them (75.23%), whereas rheumatoid factor (RF) was determined in 300 patients being positive in 264 (88.0%). The RA group did not differ from the control group with regard to any of the clinical parameters included in the study (data not shown).

The present findings do not show replicative association of individual SNPs with RA. Among 128 SNPs genotyped, 118 passed all the quality filters, after excluding SNPs with a minor allele frequency <0.01 or missingness >0.1 and those that were not in Hardy-Weinberg equilibrium (HWE) (p < 0.001) (Supplementary Table 3).

Only two markers (2%) showed significant associations (p < 0.01): rs1635567 and rs2469434 (Table 1), none of which none was confirmed in Cohort 2. When data from both cohorts were combined, rs2469434 was still significant whereas rs1635567 could not be tested because the assay failed in Cohort 2. However, the combined analysis revealed a new significant association for rs2451258 (combined p = 5 × 10⁻³; p = 0.09 after Bonferroni correction for multiple testing) (Table 1). Eighteen markers exhibited suggestive associations (p < 0.05), whereas the associations of the remainder of SNPs included in the study were not significant. The significantly-associated SNPs in our study (healthy controls vs. RA cases) were rs1635567, rs2469434, rs2451258, rs2490230, rs2490060, and rs2450506. For rs2469434 the p-value was < 0.01 and there was no association with the Chilean Genomico dataset (healthy controls vs. Chilean Genomico, p-value < 10⁻¹² and r = 0.97). However, the allele frequency in European, East Asian, Aymara and Mapuche samples showed variation compared to our cohort (r ≤ 0.70).

The sliding window test revealed several SNP blocks that were associated with RA (Table 2). The p values for the strongest sliding window (ranging from p = 9.82 × 10⁻³ to 2.04 × 10⁻²) were associated with regions around STAT4 gene. In addition to the sliding window test, we also performed case-control studies based on linkage disequilibrium (LD) haplotype block reconstruction, not revealing associations between SNPs and RA. Detailed haplotype block information and the LD plot around the STAT4 gene are shown in Supplementary Fig. 3.

**Discussion**

The present study aimed to investigate the association of SNPs markers in candidate genes and RA in the Chilean population. Our main finding was a little replication of previously reported genetic associations with RA. Indeed, only 2% of know RA loci from GWAS studies in populations of European or Asian origin were significantly associated in our LA population, and just 11% showed a suggestive association. This was unexpected because
| CHR | SNP            | POSITION | Gene | Cohort 1 | Cohort 2 | Joint analysis |
|-----|----------------|----------|------|----------|----------|----------------|
|     |                |          |      | MAF       | MAF       | MAF        |
| 1   | rs1635567       | 17683041 | PADI4| 0.33      | 0.015     | 0.02        |
| 2   | rs2469434       | 67544046 | CD228| 0.32      | 0.26      | 0.09        |
| 4   | rs2477134       | 17633572 | PADI4| 0.37      | 0.31      | 0.16        |
| 7   | rs3778753       | 128580042| IRF5 | 0.34      | 0.28      | 0.09        |
| 18  | rs9979383       | 36715761 | RUNX1| 0.31      | 0.24      | 0.01        |
| 11  | rs4409785       | 95311422 | CEBP5| 0.16      | 0.12      | 0.09        |
| 2   | rs3024903       | 19189560 | STAT4| 0.11      | 0.01      | 0.02        |
| 22  | rs3218251       | 37545505 | IL2RB| 0.18      | 0.17      | 0.03        |
| 6   | rs2451258       | 15950660 | TAGAP| 0.15      | 0.17      | 0.02        |
| 20  | rs4239762       | 44749251 | CD40 | 0.22      | 0.18      | 0.04        |
| 8   | rs998731        | 81095395 | TDP52| 0.36      | 0.31      | 0.01        |
| 10  | rs2275806       | 8095340 | GATA3| 0.38      | 0.43      | 0.09        |
| 2   | rs2843401       | 2528133  | MMELI| 0.41      | 0.46      | 0.09        |
| 4   | rs2664035       | 48220839 | TBC  | 0.39      | 0.44      | 0.06        |
| 2   | rs10175798      | 30449594 | LBB  | 0.33      | 0.37      | 0.03        |
| 11  | rs73013527      | 124896952| ETS1 | 0.36      | 0.41      | 0.04        |
| 2   | rs11889341      | 19194372 | STAT4| 0.2        | 0.34      | 0.34        |
| 14  | rs1950897       | 68760141 | RAD51B| 0.21      | 0.25      | 0.07        |
| 16  | rs4780401       | 11839326 | TXNDC1| 0.4       | 0.38      | 0.01        |
| 2   | rs11571302      | 204742934| CTLA4| 0.52      | 0.47      | 0.06        |
| 20  | rs6032662       | 44734310 | CD40 | 0.21      | 0.18      | 0.01        |
| 1   | rs2228145       | 15426970 | IL6R | 0.18      | 0.17      | 0.01        |
| 17  | rs12936409      | 38043649 | IRF5 | 0.46      | 0.42      | 0.04        |

Table 1. Association analysis of the replicated SNPs as single markers in cohort 1 and cohort 2 and the joint analysis (only SNPs with P ≤ 0.1 are shown). CHR = chromosome; SNP = single-nucleotide polymorphism; A1 = minor allele nucleotide; RA = rheumatoid arthritis; and MAF = minor allele frequency. *Failed in the replication phase (genotyping of the cohort 2).

| Gene   | Markers | P value |
|--------|---------|---------|
|        | rs3024903,rs3024896,rs4853540,rs16833220 | 3.94 × 10⁻³ |
|        | rs11889341,rs12990918,rs4643445,rs10931480 | 9.82 × 10⁻³ |
|        | rs3024903,rs3024896,rs4853540,rs16833220,rs11893432 | 6.26 × 10⁻³ |
|        | rs3024903,rs3024896,rs4853540,rs16833220,rs11893432,rs3024861 | 5.99 × 10⁻³ |
|        | rs12990918,rs4643443,rs10931480,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546 | 8.48 × 10⁻³ |
|        | rs12990918,rs4643443,rs10931480,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070 | 3.34 × 10⁻³ |
|        | rs12990918,rs4643443,rs10931480,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070,rs13227969 | 2.04 × 10⁻³ |
|        | rs2459611,rs11893432,rs12990918,rs4643443,rs10931480,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070 | 9.1 × 10⁻³ |
|        | rs4853540,rs16833220,rs11893432,rs3024861,rs15197352,rs13246947,rs2459611,rs11889341,rs12990918,rs4643443,rs10931480,rs7574865,rs6752770,rs1511440 | 7.79 × 10⁻³ |
|        | rs4853540,rs16833220,rs11893432,rs3024861,rs15197352,rs13246947,rs2459611,rs11889341,rs12990918,rs4643443,rs10931480,rs11893432,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070,rs13227969 | 4.32 × 10⁻³ |
|        | rs4853540,rs16833220,rs11893432,rs3024861,rs15197352,rs13246947,rs2459611,rs11889341,rs12990918,rs4643443,rs10931480,rs11893432,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070,rs13227969 | 5.77 × 10⁻³ |
|        | rs4853540,rs16833220,rs11893432,rs3024861,rs15197352,rs13246947,rs2459611,rs11889341,rs12990918,rs4643443,rs10931480,rs11893432,rs7574865,rs6752770,rs1511440,rs11685878,rs4853546,rs7574070,rs13227969 | 5.88 × 10⁻³ |

Table 2. Association analyses of sliding windows of 4–19 single-nucleotide polymorphisms within STAT4 (only p values < 10⁻³ are shown), using the chi-square test in Plink software (15).

SNPs in well-known RA loci were tested, such as *PAD4*, *PTPN22*, *STAT4*, *CTLA4*, *TNFAIP3*, and *CCR6* -none of which replicated. There are a number of reasons why previously GWAS-significant findings might not replicate in independent cohorts, as reviewed by Kraft et al.15. The small sample size of our study may be responsible for the
modest number of SNPs that showed associations validated in our participants. Sample sizes larger than the one used here are needed to reach high confidence levels and strong statistical power. In this regard, the low prevalence of the disease restricted the number of patients that we were able to recruit for our study. A long-term effort to progressively collect numerous patients’ samples from biobanks might allow to perform more powered genetic studies and to test for generalizability of genetic associations. Similarly, we believe that the small sample size is a main reason for the lack of differences we found between endophenotypes. Our study did not reach statistical power for one-third of the SNPs analyzed, which might provide a possible explanation, at least in part, for the lack of replication of the results in the Chilean population. However, if lack of power was the only explanation, it is expected that, overall, the OR values would follow the same trend in Chilean patients as in other populations. However, ORs in Chile show absolutely no correlation with estimates from studies with Europeans and only a very week positive association with Asians (Fig. 1). This suggests that genetic divergence between populations at these loci may be one of the reasons of the lack generalization of SNP associations.

Differences in LD patterns between populations may preclude replication of association, which can be caused by multiple factors such as different demographic history including population-specific bottlenecks, genetic drift, selection, and recent admixture, among others. Large diversity in LD among populations from different continents, including the Americas, is well documented. Furthermore, RA is a trait associated with loci responsible for the immune response, which in turn is highly associated with local adaptations and disease resistance. In support for the above interpretation of our results, although we did not find any significant SNP-wise association of \( \text{STAT4} \) with RA, we did find association for this locus when testing haplotypes instead of genotypes. Using the sliding window test revealed several haplotype associations with RA, suggesting the possible existence of untested (potentially functional) genetic variation within \( \text{STAT4} \) in the Chilean population, a result that other studies with different populations might had failed to detect or might had not shown the strongest signal. Further investigations are required to confirm these findings. The strongest association was observed for the SNP rs2451258 located upstream of the T-cell activation RhoGTPase activating protein (\( \text{TAGAP} \)) gene, although the p-value was >0.05 after Bonferroni correction for multiple testing. This variant is not within any protein-coding sequence or disrupted a non-coding functional motif, but \( \text{TAGAP} \) would be a promising biological candidate gene. \( \text{TAGAP} \) gene encodes a member of the Rho GTPase-activator protein superfamily, but little is known about their role in the immune system. Additional investigations, with higher of variants in the region are required to confirm this hypothesis.

Polygenic risk scores could be the next great stride in genomic medicine, which is generating a considerable debate regarding their use in complex phenotypes. Recently, Khera et al. proposed that it is time to contemplate the incorporation of polygenic risk prediction in clinical care, projecting these scores across a wide variety of diseases. In the present study, significant values of the previously detected SNP-wise associations were moderate and a better generalizability was found when testing association between phenotype and haplotypes rather than SNPs. Moreover, allele frequency vary between populations of different ancestries. These results suggest the existence of genomic patterns in Chilean, and probably other LA populations, that differentiate them from Europeans with regard to \( \text{loci} \) that are relevant for RA. This can be caused by different demographic histories (e.g., past population bottlenecks and migration events, or ancestries). Haplotype-based associations may capture the interacting effects among two or more potential causal variants within certain genomic region, which single-variants in non-coding DNA.
approach cannot detect. Therefore, haplotype-based approaches show a greater power to map susceptibility genes in complex traits than single-marker methods. These results support the need for GWAS in LA populations, including Chileans, to discover potentially novel loci accounting for genetic risk for RA, to investigate the contribution of genetic ancestry, and to improve performance of polygenic prediction models in these populations.

**Methods**

**Study participants.** A total of 1,340 individuals were studied as two distinct cohorts. Cohort 1 comprised 313 patients with RA and 487 healthy control subjects; cohort 2 included 250 RA patients and 290 healthy controls. The patients with RA were diagnosed following the 2010 American College of Rheumatology/European League Against Rheumatism (ACR/EULAR) classification criteria. The study was approved by the Ethical Committee of the “Servicio de Salud del Maule” (registration number 04/2014), Chile; and all individuals gave their written informed consent prior to enrolling in the study. All methods were performed in accordance with the relevant guidelines and regulations.

**SNP selection and genotyping.** A total of 128 SNPs from 73 genes were chosen for genotyping from previous GWAS in populations of diverse ethnic background. Some of them were selected as haplotype-tag-SNPs (ht-SNPs) based on LD patterns located within our candidate genes (PAD4, PTPN22, STAT4, CTLA4, TNFAIP3, and CCR6) and using the HapMap dataset. Haplotype tagging (HT)-SNPs were selected using the Tagger tool of Haploview, under the following criteria: minor allele frequency ≥ 0.01 and r² > 0.8, and based on the HapMap populations (CEU, CEU + TSI and MEX). Some of the identified associations were validated by genotyping 23 SNPs in the cohort 2. The SNPs were genotyped using the OpenArray® TaqMan platform (Applied Biosystems Inc.) in the test (Cohort 1) and replica (Cohort 2) samples. The genotyping assays were performed at the Pfizer-University of Granada-Junta de Andalucía Centre for Genomics and Oncological Research (GENYO) (Cohort 1), Granada, Spain; and at the Centro Nacional de Genotipado (Cohort 2), at the Santiago de Compostela node, Spain.

**Genotyping data from reference populations.** In order to assess ethnic differences in allelic frequencies for the SNPs evaluated in this work, we obtained genotypes for 108 AFR, 99 EUR, and 103 EAS unrelated individuals from the 1000 Genomes Project Phase 3 dataset. For Amerindian ancestry, we obtained genotypes for 85 individuals of Aymara ancestry (AYM), 54 individuals of Mapuche ancestry (MAP), and 348 of Chilean ancestry (CLG) from the ChileGenomico Project. AYM, MAP, and CLG individuals were genotyped using the Axion LAT1 Array (Affymetrix, Inc., Santa Clara, California, U.S.) and imputed using the 1000 Genomes Project phase 3.
Statistical analysis.  Power calculations were done with the GAS Power Calculator tool (http://csg.sph.umich.edu) assuming a multiplicative model, with OR = 1.5, a significance level of 0.05 and an RA prevalence of 0.5%. Only SNPs that met the quality criteria of a minor allele frequency (MAF) > 0.01, missingness < 0.1, and HWE P > 0.01 were considered for inclusion in the association analyses (Supplementary Table 3). Allele frequencies were compared between RA patients and control populations by chi-square test, and OR with 95% confidence intervals (95% CI) were calculated using PLINK software (v1.07)27. Haplotype analysis was performed using Haplovie software (v4.2)25. In addition, haplotypes based on 1-bp sliding windows of 2 to 21 SNPs each were also constructed. Association analyses were done with the chi-square test using PLINK. Pearson’s correlations and linear regression were used to evaluate differences between genetic background. The LocusZoom web-based resource was used to generate plots of association results by genomic region28.

Received: 21 August 2018; Accepted: 8 April 2020;
Published online: 12 May 2020

References
1. Tobon, G. J., Youinou, P. & Serafini, A. The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. Journal of Autoimmunity 35, 10–14 (2010).
2. Spindler, A. et al. Prevalence of rheumatoid arthritis in Tucuman, Argentina. J. Rheumatol 29, 1166–1170 (2002).
3. Rodrigues Senna, É. et al. Prevalence of Rheumatic Diseases in Brazil: A Study using the COPCORD Approach. Journal of Rheumatology 31, 594–597 (2004).
4. Bennett, K. et al. Community screening for rheumatic disease: Cross cultural adaptation and screening characteristics of the COPCORD core questionnaire in Brazil, Chile, and Mexico. Journal of Rheumatology 24, 160–168 (1997).
5. MacGregor, A. J. et al. Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins. Arthritis & Rheumatism 43, 30–37 (2000).
6. Gregersen, P. K., Silver, J. & Winchester, R. J. The shared epitope hypothesis. an approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis. Arthritis & Rheumatism 30, 1205–1213 (1987).
7. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–381 (2013).
8. Terao, C., Raychaudhuri, S. & Gregersen, P. K. Recent Advances in Defining the Genetic Basis of Rheumatoid Arthritis. Annual Review of Genomics and Human Genetics 17, 273–301 (2016).
9. Yamamoto, K., Okada, Y., Suzuki, A. & Kochi, Y. Genetics of rheumatoid arthritis in Asia—present and future. Nature Reviews Rheumatology 11, 375–379 (2015).
10. Novembre, J. & Ramachandran, S. Perspectives on human population structure at the cusp of the sequencing era. Annu. Rev. Genomics Hum. Genet 12, 245–274 (2011).
11. Herrzie, D. L. et al. Rheumatoid arthritis in Latin Americans enriched for amiodarian ancestry is associated with loci in chromosomes 1, 12, and 13, and the HLA Class II region. Arthritis and Rheumatism 65, 1457–1467 (2013).
12. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376–381 (2013).
13. Kraft, P., Zeigler, E. & Ioannidis, J. Replication in genome-wide association studies. Stat. Sci. 24, 561–573 (2010).
14. Statkin, M. Linkage disequilibrium in growing and stable populations. Genetics 137, 331–336 (1994).
15. Conrad, D. F. et al. A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. Nature Genetics 38, 1251–1260 (2006).
16. GWAS to the people. Nature Medicine 1483 (2018).
17. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nature genetics 50, 1219–1224 (2018).
18. Wang, S. et al. Genetic variation and population structure in Native Americans. PLoS Genetics 3, 2049–2067 (2007).
19. Moreno-Estrada, A. et al. The genetics of Mexico recapitulates Native American substructure and affects biomedical traits. Science 344, 1280–1285 (2014).
20. Homburger, J. B. et al. Genomic Insights into the Ancestry and Demographic History of South America. Plos Genetics 11 (2015).
21. Liu, N., Zhang, K. & Zhao, H. Haplotype-Association Analysis. Advances in Genetics, 335–405 (2008).
22. Hsieh, A. R., Hsiao, C. L., Chang, S. W., Wang, H. M. & Fann, C. S. J. On the use of multifactor dimensionality reduction (MDR) and classification and regression tree (CART) to identify haplotype-haplotype interactions in genetic studies. Genomics 97, 77–85 (2011).
23. Aaltonen, D. et al. 2010 Rheumatoid arthritis classification criteria: An American College of Rheumatology/European League Against Rheumatism collaborative initiative. Arthritis and Rheumatism 62, 2569–2581 (2010).
24. Frazer, K. A. et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 449, 811–6 (2007).
25. Barrett, J. C., Fry, B., Maller, J. & Daly, M. J. Haplovie: Analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265 (2005).
26. Aaltonen, A. et al. A global reference for human genetic variation. Nature 526, 68–74 (2015).
27. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575 (2007).
28. Prum, R. J. et al. LocusZoom: Regional visualization of genome-wide association scan results. In Bioinformatics 27, 2336–2337 (2011).

Acknowledgements
This work was supported by Fondecyt grants n° 11130198 and n° 1151048. Research by A.L. is funded by the Spanish Ministry of Economy and Competitiveness (Fondo de Investigaciones Sanitarias and Fondos FEDER [grant number PI18/00139]).

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. Study conception and design, P.C.S., L.A.Q. and R.D.P. Acquisition of data, P.C.S., R.A.A., M.A.G., J.S., J.C.A., J.O., C.P., L.A.Q. and R.D.P. Analysis and interpretation of data, P.C.S., R. A. V., A. L., L.A.Q. and R.D.P.

Competing interests
The authors declare no competing interests.
