



A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus

DOI:

[10.1136/annrheumdis-2016-209436](https://doi.org/10.1136/annrheumdis-2016-209436)

Document Version

Accepted author manuscript

[Link to publication record in Manchester Research Explorer](#)

Citation for published version (APA):

Márquez, A., Vidal-Bralo, L., Rodríguez-Rodríguez, L., González-Gay, M. A., Balsa, A., González-álvaro, I., ... Martín, J. (2017). A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. *Annals of the rheumatic diseases*, 76(1), 286-294. <https://doi.org/10.1136/annrheumdis-2016-209436>

Published in:

Annals of the rheumatic diseases

Citing this paper

Please note that where the full-text provided on Manchester Research Explorer is the Author Accepted Manuscript or Proof version this may differ from the final Published version. If citing, it is advised that you check and use the publisher's definitive version.

General rights

Copyright and moral rights for the publications made accessible in the Research Explorer are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Takedown policy

If you believe that this document breaches copyright please refer to the University of Manchester's Takedown Procedures [<http://man.ac.uk/04Y6Bo>] or contact uml.scholarlycommunications@manchester.ac.uk providing relevant details, so we can investigate your claim.



A combined large scale meta-analysis identifies *COG6* as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus

Ana Márquez^{1,2}, Laura Vidal-Bralo³, Luis Rodríguez-Rodríguez⁴, Miguel A. González-Gay⁵, Alejandro Balsa⁶, Isidoro González-Álvaro⁷, Patricia Carreira⁸, Norberto Ortego-Centeno², María M Ayala-Gutiérrez⁹, Francisco José García-Hernández¹⁰, M. Francisca González-Escribano¹¹, José Mario Sabio¹², Carles Tolosa¹³, Ana Suárez¹⁴, Antonio González³, Leonid Padyukov¹⁵, Jane Worthington¹⁶, Timothy Vyse^{17,18}, Marta E. Alarcón-Riquelme^{19,20}, Javier Martín¹.

1 Instituto de Parasitología y Biomedicina "López-Neyra," CSIC, PTS Granada, Granada, Spain. 2 Systemic Autoimmune Diseases Unit, Hospital Clínico San Cecilio, Granada, Spain. 3 Laboratorio de Investigación 10 and Rheumatology Unit, Instituto de Investigación Sanitaria - Hospital Clínico Universitario de Santiago, Santiago de Compostela, Spain. 4 Rheumatology Department and Health Research Institute (IdISSC), Hospital Clínico San Carlos, Madrid, Spain. 5 Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, IDIVAL, Santander, Spain. 6 Department of Rheumatology and Institute for Health Research (IdiPAZ), University Hospital La Paz, Madrid, Spain. 7 Rheumatology Service and Health Research Institute (IP), Hospital Universitario de La Princesa, Madrid, Spain. 8 Rheumatology Department, Hospital 12 de Octubre, Madrid, Spain. 9 Department of Internal Medicine, Hospital Carlos Haya, Málaga, Spain. 10 Department of Internal Medicine, Hospital Universitario Virgen del Rocío, Sevilla, Spain. 11 Department of Immunology, Hospital Universitario Virgen del Rocío (IBiS, CSIC, US), Sevilla, Spain. 12 Department of Internal Medicine, Hospital Virgen de las Nieves, Granada, Spain. 13 Department of Internal Medicine, Hospital Parc Taulí, Sabadell, Spain. 14 Department of Functional Biology, Immunology Area, Faculty of Medicine, University of Oviedo, Oviedo, Spain. 15 Karolinska Institutet, Rheumatology Unit, Department of Medicine, Stockholm, Sweden. 16 Arthritis Research UK Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, Manchester Academic Health Science Centre, The University of Manchester, Manchester, UK. 17 Division of Genetics and Molecular Medicine, King's College London, London, UK. 18 Division of Immunology, Infection and Inflammatory Disease, King's College London, London, UK. 19 Centro de Genómica e Investigación Oncológica (GENYO), Pfizer-Universidad de Granada-Junta de Andalucía, Granada, Spain. 20 Institute for Environmental Medicine, Karolinska Institutet, Stockholm, Sweden.

Correspondence to: Ana Márquez PhD, Instituto de Parasitología y Biomedicina López-Neyra.
Consejo Superior de Investigaciones Científicas. Parque Tecnológico Ciencias de la Salud.
Avenida del Conocimiento s/n 18016-Armilla (Granada), Spain. E-mail: anamaort@ipb.csic.es

ABSTRACT

Objectives. During the last years, genome-wide association studies (GWAS) have identified a number of common genetic risk factors for rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE). However, the genetic overlap between these two immune-mediated diseases has not been thoroughly examined thus far. The aim of the present study was to identify additional risk *loci* shared between RA and SLE.

Methods. We performed a large-scale meta-analysis of GWAS data from RA (3,911 cases and 4,083 controls) and SLE (2,237 cases and 6,315 controls). The top associated polymorphisms in the discovery phase were selected for replication in additional datasets comprising 13,641 RA cases and 31,921 controls and 1,957 SLE patients and 4,588 controls.

Results. The rs9603612 genetic variant, located nearby the *COG6* gene, an established susceptibility *locus* for RA, reached genome-wide significance in the combined analysis including both discovery and replication sets (P -value=2.95E-13). *In silico* expression quantitative trait *locus* analysis revealed that the associated polymorphism acts as a regulatory variant influencing *COG6* expression. Moreover, protein-protein interaction and gene ontology enrichment analyses suggested the existence of overlap with specific biological processes, specially the type I interferon signalling pathway. Finally, genetic correlation and polygenic risk score analyses showed cross-phenotype associations between RA and SLE.

Conclusions. In conclusion, we have identified a new risk *locus* shared between RA and SLE through a meta-analysis including GWAS datasets of both diseases. This study represents the first comprehensive large-scale analysis on the genetic overlap between these two complex disorders.

KEYWORDS: rheumatoid arthritis, systemic lupus erythematosus, gene polymorphism, meta-analysis, genome-wide association study.

INTRODUCTION

Rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) are autoimmune rheumatic diseases with a complex aetiology, in which both genetic and environmental factors are implicated in their development [1, 2]. RA is characterized by a chronic inflammation of the synovial joints, leading to damage of the articular cartilage and the underlying bone [1], while the main event in SLE is the production of antibodies against self-components of the cell nucleus which results in a variety of clinical manifestations [2].

Although both diseases present different phenotypes, several lines of evidence point to a shared genetic component between them. It has been described the existence of familial aggregation for RA and SLE [3]. In addition, genome-wide association studies (GWAS) performed during the last years have shown a genetic overlap between them, with a considerable number of *loci* implicated in both RA and SLE susceptibility [4, 5]. Moreover, gene expression studies have revealed common molecular mechanisms involved in the pathogenesis of these two conditions. In this regard, it has been described the presence of an interferon signature (expression of genes inducible by type I interferons), which is a major feature of SLE, in groups of RA patients [6]; furthermore, a percentage of SLE patients have been found to have extensive joint damage known as rhus syndrome [7].

One of the main limitations of the association studies in autoimmunity is the difficulty in identifying genetic risk variants with modest effects, given the large sample size required and the relatively low prevalence of these diseases in the general population. This limitation has been partially overcome by combining GWAS data from different pathologies as a single phenotype, thus providing the statistical power lacking in GWAS datasets of a specific disease. This approach has already been successfully applied in the study of several autoimmune diseases with common genetic backgrounds, such as Crohn's disease (CD) and celiac disease (CeD) [8], CD and psoriasis [9], CeD and RA [10, 11], and SLE and systemic sclerosis [12]. During the

last few years, several known risk *loci* for SLE have been tested for association with RA, and conversely, through candidate gene studies [13-15]; however, no comprehensive large-scale analysis of the genetic overlap between RA and SLE has been performed so far.

In order to identify novel shared risk *loci* between RA and SLE, we performed a combined meta-analysis including previously published GWAS datasets of both diseases to increase the statistical power of the study.

METHODS

Study population

A total of 17,552 RA patients, 4,194 SLE cases and 46,907 controls of European origin were enrolled in the study. Figure 1 and Supplementary Table S1 detail the cohorts included in the different stages of the study.

SLE GWAS dataset. In the discovery phase, we included GWAS data from 2,237 SLE cases and 6,315 controls from Germany, Italy, Spain, The Netherlands and The United States of America (USA), all of them included in previously published GWASs [16, 17] (see online Supplementary Table S1).

RA GWAS dataset. The RA discovery cohort was composed of 3,911 cases and 4,083 controls from Sweden and The United Kingdom (UK) [18], obtained from the Epidemiological Investigation of RA (EIRA) project (<http://www.eirasweden.se>) and the Wellcome Trust Case Control Consortium (WTCCC) data repositories (<http://www.wtccc.org.uk/>), respectively (Supplementary Table S1).

Replication cohorts. The replication phase of the study comprised 1,957 SLE patients and 4,588 controls, and 13,641 RA cases and 31,921 controls (Supplementary Table S1). The SLE replication cohort included case/control sets from UK and Spain. Genotyping data from the UK SLE patients came from a published GWAS [16], while the control individuals were obtained from the WTCCC2 (only those not overlapping with the WTCCC controls) (<http://www.wtccc.org.uk/>). Spanish SLE patients and controls came from two different cohorts; one of them was genotyped using multiplex assays based on SNaPshot single-base extension technology (Applied Biosystems) while the other one was genotyped by TaqMan® assays. These cohorts have been used and characterized in previous association studies [19, 20]. For the RA replication, Spanish patients and controls were genotyped using TaqMan® assays. Summary statistics data from six additional European case/control collections were obtained from a published GWAS meta-analysis in RA [21].

Quality control and imputation

Data quality control was performed for each sample set separately prior imputation. SNPs and subjects with success call rates lower than 95% were removed. SNPs with minor allele frequencies lower than 0.01 and those showing a deviation from Hardy-Weinberg equilibrium (HWE; $p < 0.001$) were excluded.

IMPUTE2 software was used to perform imputation as described in Howie *et al.* [22], using as reference panels the CEU+TSI HapMap phase 3 data (UCSC hg18/NCBI Build 36) with 410 phased haplotypes encompassing 1,440,616 SNPs (<http://hapmap.ncbi.nlm.nih.gov/>; <http://genome.ucsc.edu/>). All included GWAS data were imputed as described above, except those from Okada *et al.* [21] (used in the replication phase), which came from data imputed using the 1000 Genomes Phase I reference panel for European ancestry. SNP imputation showed an accuracy of 98% in the combined European cohort. Imputed data were subsequently subjected to stringent quality filters in PLINK v. 1.07 [23], i.e., individuals who generated genotypes at <90% were removed, and SNPs with call rates <90% and those that deviated from HWE in control ($p < 0.001$) were also discarded. The first five principal components (PC) were estimated and individuals showing more than four standard deviations from the cluster centroids were excluded as outliers. Duplicates and first-degree relatives were also removed.

Statistical analysis

Statistical analyses were performed with PLINK v. 1.07.

GWAS meta-analysis. First, each GWAS case/control cohort was independently analyzed by logistic regression assuming an additive model with the first five PCs as covariates, to rule out any population stratification effect. Subsequently, disease-specific meta-analyses were performed combining RA datasets, on one hand, and SLE datasets, on the other hand, by an inverse variance-weighted method. Given associations of HLA alleles with both RA and SLE have been deeply studied,

polymorphisms within this region were excluded from the subsequent analyses (chr6:20-40 Mb). Sex chromosomes were also excluded. In order to detect non-HLA genetic variants showing the same effect in both diseases (risk or protection), a combined RA-SLE meta-analysis was conducted. Those SNPs with p-values lower than 1×10^{-5} in this combined meta-GWAS and p-values lower than 0.01 in each disease meta-analysis were selected for the replication phase. On the other hand, to identify common signals with opposite effects in both diseases, the direction of association was flipped in the RA dataset (1/OR instead of OR) before performing the RA-SLE meta-analysis. Again, SNPs were selected for replication according to the above criteria. Genetic variants showing significant heterogeneity (Cochran's Q test < 0.05) in the RA or SLE meta-analysis were not considered for the validation step.

Replication analysis. Replication cohorts were also analyzed by logistic regression. For the previously selected SNPs, combined analysis of the RA and SLE replication and discovery cohorts was performed using the inverse variance method. After replication stage, polymorphisms with p-values lower than 5×10^{-8} in the RA-SLE meta-analysis (discovery and replication cohorts) and disease-specific p-values lower than 0.05 in the replication phase were considered as statistically significant.

Protein-protein interaction (PPI) and gene set enrichment analyses

To assess for interaction amongst proteins encoded by SLE and RA common risk loci, a protein-protein network were constructed using the STRING database v. 10.0 [24], which builds PPI based on direct (physical) and indirect (functional) associations.

On the other hand, Gene Ontology (GO) (<http://www.geneontology.org>) [25] was applied to perform an enrichment analysis in order to determine whether certain biological process are over-represented (or under-represented) in the common RA-SLE gene set.

Genetic pleiotropy analysis

After excluding markers within the HLA region, pleiotropy between both diseases was estimated using two different approaches:

Bivariate analysis. GCTA v1.25.0 (<http://cnsgenomics.com/software/gcta/>) was used to create a genetic relationship matrix (GRM) file containing IBD (identity by descent) relationship calculations for all pair-wise sets of individuals. Genetic correlation (rG) between both diseases was calculated by GCTA bivariate REML (restricted maximum likelihood) analysis [26] using the GRM with the first five PCs as covariates. A likelihood ratio test (LRT) was applied to determine the statistical significance of this genetic correlation.

Polygenic risk score (PRS) analysis. We used PRS to assess the genetic overlap between RA and SLE, as previously described [27]. First, we selected a filtered set of SNPs from the results of disease-specific meta-GWASs. We used the --clump algorithm in PLINK to select polymorphisms with $r^2 < 0.20$ within 500 kb windows and at a range of significance levels, specifically, we evaluated three different P thresholds (P_T), $P_T < 1 \times 10^{-4}$, $< 1 \times 10^{-3}$ and $< 1 \times 10^{-2}$. Then, for each individual from a specific-disease dataset (SLE or RA), we calculated the number of score alleles they possessed, each weighted by the log of the OR of the other disease. With the scores generated, we performed logistic regression analysis to test the relationship between the computed scores and disease status. The variance in case/control status explained by the scores was estimated as the difference in the Nagelkerke's pseudo- R^2 between a null generalized linear model, including the first five PCs and the country of origin as covariates, and an alternative model, including the same covariates and the risk scores. The significance level was estimated by the means of a LRT.

RESULTS

Meta-GWAS and replication

We performed a meta-analysis considering both diseases as a single phenotype. After quality control, the discovery cohort comprised 3,808 RA cases, 2,104 SLE patients and 10,157 controls. A total of 309,839 genetic variants outside the HLA region overlapped between the different GWAS datasets and were included in the meta-analysis.

When we combined disease-specific meta-analyses, assuming that alleles had the same effect in both diseases, 105 SNPs reached the significance threshold fixed for the combined meta-GWAS ($p\text{-value} < 1 \times 10^{-5}$). 89 of these genetic variants were located in *loci* already established as risk factors for both diseases (Figure 2 and Supplementary Table S2), *PTPN22*, *TNFAIP3*, *IRF5*, *BLK*, *ATG5*, *UBE2L3* and *ICAM3/TYK2*, and, therefore, were excluded from subsequent analyses. Regarding the remaining signals, nine met the selection criteria for the replication phase ($p\text{-value} < 1 \times 10^{-5}$ in the RA-SLE meta-GWAS and $p\text{-value} < 0.01$ in each disease-specific meta-analysis). From these nine polymorphisms, three were located next to two *loci* previously associated with RA (two of them close to *COG6* and another one near *NFKBIE*), one was located in an intergenic region close to *PTTG1* (a *locus* previously associated with SLE), and the rest of them lied within regions not associated with RA or SLE so far (three near *TGFA*, one close to *SYPL1* and another one within *RUNDC1*).

When we performed the analysis under the assumption that alleles have opposite direction in both diseases, only two polymorphisms within the *NMNAT2* gene, an established SLE risk *locus*, met the selection criteria for the replication stage (Figure 2 and Supplementary Table S3).

To confirm that these seven *loci* were associated with both diseases, the strongest associated SNP within each *locus* was selected for validation in additional sample sets (Table 1). According to the significance criteria established, the rs9603612 polymorphism, near *COG6*, achieved genome-wide significance in the combined

analysis including both discovery and replication sets ($P_{\text{Discovery+Replication}}=2.95\text{E-}13$), and also statistical significance in each disease-specific replication analysis ($P_{\text{SLE_Replication}}=0.045$; $P_{\text{RA_Replication}}=3.58\text{E-}08$).

Effect on Gene Expression

Given that rs9603612 is a noncoding variant, we assessed its potential regulatory function by means of *in silico* expression quantitative trait locus (eQTL) analysis using RegulomeDB (<http://www.regulomedb.org/>) [28]. This database includes high-throughput experimental data sets, as well as computational predictions and manual annotations to identify functional variants. According to RegulomeDB, rs9603612 showed a score 1b, indicating that this SNP likely affect transcription factor binding and is linked to expression of a gene target, specifically it may act as a cis-eqtl regulating COG6 expression in monocytes.

PPI and pathway enrichment analyses

Subsequently, we evaluated connectivity at the PPI level between the 18 common risk *loci* for RA and SLE, including COG6. One PPI network involving 13 of the 18 common proteins was formed (Figure 3). Compared with randomly selected protein datasets, a significant network connectivity was found ($p\text{-value}<0.05$). The RA-SLE common proteins were more likely to be connected to each other than expected by chance, with 32 interactions as compared with only 2 observed for random data sets. COG6 did not appear linked to any protein.

Then, we performed the analysis considering all the established risk factors for SLE ($n=44$) and RA ($n=64$) (obtained from Bentham et al. [16] and Okada et al. [21], respectively) (Supplementary Figure S1). When SLE-associated genes were input, three different networks were formed, one of them involving 37 proteins. Again, COG6 was not connected to any protein (Supplementary Figure S1A). In the case of RA, a

single PPI network comprising 52 proteins was evident, with COG6 linked to it through the AFF3 protein (Supplementary Figure S1B).

On the other hand, according to the GO enrichment analysis, several biological processes appeared over-represented amongst the RA-SLE common gene set. The most significantly over-represented pathways are related to the immune response (Table 2), especially to the type I interferon (IFN) signalling pathway (p-value=2.61E-03).

Genetic overlap between SLE and RA

After bivariate REML analysis, a significant genetic correlation between RA and SLE was evident ($r_G=0.31$, $SE=0.061$, $p\text{-value}=2.00E-07$).

Similarly, the PRS analysis showed significant differences in the score distribution between both case groups and controls (Figure 4). For each of the established scoring SNP sets, the mean score was significantly higher in patients compared with controls (Figure 4A-F), thus indicating that SLE cases had a significant enrichment of RA risk alleles, and vice-versa. For both, the most significant differences were observed when scores were calculated after applying the most stringent SNP inclusion cut-off (SLE: $P\text{-value}_{LRT}=8.93E-04$; RA: $P\text{-value}_{LRT}=8.93E-12$), explaining 0.19% and 0.78% of the variance (Nagelkerke's pseudo- R^2) in disease status for RA and SLE (Figure 4G), respectively.

DISCUSSION

In the present study, we have identified a new risk *locus* shared between RA and SLE through a meta-GWAS considering both disorders as a single phenotype. This study represents the first comprehensive large-scale analysis, including more than 22,000 cases and 47,000 controls, carried out to reach a deeper understanding of the genetic overlap between these two diseases.

According to our data, *COG6* (component of oligomeric Golgi complex 6), an established susceptibility *locus* for RA and psoriasis [21, 29], also represents a genetic risk factor influencing the SLE predisposition. This *locus* is located on chromosome 13q14.11, within a 250-kb block of linkage disequilibrium (LD) that includes only this gene. *COG6* encodes a subunit of the conserved oligomeric Golgi complex crucial for the normal structure and function of the Golgi apparatus, influencing processes such as protein sorting and glycosylation [30]. Although some hypothesis on the possible implication of *COG6* in autoimmunity have been proposed, its role in immune-mediated disorders remains unknown. It has been described that deficiency of this gene may lead to a clinical phenotype including inflammatory bowel disease and neutrophil and B and T cell dysfunction [31].

A main advantage of combining GWAS data from related diseases is to increase the statistical power in order to capture association signals that may have been undetected in previous disease-specific studies. Indeed, the *COG6* polymorphism associated with RA and SLE, rs9603612, showed a similar effect size in both disorders (OR=0.91). However, as mentioned above, this gene was previously associated with RA but not with SLE, probably due to the larger sample size of the GWASs published in the first so far. This would also explain why, in our study, the rs9603612 genetic variant reached genome-wide significance level in the RA meta-analysis, but not in the analysis of the SLE dataset. Rs9603612 shows tight LD with the *COG6* SNPs implicated in RA (rs9603616, $r^2=0.98$) [21] and psoriasis (rs7993214, $r^2=0.97$) [29]. Interestingly, whereas these show minimal evidence of acting as regulatory polymorphisms

according to RegulomeDB (score 5 and 4, respectively), rs9603612 seems to influence the *COG6* expression in monocytes, thus representing a better candidate to be the causal variant involved in the genetic predisposition to these three autoimmune conditions. Indeed, a very recent study focused on identifying functional variants for disease-associated *loci*, has confirmed this regulatory role of the rs9603612 polymorphism [32]. Specifically, the minor allele (G) was found to increase the *COG6* expression compared with the major allele, thus supporting that rs9603612 influences the development of SLE, RA and psoriasis by regulating the *COG6* levels.

The PPI analysis evidenced a high connectivity amongst proteins involved in RA and SLE, which indicates the existence of overlap between specific biological pathways implicated in these two diseases. In this sense, the type I IFN signalling pathway emerged as the most significantly over-represented biological process amongst the RA-SLE common risk *loci* set. This is consistent with the increased expression of type I IFN regulated genes observed in both disorders [6, 33] and points to a role of this interferon signature, a major feature of SLE patients, in the pathogenesis of RA. Indeed, a recent gene expression meta-analysis including Sjogren's syndrome, RA and SLE showed the interferon signature as a major over-represented shared gene profile [34]. Regarding *COG6*, no connection with any of the remaining overlapping proteins was evident in principle. However, when considering the RA-associated gene set, *COG6* was linked to the main network (which included most of the common proteins to both diseases) through *AFF3*. It should be noted that a nominal association of the RA-associated *AFF3* polymorphism with SLE was reported in a candidate gene study performed in Europeans [14]. This association was subsequently replicated in Chinese SLE patients [35]. Taking this into account, *AFF3* could represent the nexus between *COG6* and the rest of the RA-SLE common pathways. The fact that no association signals were detected within *AFF3* in our study could be due to a low SNP coverage in the region. Indeed, the reported risk variant for RA and SLE, as well as those in high linkage disequilibrium, were lacking in our GWAS dataset.

Finally, we aimed to quantify the genetic overlap between both conditions, since it has not been systematically examined using genome-wide data. As expected, genetic factors for RA and SLE were positively correlated. Similarly, PRS analysis showed that RA and SLE polymorphisms present a cross-phenotype effect, which was higher when scores were calculated from the strongest associated SNP set ($P_T=1 \times 10^{-4}$). Genetic variants with larger effect on SLE susceptibility explained a higher percentage of the RA variance, and viceversa, thus indicating that most of the genetic component shared between both disorders is driven by their main risk alleles. This is consistent with previous findings showing a genetic overlap between both diseases [13, 18], as well as a significant enrichment in carriage of SLE alleles in patients with RA compared with controls [13].

In summary, the present study adds *COG6* to the list of risk factors shared between RA and SLE. Our results highlight the existence of a relevant genetic correlation between both diseases as well as the influence of common molecular mechanisms in their pathophysiology. Since common genetic pathways are implicated in RA and SLE, a reclassification of patients from a genetic point of view will lead to more specific and effective therapeutic procedures.

ACKNOWLEDGEMENTS

We thank Sofia Vargas, Sonia Garcia and Gema Robledo for their excellent technical assistance, and all the patients and healthy controls for kindly accepting their essential collaboration. This work was supported by the following grants: SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness, P12-BIO-1395 from Consejería de Innovación, Ciencia y Tecnología, Junta de Andalucía (Spain), the European IMI BTCure Program, the EU/EFPIA Innovative Medicines Initiative Joint Undertaking PRECISESADS (ref: 115565), the Cooperative Research Thematic Network (RETICS) program, RD12/0009/0004 (RIER), from Instituto de Salud Carlos III (ISCIII, Health Ministry, Madrid, Spain) and P12/02558 from Instituto de Salud Carlos III (ISCIII, Health Ministry, Madrid, Spain). AM is recipient of a Rio Hortega fellowship (CM13/00314) from the Ministry of Economy and Competitiveness through the Instituto de Salud Carlos III (ISCIII, Health Ministry, Madrid, Spain).

Competing Interest: None declared

REFERENCES

1. Scott DL, Wolfe F, Huizinga TW. Rheumatoid arthritis. *Lancet*. 2010; **376**(9746):1094-1108.
2. Tsokos GC. Systemic lupus erythematosus. *N Engl J Med*. 2011; **365**(22):2110-2121.
3. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. *BMC Med*. 2013; **11**:73.
4. Cui Y, Sheng Y, Zhang X. Genetic susceptibility to SLE: recent progress from GWAS. *J Autoimmun*. 2013; **41**:25-33.
5. Viatte S, Plant D, Raychaudhuri S. Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol*. 2013; **9**(3):141-153.
6. Higgs BW, Liu Z, White B, et al. Patients with systemic lupus erythematosus, myositis, rheumatoid arthritis and scleroderma share activation of a common type I interferon pathway. *Ann Rheum Dis*. 2011; **70**(11):2029-2036.
7. Panush RS, Edwards NL, Longley S, et al. 'Rhupus' syndrome. *Arch Intern Med*. 1988; **148**(7):1633-1636.
8. Festen EA, Goyette P, Green T, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. *PLoS Genet*. 2011; **7**(1):e1001283.
9. Ellinghaus D, Ellinghaus E, Nair RP, et al. Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. *Am J Hum Genet*. 2012; **90**(4):636-647.
10. Gutierrez-Achury J, Zorro MM, Ricano-Ponce I, et al. Functional implications of disease-specific variants in loci jointly associated with coeliac disease and rheumatoid arthritis. *Hum Mol Genet*. 2015:pil: ddv455. [Epub ahead of print].

11. Zhernakova A, Stahl EA, Trynka G, 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**(2):e1002004.
12. Martin JE, Assassi S, Diaz-Gallo LM, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet.* 2013; **22**(19):4021-4029.
13. Orozco G, Eyre S, Hinks A, et al. Study of the common genetic background for rheumatoid arthritis and systemic lupus erythematosus. *Ann Rheum Dis.* 2011; **70**(3):463-468.
14. Ramos PS, Criswell LA, Moser KL, et al. A comprehensive analysis of shared loci between systemic lupus erythematosus (SLE) and sixteen autoimmune diseases reveals limited genetic overlap. *PLoS Genet.* 2011; **7**(12):e1002406.
15. Suarez-Gestal M, Calaza M, Dieguez-Gonzalez R, et al. Rheumatoid arthritis does not share most of the newly identified systemic lupus erythematosus genetic factors. *Arthritis Rheum.* 2009; **60**(9):2558-2564.
16. Bentham J, Morris DL, Cunninghame Graham DS, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet.* 2015; **47**(12):1457-1464.
17. Harley JB, Alarcon-Riquelme ME, Criswell LA, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nat Genet.* 2008; **40**(2):204-210.
18. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet.* 2010; **42**(6):508-514.
19. Sanchez E, Gomez LM, Lopez-Nevot MA, et al. Evidence of association of macrophage migration inhibitory factor gene polymorphisms with systemic lupus erythematosus. *Genes Immun.* 2006; **7**(5):433-436.

20. Suarez-Gestal M, Calaza M, Endreffy E, et al. Replication of recently identified systemic lupus erythematosus genetic associations: a case-control study. *Arthritis Res Ther.* 2009; **11**(3):R69.
21. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014; **506**(7488):376-381.
22. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* 2009; **5**(6):e1000529.
23. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007; **81**(3):559-575.
24. Szklarczyk D, Franceschini A, Wyder S, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015; **43**(Database issue):D447-452.
25. Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 2015; **43**(Database issue):D1049-1056.
26. Lee SH, Yang J, Goddard ME, et al. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics.* 2012; **28**(19):2540-2542.
27. Purcell SM, Wray NR, Stone JL, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature.* 2009; **460**(7256):748-752.
28. Boyle AP, Hong EL, Hariharan M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res.* 2012; **22**(9):1790-1797.
29. Liu Y, Helms C, Liao W, et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.* 2008; **4**(3):e1000041.
30. Smith RD, Lupashin VV. Role of the conserved oligomeric Golgi (COG) complex in protein glycosylation. *Carbohydr Res.* 2008; **343**(12):2024-2031.

31. Huybrechts S, De Laet C, Bontems P, et al. Deficiency of Subunit 6 of the Conserved Oligomeric Golgi Complex (COG6-CDG): Second Patient, Different Phenotype. *JIMD Rep.* 2012; **4**:103-108.
32. Cavalli M, Pan G, Nord H, et al. Allele-specific transcription factor binding to common and rare variants associated with disease and gene expression. *Hum Genet.* 2016.
33. van der Pouw Kraan TC, Wijbrandts CA, van Baarsen LG, et al. Rheumatoid arthritis subtypes identified by genomic profiling of peripheral blood cells: assignment of a type I interferon signature in a subpopulation of patients. *Ann Rheum Dis.* 2007; **66**(8):1008-1014.
34. Toro-Dominguez D, Carmona-Saez P, Alarcon-Riquelme ME. Shared signatures between rheumatoid arthritis, systemic lupus erythematosus and Sjogren's syndrome uncovered through gene expression meta-analysis. *Arthritis Res Ther.* 2014; **16**(6):489.
35. Cen H, Leng RX, Wang W, et al. Association of AFF1 rs340630 and AFF3 rs10865035 polymorphisms with systemic lupus erythematosus in a Chinese population. *Immunogenetics.* 2012; **64**(12):935-938.

Table 1. Association results obtained in the discovery and replication phases for each disease and in the rheumatoid arthritis-systemic lupus erythematosus combined meta-analysis.

					Discovery phase				Replication phase				Discovery + Replication						
	SLE		RA		SLE-RA		SLE		RA		SLE		RA		RA-SLE				
<i>Locus</i>	Chr	SNP	A1	P-value	OR	P-value	OR	P-value	OR	P-value	OR	P-value	OR	P-value	OR	P-value	OR		
COG6	13	rs9603612	G	4.41E-03	0.89	6.59E-04	0.89	9.08E-06	0.89	4.50E-02	0.92	3.58E-08	0.91	5.88E-04	0.91	1.22E-10	0.91	2.946E-13	0.91
<i>PTTG1</i>	5	rs4921283	G	8.63E-05	0.86	7.24E-03	0.91	3.99E-06	0.89	1.88E-05	0.84	0.0471	0.97	6.76E-09	0.85	3.33E-03	0.96	1.24E-07	0.93
<i>NFKBIE</i>	6	rs9395027	A	2.54E-03	0.76	7.16E-04	0.76	5.80E-06	0.76	0.2743	0.91	0.1090	0.93	4.28E-03	0.84	2.03E-03	0.88	3.46E-05	0.87
<i>NMNAT2</i>	1	rs1411393	G	3.44E-04	0.87	4.39E-03	1.10	6.62E-06	NA	2.49E-02	0.92	0.5170	1.01	3.64E-05	0.89	0.0721	1.03	4.43E-04	NA
<i>TGFA</i>	2	rs6546604	A	2.69E-03	1.14	4.11E-04	1.14	3.56E-06	1.14	0.0588	1.09	0.8863	1.00	5.20E-04	1.11	0.0889	1.03	1.75E-03	1.05
<i>RUNDC1</i>	17	rs323496	C	1.66E-03	1.15	1.81E-03	1.14	9.73E-06	1.14	0.5908	1.06	0.8169	1.00	0.0456	1.11	0.2511	1.02	0.0222	1.04
<i>SYPL1</i>	7	rs7795448	G	2.23E-05	1.18	7.24E-03	1.10	1.39E-06	1.14	0.5573	0.98	0.4204	0.98	0.0641	1.09	0.4668	1.01	0.0401	1.03

P-values and odds ratios for the minor allele are shown for the seven selected polymorphisms. The polymorphism associated at the genome-wide significance level in the RA-SLE combined meta-analysis and at the nominal significance level in the replication phase is marked in bold. *Loci* established in both diseases are not included in this table.

Chr, chromosome; SNP, single nucleotide polymorphism; A1, minor allele; SLE, systemic lupus erythematosus; RA, rheumatoid arthritis; OR, odds ratio.

Table 2. Biological processes significantly over-represented ($p < 0.05$) amongst risk *loci* shared between rheumatoid arthritis and systemic lupus erythematosus.

Biological pathway	GO term	P-value*	Genes
Defense response	GO:0006952	7.82E-04	<i>IRF5, BLK, IRAK1, TNFAIP3, IRF8, TYK2, ATG5, PTPN11, PXX, MECP2</i>
Type I interferon signaling pathway	GO:0060337	2.61E-03	<i>IRF5, IRF8, TYK2, PTPN11</i>
Positive regulation of cellular process	GO:0048522	6.03E-03	<i>ARID5B, IRF5, BLK, SPRED2, IRAK1, TNFAIP3, PTPN22, STAT4, IRF8, IKZF3, RAD51B, ATG5, PTPN11, MECP2</i>
Innate immune response	GO:0045087	6.84E-03	<i>IRF5, BLK, IRAK1, TNFAIP3, IRF8, TYK2, ATG5, PTPN11</i>
Cellular macromolecule metabolic process	GO:0044260	8.66E-03	<i>ARID5B, IRF5, BLK, IRAK1, JAZF1, TNFAIP3, PTPN22, STAT4, IRF8, IKZF3, PXX, TYK2, RAD51B, ATG5, PTPN11, MECP2</i>
Cytokine-mediated signaling pathway	GO:0019221	1.86E-02	<i>IRF5, IRAK1, STAT4, IRF8, TYK2, PTPN11</i>
Immune system process	GO:0002376	2.27E-02	<i>IRF5, BLK, IRAK1, TNFAIP3, IRF8, PTPN22, IKZF3, TYK2, ATG5, PTPN11</i>
Regulation of type I interferon production	GO:0032479	2.38E-02	<i>IRF5, TNFAIP3, PTPN22, ATG5</i>
Negative regulation of nucleotide-binding oligomerization domain containing 2 signaling pathway	GO:0070433	2.47E-02	<i>TNFAIP3, PTPN22</i>
Response to molecule of bacterial origin	GO:0002237	2.72E-02	<i>IRF5, IRAK1, TNFAIP3, IRF8, PTPN22</i>
Cellular response to lipopolysaccharide	GO:0071222	3.08E-02	<i>IRAK1, TNFAIP3, IRF8, PTPN22</i>

* P-values determined by the binomial statistic test and adjusted by the Bonferroni correction.

FIGURE LEGENDS

Figure 1. Flowchart showing the analysis strategy followed for the present study.

Figure 2. Manhattan plots representing the results obtained in the discovery phase for the SLE scan (A), the RA scan (B) and the combined RA-SLE meta-analysis (C). Genetic variants within the HLA region (chr 6: 20-40 Mb) were excluded from the analysis. In figure C, signals that met the replication criteria in the RA-SLE meta-GWAS considering opposite allelic effect appear highlighted in green; signals that reached the significance threshold established for the combined meta-GWAS ($p\text{-value} < 1 \times 10^{-05}$) but did not meet any of the remaining criteria ($p\text{-value} < 0.01$ and not significant heterogeneity in the disease-specific meta-analyses) are marked in grey; red colour denotes *loci* selected for the replication phase.

Figure 3. Interaction network formed for the RA-SLE common risk *loci*. STRING database was used to look for both direct and indirect interactions amongst genes shared by both diseases. The width of the blue lines indicates the reliability of each interaction.

Figure 4. Polygenic risk score analysis. A-F) Genetic risk score (GRS) distribution in systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and controls. Left panels: distribution curve of GRS in SLE patients and controls calculated with RA alleles showing p-values lower than 1×10^{-4} (A), 1×10^{-3} (B) and 1×10^{-2} (C). Right panels: distribution curve of GRS in RA patients and controls calculated with SLE alleles showing p-values lower than 1×10^{-4} (D), 1×10^{-3} (E) and 1×10^{-2} (F). G) Proportion of variance in the SLE set explained by risk scores from RA, and vice-versa. Variance explained, measured via Nagelkerke's pseudo- R^2 , is plotted for the three established scoring SNP sets ($P_T < 1 \times 10^{-4}$, 1×10^{-3} and 1×10^{-2}). Genetic variants within the HLA region were excluded from this analysis.