Cross-disease Meta-analysis of Genome-wide Association Studies for Systemic Sclerosis and Rheumatoid Arthritis Reveals *IRF4* as a New Common Susceptibility *Locus*

Elena López-Isac[1*], Jose-Ezequiel Martín[1], Shervin Assassi[2], Carmen Pilar Simeón[3], Patricia Carreira[4], Norberto Ortego-Centeno[5], Mayka Freire[6], Emma Beltrán[7], Francisco J Narváez[8], Juan José Alegre-Sancho[9], the Spanish Scleroderma Group[10], Benjamín Fernández-Gutiérrez[11], Alejandro Balsa[12], Isidoro González-Álvaro[13], Miguel A González-Gay[14], Lorenzo Beretta[15], Alessandro Santaniello[15], Chiara Bellocchi[15], Claudio Lunardi[16], Gianluca Moroncini[17], Armando Gabrielli[17], Torsten Witte[18], Nicolas Hunzelmann[19], Jörg H.W. Distler[20], Gabriella Riekemasten[21], Annete HM van der Helm[22], Jeska de Vries-Bouwstra[22], Cesar Magro[22], Alexandre E. Voskuyl[23], Madelon C Vonk[24], Øyvind Molberg[25], Benedicte Lie[25], Tony Merriman[26], Roger Hesselstrand[27], Annika Nordin[28], Leonid Padyukov[28], Ariane Herrick[29], Steve Eyre[29], Bobby P Koeleman[30], Christopher P. Denton[31], Carmen Fonseca[31], Timothy RDJ Radstake[32], Jane Worthington[29], Maureen D. Mayes[2], Javier Martín[1*]

1. Institute of Parasitology and Biomedicine López-Neyra, IPBLN-CSIC, PTS Granada, Granada, Spain.
2. The University of Texas Health Science Center–Houston, Houston, USA.
3. Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain.
4. Department of Rheumatology, 12 de Octubre University Hospital, Madrid, Spain.
5. Department of Internal Medicine, Clinic University Hospital, Granada, Spain.
6. Department of Internal Medicine, Thrombosis and Vasculitis Unit, Complexo Hospitalario Universitario de Vigo, Vigo, Spain.
7. Department of Rheumatology, Hospital General Universitario de Valencia, Valencia, Spain.
8. Department of Rheumatology, Hospital Universitari de Bellvitge, Barcelona, Spain.
9. Department of Rheumatology, Hospital del Doctor Peset, Valencia, Spain.
10. See supplementary note.
11. Rheumatology Service, Hospital Clínico San Carlos, Madrid, Spain.
12. Rheumatology Service, Hospital Universitario La Paz, Madrid, Spain.
13. Rheumatology Service, Hospital Universitario La Princesa, Instituto de Investigación Sanitaria La Princesa, Madrid, Spain.
14. Rheumatology Service, Hospital Universitario Marqués de Valdecilla, IFIMAV, Santander, Spain.
15. Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan Italy.
16. Department of Medicine, Università degli Studi di Verona, Verona, Italy.

17. Clinica Medica, Department of Clinical and Molecular Science, Università Politecnica delle Marche and Ospedali Riuniti, Ancona, Italy.
18. Department of Clinical Immunology, Hannover Medical School, Hannover, Germany.
19. Department of Dermatology, University of Cologne, Cologne, Germany.
20. Department of Internal Medicine, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen, Germany.
21. Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany German Rheumatism Research Center (DRFZ), a Leibniz Institute, Berlin, Germany.
22. Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.
23. Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands.
24. Department of Rheumatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands.
25. Department of Immunology, Oslo University Hospital, Rikshospitalet, Oslo, Norway.
26. Department of Biochemistry, University of Otago, New Zealand.
27. Department of Rheumatology, Lund University, Lund, Sweden.
28. Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden.
29. Centre for Musculoskeletal Research and NIHR Manchester Musculoskeletal Biomedical Research Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK.
30. Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands.
31. Centre for Rheumatology, Royal Free and University College Medical School, London, United Kingdom.
32. Department of Rheumatology & Clinical Immunology, Laboratory of Translational Immunology, department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.

***Correspondence to:**

Javier Martin MD, PhD

Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC. Parque Tecnológico Ciencias de la Salud. Avenida del Conocimiento s/n 18016-Armilla (Granada), Spain.

E-mail: martin@ipb.csic.es

Tel: 34 958181669

Fax: 34 958181632

Elena López-Isac

Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC. Parque Tecnológico Ciencias de la Salud. Avenida del Conocimiento s/n 18016-Armilla (Granada), Spain.

E-mail: elenalopezisac@ipb.csic.es

Tel: 958181621

Fax: 958181632

ABSTRACT

Objectives: Systemic sclerosis (SSc) and rheumatoid arthritis (RA) are autoimmune diseases that share clinical and immunological characteristics. To date, several shared SSc-RA loci have been identified independently. In this study, we aimed to systematically search for new common SSc-RA loci through an inter-disease meta-GWAS strategy.

Methods: We performed a meta-analysis combining GWAS datasets of SSc and RA. The top single-nucleotide polymorphisms (SNPs) were followed-up in independent SSc and RA case-control cohorts. This allowed us to increase the sample size to a total of 8,830 SSc patients, 16,870 RA patients and 43,393 controls.

Results: The cross-disease meta-analysis of the GWAS datasets identified several loci with nominal association signals (*P*-value < 5×10^{-6}), which also showed evidence of association in the disease-specific GWAS scan. These loci included several genomic regions not previously reported as shared loci, besides risk factors associated with both diseases in previous studies. The follow-up of the putatively new SSc-RA loci confirmed *IRF4* as a shared risk factor for these two diseases ($P_{combined} = 3.29 \times 10^{-12}$). In addition, the analysis of the biological relevance of the known SSc-RA shared loci pointed to the type I interferon and the interleukin 12 signaling pathways as the main common aetiopathogenic factors.

Conclusions: Our study have identified a novel shared locus, *IRF4*, for SSc and RA and highlighted the usefulness of cross-disease GWAS meta-analysis in the identification of loci with pleiotropic effects.

INTRODUCTION

Autoimmune diseases (ADs) comprise a group of heterogeneous disorders affecting around 5-10% of individuals of European origin. Most ADs are complex traits and the current knowledge of the aetiopathogenic factors underlying these diseases is still limited. (1) Genome-wide association studies (GWASs) and immune-focused fine-mapping studies have revolutionized the understanding of the genetic component of complex ADs by the identification of thousands of susceptibility loci associated with autoimmunity, that have provided an unbiased support for possible pathogenic pathways, including antigen presentation, cytokine signaling, and B and T cells differentiation.(2-3) The vast majority of these loci are shared risk factors for at least two or more ADs, pointing to a common genetic background and shared molecular pathways underlying autoimmune processes. These genetic overlapping has been suspected some time ago, given the high rate of co-occurrence of ADs in patients affected by another immune-mediated disease and the well-established familial aggregation reported for these immune disorders. (4)

Systemic sclerosis (SSc) and rheumatoid arthritis (RA) are complex ADs which share clinical and immunological features. Both diseases are rheumatic connective tissue disorders, more frequent in women than men, characterized by an exacerbated inflammatory response, deregulation of innate and adaptive immunity, including autoantibodies production, and systemic complications. (5-6) Thank to the establishment of great consortiums and international collaborations, the number of confirmed RA susceptibility factors have achieved a total of 101 loci associated with the disease at the genome-wide significance level. (7) In regard to SSc, GWASs, Immunochip and candidate gene studies have clearly identified several genetic regions involved in SSc susceptibility (8-10). However, the knowledge of the genetic predisposition to this disease is still

relatively limited, in part due to its low prevalence, which makes difficult the recruitment of large cohorts that allow reaching a high statistical power to effectively detect association signals. Interestingly, a considerable proportion of the SSc susceptibility factors constitute also RA risk loci, such as *STAT4* (MIM 600558), *IRF5* (MIM 607218), *IRF8* (MIM 601565), *TNFAIP3* (MIM 191163), *CD247* (MIM 186780), *DNASE1L3* (MIM 602244), *ATG5* (MIM 604261), *BLK* (MIM 191305) and *PTPN22* (MIM 600716), suggesting shared aetiopathogenic pathways for both ADs (7, 10). In addition, although not very common, co-familiality and co-occurrence between these two rheumatic conditions has been observed, with a weighted estimated prevalence around 5% for patients with SSc and RA according to Elhai *et al*. (2013). (11) These observations provide evidence of a genetic overlapping for both diseases, thus it is expected that additional shared risk factors remain to be discovered.

One approach that has been developed for the identification of common genetic loci in a cost-effective manner is to perform a combined-phenotype GWAS, that is, to combine genome-wide genotype data from two autoimmune diseases regarding them as a single phenotype. This strategy has been successfully applied not only in the study of closely related phenotypes, such as Crohn disease (CD) and ulcerative colitis (UC), but also to non-related phenotypes, showing encouraging results. (12-16)

Considering all the above, the goal of the present study was to systematically identify new pleiotropic risk loci for SSc and RA by applying the combined-phenotype GWAS strategy, followed by replication testing in independent case-control datasets in both diseases.

METHODS

Study population

The stage I of the present study included 6,537 SSc/RA patients and 8,741 healthy controls. SSc GWAS panel comprised four case-control sets from Spain, Germany, The Netherlands and US (2,716 cases and 5,666 controls) which were obtained from previous studies. (8, 15, 17) RA case-control GWA study included two previously published RA GWAS cohorts (WTCCC, EIRA) from UK and Sweden (3,821 cases and 3,075 controls). (18)

The replication stage was drawn in independent SSc and RA case-control sets of European ancestry. The SSc replication cohort included 6,114 cases and 8,744 controls from 8 different countries (Spain, Germany, Italy, UK, The Netherlands, Sweden, Norway and US). The healthy controls from UK and US partially overlapped with control sets of previously published cohorts (WTCCC and NARAC2). (18) The RA replication cohort included 9 case-control collections from North America (US, Canada), Spain, The Netherlands, UK, Sweden, France and New Zealand, and comprised a total of 13,049 RA cases and 25,908 healthy controls. Of these, 9,711 cases and 24,253 healthy controls were obtained from several previously published studies (BRASS, NARAC1, CANADA, Rheumatoid Arthritis Consortium International for Immunochip (RACI)-US, RACI-i2b2, CORRONA, Vanderbilt, RACI-UK, RACI-SE-U, RACI-NL, Dutch (AMC, BeSt, LUMC, and DREAM), ReAct, and the anti-TNF response to therapy collection (ACR-REF)) (7). SSc patients fulfilled the 1980 American College of Rheumatology (ACR) classification criteria for this disease or the criteria proposed by LeRoy and Medsger for early-SSc. (19-20). All RA patients fulfilled the 1987 criteria of the ACR for RA diagnosis (21) or were diagnosed as RA by a professional rheumatologist. All individuals enrolled in the present

study provided written informed consent and approval from the local ethical committees was obtained from all the centers that recruited the participating individuals.

Study design

Stage I. In the present study, we performed a two-stage study to systematically identify shared risk factors for SSc and RA (Figure 1). In the stage I, we performed GWAS analysis for each disease separately and a combined-phenotype GWAS analysis. Two different tests that have been successfully applied in previous studies were considered for the combined analysis (12, 15):

1) In order to detect common signals for SSc and RA with same-direction allelic effects, the meta-analysis considering both diseases was performed as usual. Those SNPs that showed a *P*-value $< 5 \times 10^{-6}$ in the combined-phenotype analysis and nominal significance in the association study for each disease (*P*-value < 0.05) were selected for follow-up in the replication stage.

2) To identify common signals for SSc and RA with opposite-direction allelic effects, we flipped the direction of association (1/OR) in the RA dataset for the combined-disease meta-analysis. To select SNPs for replication, the same selection criteria stated above was followed.

For both sorts of meta-analysis, we only considered for follow-up those SNPs that had not been previously reported as genetic risk factors for SSc and RA, or those that had been reported for one of the disease but not reported for the other. We also did not consider markers from the extended HLA region on chromosome six.

Stage II and meta-analysis. The selected SNPs were followed-up in independent replication cohorts recruited for both SSc and RA studies. After replication stage, we performed a meta-analysis of the initial GWAS screening stage and replication stage. The

SNP signals that reached (1) genome-wide significance level (P-value < 5 x 10^{-8}) in the combined-phenotype meta-analysis (GWAS + Replication stage), and (2) showed , for each disease separately, nominally significant associations (P-value < 0.05) in the replication step as well as $P_{GWAS+Repl}$ < 5 x 10^{-3} were considered shared risk factors for the two analyzed diseases.

Quality control and genotype imputation of GWAS data

We applied stringent QC criteria in all the GWAS datasets. Cutoff values for sample call rate and SNP call rate was set at 90%. Markers with allele distributions deviating from Hardy-Weinberg equilibrium (HWE) (P-value < 0.001) in controls of any of the populations analyzed separately were excluded. Markers with minor allele frequencies (MAFs) lower than 1% were also filtered out. The top 5 principal components were estimated and individuals deviating more than six standard deviations (SDs) from the cluster centroids in the first three principal components were considered outliers. In addition, duplicate pairs or highly related individuals among datasets were also removed on the basis of pairwise comparisons by using the Genome function in PLINK (Pi-HAT threshold of 0.5) (22). After applying QC criteria, we performed whole-genome genotype imputation using HapMap CEU and TSI populations as reference panels.

Follow-up genotyping

The genotyping of the replication cohorts was performed with either (1) the TaqMan SNP genotyping technology in a LightCycler® 480 Real-Time PCR System (Roche Applied Science, Mannheim, Germany), or (2) the GWAS and Immunochip platforms.

For the SSc study, all cases were genotyped by TaqMan genotyping system using TaqMan 5' allele discrimination predesigned assays from Applied Biosystems. Genotyping

call rate was > 95% for the three SNPs. The control samples were also genotyped by this technology, with the exception of the UK and USA cohorts. For these two control cohorts, genotyping data were obtained from previously published genome-wide genotyping datasets (WTCCC and NARAC2). (18)

RA cases from Spain and New Zealand and Spanish controls were genotyped by TaqMan technology. Genotyping data for New Zealander healthy controls partially overlapped with those from a previous GWAS report (23). For the remaining RA case-control sets, genotype frequencies and association data for the three replication SNPs were obtained from a previously published study (7). Genotype methods of these studies were described in detail in (7). For those cohorts that were genotyped with the Illumina Immunochip platform, only data for *IRF4* rs9328192 were available.

Data analysis

All data were analyzed using PLINK V.1.07 software (22). To test for SNPs association, we performed logistic regression analysis in each of the SSc and RA GWAS cohorts separately. The top three PCs were included as covariates to control for any potential population stratification effects. The disease-specific meta-analyses were performed with inverse-variance weighting under a fixed-effects model. The meta-analysis for the combined-phenotype GWAS analysis was performed considering SSc and RA datasets as a single phenotype. The replication cohorts were also analyzed by logistic regression analysis. The combined analysis of the discovery and replication cohorts was carried out using the inverse variance method based on population specific logistic regression results. Heterogeneity of the ORs across studies was assessed using Cochran's Q test. Hardy-Weinberg equilibrium was tested for all the validation cohorts genotyped by TaqMan technology (HWE *P*-values < 0.01 were considered to show significant deviation

from the equilibrium). None of the included control cohorts showed significant deviation from HWE for all the genotyped SNPs, with the exception of *HNF1A* rs10774577. The cohorts that failed HWE were excluded for the analysis of this specific SNP. The Manhattan plots were obtained with an in-house modification of the R script written by Stephen Turner.

Functional annotations and enrichment analysis

We performed functional protein association analyses considering all the shared risk factors described to date between SSc and RA in the European population. For this purpose, we only included firmly associated loci for both diseases, which were selected on the basis of the following criteria: (1) loci associated at the genome-wide significance level (*P*-value 5 x 10^{-8}), (2) those that reached second tier level associations (*p*-value < 5x10^{-5}), or (3) that have been replicated in independent studies. In total, 14 well-established SSc-RA loci were included (Figure 3).

Protein-protein interaction (PPI) analysis. PPIs among the 14 SSc-RA loci were interrogated using STRING V.10 that provides a critical integration of protein–protein interactions, including direct (physical) as well as indirect (functional) associations. (24) We used the *Homo sapiens* STRING database and applied a confidence score of 0,400.

Molecular pathway enrichment analysis. We conducted molecular pathway enrichment analyses using the Gene Set Enrichment Analysis (GSEA) and DAVID approaches. (25-27) These bioinformatics tools evaluate the overlap of a specific gen set with gen sets from the MSigDB collections. The statistical significance of the overrepresentation of functional annotation terms is calculated on the basis of a hypergeometric testing. Two MSigDB collections (Biocarta and Reactome collections) were used in our analysis. A False

Discovery Rate (FDR) correction was applied for the results from GSEA-based results, and Bonferroni correction was applied for DAVID-based results.

RESULTS

Discovery analysis.

In the stage I of this study we conducted a cross-disease meta-analysis using an inverse variance-weighted method in order to identify new putatively shared loci between SSc and RA. The SSc and RA genome-wide datasets used for discovery comprised 2,716 cases/5,666 controls and 3,821 cases/3,075 controls, respectively. The overall workflow of the study is illustrated in Figure 1.

The meta-analysis combining both datasets identified several SNPs from seven distinct genomic regions showing a *P*-value < 5 x 10^{-6}, as well as a nominal signal of association (*P*-value < 0.05) in the disease-specific analyses (Table 1, Supplementary Table 2). The strongest association was found in the well accepted SSc and RA associated locus *IRF5* ($P_{combined}$ = 8.44 x 10^{-17}; SSc P_{GWAS} = 1.14 x 10^{-16}, OR = 1.52; RA P_{GWAS} = 7.86 x 10^{-4}, OR = 1.20). Three additional SSc-RA known loci, namely *PTPN22*, *ATG5* and *BLK*, were also identified at this stage. The remaining SNPs were located in three different loci: *FBN2* and *HNF1A* that had not been previously reported as genetic risk factors for SSc and RA; and *IRF4*, associated with RA in previous studies (Figure 2, Supplementary Table 2). These putatively new shared SNPs were selected for follow-up in additional SSc and RA replication cohorts. For *IRF4*, three SNPs met our criteria for being selected for validation in the replication step. In this case, we selected the SNP with the lowest *P*-value.

Replication Phase and meta-analysis.

According to the established thresholds (see Material and Methods section for more details), we identified one new association signal shared between SSc and RA at *IRF4* for SNP rs9328192 ($P_{combined}$ = 3.29 x 10-^{12}). Furthermore, this *IRF4* SNP almost reached

genome-wide significance in the meta-analysis for each disease separately (SSc $P_{GWAS+Repl}$ = 2.78 x 10^{-7}, OR = 0.90; RA $P_{GWAS+Repl}$ = 1.44 x 10^{-6}, OR = 1.08) (Table 1). We also observed significant association signals for *HNF1A* rs10774577 in the SSc analysis (SSc P_{Repl} = 0.036, OR = 0.94; SSc $P_{GWAS+Repl}$ = 1.64 x 10^{-4}, OR = 0.91), although no evidence of association was observed in the RA analysis. The initial signal observed for *FBN2* rs6897611 in the original scan showed no significant associations in the SSc and RA replication cohort datasets.

Genetic connections across the SSc-RA shared loci

In order to gain insight into the common aetiopathogenic factors that underlie SSc and RA, we performed functional protein association analyses considering the 14 shared risk factors described to date between these two diseases (For a more detailed description of the included loci, see the Methods section).

PPI analysis across the 14 well-established SSc-RA shared risk factors showed that our network was significantly enriched in interactions (*P*-value < 5 x 10^{-10}), which implies that the proteins encoded by the SSc-RA risk loci interact with each other more than expected by chance (Figure 3). This enrichment in interactions suggests common altered pathways in SSc and RA. Therefore, we conducted molecular pathway enrichment analyses using GSEA and DAVID. These analyses identified significant overrepresentations of several gen sets that were mainly involved in the immune responses, the type I interferon (IFN) signaling pathway and the interleukin 12 (IL-12) signaling pathway (Table 2).

DISCUSSION

In the present study we have identified one new non-HLA susceptibility loci shared between SSc and RA, namely *IRF4*, by a combined-phenotype GWAS strategy in large case-control cohorts of SSc and RA. This locus, *IRF4*, was already known to be involved in RA susceptibility, but had not been previously associated with SSc. (7)

The cross-disease meta-analysis performed with the SSc and RA GWAS datasets identified several SNPs from seven different loci that met our stringent selection criteria for the replication phase ($P_{combined} < 5 \times 10^{-6}$; SSc $P_{GWAS} < 0.05$; RA $P_{GWAS} < 0.05$). Four of them were already SSc and RA known risk factors (*PTPN22*, *ATG5*, *IRF5* and *BLK*), thus providing support for the effectiveness of this strategy in the identification of shared risk loci (10, 28). It is worth mentioning that these loci were detected by the two different tests used in the first stage, which were performed in order to detect both same-direction and opposite-direction allelic effects. In fact, the new pleiotropic *IRF4* SNP identified in this study showed opposite effect for SSc and RA (protection and risk effect, respectively).

Interestingly, the signal of association found in *IRF4* almost reached the genome-wide significance level in the meta-analysis for each disease separately (Table 1), besides reaching a *P*-value $< 5 \times 10^{-8}$ in the combined-phenotype meta-analysis. The associated *IRF4* SNP (rs9328192) showed modest effect sizes for SSc and RA. However, we were able to capture this association in our meta-analysis thank to the large cohort used in this study together with the combined-phenotype approach, which allow us to increase the statistical power. This highlights the capability of the combined-phenotype approach in the identification of shared variants with low penetrance, whose associations might have been missed in disease-specific GWASs due to a lack of power.(13)

Interferon regulatory factor 4 (IRF4) belongs to the IRF family of transcription factors and plays a pivotal role in the development and function of several autoimmune-associated cells, such as B cells, T regulatory cells (Treg), T helper cells (Th) and plasma cells. (29-32) Various genetic and functional studies have pointed to IRF4 as a master regulator for autoimmunity.(33-34) It has been demonstrated that IRF4 is a crucial factor for the editing and L-chain rearrangements of the B cell receptor, and the pre-B cell expansion, which are processes directly related with the development of autoimmunity. (35-36) In addition, IRF4 is a critical controller of the T helper 17 cells (Th17) differentiation (32, 37) and the production of interleukin (IL) 17 and 21 (33, 38), which are immune system components that play a key role in the pathogenesis of SSc and RA. (39-42)

The results of the present study add another IRF to the list of IRFs associated with SSc (*IRF4*, *IRF5*, *IRF7* and *IRF8*) and RA (*IRF4*, *IRF5* and *IRF8*) (7, 10). IRFs are transcriptional regulators of type I IFNs and IFN-inducible genes (43). Therefore, these findings provide genetic support for the IFN signature described for SSc and RA patients (44-45). Moreover, our pathway enrichment analysis also identified the type I IFN signaling pathway as one of the most relevant common pathways between SSc and RA on the basis of their common genetic background. Therefore, deregulation of this signaling pathway might be a biological process underlying the onset of these two autoimmune rheumatic conditions.

Regarding *HNF1A* and *FBN2* genetic variants, despite the initial suggestive association signals found in the first step of the present study, these *loci* did not showed genome-wide significance in our combined-phenotype meta-analysis. Nevertheless, *HNF1A* showed suggestive evidence of association in the meta-analysis performed in SSc alone. Thus,

further confirmations through independent studies are needed to confirm and establish this locus as a true susceptibility factor for SSc.

In summary, through a cross-disease meta-analysis of GWASs for SSc and RA, we were able to identify a new shared locus for SSc and RA, namely *IRF4*. The present study, together with previous reports, reinforces the idea of the common genetic background for SSc and RA. The identification of these pleiotropic autoimmunity loci may point to common pathogenic pathways, which ultimately may represent a clinical advantage, thus providing support for drug repositioning on the basis of the true understanding of pathogenic mechanisms.

FIGURES

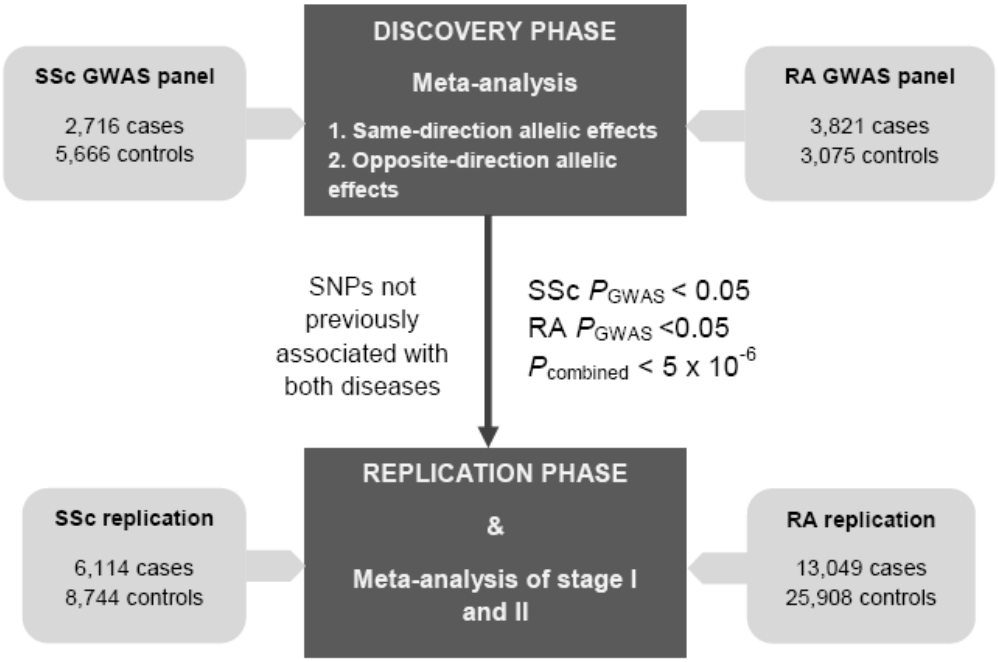

Figure 1. Overall workflow of the present study.

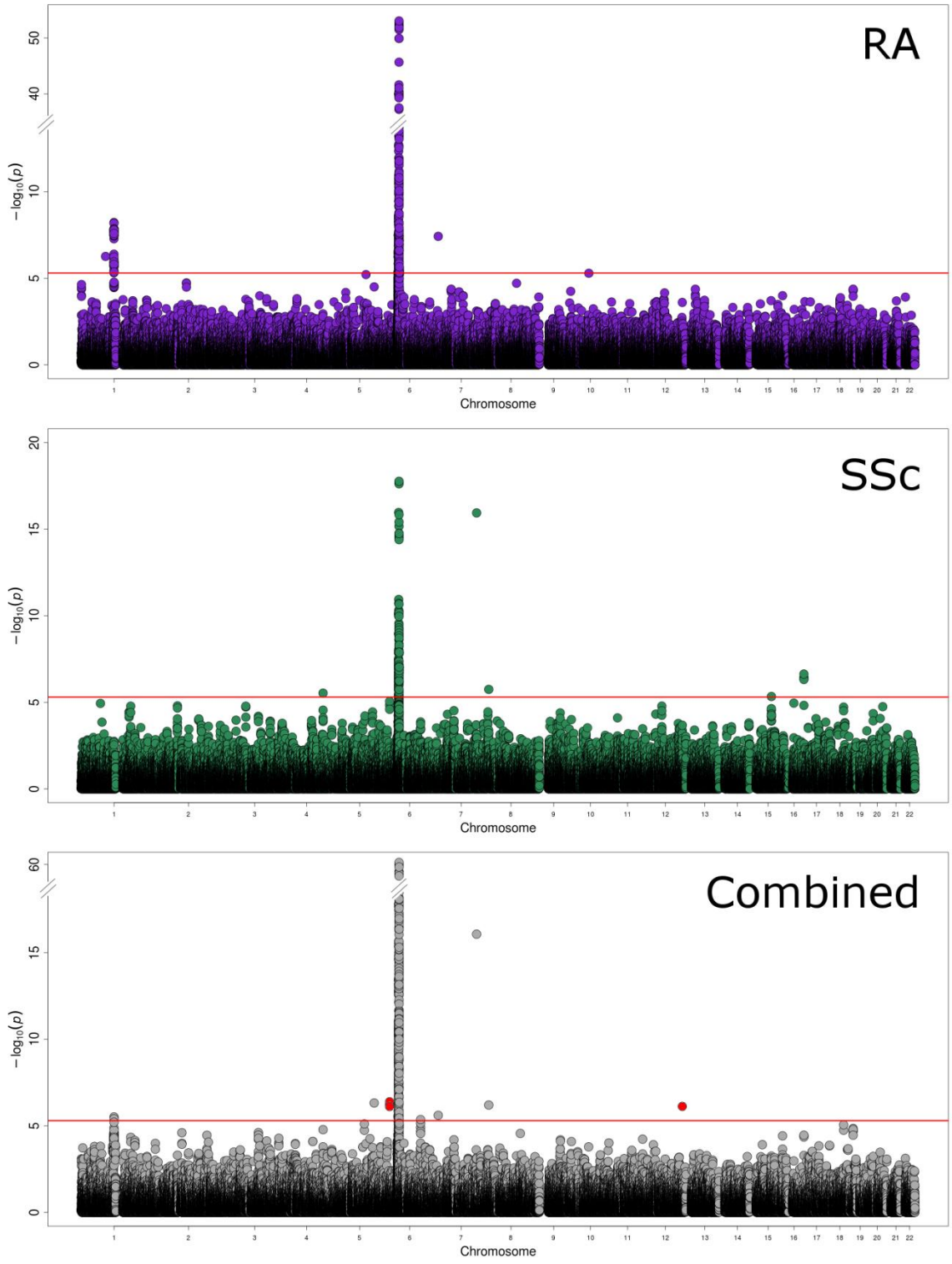

Figure 2. Manhattan plot showing the results of the SSc GWAS and RA GWAS panels, and the results of the cross-disease meta-GWAS.

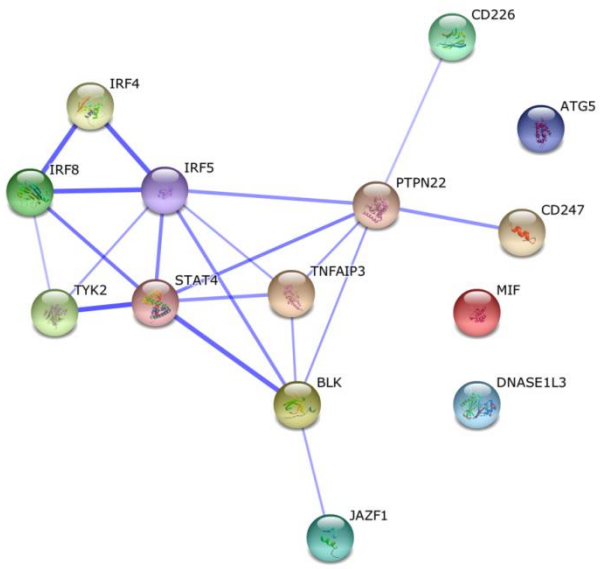

Figure 3. Protein protein interaction (PPI) network accros the 14 well-established SSc-RA *loci* in STRING. The plot shows the 'confidence' view. Thicker lines represent stronger associations.

TABLES (excel file)

Table 1. Association results of the cross-disease meta-GWAS for three selected SNPs.

Table 2. Gene sets that showed a significant enrichment P-value (P-value < 0.05) after False Discovery Rate correction in GSEA-based analysis or Bonferroni correction in DAVID-based analysis

Table Supp. 2. List of SNPs with a combined meta-analysis *P* value lower than $5x10^{-6}$, SSc *P* < 0.05 and RA *P* < 0.05 in the discovery phase.

ACKNOWLEDGMENTS

We thank Sofia Vargas, Sonia García and Gema Robledo for her excellent technical assistance and all the patients and control donors for their essential collaboration. We thank National DNA Bank Carlos III (University of Salamanca, Spain) who supplied part of the control DNA samples. We also would like to thank the following organizations: The EULAR Scleroderma Trials and Research group (EUSTAR), the German Network of Systemic Sclerosis, The Scleroderma Foundation (USA) and RSA (Raynauds & Scleroderma Association).

COMPETING INTERESTS.

None

FUNDING

This work was supported by the following grants: JM was funded by SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness, the EU/EFPIA Innovative Medicines Initiative Joint Undertaking PRECISESADS (ref: 115565) and BIO-1395 from Junta de Andalucía. NO was funded by PI-0590-2010, from Consejería de Salud y Bienestar Social, Junta de Andalucía, Spain. ELI was supported by Ministerio de Educación, Cultura y Deporte through the program FPU. TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). Study on USA samples were supported by the Institutes of Health (NIH) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Centers of Research Translation (CORT) grant P50AR054144 (MDM), the NIH-NIAMS SSc Family Registry and DNA Repository (N01-AR-0-2251) (MDM), NIH-KL2RR024149-04 (SA), NIH-NCRR 3UL1RR024148, US NIH NIAID UO1 1U01AI09090, K23AR061436 (SA), Department of Defense PR1206877 (MDM) and NIH/NIAMS-RO1- AR055258 (MDM).

REFERENCES

1. Cho JH, Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med* 2011;365:1612-23.
2. Welter D, MacArthur J, Morales J, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res* 2014;42:D1001-6.
3. Burton PR, Clayton DG, Cardon LR, et al. Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* 2007;39:1329-37.
4. Parkes M, Cortes A, van Heel DA, et al. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet* 2013;14:661-73.
5. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009;360:1989-2003.
6. McInnes IB, Schett G. The pathogenesis of rheumatoid arthritis. *N Engl J Med* 2011;365:2205-19.
7. Okada Y, Wu D, Trynka G, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* 2014;506:376-81.
8. Radstake TR, Gorlova O, Rueda B, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet* 2010;42:426-9.
9. Mayes MD, Bossini-Castillo L, Gorlova O, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet* 2014;94:47-61.
10. Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: Defining heritability, functional variants and shared-autoimmunity pathways. *J Autoimmun* 2015 Jul 23.
11. Elhai M, Avouac J, Kahan A, et al. Systemic sclerosis at the crossroad of polyautoimmunity. *Autoimmun Rev* 2013;12:1052-7.
12. 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:e1002004.
13. 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:e1001283.
14. 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;90:636-47.
15. 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;22:4021-9.
16. Li YR, Li J, Zhao SD, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 2015;21:1018-27.

17. Martin JE, Broen JC, Carmona FD, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. *Hum Mol Genet* 2012;21:2825-35.

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;42:508-14.

19. Anon MC. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum* 1980;23:581-90.

20. LeRoy EC, Medsger TA, Jr. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001;28:1573-6.

21. Arnett FC, Edworthy SM, Bloch DA, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315-24.

22. 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;81:559-75.

23. Consortium TAaNZMSG. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet* 2009;41:824-8.

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:D447-52.

25. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.

26. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44-57.

27. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1-13.

28. Messemaker TC, Huizinga TW, Kurreeman F. Immunogenetics of rheumatoid arthritis: Understanding functional implications. *J Autoimmun* 2015.

29. Mittrucker HW, Matsuyama T, Grossman A, et al. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* 1997;275:540-3.

30. Klein U, Casola S, Cattoretti G, et al. Transcription factor IRF4 controls plasma cell differentiation and class-switch recombination. *Nat Immunol* 2006;7:773-82.

31. Zheng Y, Chaudhry A, Kas A, et al. Regulatory T-cell suppressor program co-opts transcription factor IRF4 to control T(H)2 responses. *Nature* 2009;458:351-6.

32. Brustle A, Heink S, Huber M, et al. The development of inflammatory T(H)-17 cells requires interferon-regulatory factor 4. *Nat Immunol* 2007;8:958-66.

33. Xu WD, Pan HF, Ye DQ, et al. Targeting IRF4 in autoimmune diseases. *Autoimmun Rev* 2012;11:918-24.

34. Biswas PS, Gupta S, Stirzaker RA, et al. Dual regulation of IRF4 function in T and B cells is required for the coordination of T-B cell interactions and the prevention of autoimmunity. *J Exp Med* 2012;209:581-96.

35. Zouali M. Transcriptional and metabolic pre-B cell receptor-mediated checkpoints: implications for autoimmune diseases. *Mol Immunol* 2014;62:315-20.

36. Zouali M. Receptor editing and receptor revision in rheumatic autoimmune diseases. *Trends Immunol* 2008;29:103-9.

37. Ciofani M, Madar A, Galan C, et al. A validated regulatory network for Th17 cell specification. *Cell* 2012;151:289-303.

38. Biswas PS, Gupta S, Chang E, et al. Phosphorylation of IRF4 by ROCK2 regulates IL-17 and IL-21 production and the development of autoimmunity in mice. *J Clin Invest* 2010 Sep;120(9):3280-95.

39. Fenoglio D, Bernuzzi F, Battaglia F, et al. Th17 and regulatory T lymphocytes in primary biliary cirrhosis and systemic sclerosis as models of autoimmune fibrotic diseases. *Autoimmun Rev* 2012;12:300-4.

40. Yang X, Yang J, Xing X, et al. Increased frequency of Th17 cells in systemic sclerosis is related to disease activity and collagen overproduction. *Arthritis Res Ther* 2014;16:R4.

41. McInnes IB, Schett G. Cytokines in the pathogenesis of rheumatoid arthritis. *Nat Rev Immunol* 2007;7:429-42.

42. Toh ML, Miossec P. The role of T cells in rheumatoid arthritis: new subsets and new targets. *Curr Opin Rheumatol* 2007;19:284-8.

43. Ivashkiv LB, Donlin LT. Regulation of type I interferon responses. *Nat Rev Immunol* 2014;14:36-49.

44. 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;70:2029-36.

45. Ronnblom L, Eloranta ML. The interferon signature in autoimmune diseases. *Curr Opin Rheumatol* 2013;25:248-53.

Supplementary note

Spanish Scleroderma Group

Raquel Ríos and **Jose Luis Callejas**, Unidad de Enfermedades Sistémicas Autoinmunes, Department of Internal Medicine, Hospital Clínico Universitario San Cecilio, Granada; **José Antonio Vargas Hitos**, Department of Internal Medicine, Hospital Virgen de las Nieves, Granada; **Rosa García Portales**, Department of Rheumatology, Hospital Virgen de la Victoria, Málaga; **María Teresa Camps**, Department of Internal Medicine, Hospital Carlos Haya, Málaga; **Antonio Fernández-Nebro**, Department of Rheumatology, Hospital Carlos Haya, Málaga; **María F. González-Escribano**, Department of Immunology, Hospital Virgen del Rocío, Sevilla; **Francisco José García-Hernández** and **Mª Jesús Castillo**, Department of Internal Medicine, Hospital Virgen del Rocío, Sevilla; **Mª Ángeles Aguirre** and **Inmaculada Gómez-Gracia**, Department of Rheumatology, Hospital Reina Sofía/IMIBIC, Córdoba; **Luis Rodríguez-Rodríguez**, Department of Rheumatology, Hospital Clínico San Carlos, Madrid; **Paloma García de la Peña,** Department of Rheumatology, Madrid Norte Sanchinarro Hospital, Madrid; **Esther Vicente,** Department of Rheumatology, Hospital La Princesa, Madrid; **José Luis Andreu** and **Mónica Fernández de Castro**, Department of Rheumatology, Hospital Puerta de Hierro Majadahonda, Madrid; **Francisco Javier López-Longo** and **Lina Martínez**, Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Madrid; **Vicente Fonollosa and Alfredo Guillén**, Department of Internal Medicine, Hospital Valle de Hebrón, Barcelona; **Iván Castellví,** Department of Rheumatology, Santa Creu i Sant Pau University Hospital, Barcelona; **Gerard Espinosa**, Department of Internal Medicine, Hospital Clinic, Barcelona; **Carlos Tolosa**, Department of Internal Medicine, Hospital Parc Tauli, Sabadell; **Anna Pros**, Department of Rheumatology, Hospital Del Mar, Barcelona; **Mónica Rodríguez Carballeira**, Department of Internal Medicine, Hospital Universitari Mútua Terrasa, Barcelona; **Francisco Javier Narváez**, Department of Rheumatology, Hospital Universitari de *Bellvitge*, Barcelona; **Manel Rubio Rivas**, Department of Internal Medicine, Hospital Universitari de *Bellvitge*, Barcelona; **Vera Ortiz Santamaría**, Department of Rheumatology, Hospital General de Granollers, Granollers; **Ana Belén Madroñero,** Department of Internal Medicine, Hospital General San Jorge, Huesca; **Bernardino Díaz** and **Luis Trapiella**, Department of Internal Medicine, Hospital Central de Asturias, Oviedo; **Adrián Sousa**, Unidad de Trombosis y Vasculitis, Department of Internal Medicine, Hospital Xeral-Complexo

Hospitalario Universitario de Vigo, Vigo; **María Victoria Egurbide**, Department of Internal Medicine, Hospital Universitario Cruces, Barakaldo; **Patricia Fanlo Mateo**, Department of Internal Medicine Hospital Virgen del Camino, Pamplona; **Luis Sáez-Comet**, Unidad de Enfermedades Autoinmunes Sistémicas, Department of Internal Medicine, Hospital Universitario Miguel Servet, Zaragoza; **Federico Díaz** and **Vanesa Hernández**, Department of Rheumatology, Hospital Universitario de Canarias, Tenerife; **Emma Beltrán**, Department of Rheumatology, Hospital General Universitario de Valencia, Valencia; **José Andrés Román-Ivorra and Elena Grau**, Department of Rheumatology, Hospital Universitari i Politecnic La Fe, Valencia. **Juan José Alegre Sancho**, Department of Rheumatology, Hospital del Doctor Peset, Valencia. **Francisco J. Blanco García** and **Natividad Oreiro**, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña.