Running head: **IRF4**, a new SSc-RA shared locus

**Title: 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-Isac¹*, Jose-Ezequiel Martín¹, Shervin Assassi², Carmen Pilar Simeón³, Patricia Carreira⁴, Norberto Ortego-Centeno⁵, Mayka Freire⁶, Emma Beltrán⁷, Javier Narváez⁸, Juan José Alegre-Sancho⁹, the Spanish Scleroderma Group¹⁰, Benjamín Fernández-Gutiérrez¹¹, Alejandro Balsa¹², Ana M Ortiz¹³, Miguel A González-Gay¹⁴, Lorenzo Beretta¹⁵, Alessandro Santaniello¹⁵, Chiara Bellocchi¹⁵, Claudio Lunardi¹⁶, Gianluca Moroncini¹⁷, Armando Gabrielli¹⁷, Torsten Witte¹⁸, Nicolas Hunzelmann¹⁹, Jörg H.W. Distler²⁰, Gabriella Riekemasten²¹, Annete H van der Helm-van Mil²², Jeska de Vries-Bouwstra²², Cesar Magro-Checa²², Alexandre E. Voskuyl²³, Madelon C Vonk²⁴, Øyvind Molberg²⁵, Tony Merriman²⁶, Roger Hesselstrand²⁷, Annika Nordin²⁸, Leonid Padyukov²⁸, Ariane Herrick²⁹, Steve Eyre²⁹, Bobby PC Koeleman³⁰, Christopher P. Denton³¹, Carmen Fonseca³¹, Timothy RDJ Radstake³², Jane Worthington²⁹, Maureen D. Mayes², Javier Martín¹*

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 Universitario Doctor Peset, Valencia, Spain.
10. See supplementary note.
11. Rheumatology Service, Hospital Clínico San Carlos, Madrid, Spain.
12. Rheumatology Unit, Hospital Universitario La Paz, Instituto de Investigación Sanitaria La Paz. IdiPAZ. Madrid, Spain.
13. Rheumatology Service, Hospital Universitario La Princesa, Instituto de Investigación Sanitaria La Princesa, Madrid, Spain.
14. Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, DIVAL, University of Cantabria, 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. Clinic of Rheumatology, University of Lübeck, Lübeck, 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. Rheumatology Unit, Oslo University Hospital Rikshospitalet and Institute of Clinical Medicine, University of Oslo, 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

**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 U011U01AI09090, K23AR061436 (SA), Department of Defense PR1206877 (MDM) and NIH/NIAMS-RO1-AR055258 (MDM).
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 INTEREST.

None.
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 using a strategy that allowed identification of loci with both same-direction and opposing-direction allelic effects. 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 x 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 identified 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 etiopathogenic factors.

Conclusions: Our study has identified a novel shared locus, IRF4, for SSc and RA and highlighted the usefulness of cross-disease GWAS meta-analysis in the identification of common risk loci.

KEYWORDS

Systemic sclerosis, rheumatoid arthritis, genome-wide association study, shared loci.
INTRODUCTION

Genome-wide association studies (GWASs) and immune-focused fine-mapping studies have revolutionized our understanding of the genetic component of complex autoimmune diseases (ADs) by the identification of thousands of susceptibility loci associated with autoimmunity (1). The vast majority of these loci are shared risk factors for at least two or more ADs, pointing to a common genetic background underlying autoimmune processes. This genetic overlap has been suspected some time ago, given the high rate of co-occurrence of ADs and the well-established familial aggregation reported for these immune disorders (1).

Systemic sclerosis (SSc) and rheumatoid arthritis (RA) are complex ADs which share clinical and immunological features. Both diseases are rheumatic connective tissue disorders, characterized by an exacerbated inflammatory response, deregulation of innate and adaptive immunity, including autoantibody production, and systemic complications. Thanks to the establishment of large consortiums and international collaborations, the number of confirmed RA susceptibility factors has increased up to a total of 101 loci associated with the disease at the genome-wide significance level (2). In regard to SSc, GWASs, Immunochip and candidate gene studies have clearly identified various genetic regions involved in SSc susceptibility (3). However, the knowledge of the genetic predisposition to this disease is relatively limited, in part due to its low prevalence, which impairs the recruitment of large cohorts required to reach a high statistical power and to effectively detect association signals. Interestingly, a considerable proportion of the SSc susceptibility factors also represent RA risk loci (2-3). In addition, although not very common, co-familiarity and co-occurrence between these two rheumatic conditions have
been observed (4). These observations provide evidence of a genetic overlap 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 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. This strategy has been successfully applied not only to the study of closely related phenotypes but also to non-related phenotypes, showing encouraging results (5). Taking into account all these considerations, the purpose of the present study was to systematically identify new common risk loci for SSc and RA by applying the combined-phenotype GWAS strategy, followed by replication testing in independent case-control datasets.
METHODS

Study population

The stage I of the present study included 6,537 SSc/RA patients and 8,741 healthy controls. The 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 (5-7). The 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) (8).

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) (8). 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, RACI-US, RACI-i2b2, CORRONA, Vanderbilt, RACI-UK, RACI-SE-U, RACI-NL, Dutch (AMC, BeSt, LUMC, and DREAM), ReAct, ACR-REF) (2). All SSc and RA patients fulfilled previously described classification criteria for SSc and RA respectively (2, 5). All individuals enrolled in the present study provided written informed consent and approval from the local ethical committees was obtained from all the centers in accordance with the tenets of the Declaration of Helsinki.

Study design
In the present study, we performed a two-stage study to systematically identify SSc-RA shared risk factors (Figure 1).

**Stage I.** We performed GWAS analysis for each disease separately and a combined-phenotype GWAS analysis. Two different tests were considered for the combined analysis (5):

1) 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 x 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 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 were followed.

For both sorts of meta-analyses, 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 disease but not reported for the other.

**Stage II.** The selected SNPs were followed-up in independent replication cohorts. Subsequently, we performed a meta-analysis of the initial GWAS screening and replication stages. 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 quality control (QC) criteria in all the GWAS datasets. Cutoff values for sample call rate and SNP call rate were set up at 95%. Markers with allele distributions deviating from Hardy-Weinberg equilibrium (HWE) ($P$-value < 0.001) in controls from any of the populations analyzed separately were excluded. Markers with minor allele frequencies (MAF) lower than 1% were filtered out. After QC, we performed whole-genome genotype imputation with IMPUTE2 software (9) using the CEU and TSI populations of the HapMap Phase 3 project as reference panels (http://www.hapmap.org). Imputed SNP quality was assessed by establishing a probability threshold for merging genotypes at 0.9. Subsequently, stringent QC was applied to the imputed data using the same criteria stated above. After that, genome-wide genotyping data were available for a total of 219,756 SNPs. The first 5 principal components (PC) were estimated and individuals deviating more than six standard deviations (SDs) from the cluster centroids 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 v1.7 (http://pngu.mgh.harvard.edu/purcell/plink/) (Pi-HAT threshold of 0.5).

**Follow-up genotyping**

The genotyping of the replication cohorts was performed with either (1) 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) (8).

RA cases from Spain and New Zealand and Spanish controls were genotyped by TaqMan technology. Genotype data for New Zealand healthy controls partially overlapped with those from a previous GWAS report (10). For the remaining RA case-control sets, genotype frequencies and association data were obtained from a previously published study (2). Genotype methods of these studies were described in detail in Okada et al. (2). 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. To test for association, we performed logistic regression analysis in each of the SSc and RA GWAS cohorts separately. The first 5 PC were included as covariates to control for any potential population stratification effects. The replication cohorts were also analyzed by logistic regression analysis. The meta-analyses were performed with inverse-variance method based on population specific logistic regression results. Heterogeneity of the ORs across studies was assessed using Cochran’s Q test. HWE was tested for all the validation cohorts genotyped by TaqMan technology (HWE P-values < 0.01 were considered to show significant deviation from equilibrium). None of the included control cohorts showed significant deviation from HWE, with the exception of HNF1A rs10774577. The cohorts that failed HWE were excluded for the analysis of this specific SNP. The statistical power of the combined analysis is shown in Supp. Table 1.

RESULTS

Discovery analysis.
In the stage I of this study we conducted a cross-disease meta-analysis in order to systematically identify new putatively shared loci between SSc and RA. The overall workflow of the study is illustrated in Figure 1.

The meta-analysis combining both datasets identified various SNPs from seven distinct genomic regions showing a $P$-value $< 5 \times 10^{-6}$, as well as a nominal signal of association ($P$-value $< 0.05$) in the disease-specific analyses. The strongest association was found in the well accepted SSc and RA associated locus $IRF5$ ($P_{\text{combined}} = 8.44 \times 10^{-17}$; SSc $P_{\text{GWAS}} = 1.14 \times 10^{-16}$; RA $P_{\text{GWAS}} = 7.86 \times 10^{-4}$). Three additional SSc-RA known loci, namely $PTPN22$, $ATG5$ and $BLK$, were also identified at this stage (Figure 2, Supp. Table 2 and Supp. Figure 1). 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 (Table 1, Figure 2, and Supp. Figure 2). Interestingly, the regional association plots of $FBN2$, $IRF4$ and $HNF1A$ loci showed that the top SNPs of the combined analysis were also the top SNPs in the analyses for SSc and RA separately, or at least were in high linkage disequilibrium with the top signal observed for each disease (Supp. Figure 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 (Supp. Table 2).

**Replication Phase and meta-analysis.**

According to the established thresholds (see Methods section for more details), we identified one new association signal shared between SSc and RA at $IRF4$ for SNP rs9328192 ($P_{\text{combined}} = 3.29 \times 10^{-12}$). Furthermore, this $IRF4$ SNP almost reached genome-
wide significance in the meta-analysis for each disease separately (SSc $P_{GWAS+Repl} = 2.78 \times 10^{-7}$, OR = 0.90; RA $P_{GWAS+Repl} = 1.44 \times 10^{-6}$, OR = 1.08) (Table 1).

Regarding $HNF1A$ and $FBN2$ genetic variants, despite the initial suggestive association signals found in the first stage, these loci did not show genome-wide significance in our combined-phenotype meta-analysis. Nevertheless, $HNF1A$ rs10774577 showed suggestive evidence of association in the meta-analysis performed in SSc alone (SSc $P_{Repl} = 0.036$, OR = 0.94; SSc $P_{GWAS+Repl} = 1.64 \times 10^{-4}$, OR = 0.91), and a $P$-value of $1.59 \times 10^{-6}$ in the combined-phenotype meta-analysis. Considering that this SNP was not included in those cohorts that were genotyped with Immunochip, the present study had a lower statistical power for the analysis of this genomic region. Therefore, a slight or modest genetic effect of $HNF1A$ rs10774577 cannot be ruled out and further studies will be required to establish whether this locus is a shared SSc–RA risk factor.
DISCUSSION

In the present study we have identified a novel non-HLA susceptibility locus 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 reported to be involved in RA susceptibility, but had not been previously associated with SSc (2).

The cross-disease meta-analysis performed with the SSc and RA GWAS datasets identified various SNPs from seven different loci that met our stringent selection criteria for the replication phase ($P_{\text{combined}} < 5 \times 10^{-6}$; SSc $P_{\text{GWAS}} < 0.05$; RA $P_{\text{GWAS}} < 0.05$). Four of them were already SSc and RA risk factors ($\text{PTPN}22$, $\text{ATG}5$, $\text{IRF}5$ and $\text{BLK}$), thus providing support for the effectiveness of this strategy in the identification of shared risk loci (2-3). 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 shared IRF4 SNP identified in this study showed opposite effects for SSc and RA (protection and risk effects, respectively). This discrepancy might be due to the fact that the actual causal variants for the associations in each disease could be different and IRF4 rs9328192 is tagging them. This discordant phenomenon is particularly common between ADs. (1). However, to completely understand these discordant effects, the interaction with other genetic variants contributing to disease susceptibility should be considered, besides analyze the precise biological impact of the associations.

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 thanks to the large cohort used in this study together with the combined-phenotype approach, which allowed 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 (11).

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 (12). Various genetic and functional studies have pointed to IRF4 as a master regulator for autoimmunity (12-13). 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 (14). In addition, IRF4 is a critical controller of the T helper 17 cells (Th17) differentiation and the production of interleukin (IL) 17 and 21(12), which are immune system components that play a key role in the pathogenesis of SSc and RA.

The results of the present study add another interferon regulatory factor to the list of IRFs associated with SSc (IRF4, IRF5, IRF7 and IRF8) and RA (IRF4, IRF5 and IRF8) (2-3), thus providing genetic support for the IFN signature described for SSc and RA patients(15). 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 (Supp. Methods, Supp. Figure 3, Supp. Table 3). Therefore, deregulation of this signaling pathway might be a biological process underlying the onset of these two autoimmune rheumatic conditions.

In summary, through a cross-disease meta-analysis of GWASs for SSc and RA, we were able to identify IRF4 as a new shared susceptibility locus for SSc and RA. The present study, together with previous reports, reinforces the idea of a common genetic background between 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 the pathogenic mechanisms.
REFERENCES


## Table

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

<table>
<thead>
<tr>
<th>Locus (Chr)</th>
<th>SNP</th>
<th>Ref. Allele</th>
<th>SSc_GWAS P</th>
<th>SSc_GWAS OR⁺</th>
<th>SSc_Repl P</th>
<th>SSc_Repl OR⁺</th>
<th>SSc_GWAS+Repl P</th>
<th>SSc_GWAS+Repl OR⁺</th>
<th>RA_GWAS P</th>
<th>RA_GWAS OR⁺</th>
<th>RA_Repl P</th>
<th>RA_Repl OR⁺</th>
<th>RA_GWAS+Repl P</th>
<th>RA_GWAS+Repl OR⁺</th>
<th>SSc-RA GWAS+Repl Status now</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FBN2</strong></td>
<td>(5)</td>
<td>rs6897611</td>
<td>T</td>
<td>4.79E-07</td>
<td>2.85E-03</td>
<td>1.16</td>
<td>0.641</td>
<td>0.98</td>
<td>0.165</td>
<td>1.04</td>
<td>3.15E-05</td>
<td>1.24</td>
<td>0.684</td>
<td>0.99</td>
<td>0.650*</td>
</tr>
<tr>
<td><strong>IRF4</strong></td>
<td>(6)</td>
<td>rs9328192</td>
<td>G</td>
<td>4.06E-07</td>
<td>8.86E-06</td>
<td>0.86</td>
<td>1.89E-03</td>
<td>0.93</td>
<td>2.78E-07</td>
<td>0.90</td>
<td>7.26E-03</td>
<td>1.10</td>
<td>5.22E-05</td>
<td>1.07</td>
<td>1.44E-06</td>
</tr>
<tr>
<td><strong>HNF1A</strong></td>
<td>(12)</td>
<td>rs10774577**</td>
<td>T</td>
<td>7.53E-07</td>
<td>8.62E-04</td>
<td>0.89</td>
<td>0.036</td>
<td>0.94</td>
<td>1.64E-04</td>
<td>0.91</td>
<td>2.50E-04</td>
<td>1.14</td>
<td>0.290</td>
<td>1.03</td>
<td>0.208*</td>
</tr>
</tbody>
</table>

* Odds ratio for the reference allele.
* P-value and OR from meta-analysis under random effects due to heterogeneity of the ORs among cohorts.
** The RA and SSc replication cohorts from Spain, and the SSc replication cohorts from Italy and The Netherlands were excluded from the analysis of rs10774577 due to HWE issues.

Chr, chromosome; GWAS, genome-wide association study; OR, odds ratio; RA, rheumatoid arthritis; Repl, replication; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.
FIGURE LEGENDS

Figure 1. Overall workflow of the present study.

Figure 2. Manhattan plot showing the results of the cross-disease meta-GWAS. The $-\log_{10}$ of the combined-phenotype meta-analysis $P$-values are plotted against their physical chromosomal position. The plot displays the $-\log_{10}$ $P$-values from the same-direction meta-analysis of SSc and RA. The signals from the opposite-direction meta-analysis that reached the selection criteria are also plotted (red points). The red line represents the threshold at $5 \times 10^{-6}$. Those loci with SNPs that reached the selection criteria for the replication phase are plotted (loci selected for follow-up are highlighted in pink).
SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

Biological connection across SSc-RA shared loci

In order to gain insight into the common etiopathogenic factors that underlie SSc and RA, we performed functional protein association analyses considering the shared risk factors described to date between SSc and RA. 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 $5 \times 10^{-8}$), (2) those that reached second tier level associations ($p$-value $< 5 \times 10^{-5}$), or (3) that have been replicated in independent studies. In total, 14 well-established SSc-RA loci were included (Supp. 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 PPIs, including direct (physical) as well as indirect (functional) associations. (1) A confidence score of 0.400 was applied.

Molecular pathway enrichment analysis. We conducted molecular pathway enrichment analyses using the Gene Set Enrichment Analysis (GSEA) and DAVID approaches (2-4). These 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) were used. A False Discovery Rate (FDR) correction was applied for the results from GSEA-based results, and Bonferroni correction was applied for DAVID-based results.
LEGENDS OF SUPPLEMENTARY FIGURES

Supp. Figure 1. Manhattan plots summarizing the results of the genome-wide association studies of systemic sclerosis (SSc) and rheumatoid arthritis (RA).

Supp. Figure 2. Regional association plots of the three loci selected for replication. The selected SNPs are represented as a purple diamond.
A) FBN2 associations in the combined-phenotype meta-analysis (same-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).
B) IRF4 associations in the combined-phenotype meta-analysis (opposite-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).
C) HNF1A associations in the combined-phenotype meta-analysis (opposite-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).

Supp. Figure 3. Protein protein interaction (PPI) network across the 14 well-established SSc-RA loci in STRING. The network was significantly enriched in interactions (P-value < 5 x 10^{-10}). Thus the proteins encoded by the SSc-RA risk loci interact with each other more than expected by chance, suggesting common altered pathways in SSc and RA. The plot shows the ‘confidence’ view. Thicker lines represent stronger associations.
SUPPLEMENTARY TABLES

Supplementary Table 1. Overall statistical power of the study in both the combined analysis and the analysis for each disease separately.

<table>
<thead>
<tr>
<th>Study</th>
<th>Cases/Controls</th>
<th>OR 1.2 MAF 0.40</th>
<th>OR 1.2 MAF 0.30</th>
<th>OR 1.2 MAF 0.20</th>
<th>OR 1.2 MAF 0.10</th>
<th>OR 1.3 MAF 0.40</th>
<th>OR 1.3 MAF 0.30</th>
<th>OR 1.3 MAF 0.20</th>
<th>OR 1.3 MAF 0.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSc+RA</td>
<td>23,240/45,853</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>SSc</td>
<td>8,830/14,410</td>
<td>1.00</td>
<td>0.99</td>
<td>0.93</td>
<td>0.42</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>RA</td>
<td>16,870/28,983</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.50</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.99</td>
</tr>
</tbody>
</table>

MAF, minor allele frequency; OR, odds ratio. Statistical power calculated according to Power Calculator for Genetic Studies 2006 software under an additive model (5).
Supplementary Table 2. List of SNPs with a combined meta-analysis \( P \) value lower than \( 5 \times 10^{-6} \), SSc \( p < 0.05 \) and RA \( p < 0.05 \) in the discovery phase.

<table>
<thead>
<tr>
<th>Chr</th>
<th>Bp</th>
<th>Locus</th>
<th>SNP</th>
<th>Ref. Allele</th>
<th>( P_{\text{combined}} )</th>
<th>( OR^* )</th>
<th>( P ) SSc</th>
<th>OR SSc*</th>
<th>( P ) RA</th>
<th>OR RA*</th>
<th>Previous Association</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>114230863</td>
<td>PTPN22 / C1orf178</td>
<td>rs1746860</td>
<td>C</td>
<td>3.06E-06</td>
<td>0.88</td>
<td>1.79E-03</td>
<td>0.89</td>
<td>4.94E-04</td>
<td>0.88</td>
<td>SSc, RA</td>
</tr>
<tr>
<td>1</td>
<td>114240474</td>
<td>PTPN22 / AP4B1</td>
<td>rs1217401</td>
<td>G</td>
<td>3.90E-06</td>
<td>0.89</td>
<td>2.13E-03</td>
<td>0.89</td>
<td>5.31E-04</td>
<td>0.88</td>
<td>SSc, RA</td>
</tr>
<tr>
<td>5</td>
<td>128011704</td>
<td>FBN2</td>
<td>rs6897611</td>
<td>T</td>
<td>4.79E-07</td>
<td>1.20</td>
<td>2.85E-03</td>
<td>1.16</td>
<td>3.15E-05</td>
<td>1.24</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>106833908</td>
<td>ATG5</td>
<td>rs3804333</td>
<td>T</td>
<td>4.36E-06</td>
<td>1.15</td>
<td>3.48E-04</td>
<td>1.16</td>
<td>3.66E-03</td>
<td>1.13</td>
<td>SSc, RA</td>
</tr>
<tr>
<td>7</td>
<td>128404702</td>
<td>TNPO3/IRF5</td>
<td>rs12531711</td>
<td>G</td>
<td>8.44E-17</td>
<td>1.36</td>
<td>1.14E-16</td>
<td>1.52</td>
<td>7.86E-04</td>
<td>1.20</td>
<td>SSc, RA</td>
</tr>
<tr>
<td>8</td>
<td>11381382</td>
<td>BLK</td>
<td>rs2736340</td>
<td>T</td>
<td>6.29E-07</td>
<td>1.15</td>
<td>1.79E-06</td>
<td>1.20</td>
<td>2.51E-02</td>
<td>1.09</td>
<td>SSc, RA</td>
</tr>
<tr>
<td>6</td>
<td>379364</td>
<td>IRF4</td>
<td>rs2048698</td>
<td>T</td>
<td>5.83E-07</td>
<td>0.89</td>
<td>1.48E-05</td>
<td>0.86</td>
<td>6.77E-03</td>
<td>1.10</td>
<td>RA</td>
</tr>
<tr>
<td>6</td>
<td>379364</td>
<td>IRF4</td>
<td>rs9328192</td>
<td>G</td>
<td>4.06E-07</td>
<td>0.88</td>
<td>8.86E-06</td>
<td>0.86</td>
<td>7.26E-03</td>
<td>1.10</td>
<td>RA</td>
</tr>
<tr>
<td>6</td>
<td>379915</td>
<td>IRF4</td>
<td>rs4256472</td>
<td>A</td>
<td>7.68E-07</td>
<td>0.89</td>
<td>2.54E-05</td>
<td>0.87</td>
<td>5.91E-03</td>
<td>1.10</td>
<td>RA</td>
</tr>
<tr>
<td>12</td>
<td>119848707</td>
<td>HNF1A</td>
<td>rs10774577</td>
<td>T</td>
<td>7.53E-07</td>
<td>0.88</td>
<td>8.62E-04</td>
<td>0.89</td>
<td>2.50E-04</td>
<td>1.14</td>
<td>-</td>
</tr>
</tbody>
</table>

\( ^* \) Odds ratio for the reference allele.

Grey shading indicates the signals from the same-direction allelic effects meta-analysis.

Bp, base pair; Chr, chromosome; OR, odds ratio; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.
Supplementary Table 3. Gene sets that showed a significant enrichment \( P \)-value \( (P\text{-value} < 0.05) \) after False Discovery Rate correction in GSEA-based analysis or Bonferroni correction in DAVID-based analysis.

<table>
<thead>
<tr>
<th>Database</th>
<th>Gene Set</th>
<th>( K )</th>
<th>( k )</th>
<th>( k/K )</th>
<th>( P\text{-value} )</th>
<th>FDR ( P\text{-value} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reactome</td>
<td>Genes involved in Immune System</td>
<td>933</td>
<td>9</td>
<td>0.010</td>
<td>1.03E-12</td>
<td>9.19E-10</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Interferon alpha/beta signaling*</td>
<td>64</td>
<td>4</td>
<td>0.063</td>
<td>3.39E-09</td>
<td>1.51E-06</td>
</tr>
<tr>
<td>Biocarta</td>
<td>NO2-dependent IL 12 Pathway in NK cells*</td>
<td>17</td>
<td>3</td>
<td>0.177</td>
<td>1.53E-08</td>
<td>4.53E-06</td>
</tr>
<tr>
<td>Biocarta</td>
<td>IL12 and Stat4 Dependent Signaling Pathway in Th1 Development*</td>
<td>23</td>
<td>3</td>
<td>0.130</td>
<td>3.97E-08</td>
<td>8.85E-06</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Interferon Signaling*</td>
<td>159</td>
<td>4</td>
<td>0.025</td>
<td>1.34E-07</td>
<td>2.40E-05</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Interferon gamma signaling*</td>
<td>63</td>
<td>3</td>
<td>0.048</td>
<td>8.84E-07</td>
<td>1.31E-04</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Cytokine Signaling in Immune system</td>
<td>270</td>
<td>4</td>
<td>0.015</td>
<td>1.11E-06</td>
<td>1.42E-04</td>
</tr>
<tr>
<td>Biocarta</td>
<td>IL22 Soluble Receptor Signaling Pathway</td>
<td>16</td>
<td>2</td>
<td>0.125</td>
<td>1.03E-05</td>
<td>1.15E-03</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Negative regulators of RIG-I/MDA5 signaling</td>
<td>31</td>
<td>2</td>
<td>0.065</td>
<td>3.99E-05</td>
<td>3.95E-03</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell</td>
<td>70</td>
<td>2</td>
<td>0.029</td>
<td>2.06E-04</td>
<td>1.81E-02</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways</td>
<td>73</td>
<td>2</td>
<td>0.027</td>
<td>2.24E-04</td>
<td>1.81E-02</td>
</tr>
<tr>
<td>Reactome</td>
<td>Genes involved in Adaptive Immune System</td>
<td>539</td>
<td>3</td>
<td>0.006</td>
<td>5.30E-04</td>
<td>3.94E-02</td>
</tr>
</tbody>
</table>

*Significant in DAVID-based analysis

FDR \( P\)-value, false discovery rate corrected \( P\)-value; \( K \), genes in Gene Set; \( k \), genes in overlap
SUPPLEMENTARY REFERENCES


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 Sofia/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 Taulí, 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 Universitari Doctor Peset, Valencia. Francisco J. Blanco García and Natividad Oreiro, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña.