

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-Isac^{1*} MSc, Jose-Ezequiel Martín¹ PhD, Shervin Assasi² MD, Carmen P Simeón³ MD PhD, Patricia Carreira⁴ MD, Norberto Ortego-Centeno⁵ MD, Mayka Freire⁶ MD PhD, Emma Beltrán⁷ MD, Javier Narváez⁸ MD PhD, Juan J Alegre-Sancho⁹ MD PhD, the Spanish Scleroderma Group¹⁰, Benjamín Fernández-Gutiérrez¹¹ MD PhD, Alejandro Balsa¹² MD PhD, Ana M Ortiz¹³ MD PhD, Miguel A González-Gay¹⁴ MD PhD, Lorenzo Beretta¹⁵ MD, Alessandro Santaniello¹⁵ MD, Chiara Bellocchi¹⁵ MD, Claudio Lunardi¹⁶ MD, Gianluca Moroncini¹⁷ MD PhD, Armando Gabrielli¹⁷ MD PhD, Torsten Witte¹⁸ MD PhD, Nicolas Hunzelmann¹⁹ MD, Jörg HW Distler²⁰ MD, Gabriella Riekemasten²¹ MD PhD, Annete H van der Helm-van Mil²² MD PhD, Jeska de Vries-Bouwstra²² MD PhD, Cesar Magro-Checa²² MD, Alexandre E Voskuyl²³ MD, Madelon C Vonk²⁴ MD PhD, Øyvind Molberg²⁵ MD PhD, Tony Merriman²⁶ PhD, Roger Hesselstrand²⁷ MD PhD, Annika Nordin²⁸ MD PhD, Leonid Padyukov²⁸ PhD, Ariane Herrick²⁹ MD PhD, Steve Eyre²⁹ MD PhD, Bobby PC Koeleman³⁰ PhD, Christopher P Denton³¹ PhD FRCP, Carmen Fonseca³¹ MD PhD, Timothy RDJ Radstake³² MD PhD, Jane Worthington²⁹ MD PhD, Maureen D Mayes² MD MPH, Javier Martín^{1*} MD PhD

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, Complejo 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 Universitari Doctor Peset, Valencia, Spain.
10. See supplementary note.
11. Rheumatology Service, Hospital Clínico San Carlos, Madrid, Spain.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1002/art.39730

© 2016 American College of Rheumatology

Received: Dec 11, 2015; Revised: Mar 30, 2016; Accepted: Apr 19, 2016

This article is protected by copyright. All rights reserved.

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 U01 1U01AI09090, 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 $< 5 \times 10^{-6}$), 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 ($P_{\text{combined}} = 3.29 \times 10^{-12}$). 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 $< 5 \times 10^{-6}$ 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 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 $< 5 \times 10^{-8}$) 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 $P_{\text{GWAS+Repl}} < 5 \times 10^{-3}$ 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 *HNFI1A* 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 $P_{\text{GWAS+Repl}} = 2.78 \times 10^{-7}$, OR = 0.90; RA $P_{\text{GWAS+Repl}} = 1.44 \times 10^{-6}$, 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 $P_{\text{Repl}} = 0.036$, OR = 0.94; SSc $P_{\text{GWAS+Repl}} = 1.64 \times 10^{-4}$, OR = 0.91), and a P -value of 1.59×10^{-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 (*PTPN22*, *ATG5*, *IRF5* and *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.

Accepted Article

REFERENCES

1. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet.* 2013;14:661-73.
2. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature.* 2014;506:376-81.
3. 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.
4. Elhai M, Avouac J, Kahan A, Allanore Y. Systemic sclerosis at the crossroad of polyautoimmunity. *Autoimmun Rev.* 2013;12:1052-7.
5. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet.* 2013;22:4021-9.
6. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet.* 2010;42:426-9.
7. Martin JE, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, 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.
8. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet.* 2010;42:508-14.
9. 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:e1000529.
10. Consortium TAaNZMSG. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. *Nat Genet.* 2009;41:824-8.
11. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, 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.

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

13. Biswas PS, Gupta S, Stirzaker RA, Kumar V, Jessberger R, Lu TT, 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.

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

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

TABLE

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

Locus (Chr)	SNP	Ref. Allele	SSc-GWAS		SSc Repl		SSc GWAS+Repl		RA GWAS		RA Repl		RA GWAS+Repl		SSc-RA GWAS+Repl		Status now
			P GWAS	OR ⁺ GWAS	SSc P Repl	OR ⁺ Repl	SSc P Meta	OR ⁺ Meta	RA P GWAS	OR ⁺ GWAS	RA P Repl	OR ⁺ Repl	RA P Meta	OR ⁺ Meta	SSc-RA P Meta		
<i>FBN2</i>																	
(5)	rs6897611	T	4.79E-07	2.85E-03	1.16	0.641	0.98	0.165	1.04	3.15E-05	1.24	0.684	0.99	0.650*	1.02*	0.018	-
<i>IRF4</i>																	
(6)	rs9328192	G	4.06E-07	8.86E-06	0.86	1.89E-03	0.93	2.78E-07	0.90	7.26E-03	1.10	5.22E-05	1.07	1.44E-06	1.08	3.29E-12	SSc-RA, RA
<i>HNF1A</i>																	
(12)	rs10774577**	T	7.53E-07	8.62E-04	0.89	0.036	0.94	1.64E-04	0.91	2.50E-04	1.14	0.290	1.03	0.208*	1.05*	1.59E-06	-

⁺ Odds ratio for the reference allele.

* *P*-value and OR from meta-analysis under random effects due to heterogeneity of the ORs among cohorts.

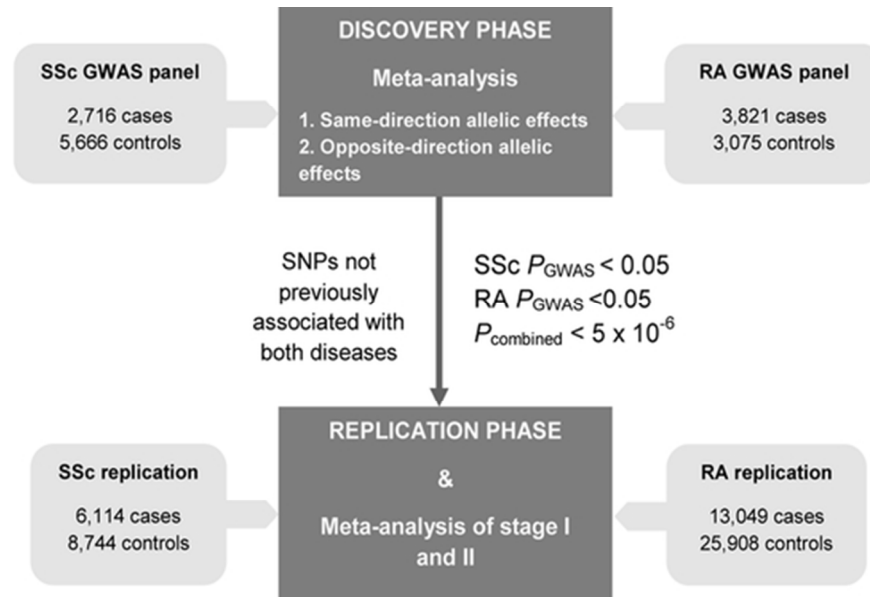
**The RA and SSc replication cohorts from Spain, and the SSc replication cohorts from Italy and The Netherlands were excluded from the analysis of rs10774577 due to HWE issues.

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

FIGURE LEGENDS

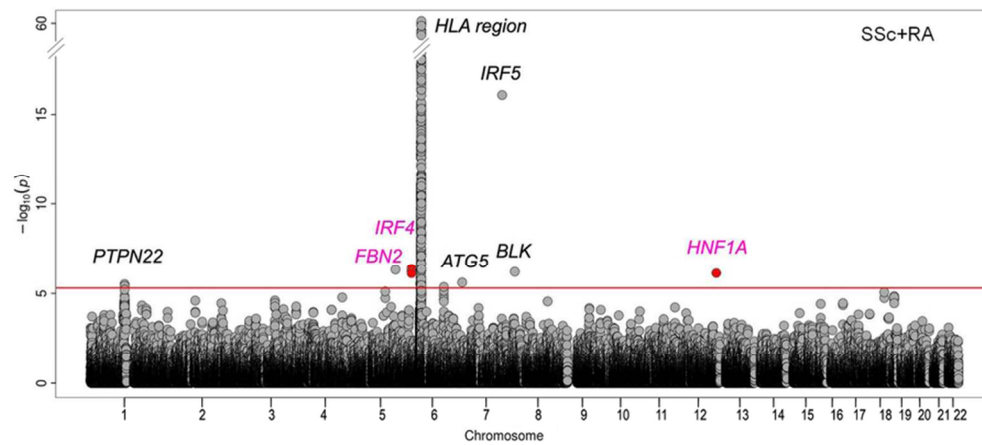
Figure 1. Overall workflow of the present study.

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 5×10^{-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).



53x31mm (300 x 300 DPI)

Accepted



77x34mm (300 x 300 DPI)

Accepted

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 5×10^{-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.