Published in final edited form as:

Ann Rheum Dis. 2012 January ; 71(1): 114–119. doi:10.1136/annrheumdis-2011-200275.

Novel identification of the *IRF7* region as an anticentromere autoantibody propensity locus in systemic sclerosis

F David Carmona¹, Ramana Gutala², Carmen P Simeón³, Patricia Carreira⁴, Norberto Ortego-Centeno⁵, Esther Vicente-Rabaneda⁶, Francisco J García-Hernández⁷, Paloma García de la Peña⁸, Mónica Fernández-Castro⁹, Lina Martínez-Estupiñán¹⁰, María Victoria Egurbide¹¹, Betty P Tsao¹², Pravitt Gourh², Sandeep K Agarwal², Shervin Assassi², Maureen D Mayes², Frank C Arnett², Filemon K Tan², Javier Martín¹, and Spanish Scleroderma Group

¹Instituto de Parasitología y Biomedicina López-Neyra, CSIC, Granada, Spain

²Department of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center, Houston, Texas, USA

³Department of Internal Medicine, Hospital Valle de Hebron, Barcelona, Spain

⁴Department of Rheumatology, Hospital 12 de Octubre, Madrid, Spain

⁵Department of Internal Medicine, Hospital Clínico San Cecilio, Granada, Spain

⁶Department of Rheumatology, Hospital de la Princesa, Madrid, Spain

⁷Department of Internal Medicine, Hospital Virgen del Rocío, Sevilla, Spain

⁸Department of Rheumatology, Hospital Universitario Madrid Norte Sanchinarro, Madrid, Spain

⁹Department of Rheumatology, Hospital Puerta de Hierro, Madrid, Spain

¹⁰Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Madrid, Spain

¹¹Department of Internal Medicine, Hospital de Cruces, Barakaldo, Spain

¹²Division of Rheumatology, University of California, Los Angeles, California, USA

Abstract

Objective—Systemic sclerosis (SSc) and systemic lupus erythematosus (SLE) are related chronic autoimmune diseases of complex aetiology in which the interferon (IFN) pathway plays a key role. Recent studies have reported an association between *IRF7* and SLE which confers a risk to autoantibody production. A study was undertaken to investigate whether the *IRF7* genomic region is also involved in susceptibility to SSc and the main clinical features.

Correspondence to: F David Carmona, Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18100-Armilla, Granada, Spain; dcarmona@ipb.csic.es.

FDC and RG contributed equally to this work.

FKT and JM are joint senior authors.

Competing interests None.

Ethics approval Ethical approval was obtained from the Committee for the Protection of Human Subjects of the University of Texas Health Science Center at Houston (USA) and the Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas (Granada, Spain) and informed written consent was obtained from all participants in accordance with the tenets of the Declaration of Helsinki.

Provenance and peer review Not commissioned; internally peer reviewed. Members of the Spanish Scleroderma Group are shown in the online supplement.

Methods—Two case-control sets of Caucasian origin from the USA and Spain, comprising a total of 2316 cases of SSc and 2347 healthy controls, were included in the study. Five single nucleotide polymorphisms (SNPs) in the *PHRF1-IRF7-CDHR5* locus were genotyped using TaqMan allelic discrimination technology. A meta-analysis was performed to test the overall effect of these genetic variants on SSc.

Results—Four out of five analysed SNPs were Significantly associated with the presence of anticentromere autoantibodies (ACA) in the patients with SSc in the combined analysis (rs1131665: $p_{FDR}=6.14 \times 10^{-4}$, OR=0.78; rs4963128: $p_{FDR}=6.14 \times 10^{-4}$, OR=0.79; rs702966: $p_{FDR}=3.83 \times 10^{-3}$, OR=0.82; and rs2246614: $p_{FDR}=3.83 \times 10^{-3}$, OR=0.83). Significant p values were also obtained when the disease was tested globally; however, the statistical significance was lost when the ACA-positive patients were excluded from the study, suggesting that these associations rely on ACA positivity. Conditional logistic regression and allelic combination analyses suggested that the functional *IRF7* SNP rs1131665 is the most likely causal variant.

Conclusions—The results show that variation in the *IRF7* genomic region is associated with the presence of ACA in patients with SSc, supporting other evidence that this locus represents a common risk factor for autoantibody production in autoimmune diseases.

INTRODUCTION

Systemic sclerosis (SSc) is a chronic fibrotic autoimmune disease in which autoantibodies against several nuclear and/or nucleolar antigens are commonly produced; however, each SSc-associated antibody specificity tends to be mutually exclusive in distinct clinical subsets of the disease. Thus, they are important diagnostic and prognostic markers in clinical practice. Although antinuclear autoantibodies are detected in different connective tissue autoimmune diseases, SSc shows its own particular autoantibody profile that tends not to overlap with that of other related diseases. In SSc the two major subclasses of specific autoantibodies are the anticentromere autoantibodies (ACA), which are related to limited skin involvement and an increased risk of pulmonary arterial hypertension, and the antitopoisomerase autoantibodies (ATA), which confer susceptibility to diffuse skin and pulmonary fibrosis with an increased mortality.¹⁻³

SSc has a complex aetiology with multiple susceptibility genes interacting for the development of the disease in concert with epigenetic and environmental factors. It is likely that an imbalance between risk and protective loci is a key factor contributing to the predisposition and clinical phenotype of SSc.⁴ Recent candidate gene and genome-wide association studies (GWAS) have identified several markers that are clearly associated with SSc.⁵ Noteworthy are the associations reported for *STAT4* and *IRF5*, since polymorphisms of these genes showed the strongest signals outside the HLA region in a recent GWAS of SSc.⁶ These two genes are representative regulators in the interferon (IFN) pathway, which comprise a large number of cytokines with different modulatory effects on innate and adaptive immunity. In this regard, type I IFNs were reported to have a central aetiopathogenic role in the development and progression of systemic lupus erythematosus (SLE).⁷ Interestingly, a type I IFN signature similar to that described in SLE has also been observed in microarray analyses of peripheral blood and skin cells of patients with SSc.⁸⁹

Interferon regulatory factor 7 (*IRF7*) has recently been described as a susceptibility locus for SLE.¹⁰⁻¹² This gene encodes a member of the IFN regulatory transcription factor family, which plays a key role in the IFN-inducible pathway by activating type I IFN genes in response to viral infection or DNA/RNA-containing immune complexes.¹³ Hence, tight regulation of *IRF7* expression and activity is crucial for appropriate IFN-mediated physiological functions.¹⁴ Taking into account the genetic similarities between SSc and

SLE,⁴⁹¹⁵ we aimed to investigate whether variation within this genomic region is also involved in SSc susceptibility and/or its major clinical and autoantibody manifestations.

METHODS

Study population

Two independent Caucasian populations, a discovery cohort from the USA and a replication cohort from Spain, were analysed in this study, comprising a total of 2316 SSc cases and 2347 unrelated healthy individuals recruited in the same geographical areas and matched by age, sex and ethnicity. The US cohort was composed of 1282 cases of SSc and 875 controls. Samples from patients in the USA came from the Scleroderma Registry and DNA Repository, Genetics versus Environment in Scleroderma Outcomes Study (GENISOS) and the rheumatology divisional collection evaluated at the University of Texas Health Science Center at Houston. The Spanish cohort consisted of 1034 cases of SSc and 1472 controls from previously established collections with nationally representative recruitment. Clinical features of the patients from both cohorts are summarised in table 1.

All patients with SSc fulfilled the 1980 American College of Rheumatology classification criteria for this disease¹⁶ or had at least three of the five CREST (Calcinosis, Raynaud's phenomenon, Esophageal dysmotility, Sclerodactyly, Telangiectasias) features.¹⁷ Case sets were further subdivided based on their skin involvement into limited cutaneous scleroderma (lcSSc) and diffuse cutaneous scleroderma (dcSSc) subgroups,¹⁸ and by autoantibody status according to the presence of ACA or ATA. ACAs were determined by their characteristic distinctive pattern on HEP 2 cells, and ATAs were detected by passive immunodiffusion against calf thymus extract (Inova Diagnostics, Davis, California, USA).

SNP selection and genotyping

Four single nucleotide polymorphisms (SNPs), rs12286521, rs4963128, rs702966 and rs2246614, which span a 48 kb genomic region including PHD and ring finger domains 1 (*PHRF1*, also known as *KIAA1542*), cadherin-related family member 5 (*CDHR5*) and *IRF7* genes were selected to tag haplotype blocks present in the CEU HapMap reference dataset, as described previously.¹⁰ A fifth SNP within the *IRF7* gene, rs1131665, was also included in the study because it produces a non-synonymous change in the DNA sequence (Q412R) and has recently been associated with SLE susceptibility.¹¹

Genomic DNA was extracted from peripheral white blood cells following standard procedures. Samples were genotyped for the above *PHRF1-IRF7-CDHR5* genetic variants using TaqMan 5' allele discrimination assays (rs12286521, rs4963128, rs702966 and rs2246614 were predesigned assays with IDs: C_26650291_10, C_1611594_10, C_7470754_10 and C_16061601_10, and rs1131665 was designed as a custom assay) in a 7900HT Fast Real-Time PCR System (Applied Biosystems, Foster City, California, USA).

Statistical analysis

The statistical power of the combined analysis was >99% for all the SNPs to detect associations with OR=1.3 at the 5% significance level, according to Power Calculator for Genetic Studies 2006 software which uses the methods described by Skol *et al.*¹⁹

All statistical analyses of the allele frequencies were carried out using the Linux software PLINK V.1.07 (<http://pngu.mgh.harvard.edu/purcell/plink/>).²⁰ To test for association, 2×2 contingency tables and χ^2 and/or Fisher exact tests, when appropriate, were used to calculate p values. OR and 95% CIs were obtained according to the Woolf method. Mantel-Haenszel tests under fixed effects or random effects (DerSimonian-Laird), when necessary,

were performed to submit the combined data to meta-analysis and the Breslow–Day method was used to estimate the homogeneity among the two cohorts. The Benjamini and Hochberg step-up false discovery rate (FDR) control correction²¹ for multiple testing was applied to the p values in the analyses of the discovery cohort and in the combined meta-analyses; p values <0.05 were considered statistically Significant.

Dependency of association between each SNP and every studied genetic variant was determined by conditional logistic regression analysis as implemented in PLINK. We also analysed the different allelic combinations using PLINK and Haploview (V.4.2).²² Allelic combinations with a frequency <5% in control groups were excluded from the analysis. Meta-analyses of the different allelic combinations were performed using PLINK and StatsDirect V.2.6.6 (StatsDirect, Altrincham, UK).

RESULTS

In all cohorts the genotyping success rate was >95% and there was no Significant departure from Hardy–Weinberg equilibrium at the 5% significance level. In addition, control allelic frequencies of rs4963128, rs702966, rs1131665 and rs12286521 were similar to those reported for the CEU population in the HapMap project (<http://hapmap.ncbi.nlm.nih.gov/>). The genetic variant rs1131665 was not genotyped in HapMap; however, the minor allele frequency that we have observed for this SNP in the control population is in agreement with previous published data.¹¹

Independent analyses

We first investigated whether these five representative polymorphisms within the genomic region around the *IRF7* locus were associated with SSc and clinical/autoantibody phenotypes in a large Caucasian US cohort (see table S1 in online supplement). This preliminary analysis showed Significant differences in allele frequencies between the ACA-positive subgroup and the control set for rs1131665 ($p_{\text{FDR}}=3.09 \times 10^{-3}$, OR=0.70, CI 95% 0.57 to 0.87), rs702966 ($p_{\text{FDR}}=3.09 \times 10^{-3}$, OR=0.71, CI 95% 0.58 to 0.88), rs4963128 ($p_{\text{FDR}}=0.028$, OR=0.79, CI 95% 0.65 to 0.96) and rs2246614 ($p_{\text{FDR}}=0.046$, OR=0.82, CI 95% 0.68 to 0.99). To further explore the possible involvement of the *IRF7* region in ACA susceptibility, a second large and well-defined Caucasian cohort from Spain was genotyped (see table S2 in online supplement). Three of these associations were replicated in the Spanish cohort—namely, rs1131665 ($p=0.037$, OR=0.83, CI 95% 0.70 to 0.99), rs4963128 ($p=5.42 \times 10^{-3}$, OR=0.79, CI 95% 0.67 to 0.93) and rs2246614 ($p=0.035$, OR=0.84, CI 95% 0.71 to 0.99). Although the analysis of rs702966 did not reach statistical significance in the comparison of the ACA-positive group and controls, a protective effect was also suggested for the minor allele ($p=0.208$, OR=0.90, CI 95% 0.76 to 1.06).

Meta-analysis

Since no heterogeneity of the ORs among the US and Spanish populations was observed, a combined meta-analysis was performed to increase the statistical power and hence to obtain more accurate estimates of susceptibility (table 2). The Mantel–Haenszel test under an allelic model revealed strong association signals in the ACA analysis for rs1131665 ($p_{\text{FDR}}=6.14 \times 10^{-4}$, OR=0.78, CI 95% 0.68 to 0.89), rs4963128 ($p_{\text{FDR}}=6.14 \times 10^{-4}$, OR=0.79, CI 95% 0.70 to 0.90), rs702966 ($p_{\text{FDR}}=3.83 \times 10^{-3}$, OR=0.82, CI 95% 0.72 to 0.93) and rs2246614 ($p_{\text{FDR}}=3.83 \times 10^{-3}$, OR=0.83, CI 95% 0.73 to 0.94).

Significant p values were also obtained when these four SNPs were tested for the global disease (rs1131665: $p_{\text{FDR}}=0.024$, OR=0.88, CI 95% 0.80 to 0.97; rs4963128: $p_{\text{FDR}}=0.017$, OR=0.87, CI 95% 0.80 to 0.96; rs702966: $p_{\text{FDR}}=0.046$, OR=0.90, CI 95% 0.82 to 0.99;

rs2246614: $p_{FDR}=0.031$, $OR=0.90$, $CI\ 95\% \ 0.82\ to\ 0.98$). On the other hand, rs4963128 showed a weak association with dcSSc ($p_{FDR}=0.034$, $OR=0.84$, $CI\ 95\% \ 0.74\ to\ 0.95$) and trends of association were observed in the rest of the subtype analyses for these SNPs. However, the statistical significance was lost when the ACA-positive patients were excluded from all the above analyses, suggesting that these associations rely on ACA positivity.

Additionally, to obtain further evidence of the association between the *IRF7* region and ACA production, we compared the patients with SSc positive for ACA with those with SSc but without this type of autoantibody. The meta-analysis of the comparison of ACA-positive and ACA-negative patients (see table S3 in online supplement) showed statistically significant differences for rs1131665 ($p_{FDR}=0.015$, $OR=0.80$, $CI\ 95\% \ 0.69\ to\ 0.93$), rs4963128 ($p_{FDR}=0.035$, $OR=0.85$, $CI\ 95\% \ 0.74\ to\ 0.98$) and rs702966 ($p_{FDR}=0.029$, $OR=0.83$, $CI\ 95\% \ 0.72\ to\ 0.96$), which strengthens the consistency of our results.

Conditional logistic regression

The linkage disequilibrium (LD) analysis showed that rs702966 was in almost complete LD with the non-synonymous *IRF7* SNP rs1131665 in the studied populations ($r^2=0.93$ in the USA and 0.97 in Spain). It was therefore not possible to determine whether one of these two genetic variants had an independent effect of the haplotype background. However, pairwise conditioning showed that the association of rs2246614 with the presence of ACA could be explained by those of rs1131665/rs702966 and rs4963128, since the latter SNPs remained significant when conditioned to rs2246614, taking into account the two populations as covariate (table 3). Nevertheless, we could not establish whether the rs1131665 and rs4963128 signals were independent.

Allelic combinations

We performed several allelic combination tests to analyse the possible interaction between the SNPs of this genomic region. Our results indicated that the association with the presence of ACA could be narrowed down to allelic combinations of the two most associated SNPs, rs4963128 and rs1131665 (table 4), as none of the remaining allelic combinations between the four associated SNPs increased their statistical significance (see table S4 in online supplement). However, the strength of the ACA association observed in the allelic combinations was not higher than that observed for rs4963128 and rs1131665 in the independent analysis (table 2).

DISCUSSION

IRF7 has been shown to be a master factor in the transcriptional activation of type I IFN genes which encode proteins that are key regulators of the immune system.¹³ Upregulated expression levels of type I IFN-inducible genes in peripheral blood cells have been correlated with increased disease activity and specific autoantibody profiles in patients with SLE.^{23–25} Recent studies have also shown a dysregulation of the type I IFN pathways in patients with SSc compared with healthy controls.^{26–29} In this regard, it is likely that type I IFN molecules increase the inflammatory potential of dermal fibroblasts through the upregulation of Toll-like receptors.³⁰ In addition, several genetic variants in the IFN pathway, including polymorphisms within *IRF5* and *STAT4* loci, have been reported to be associated with both SLE and SSc susceptibility.⁶³¹ Hence, it is likely that the presence of a type I IFN signature is also of special relevance in the pathophysiology of SSc.

Despite the fact that no functional or mechanistic data have been obtained, which is certainly an important limitation of this study, our results clearly indicate that the *IRF7* genomic region is a susceptibility locus for SSc and may confer disease risk through a role in ACA

production. This is supported by the fact that the statistical significance observed when the global disease and the limited/diffuse subsets were tested was lost when the ACA-positive patients were removed from the analyses. Four of the five polymorphisms analysed showed strong evidence of association with ACA production. In relation to rs2246614, this genetic variant is specifically located within the *IRF7* adjacent gene *CDHR5*, a novel mucin-like gene member of the cadherin superfamily whose specific function has not yet been determined.³² Interestingly, rs2246614 corresponds to a non-synonymous modification (R357S), although its possible functional implication is still unknown. Our logistic regression data indicate that this SNP may be tagged by rs1131665 and/or rs4963128. However, since the remaining p values of the latter variants were weak after the analysis, further studies should be carried out to investigate whether *CDHR5* rs2246614 is indeed involved in SSc susceptibility or simply reflects association signals within *IRF7*.

In relation to the analysed *IRF7-PHRF1* genetic variants, rs4963128 and rs702966 were highly associated with the presence of ACA in our study. Interestingly, GWAS data showed a strong association between rs4963128 and SLE,¹⁵ and subsequent studies suggested that this association is probably specific for autoantibody production.¹⁰¹² On the other hand, the 3'-untranslated region *PHRF1* SNP rs702966 was also found to be associated with the presence of anti-double-stranded DNA (anti-dsDNA) antibodies in patients with SLE.¹⁰ Owing to the LD structure of the *PHRF1-IRF7* genomic region, the associations of both SNPs—which are located 23 kb and 0.6 kb telomeric to *IRF7*, respectively—have been proposed to be representative of the SLE association signals within the *IRF7* gene.¹¹¹⁵³³ Moreover, autoantibody-positive SLE patients carrying rs4963128 and/or rs702966 risk genotypes exhibited increased serum levels of IFN α ,¹⁰ thus suggesting that these polymorphisms may be tagged by a causal *IRF7* SNP. Supporting this notion, the analysis of the LD structure of our cohorts showed that rs1131665 is in strong LD with rs4963128 and in nearly complete LD with rs702966. This strong correlation between the alleles of these SNPs made it impossible to discern the true susceptibility variant. Nevertheless, we speculate that the best candidate should be rs1131665, since it produces a non-synonymous change (Q412R) located in exon 5 of the *IRF7* gene. It has been reported that the risk allele of this SNP leads to increased activation of *IRF7* in vitro, correlated with SLE susceptibility in populations of different ethnicity.¹¹ This functional *IRF7* variant showed the highest association signal with ACA production in our meta-analysis, and the genetic combinations that were associated with ACA susceptibility did not reach higher statistical significance than that observed in the independent analysis of rs1131665. The LD pattern of rs1131665 is not yet clear because it was not genotyped in the reference population of the HapMap project. However, since this *IRF7* variant has been shown to have functional consequences in the downstream interferon pathway, it may represent the causal SNP of this association. In any case, this genomic region should be analysed in more detail to determine definitively whether rs1131665 is the tagger SNP or whether there are other independent signals.

In summary, together with previous findings,^{26–30} our results strongly suggest that the *IRF7* genomic region plays an important role in ACA production among patients with SSc. The fact that this same locus is also associated with a specific autoantibody (dsDNA) in SLE¹⁰¹² indicates that common immunological pathways may underlie both diseases. Supporting this idea, previous studies have reported that SLE showed the strongest familial accumulation in patients with SSc and that ACA positivity was associated with polyautoimmunity in SSc families.³⁴

However, the main limitation of this study is the lack of functional data that would be essential for a better understanding of how *IRF7* leads to a specific autoantibody profile in autoimmunity. More comprehensive studies therefore need to be performed to elucidate the causal polymorphism(s) of this association and to identify the molecular mechanisms

involved in the production of autoantibodies, with the aim of developing more effective therapeutic strategies in autoimmune diseases.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Sofía Vargas, Sonia García and Gema Robledo (from Instituto de Parasitología y Biomedicina 'López-Neyra', CSIC, Spain), Julio Charles and Marilyn Perry (University of Texas) for their excellent technical assistance and all the patients and healthy controls for kindly agreeing to their essential collaboration. They also thank Banco Nacional de ADN (University of Salamanca, Spain) for supplying part of the control material.

Funding This work was supported by the following grants: JM was funded by GENFER from the Spanish Society of Rheumatology, SAF2009–11110 from the Spanish Ministry of Science, CTS-4977 and CTS-180 from Junta de Andalucía, RETICS Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain, within the VI PN de I+D+i 2008–2011 (FEDER), and is sponsored by the Orphan Disease Program grant from EULAR. NO and JM are funded by Consejería de Salud, Junta de Andalucía through PI-0590–2010. FDC was supported by Consejo Superior de Investigaciones Científicas (CSIC) through the program JAE-DOC. The US analyses were supported by US National Institutes of Health and the National Institute of Arthritis and Musculoskeletal Diseases (NIH-NIAMS) R01-AR-055258, Two-Stage Genome Wide Association Study in Systemic Sclerosis (MDM) and by the NIH-NIAMS Center of Research Translation (CORT) in SSc (P50AR054144) (MDM, FCA, FKT), the NIH-NIAMS SSc Family Registry and DNA Repository (N01-AR-0-2251) (MDM), NIH-NIAMS K08 Award (K08AR054404) (SKA), SSc Foundation New Investigator Award (SKA), NIH-KL2RR024149-04 (SA) and the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111) (MDM).

REFERENCES

- Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med*. 2009; 360:1989–2003. [PubMed: 19420368]
- Koenig M, Dieudé M, Senécal JL. Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. *Autoimmun Rev*. 2008; 7:588–593. [PubMed: 18617021]
- Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol*. 2011; 6:509–537. [PubMed: 21090968]
- Agarwal SK, Reveille JD. The genetics of scleroderma (systemic sclerosis). *Curr Opin Rheumatol*. 2010; 22:133–138. [PubMed: 20090527]
- Martin J, Fonseca C. The genetics of scleroderma. *Curr Rheumatol Rep*. 2011; 13:13–20. [PubMed: 20960081]
- 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; 42:426–429. [PubMed: 20383147]
- Sozzani S, Bosio D, Scarsi M, et al. Type I interferons in systemic autoimmunity. *Autoimmunity*. 2010; 43:196–203. [PubMed: 20298124]
- Milano A, Pendergrass SA, Sargent JL, et al. Molecular subsets in the gene expression signatures of scleroderma skin. *PLoS ONE*. 2008; 3:e2696. [PubMed: 18648520]
- Assassi S, Mayes MD, Arnett FC, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum*. 2010; 62:589–598. [PubMed: 20112391]
- Salloum R, Franek BS, Kariuki SN, et al. Genetic variation at the IRF7/PHRF1 locus is associated with autoantibody profile and serum interferon-alpha activity in lupus patients. *Arthritis Rheum*. 2010; 62:553–561. [PubMed: 20112359]
- Fu Q, Zhao J, Qian X, et al. Association of a functional IRF7 variant with systemic lupus erythematosus. *Arthritis Rheum*. 2011; 63:749–754. [PubMed: 21360504]
- Chung SA, Taylor KE, Graham RR, et al. Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet*. 2011; 7:e1001323. [PubMed: 21408207]
- Honda K, Yanai H, Negishi H, et al. IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature*. 2005; 434:772–777. [PubMed: 15800576]

14. Ning S, Pagano JS, Barber GN. IRF7: activation, regulation, modification and function. *Genes Immun.* 2011; 12:399–414. [PubMed: 21490621]
15. Harley JB, Alarcón-Riquelme ME, Criswell LA, et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PTK, KIAA1542 and other loci. *Nat Genet.* 2008; 40:204–210. [PubMed: 18204446]
16. Anon. 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; 23:581–590. [PubMed: 7378088]
17. Rodnan G, Medsger T, Buckingham R. Progressive systemic sclerosis – CREST syndrome: observations on natural history and late complications in 90 patients. *Arthritis Rheum.* 1975; 18:423.
18. LeRoy EC, Black C, Fleischmajer R, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol.* 1988; 15:202–205. [PubMed: 3361530]
19. Skol AD, Scott LJ, Abecasis GR, et al. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006; 38:209–213. [PubMed: 16415888]
20. 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; 81:559–575. [PubMed: 17701901]
21. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B.* 1995; 57:289–300.
22. Barrett JC, Fry B, Maller J, et al. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005; 21:263–265. [PubMed: 15297300]
23. Feng X, Wu H, Grossman JM, et al. Association of increased interferon-inducible gene expression with disease activity and lupus nephritis in patients with systemic lupus erythematosus. *Arthritis Rheum.* 2006; 54:2951–2962. [PubMed: 16947629]
24. Baechler EC, Batliwalla FM, Karypis G, et al. Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA.* 2003; 100:2610–2615. [PubMed: 12604793]
25. Bennett L, Palucka AK, Arce E, et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med.* 2003; 197:711–723. [PubMed: 12642603]
26. Tan FK, Zhou X, Mayes MD, et al. Signatures of differentially regulated interferon gene expression and vasculotrophism in the peripheral blood cells of systemic sclerosis patients. *Rheumatology (Oxford).* 2006; 45:694–702. [PubMed: 16418202]
27. York MR, Nagai T, Mangini AJ, et al. A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type I interferons and toll-like receptor agonists. *Arthritis Rheum.* 2007; 56:1010–1020. [PubMed: 17328080]
28. Duan H, Fleming J, Pritchard DK, et al. Combined analysis of monocyte and lymphocyte messenger RNA expression with serum protein profiles in patients with scleroderma. *Arthritis Rheum.* 2008; 58:1465–1474. [PubMed: 18438864]
29. Gardner H, Shearstone JR, Bandaru R, et al. Gene profiling of scleroderma skin reveals robust signatures of disease that are imperfectly reflected in the transcript profiles of explanted fibroblasts. *Arthritis Rheum.* 2006; 54:1961–1973. [PubMed: 16736506]
30. Agarwal SK, Wu M, Livingston CK, et al. Toll-like receptor 3 upregulation by type I interferon in healthy and scleroderma dermal fibroblasts. *Arthritis Res Ther.* 2011; 13:R3. [PubMed: 21223583]
31. Moser KL, Kelly JA, Lessard CJ, et al. Recent insights into the genetic basis of systemic lupus erythematosus. *Genes Immun.* 2009; 10:373–379. [PubMed: 19440199]
32. Yagi T, Takeichi M. Cadherin superfamily genes: functions, genomic organization, and neurologic diversity. *Genes Dev.* 2000; 14:1169–1180. [PubMed: 10817752]
33. Crow MK. Collaboration, genetic associations, and lupus erythematosus. *N Engl J Med.* 2008; 358:956–961. [PubMed: 18204099]
34. Arora-Singh RK, Assassi S, del Junco DJ, et al. Autoimmune diseases and autoantibodies in the first degree relatives of patients with systemic sclerosis. *J Autoimmun.* 2010; 35:52–57. [PubMed: 20223638]

Table 1

Main clinical features of patients with systemic sclerosis (SSc) included in the study

Feature	Spain, n (%)	USA, n (%)
Women	920 (88.99)	1137 (88.69)
Men	114 (11.01)	145 (11.31)
lcSSc	710 (68.67)	789 (61.54)
dcSSc	324 (31.33)	493 (38.46)
ACA+	478 (46.23)	357 (27.85)
ACA–	528 (51.06)	778 (60.69)
ATA+	226 (21.86)	205 (15.99)
ATA–	757 (73.21)	928 (72.39)

Data are referred to the total analysed individuals.

ACA, anticentromere antibodies; ATA, antitopoisomerase antibodies; dcSSc, diffuse cutaneous systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis.

Genotype and minor allele frequencies of SNPs located around the *IRF7* locus in Caucasian patients with systemic sclerosis (SSc) and healthy controls from Spain and the USA

Table 2

SNP	1/2	Subgroup (N)	Genotype, N (%)			M-H allele test			
			1/1	1/2	2/2	MAF (%)	p Value*	pFDR [†]	OR (CI 95%) [‡]
rs1131665	G/A	Controls (n=2260)	165 (7.30)	939 (41.55)	1156 (51.15)	28.08			
		SSc (n=2264)	158 (6.98)	840 (37.10)	1266 (55.92)	25.53	0.010	0.024	0.88 (0.80 to 0.97)
		lcSSc (n=1467)	111 (7.57)	532 (36.26)	824 (56.17)	25.70	0.031	0.052	0.89 (0.80 to 0.99)
		dcSSc (n=797)	47 (5.90)	308 (38.64)	442 (55.46)	25.22	0.035	0.086	0.87 (0.76 to 0.99)
		ACA+ (n=821)	50 (6.09)	282 (34.35)	489 (59.56)	23.26	1.93E-04	6.14E-04	0.78 (0.68 to 0.89)
		ATA+ (n=417)	31 (7.43)	164 (39.33)	222 (53.24)	27.10	0.559	0.713	0.95 (0.81 to 1.12)
rs4963128	A/G	Controls (n=2268)	239 (10.54)	1022 (45.06)	1007 (44.40)	33.07			
		SSc (n=2254)	221 (9.80)	927 (41.13)	1106 (49.07)	30.37	3.48E-03	0.017	0.87 (0.80 to 0.96)
		lcSSc (n=1461)	149 (10.20)	604 (41.34)	708 (48.46)	30.87	0.031	0.052	0.89 (0.81 to 0.99)
		dcSSc (n=793)	72 (9.08)	323 (40.73)	398 (50.19)	29.45	6.87E-03	0.034	0.84 (0.74 to 0.95)
		ACA+ (n=818)	75 (9.17)	311 (38.02)	432 (52.81)	28.18	2.46E-04	6.14E-04	0.79 (0.70 to 0.90)
		ATA+ (n=412)	41 (9.95)	169 (41.02)	202 (49.03)	30.46	0.129	0.322	0.88 (0.75 to 1.04)
rs702966	C/G	Controls (n=2258)	164 (7.26)	920 (40.74)	1174 (51.99)	27.64			
		SSc (n=2263)	161 (7.11)	838 (37.03)	1264 (55.86)	25.63	0.037	0.046	0.90 (0.82 to 0.99)
		lcSSc (n=1462)	110 (7.52)	535 (36.59)	817 (55.88)	25.82	0.092	0.115	0.91 (0.82 to 1.02)
		dcSSc (n=801)	51 (6.37)	303 (37.83)	447 (55.81)	25.28	0.067	0.111	0.88 (0.77 to 1.01)
		ACA+ (n=816)	52 (6.37)	283 (34.68)	481 (58.95)	23.71	2.33E-03	3.83E-03	0.82 (0.72 to 0.93)
		ATA+ (n=418)	33 (7.89)	162 (38.76)	223 (53.35)	27.27	0.801	0.801	0.98 (0.83 to 1.16)
rs2246614	A/C	Controls (n=2257)	282 (12.49)	992 (43.95)	983 (43.55)	34.47			
		SSc (n=2245)	256 (11.40)	954 (42.49)	1035 (46.10)	32.65	0.019	0.031	0.90 (0.82 to 0.98)
		lcSSc (n=1448)	161 (11.12)	608 (41.99)	679 (46.89)	32.11	0.012	0.052	0.88 (0.80 to 0.97)
		dcSSc (n=797)	95 (11.92)	346 (43.41)	356 (44.67)	33.63	0.283	0.354	0.93 (0.83 to 1.06)
		ACA+ (n=805)	86 (10.68)	320 (39.75)	399 (49.57)	30.56	3.06E-03	3.83E-03	0.83 (0.73 to 0.94)
		ATA+ (n=416)	40 (9.62)	180 (43.27)	196 (47.12)	31.25	0.047	0.233	0.85 (0.73 to 1.00)
rs12286521	G/A	Controls (n=2151)	256 (11.90)	998 (46.40)	897 (41.70)	35.10			
		SSc (n=2056)	236 (11.48)	916 (44.55)	904 (43.97)	33.75	0.268	0.268	0.95 (0.87 to 1.04)

SNP	1/2	Subgroup (N)	Genotype, N (%)				M-H allele test			
			1/1	1/2	2/2	MAF (%)	p Value *	pFDR [†]	OR (CI 95%) [‡]	
		lcSSc (n=1350)	158 (11.70)	591 (43.78)	601 (44.52)	33.59	0.246	0.246	0.94 (0.85 to 1.04)	
		dcSSc (n=706)	78 (11.05)	325 (46.03)	303 (42.92)	34.07	0.605	0.605	0.97 (0.85 to 1.10)	
		ACA+ (n=800)	98 (12.25)	325 (40.63)	377 (47.13)	32.56	0.075	0.075	0.90 (0.79 to 1.01)	
		ATA+ (n=405)	39 (9.63)	197 (48.64)	169 (41.73)	33.95	0.570	0.713	0.95 (0.81 to 1.12)	

* All p values have been calculated for the allelic model.

[†]Benjamini and Hochberg (1995) step-up FDR control.

[‡]OR for the minor allele.

ACA, antinuclear antibodies; ATA, anti-topoisomerase antibodies; dcSSc, diffuse cutaneous systemic sclerosis; FDR, false discovery rate; lcSSc, limited cutaneous systemic sclerosis; MAF, minor allele frequency; M-H, Mantel-Haenszel test under fixed effect; SNP, single nucleotide polymorphism; SSC, systemic sclerosis.

Conditional logistic regression analysis for the *PHRF1-IRF7-CDHR5* SNPs in ACA data considering the two populations as covariate

Table 3

SNP	MAF (cases)	MAF (controls)	p Value: add to rs1131665	rs1131665 p value: add to SNP	r ² with rs1131665		p Value: add to rs4963128	rs4963128 p value: add to SNP	r ² with rs4963128	
					USA	Spain			USA	Spain
rs1131665	0.23	0.28	NA	NA	NA	NA	0.2595	0.2427	0.62	0.70
rs4963128	0.28	0.33	0.2427	0.2595	0.62	0.70	NA	NA	NA	NA
rs702966	0.24	0.28	0.5861	0.2446	0.93	0.97	0.4099	0.2185	0.66	0.70
rs2246614	0.31	0.34	0.2980	0.0283	0.33	0.41	0.4222	0.0283	0.41	0.39

ACA, antinuclear antibodies; MAF, minor allele frequency; SNP, single nucleotide polymorphism.

Table 4
Pooled analysis of rs4963128–rs1131665 allelic combinations according to disease and ACA status

Allelic combination	Controls, n (%)	SSc, n (%)	PMH	OR (95% CI)	PBD	ACA+, n (%)	PMH	OR (95% CI)	PBD
GA	2920 (65.53)	3048 (68.19)	4.63E-03	1.14 (1.04 to 1.25)	0.72	1151 (70.70)	2.02E-04	1.27 (1.12 to 1.43)	0.58
AG	1180 (26.48)	1077 (24.09)	0.015	0.89 (0.80 to 0.98)	0.64	362 (22.24)	8.43E-04	0.79 (0.69 to 0.91)	0.46
AA	294 (6.60)	283 (6.33)	0.321	0.91 (0.77 to 1.08)	0.12	98 (6.02)	0.447	0.91 (0.72 to 1.15)	0.23
GG	62 (1.40)	62 (1.39)	0.962	0.98 (0.68 to 1.40)	0.99	17 (1.04)	0.559*	0.64 (0.14 to 2.84)	0.02

Order of SNPs: rs4963128–rs1131665.

* DerSimonian–Laird random effects model p value.

PMH: allelic Mantel–Haenszel fixed effects model p value; PBD: Breslow–Day p value.

ACA, anticentromere antibodies; SSc, systemic sclerosis.