TRANSLATIONAL SCIENCE

Genomic Risk Score impact on susceptibility to systemic sclerosis

Lara Bossini-Castillo,1 Gonzalo Villanueva-Martin,2 Martin Kerick,2 Marialbert Acosta-Herrera,2 Elena López-Isac,2 Carmen P Simeón,3 Norberto Ortego-Centeno,4 Shervin Assassi,5 International SSc Group, Australian Scleroderma Interest Group (ASIG), PRECISESADS Clinical Consortium, PRECISESADS Flow Cytometry study group, Nicolas Hunzelmann,6 Armando Gabrielli,7 J K de Vries-Bouwstra,8 Yannick Allanoare,9 Carmen Fonseca,10 Christopher P Denton,9,10 Timothy RDJ Radstake,11 Marta Eugenia Alarcón-Riquelme,1,12 Lorenzo Beretta,1,13 Maureen D Mayes,5 Javier Martin2

ABSTRACT

Objectives Genomic Risk Scores (GRS) successfully demonstrated the ability of genetics to identify those individuals at high risk for complex traits including immune-mediated inflammatory diseases (IMIDs). We aimed to test the performance of GRS in the prediction of risk for systemic sclerosis (SSc) for the first time.

Methods Allelic effects were obtained from the largest SSc Genome-Wide Association Study (GWAS) to date (9 095 SSc and 17 584 healthy controls with European ancestry). The best-fitting GRS was identified under the additive model in an independent cohort that comprised 400 patients with SSc and 571 controls. Additionally, GRS for clinical subtypes (limited cutaneous SSc and diffuse cutaneous SSc) and serological subtypes (anti-topoisomerase positive (ATA+ ) and auto-centromere positive (ACA+)) were generated. We combined the estimated GRS with demographic and immunological parameters in a multivariate generalised linear model.

Results The best-fitting SSc GRS included 33 single nucleotide polymorphisms (SNPs) and discriminated between patients with SSc and controls (area under the receiver operating characteristic (ROC) curve (AUC)=0.673). Moreover, the GRS differentiated between SSc and other IMIDs, such as rheumatoid arthritis and Sjögren’s syndrome. Finally, the combination of GRS with age and immune cell counts significantly increased the performance of the model (AUC=0.787). While the SSc GRS was not able to discriminate between ATA+ and ACA+ patients (AUC<0.5), the serological subtype GRS, which was based on the allelic effects observed for the comparison between ACA+ and ATA+ patients, reached an AUC=0.693.

Conclusions GRS was successfully implemented in SSc. The model discriminated between patients with SSc and controls or other IMIDs, confirming the potential of GRS to support early and differential diagnosis for SSc.

INTRODUCTION

Complex diseases are a devastating consequence of usually unknown environmental factors and the combined effects of tens to thousands of genetic variants that are spread throughout the genome. The advanced use of bioinformatic tools will provide a better understanding of the intricate network of multiple genetic effects that shapes the architecture of complex diseases.

Immune-mediated inflammatory diseases (IMIDs) comprise a variety of complex diseases characterised by the loss of self-tolerance, the maintenance of chronic inflammation and an aberrant immune response. Genome-wide association studies (GWAS) have largely increased our understanding of the aetiology of complex diseases, providing new data about the genome and lighting the way to the identification of genes and pathways that contribute to disease susceptibility and prognosis. Many susceptibility loci have been discovered for IMIDs, and several are shared between diseases, adding a common genetic background to their overlapping clinical and immunological characteristics. Additionally, GWAS findings have also confirmed that the contribution of

Key messages

What is already known about this subject?

► Systemic sclerosis (SSc) is a complex immune-mediated disease (IMID) for which a Genomic Risk Score (GRS) has never been implemented.

What does this study add?

► A SSc GRS discriminates between patients with SSc and healthy controls with a remarkable predictive value.

► Clinical information, such as serologic subtype and immune cell counts, adds accuracy to the GRS model.

► The SSc GRS is capable of discriminating between SSc and other IMIDs.

How might this impact clinical practice or future developments?

► This SSc GRS is a promising tool to improve the diagnosis and prognosis of patients with SSc.

Correspondence to
Dr Lara Bossini-Castillo, Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Centro de Investigación Biomédica (CIBER), Parque Tecnológico de la Salud, Avenida del Conocimiento, s/n, 18016, Armilla (Granada), Andalucía, Spain; lbossinicastillo@ugr.es and Dr Javier Martin, Institute of Parasitology and Biomedicine López-Neyra, IPB, Consejo Superior de Investigaciones Científicas (CSIC, Parque Tecnológico de las Ciencias de la Salud, Avenida del Conocimiento, 17, 18016, Armilla, Granada, Andalucía, Spain; javiermartin@ipb.csic.es

LB-C and GV-M contributed equally.

LB-C and GV-M are joint first authors.

Received 8 July 2020
Revised 27 August 2020
Accepted 30 August 2020
Published Online First 1 October 2020

Systemic sclerosis

Systemic sclerosis (SSc) or scleroderma is a complex chronic autoimmune disease. It belongs to the group of IMIDs and it has one of the highest mortality rates among them. SSc affects the connective tissue and shows complex and varied clinical manifestations. Raynaud’s phenomenon and gastro-oesophageal reflux are two common onset symptoms, but they are not exclusive to SSc. Conversely, the disease can manifest in different ways, such as affection of the skin (inflammatory skin disease, extensive fibrosis), musculoskeletal inflammation and vascular damage. Furthermore, SSc also shows organ-specific manifestations, such as lung fibrosis, pulmonary arterial hypertension, renal failure and gastrointestinal complications. Notably, the involvement of the lungs, with pulmonary hypertension and/or pulmonary fibrosis, is the leading cause of death in SSc.

Patients with SSc can be classified into different subgroups according to clinical outcome: limited cutaneous scleroderma (lcSSc) or diffuse cutaneous scleroderma (dcSSc), depending on how widespread fibrosis is. On the other hand, they can also be classified depending on their serological status, considering the presence of the mutually exclusive anti-centromere or anti-topoisomerase autoantibodies (ie, ACA+ or ATA+). Since the first SSc GWAS in European populations was carried out 10 years ago, our recently published meta-GWAS is the largest effort to decipher the genetic component of SSc. In addition to the extensively known association of the human leucocyte antigen (HLA) region with the disease, non-HLA GWAS level associations and 3 suggestive loci were identified.

Considering the heterogeneity and variable prognosis of patients with SSc, GRS could be a powerful tool in clinical diagnosis to identify patients in the early stages of the disease and to differentiate them from patients with confounding diseases. By taking advantage of the summary statistics of this large meta-GWAS, we generated an accurate SSc GRS through the use of an independent and unique dataset comprising patients with SSc and with other IMIDs (figure 1). We generated subtype-specific GRS for the clinical and serological SSc subgroups of patients, and we tested the clinical implications of GRS in SSc. Finally, the GRS was complemented with additional demographic and immunological information.
Systemic sclerosis

METHODS

GRS calculation

GRS was developed as implemented in PRSice-2,26 using summary statistics and assuming an additive effect for the effective allele. Briefly, PRSice-2 calculated the product of the number of effect alleles per individual and the respective SNP weights. The score was averaged by the number of alleles included in the GRS per individual (argument --score avg). We used the minor allele frequency in the PRECISESADS cohort as the genotype for the samples with missing genotype. We applied a 10 000 permutation procedure to calculate the empirical p value (--perm 10000).

PRSice-2 allowed us to fit different GRS models by selecting only the variants that passed a number of different p value thresholds in the GWAS summary statistics (argument --bar-levels 5e-11, 5e-10, 5e-09, 5e-08, 5e-07, 5e-06, 5e-05, 0.0001, 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, but GRS calculated at all intermediate p value thresholds, high resolution parameters, were calculated) using sex (female/male) as covariate. Therefore, the model fit is defined as: $R^2$ of the full model (SSC case or control ~ GRS + Sex) - $R^2$ of the null model (SSC case or control ~ Sex).

Multivariate model

In order to test if a combination of GRS with demographic factors and the counts of immune cell subpopulations in peripheral blood would improve the predictive value of our model, we divided our score development cohort into an initial set, comprising the non-Spanish individuals in the PRECISESADS study ($n=518$), in which we developed a multivariate model and a testing set that comprised all the Spanish individuals in this study ($n=339$).

First, we built several generalised linear models that included GRS and each demographic and immune parameter in online supplemental table 1 individually, then we compared them to the null model that included only GRS and sex as covariates. Improvement over the null model was defined by an LRT ($p<0.05$).

Second, we generated a multivariate model that included the 13 phenotypic variables that had been identified as informative in the previous step. Using leave-one-out prediction (ie, including all variables but one in the model) and comparing to the full model, we calculated the contribution of all variables to the multivariate model. This model was applied to the testing set of individuals.

Details about the cohorts, linkage disequilibrium (LD) clumping, GRS additive model, the model fitting analyses and the effects of including country of origin as covariates are shown in the online supplemental methods section.

RESULTS

A 33-variant GRS discriminates between patients with SSC and controls

We calculated GRS in an independent score development cohort comprising 400 patients with SSC and 571 healthy controls.27 We observed that the best-fitting GRS (GRS $R^2=0.13$; $p=1.27\times10^{-12}$; permutation p value=$9.99\times10^{-10}$) included 33 independent SNPs that had a $p<2.215\times10^{-3}$ (figure 2A).

Sex, which was included as a covariate, contributed very modestly to the explained variance ($R^2=0.01$).

As expected, the SSC cases and controls showed significantly different GRS distributions (figure 2B, control group mean=$-8.35\times10^{-3}$ and SSC group mean=$-1.91\times10^{-3}$, t-test $p=2.2\times10^{-13}$). Reassuringly, individuals with GRS in the 95th percentile showed a fivefold higher relative risk ($OR=7.89$, 95% CI 3.44 to 18.08) than the reference quantile (40th–60th percentiles) (figure 2C).

Reassuringly, the 33 variant GRS had a 67% chance of accurately predicting if an individual was a patient with SSC or an unaffected control ($AUC=0.673$, 95% CI 0.64 to 0.71, $p=3.90\times10^{-23}$, figure 2D). We determined a best-fitting GRS threshold (GRS controls<-$1.86\times10^{-3}$<$GRS$ cases, details in online supplemental methods) and reached a moderate discrimination between cases and controls (specificity=$0.76$; sensitivity=$0.51$; accuracy=$0.66$, figure 2D).

We observed that if the receiver operating characteristic (ROC) curves were calculated separately for each country of origin, the AUC determined by the 33 variant GRS ranged from 0.60 to 0.73 (online supplemental figure 2A). However, variability of the AUC did not correlate with either country longitude, latitude or distance to 1000 genomes GBR and CEU populations (see online supplemental methods, online supplemental figure 2B-D).

Subtype stratified SSC GWAS summary stats discriminate between clinical and serological subtypes

The 33 variant GRS previously described distinguished between patients with SSC and healthy controls. However, SSC is a heterogeneous disease with both clinical and serological subtypes that influence the progression of the disease, and the prediction of these subtypes is a major clinical demand. The 33 SNP SSc GRS showed no predictive value for clinical subtypes (dcSSc vs lcSSC $AUC=0.496$, 95% CI 0.40 to 0.59, $p=0.93$, online supplemental figure 3) and serological subtypes (ATA+ vs ACA+ $AUC=0.464$, 95% CI 0.37 to 0.56, $p=0.45$, online supplemental figure 3). Furthermore, this SSc GRS was not able to predict the development of pulmonary fibrosis in patients with SSC with pulmonary fibrosis vs SSC without pulmonary fibrosis $AUC=0.479$, 95% CI 0.38 to 0.57, $p=0.66$, online supplemental figure 3).

Therefore, we used the allelic effects obtained in the GWAS comparison between dcSSc and lcSSc and between ATA+ and ACA+ patients to build subtype-specific GRS. The best-fitting GRS p value threshold for the variants in the dcSSc versus lcSSc comparison, clinical subtype GRS, comprised up to 9780 SNPs ($SNP$ p value threshold for the best- fitting dcSSc vs lcSSc GRS $<9.99\times10^{-2}$, figure 3A). This clinical subtype GRS was not limited to highly significant variants but it also included thousands of additional SNPs with very low significance. The GRS for the variants in the ATA+ vs ACA+ comparison, serological subtype GRS, required up to 35 058 SNPs ($SNP$ p value threshold for the best-fitting ATA+ vs ACA+ GRS $<3.48\times10^{-1}$, figure 3A). The clinical subtype GRS did not explain much of the phenotypic variance between dcSSc and lcSSc ($R^2=0.053$), while the explained variance between them using the serological subtype GRS was comparable with the SSC GRS ($R^2=0.115$). In this context, the subtype-specific GRS distributions (mean dcSSc GRS=$2.46\times10^{-1}$; mean lcSSc GRS=$2.16\times10^{-1}$; t-test p value=$1.21\times10^{-2}$, figure 3B), and AUC based on the clinical subtype GRS led to a modest classification of the patients into the dcSSc or lcSSc groups ($AUC=0.604$, 95% CI 0.51 to 0.70, $p=2.59\times10^{-2}$, figure 3C). However, the serological subtype GRS (comprising 35 058 SNPs) showed more distinctive GRS distributions between ATA+ and ACA+ patients (mean ATA+ GRS = $1.39\times10^{-3}$ and mean ACA+ GRS=$1.11\times10^{-3}$, t-test p value=$1.12\times10^{-4}$, figure 3B), and best classification results for the ATA+ or ACA+ subgroups of patients ($AUC=0.693$, 95% CI 0.61 to 0.78, $p=7.58\times10^{-10}$, figure 3C).

Considering the clinical relevance of pulmonary fibrosis for the prognosis of patients with SSC, we tested the predictive value of both the clinical and the serological GRS on the development of lung fibrosis. Interestingly, we observed that the serological GRS was marginally able to discriminate between patients with

120

and without lung fibrosis but the model did not reach statistical significance (AUC = 0.575, 95% CI 0.48 to 0.67, p value = 0.11, online supplemental figure 3).

**GRS separates SSc from other IMIDs**

Considering the shared genetic component of IMIDs, the implementation of the proposed GRS might help to identify high-risk individuals not only for SSc but also for other immune-related traits. Regarding the accuracy of the 33 variant SSc GRS in other IMIDs, we observed that the SSc GRS was able to separate patients with RA (RA group mean = −4.46×10⁻³; t-test p value < 2.8×10⁻⁴), Sjögren syndrome, SJS (SJS group mean = −1.78×10⁻¹; t-test p value < 3.54×10⁻⁶) and SLE (SLE group mean = −3.67×10⁻¹; t-test p value < 8.51×10⁻¹³) from the non-affected individuals. However, as expected, the GRS differences between patients with RA, SJS and SLE and controls were less significant than between SSc cases and controls (figure 4A). Furthermore, using the SSc GRS in these three additional IMIDs, the AUCs showed a modest predictive value (AUC RA = 0.608, 95% CI 0.57 to 0.64, p value = 6.58×10⁻⁴; SJS = 0.590, 95% CI 0.55 to 0.63, p value = 1.58×10⁻⁵; AUC SLE = 0.623, 95% CI 0.59 to 0.66, p value = 3.94×10⁻¹², figure 4B).

A key point toward GRS being implemented from bench-to-bedside is not only the ability to identify individuals at high risk of developing SSc in the general population, but also to help in the differential diagnosis between SSc and other IMIDs. In the pursuit of this objective, we tested the effectiveness of our SSc GRS to correctly classify between patients with SSc and those affected by other IMIDs. We report statistical differences between the GRS distributions for SSc and rheumatoid arthritis (RA) (t-test p value < 3.78×10⁻⁴) or SJS (t-test p value < 3.70×10⁻⁶), but only nominally significant differences in the case of SLE (t-test p value < 1.37×10⁻²) (figure 4A). These results were aligned with the predictive capacity of the GRS in the separation between patients with SSc and other IMIDs. The greatest AUC was observed for the classification of patients with SSc versus patients with SJS (SJS AUC = 0.585, 95% CI 0.55 to 0.62, p value = 2.22×10⁻⁴), and decreased in more closely related IMIDs, such as RA (AUC RA = 0.568, 95% CI 0.53 to 0.61, p value = 8.84×10⁻⁴) and, especially, SLE (SLE AUC = 0.553, 95% CI 0.51 to 0.59, p value = 1.19×10⁻²) (figure 4C).

**Age and immune cell counts improve the prediction accuracy**

The score development cohort recruited in the PRECISESADS study was comprehensively phenotyped and allowed us to complement our GRS with additional demographic (age, sex) and immunological (immune cell counts in peripheral blood estimated using a large flow cytometry panel) parameters (online supplemental table 1). We divided our score development cohort into an initial set (n=518) and a testing subgroup (n=339). The

---

**Figure 2** Systemic sclerosis Genomic Risk Scores (SSc GRS). (A) Identification of the best-fitting GRS in the score development cohort. Tested p value thresholds for the SNPs included in the GWAS summary statistics are presented in the x-axis. The number of SNPs included in the models corresponding to each p value threshold is shown on the left y-axis. Model fit (R²) is represented in the right y-axis. (B) Distribution of GRS for patients with SSc and healthy controls in the score development cohort. (C) Relative risk for individuals in different quantiles of the GRS distribution. (D) receiver operating characteristic (ROC) curve for the 33 SNP SSc GRS. AUC, area under the ROC curve; GWAS, genome-wide association studies; SNPs, single nucleotide polymorphisms.
First, we identified the demographic and immunological parameters which improved the GRS model (LRT p value < 0.05) (online supplemental table 1). Twelve immune cell subtypes in the initial set allowed us to test the relevance of the different parameters in a combined GRS and phenotypic model. On the other hand, the testing set confirmed these findings.

Figure 3 Characteristics of clinical subtype-specific Genomic Risk Scores (GRS) (left) and serological subtype-specific GRS (right). (A) Identification of the best-fitting GRS in the score development cohort. Tested p value thresholds for the SNPs included in the GWAS summary statistics are presented in the x-axis. The number of SNPs included in the models corresponding to each p value threshold is shown on the left y-axis. Model fit ($R^2$) is represented in the right y-axis. (B) Distribution of GRS for patients with systemic sclerosis (SSc) in each subtype group. (C) Receiver operating characteristic (ROC) curves for the 9,780 SNP clinical subtype-specific GRS and 35,058 SNP serological subtype-specific GRS. AUC, area under the ROC curve; SNPs, single nucleotide polymorphisms.
Systemic sclerosis

peripheral blood showed a significant contribution to the model, but the most significant contribution among the phenotypic variables corresponded to age (LRT p value=3.47×10^{-20}, online supplemental table 2).

When we combined only the informative variables into the same model, multivariate GLM, in addition to GRS and age, only 4 out of the 12 immune cell types remained as independently associated in the multivariate model: resting NK cells, M0 macrophages, activated dendritic cells and memory B cells (online supplemental table 3). The contribution of sex to the model did not remain significant when considering all the independent variables together and GRS score distributions between male and female patients did not show significant information (t-test p value=0.24, online supplemental table 3). Using leave-one-out prediction, we identified age as the most informative variable, followed by GRS (online supplemental table 4).

We observed that the contribution of GRS to the model was comparable with the contribution of all significant parameters of immune cell count together (GRS LRT p value=2.59×10^{-12}; GRS LRT p value=1.26×10^{-12}, online supplemental table 4). The multivariate GLM described above (SSc status ~GRS+Age+Memory B cells+Resting NK cells+M0 Macrophages+Activated dendritic cells) greatly outperformed the GRS and sex only model both in the initial (AUC discovery=0.847, 95% CI 0.81 to 0.88, p value=1.10×10^{-90}) and in the testing set (AUC=0.787, 95% CI 0.73 to 0.84, p value=1.31×10^{-24}),

Figure 4 Impact of the 33 SNP systemic sclerosis (SSc) Genomic Risk Scores (GRS) on the differential classification with other immune-mediated inflammatory diseases (IMIDs). (A) Distribution of GRS for healthy controls and patients with SSc, systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and Sjögren syndrome (SJS). (B) Receiver operating characteristic (ROC) curves for the predictive value of the SSc GRS to distinguish between patients with SSc, SLE, RA or SJS and healthy controls. (C) ROC curves for the predictive value of the SSc GRS to distinguish between patients with SLE, RA or SJS and patients with SSc. AUC, area under the ROC curves.
Systemic sclerosis

as illustrated in figure 5. Moreover, the multivariate GLM outperformed the models that did not include age, GRS or both (figure 5).

DISCUSSION

We generated a GRS based on the allelic effects identified in the largest GWAS in SSc to date. 25 We obtained a predictive GRS model comprising 33 genetic polymorphisms, which allowed us to differentiate between SSc and controls in an independent SSc patient cohort. 27 A serological subtype-specific GRS (based on the GWAS comparison between ATA+ and ACA+ SSc patients) showed the best predictive value to classify patients based on the presence of different autoantibodies. Furthermore, we demonstrated the accuracy of the model in the differentiation between SSc and other IMIDs, such as RA and SJS. Finally, we complemented the SSc GRS with demographic data and peripheral blood immune cell counts in a multivariate model which reached a very significant recall rate.

The reported SSc GRS showed good predictive value (AUC=0.673), in line with the GRS developed for other IMIDs. For example, a similar AUC was reported for inflammatory bowel disease with a GRS based on the allelic effects observed for 12 882 cases and 132 532 healthy controls (AUC=0.7229) and in SLE (AUCs ranging 0.62–0.78.16 17 Moreover, Stahl et al implemented a Bayesian inference model in a GWAS that comprised 5 485 cases of RA and 22 609 healthy controls, and the model explained 18% of the total variance, which is comparable to the variance explained by our model (R²=0.13). 30 We would like to note that the previously conducted GWAS comprised 9 095 SSc cases and 17 584 controls, and the SSc GRS was developed in an independent cohort of 400 patients with SSc and 571 non-affected controls recruited for the PRECISESADS project. 27 Since sample size is key in the identification of reliable genetic association signals and in the accurate estimation of allelic effects in GWAS, 6 7 31 the presented SSc GRS represents a robust model supported by substantial statistical power. Nevertheless, despite the promising results of the described SSc GRS, the sensitivity and specificity of the model are still far from clinical use and it will require the addition of extra information and/or the development of well-powered phenotype-specific GWAS to identify cases with specific phenotypes with higher statistical power.

Furthermore, we consider that the SSc GRS is not heavily influenced by LD clumping, since we included only the top HLA SNP association in the GRS in order to avoid an overrepresentation of HLA polymorphisms without discarding completely the potential of this region in GRS. Nevertheless, it should be noted that all the samples included in the GWAS summary stats and in the score development cohorts for the SSc GRS had European ancestry 25 27 (online supplemental figure 1). One of the major limitations of GRS implementation is the bias toward populations with a similar ethnic origin to the discovery sample, that is, the GRS shows better accuracy in closely related populations. 7 32 As we illustrated in online supplemental figure 2, we found differences in the AUCs reached by the SSc GRS in the score development cohort depending on the origin of the individuals. Consequently, the performance of the SSc GRS in non-European or mixed populations should be taken with caution. 7 33

A possible confounding factor for GRS in IMIDs is the shared genetic and immunological component that makes diagnosis complex and a slow clinical process especially in the early stages of these diseases. 4–6 As a clinical tool, a robust GRS improves

Figure 5  Receiver operating characteristic (ROC) curves for the predictive value of the multivariate generalised linear model (GLM), (SSc status ~GRS+Age+Memory B cells+Resting NK cells+M0 Macrophages+Activated dendritic cells) to distinguish between patients with SSc and healthy controls in the initial and replication cohorts depending on the variables included in the models. GRS, Genomic Risk Scores; NK cells, natural killer cells; SSc, systemic sclerosis.
In summary, we developed a GRS based on the largest GWAS in SSc, resulting in a sensitive model to differentiate between SSc cases and non-affected controls, but also to differentiate within the different SSc serological subtypes (ATA+ and ACA+). Additionally, the GRS was also useful to differentiate patients with SSc from those affected by RA and SJS. We have shown that the GRS strategy in SSc has great potential to contribute to the field. However, several limitations and challenges, such as non-European ancestry or sample size, must be overcome to implement this strategy in clinical management.

Author affiliations
1Department of Genetic e Instituto de Biotecnologia, Universidad de Granada, Granada, Andalucía, Spain
2Instituto de Parasitología y Biomedicina López-Neyra, Granada, Andalucía, Spain
3Department of Internal Medicine, Hospital Vall d’Hebron, Barcelona, Catalonia, Spain
4Department of Internal Medicine, Hospital Universitario San Cecilio, Granada, Andalucía, Spain
5Division of Rheumatology and Clinical Immunogenetics, University of Texas Health Science Center at Houston, Houston, Texas, USA
6Department of Dermatology, University of Cologne, Köln, Nordrhein-Westfalen, Germany
7Istituto di Clinica Medica Generale, Ematologia ed Immunologia Clinica, Università Politecnica delle Marche, Ancona, Marche, Italy
8Rheumatology, Leiden University Medical Center, Leiden, Zuid-Holland, Netherlands
9Rheumatology A Department, Hospital Cochin, Paris, Ile-de-France, France
10Centre for Rheumatology, Royal Free and University College Medical School, London, UK
11Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, Utrecht, Netherlands
12Centre for Genomics and Oncological Research (GENO), Pfizer-University of Granada-Andalusian Regional Government, Granada, Spain
13Scierderoma Unit, Referral Center for Systemic Autoimmune Diseases, La Fondazione IRCCS Ca’ Granda Ospedale Maggiore di Milano Policlinico, Milano, Lombardia, Italy
14Twitter Shevin Assassi @ShevinAssassi

Acknowledgements
This work is part of the PhD thesis entitled Deciphering the genetic basis of systemic sclerosis by GV-M. We would like to thank Soñia Vargas and Gemma Robledo for their excellent technical assistance. We also appreciate the controls and the affected individuals who generously provided the samples for these studies.

Collaborators
International SSc Group: P. Carreira, Department of Rheumatology, 12 de Octubre University Hospital, Madrid, Spain; I. Castellvi, Department of Rheumatology, Santa Creu i Sant Pau University Hospital, Barcelona, Spain; R. Rios, Department of Internal Medicine, San Cecilio Clinic University Hospital, Granada, Spain; J. L. Callejas, Department of Internal Medicine, San Cecilio Clinic University Hospital, Granada, Spain; R. García Portales, Department of Rheumatology, Virgen de la Victoria Hospital, Málaga, Spain; A. Fernández-Nebro, Department of Rheumatology, Carlos Haya Hospital, Málaga, Spain; F. J. García-Hernández, Department of Internal Medicine, Virgen del Rocio Hospital, Sevilla, Spain; M. A. Aguirre, Department of Rheumatology, Reina Sofia/IMIBIC Hospital, Córdoba, Spain; B. Fernández-Gutiérrez, Department of Rheumatology, San Carlos Clinical Hospital, Madrid, Spain; L. Rodríguez-Rodriguez, Department of Rheumatology, San Carlos Clinical Hospital, Madrid, Spain; P. Garcia de la Peña, Department of Rheumatology, Madrid Norte Sanchinarro Hospital, Madrid, Spain; E. Vicente, Department of Rheumatology, La Princesa Hospital, Madrid, Spain; J. L. Andreu, Department of Rheumatology, Puerta de Hierro Hospital-Majadahonda, Madrid, Spain; M. Fernández de Castro, Department of Rheumatology, Puerta de Hierro Hospital-Majadahonda, Madrid, Spain; F. J. López-Longo, Department of Rheumatology, Gregorio Marañón University Hospital, Madrid, Spain; G. Espinosa, Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain; A. Guilén, Department of Internal Medicine, Vall d’Hebron Hospital, Barcelona, Spain; C. Tolsós, Department of Internal Medicine, Parc Taulí Hospital, Sabadell, Spain; A. Pros, Department of Rheumatology, Hospital Del Mar, Barcelona, Spain; M. Rodríguez Carballeira, Department of Internal Medicine, Hospital Universitari Mutua Terrassa, Barcelona, Spain; F. J. Narváez, Department of Rheumatology, Bellvitge University Hospital, Barcelona, Spain; M. Rubio Riva, Department of Internal Medicine, Bellvitge University Hospital, Barcelona, Spain; V. Ortiz-Santamaria, Department of Rheumatology, Granollers Hospital, Granollers, Spain; A. B. Madroñero, Department of Internal Medicine, Hospital General San Jorge, Huesca, Spain; M. A. González-Gay, Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, DIVAL, University of Cantabria, Santander, Spain; B. Díaz,

Systemic sclerosis
Department of Internal Medicine, Hospital Central de Asturias, Oviedo, Spain; L. Triapeila, Department of Internal Medicine, Hospital Central de Asturias, Oviedo, Spain; M. V. Eguardh, Department of Internal Medicine, Hospital Universitario Cruces, Barakaldo, Spain; P. Fanlo-Mateo, Department of Internal Medicine, Hospital Virgen del Camino, Pamplona, Spain; L. Saez-Comet, Department of Internal Medicine, Hospital Universitario Miguel Servet, Zaragoza, Spain; F. Diaz, Department of Rheumatology, Hospital Universitario de Canarias Tenerife, Spain; E. Bltearn, Department of Rheumatology, Hospital General Universitario de Valencia, Valencia, Spain; J. A. Roman-Ivana, Department of Rheumatology, Hospital Universitario Policlínico de Valencia, Valencia, Spain; Brian Muchmore, GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain; A. Hoffmann- Bossini-Castillo L, Department of Rheumatology, Hospital Universitario de Canarias Tenerife, Spain; Nieves Varela, GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain; Aleksandru Dufou, Immunology & Allergy, University Hospital and School of Medicine, Geneva, Switzerland; Montserrat Alvarez, Immunology & Allergy, University Hospital and School of Medicine, Geneva, Switzerland; Jonathan Cremers, Laboratory of Clinical Immunology, Department of Microbiology and Immunology, KU Leuven, Leuven, Belgium; Nuria Barborracho, IMIBIC, Reina Sofia Hospital, University of Cordoba, Cordoba, Spain; Veldia Ger, Department of Rheumatology and Clinical Immunology, Centre Hospitalier Universitaire de Montréal, Montreal, Canada; Christophe Jamin, Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany; Yingyu Cheng, Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany; Anne Buttgereit, Bayer AG, Berlin, Germany; Aurelie De Groof, Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium; Julie Druceue, Pôle de Pathologies Rhumatismales Inflammatoires et Systémiques, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium; Elena Trombetta, Laboratorio di Analisi Chimico Cliniche e Microbiologia - Servizio di Citofluorimetria, Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico di Milano, Milano, Italy; Tianlu Li, Chronatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain; Daniama Alvarez-Enrico, Chronatin and Disease Group, Bellvitge Biomedical Research Institute (IDIBELL), Barcelona, Spain; Torsten Witte, Klinik für Immunologie und Rheumatologie, Medical University Hannover, Hannover, Germany; Katja Kniech, Klinik für Immunologie und Rheumatologie, Medical University Hannover, Hannover, Germany; Nancy Azevedo, GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain; Esmeralda Neves, IMIBIC, Reina Sofia Hospital, University of Cordoba, Cordoba, Spain and Servio de Imunologia EX-CICAP, Centro Hospitalario Universitario de Porto, Porto, Portugal, Belgium; Sambasiva Rao, Sanofi Genzyme, Framingham, MA, USA; Pierre-Emmanuel Jouve, Atribio SAS, Lyon, France.

Contributors LB-C: data analysis, manuscript drafting, revision and approval; GV-M: data analysis, manuscript drafting, revision and approval; MIK: interpretation of data, manuscript revision and approval; MA-H: interpretation of data, manuscript revision and approval; ELI: data interpretation, manuscript revision and approval; PRECISESADS Clinical Consortium: data acquisition, manuscript revision and approval; MEAI-R: data interpretation, manuscript revision and approval; LB: data interpretation, manuscript revision and approval; JM: study design, manuscript drafting, revision and approval.

Funding This work was supported by the Spanish Ministry of Science and Innovation (grant ref. RTI201810332-B-100, Red de Investigacion en Inflamacion y Enfermedades Reumaticas (RIER) from Instituto de Salud Carlos III (RD16/0012/0013), and EU/EFPIA/Innovative Medicines Initiative Joint Undertaking PRECISESADS grant no. 111565, LBC and MAH were funded by the Spanish Ministry of Science and Innovation through the Juan de la Cierva incorporation program (ref. UC2018-030826-I and UC2018-035131-L, respectively). GV-M was funded by the Spanish Ministry of Science and Innovation through the Ayudas para contratos predoctorales para la formacion de doctores 2019 program (ref RTI2018-101332-B-I00).


Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval An ethical protocol was prepared, consensus was reached across all partners, academic and industrial, translated into all participants’ languages and approved by each of the local ethical committees of the clinical recruitment centres. The studies adhered to the standards set by the International Conference on Harmonization and Good Clinical Practice (ICH-GCP), and to the ethical principles that have their origin in the Declaration of Helsinki (2013). The protection of the confidentiality of records that could identify the included subjects is ensured as defined by the EU Directive 2001/20/EC and the applicable national and international requirements relating to data protection in each participating country. The CS study is registered with number NCT02890134, and the inception study with number NCT02890121, and the inception study with number NCT02890134 in ClinicalTrials.gov. The study (PRECISESADS cross-sectional study) was approved by the following ethical committees: Comitato Etico Area 2 (Fondazione IRCCS Ca’ Granda Ospedale Maggiore Policlinico di Milano and University of Milan); approval no. 4256b 19 November 2014, and no. 671_2018 19 September 2018; Klinikum der Universität zu Koeln, Cologne, Germany; Geschäftsstelle Ethikkommission, Klinik für Innere Medizin, Rheumatologie und Immunologie, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. Comité d’Ethique Hospitalo-Facultaire, University of Szeged, Szeged, Hungary. Consortium Meygai Kornamviah; Hospital

Systemic sclerosis

REFERENCES


