

A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci

Jose-Ezequiel Martin^{1,†,*}, Shervin Assassi^{2,†}, Lina-Marcela Diaz-Gallo¹, Jasper C. Broen^{3,4}, Carmen P. Simeon⁵, Ivan Castellvi⁶, Esther Vicente-Rabaneda⁷, Vicente Fonollosa⁵, Norberto Ortego-Centeno⁸, Miguel A. González-Gay⁹, Gerard Espinosa¹⁰, Patricia Carreira¹¹, Spanish Scleroderma Group[†], SLEGEN consortium[§], U.S. Scleroderma GWAS group^{**}, BIOLUPUS^{††}, Mayte Camps¹², Jose M. Sabio¹³, Sandra D'alfonso¹⁴, Madelon C. Vonk³, Alexandre E. Voskuyl¹⁵, Annemie J. Schuerwegh¹⁶, Alexander Kreuter¹⁷, Torsten Witte¹⁸, Gabriella Riemekasten¹⁹, Nicolas Hunzelmann²⁰, Paolo Airo²¹, Lorenzo Beretta²², Raffaella Scorza²², Claudio Lunardi²³, Jacob Van Laar²⁴, Meng May Chee²⁵, Jane Worthington²⁶, Arianne Herrick²⁶, Christopher Denton²⁷, Carmen Fonseca²⁷, Filemon K. Tan², Frank Arnett², Xiaodong Zhou², John D. Reveille², Olga Gorlova²⁸, Bobby P.C. Koeleman²⁹, Timothy R.D.J. Radstake^{4,†}, Timothy Vyse^{30,†}, Maureen D. Mayes^{2,†}, Marta E. Alarcón-Riquelme^{31,†} and Javier Martin^{1,†}

¹Instituto de Parasitología y Biomedicina Lopez-Neyra, CSIC, Granada, Spain, ²The University of Texas Health Science Center–Houston, Houston, TX, USA, ³Department of Rheumatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands, ⁴Department of Rheumatology, Clinical Immunology and Translational Immunology,

*To whom correspondence should be addressed. Email: cebercoto@ipb.csic.es

†These authors contributed equally to this work.

[‡]The Spanish Scleroderma Group members are: Rosa García Portales, Hospital Virgen de la Victoria, Málaga. Enrique de Ramon and Antonio Fernández-Nebro, Hospital Carlos Haya, Málaga. Jose Román-Ivorra and Emma Beltran, Hospital Dr. Peset, Valencia. M^a Jesús Castillo and Julio Sánchez-Román, Hospital Virgen del Rocío, Sevilla. Esther Vicente, Hospital La Princesa, Madrid. Bernardino Díaz, Luis Trapiella and María Gallego, Hospital Central de Asturias, Oviedo. M^a Angeles Aguirre, Hospital Reina Sofía, Córdoba. Jose Luis Callejas-Rubio and Raquel Ríos, Hospital San Cecilio, Granada. Iván Castellvi, Hospital Sant Pau, Barcelona. M^a Victoria Egurbide, Hospital de Cruces, Vizcaya. Gerard Espinosa, Hospital Clinic, Barcelona. Anna Pros, Hospital del Mar, Barcelona. Nuria Navarrete, Hospital Virgen de las Nieves, Granada. Federico Díaz-González and Vanesa Hernández, Hospital Universitario de Canarias, Tenerife. José Luis Andréu and Mónica Fernández-Castro, Hospital Puerta del Hierro, Madrid. Luis Sáez Comet, Hospital Miguel Servet, Zaragoza. Francisco J. López-Longo and Lina Martínez-Estupiñan, Hospital Gregorio Marañón, Madrid. Mónica Rodríguez-Carballeira, Hospital Universitari Mútua Terrassa, Barcelona. Francisco Javier Narváez, Servicio de Reumatología, Hospital Universitari de Bellvitge, Barcelona. María del Carmen Freire and Inés Vaqueiro, Hospital Xeral-Complejo Hospitalario, Vigo.

[§]The SLEGEN Consortium members are: Marta E. Alarcón-Riquelme (Oklahoma Medical Research Foundation and Centro de Genómica e Investigación Oncológica Pfizer-Universidad de Granada-Junta de Andalucía), Lindsey A. Criswell (University of California, San Francisco), John B. Harley (Cincinnati Children's Hospital Medical Center), Chaim O. Jacob (University of Southern California), Robert P. Kimberly (University of Alabama at Birmingham), Carl D. Langefeld (Wake Forest University Health Sciences), Kathy L. Moser (Oklahoma Medical Research Foundation), Betty P. Tsao (University of California, Los Angeles), and Timothy J. Vyse (King's College London).

^{**}The United States Scleroderma GWAS Group members are: Jun Ying (Department of Epidemiology, M.D. Anderson Cancer Center, Houston, TX, USA), John Varga, Monique Hinchcliff (Northwestern University Feinberg School of Medicine, Chicago, IL USA), Peter Gregersen, Annette T. Lee (Feinstein Institute of Medical Research, Manhasset, NY, USA), Fredrick M. Wigley, Laura Hummers (Johns Hopkins University Medical Center, Baltimore, MD, USA).

^{††}The following samples were obtained in via the BIOLUPUS network coordinated by Marta Alarcón-Riquelme (GENYO, OMRF) Belgium: Bernard Lawerys and Fredric Houssiau (Université Catholique de Louvain) Denmark: Søren Jacobsen (University of Copenhagen), Peter Junker, Helle Lastrup (Odense University Hospital). Germany: Torsten Witte (Medizinische Hochschule Hannover). Greece: Haralampos Moutsopoulos, Etstathia K Kapsogeorgou (National University of Athens). Hungary: Emöke Endreffy and Laszlo Kovacs (Albert Szent-Györgyi Medical University). Iceland: Kristján Steinsson (Landspítali National University Hospital). Italy: Andrea Doria (University of Padova), Pier Luigi Meroni (IRCCS Istituto Auxologico Italiano), Raffaella Scorza (University of Milan), Sandra D'Alfonso (providing samples from Rome, Naples and Siena, Università del Piemonte Orientale). Netherlands: Marc Bijl, Cees Kallenberg (University of Groningen). Portugal: Carlos Vasconcelos (Hospital Santo António, Porto), Berta Martins Silva (University of Porto). Spain: Javier Martín, Jose-Ezequiel Martín (Instituto de Parasitología y Biomedicina "Lopez-Neyra"), Ana Suárez (Hospital Universitario Central de Asturias), Iñigo Rúa Figueroa (Hospital Dr Negrín, Gran Canaria), Guillermo Pons-Estel (Hospital Clinic, Barcelona). Sweden: Johan Frostegård (Karolinska Institutet). From the GENLES collaboration: Argentina: Bernardo Pons-Estel (Hospital Provincial de Rosario).

University Utrecht Medical Center, Utrecht, The Netherlands, ⁵Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona, Spain, ⁶Servicio de Reumatología, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain, ⁷Servicio de Reumatología, Hospital La Princesa, Madrid, Spain, ⁸Servicio de Medicina Interna, Hospital Clínico Universitario, Granada, Spain, ⁹Servicio de Reumatología, Hospital Marqués de Valdecilla, IFIMAV, Santander, Spain, ¹⁰Department of Systemic Autoimmune Diseases, Hospital Clinic de Barcelona, Barcelona, Spain, ¹¹Hospital 12 de Octubre, Madrid, Spain, ¹²Servicio de Medicina Interna, Hospital Carlos Haya, Málaga, Spain, ¹³Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital Universitario Virgen de las Nieves, Granada, Spain, ¹⁴Department of Medical Sciences and Institute of Research in Chronic Autoimmune Diseases (IRCAD), University of Eastern Piedmont, Novara, Italy, ¹⁵VU University Medical Center, Amsterdam, The Netherlands, ¹⁶Department of Rheumatology, University of Leiden, Leiden, The Netherlands, ¹⁷Department of Dermatology, Josefs-Hospital, Ruhr University Bochum, Bochum, Germany, ¹⁸Hannover Medical School, Hannover, Germany, ¹⁹Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany, ²⁰Department of Dermatology, University of Cologne, Cologne, Germany, ²¹Rheumatology Unit and Chair, Spedali Civili, Università de Gli Studi, Brescia, Italy, ²²Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca'Granda Ospedale Ma Repiore Policlinico and University of Milan, Milan, Italy, ²³Department of Medicine, Policlinico GB Rossi, University of Verona, Verona, Italy, ²⁴Institute of Cellular Medicine, Newcastle University, Newcastle Upon Tyne, UK, ²⁵Centre for Rheumatic Diseases, Glasgow Royal Infirmary Glasgow, Glasgow, UK, ²⁶Department of Rheumatology and Epidemiology, University of Manchester, Manchester Academic Health Science Centre, Manchester, UK, ²⁷Centre for Rheumatology, Royal Free and University College School, London, ²⁸Department of Epidemiology, MD Anderson Cancer Center, Houston, TX, USA, ²⁹Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands, ³⁰Divisions of Genetics and Molecular Medicine and Division of Immunology, Infection and Inflammatory Disease, King's College London, Guy's Hospital, London, UK and ³¹Centro de Genómica e Investigación Oncológica (GENYO) Pfizer-Universidad de Granada-Junta de Andalucía, Granada, Spain

Received March 22, 2013; Revised and Accepted May 28, 2013

Systemic sclerosis (SSc) and systemic lupus erythematosus (SLE) are two archetypal systemic autoimmune diseases which have been shown to share multiple genetic susceptibility loci. In order to gain insight into the genetic basis of these diseases, we performed a pan-meta-analysis of two genome-wide association studies (GWASs) together with a replication stage including additional SSc and SLE cohorts. This increased the sample size to a total of 21 109 (6835 cases and 14 274 controls). We selected for replication 19 SNPs from the GWAS data. We were able to validate *KIAA0319L* ($P = 3.31 \times 10^{-11}$, OR = 1.49) as novel susceptibility loci for SSc and SLE. Furthermore, we also determined that the previously described SLE susceptibility loci *PXK* ($P = 3.27 \times 10^{-11}$, OR = 1.20) and *JAZF1* ($P = 1.11 \times 10^{-8}$, OR = 1.13) are shared with SSc. Supporting these new discoveries, we observed that *KIAA0319L* was overexpressed in peripheral blood cells of SSc and SLE patients compared with healthy controls. With these, we add three (*KIAA0319L*, *PXK* and *JAZF1*) and one (*KIAA0319L*) new susceptibility loci for SSc and SLE, respectively, increasing significantly the knowledge of the genetic basis of autoimmunity.

INTRODUCTION

Most autoimmune disorders are genetically complex and clinically heterogeneous. Classification permits physicians to distinguish individual autoimmune criteria based on the typical clinical features. However, underlying these separate clinical features, there is a genetic continuum of susceptibility factors and molecular pathways leading to autoimmunity. This becomes clear when analyzing existing genetic data according to clinical subphenotypes, reducing the overall genetic heterogeneity in the analyses (1–8). This has been clearly demonstrated in rheumatoid arthritis (RA), in which separating anti-citrullinated protein antibody positive and negative cases resulted in a far more homogeneous genetic analysis revealing

genetic subgroups (1,3–6). The aim of personalized medicine is thus to accomplish a degree of resolution that allows us to distinguish genetically and molecularly such groups and provide more targeted therapeutic interventions.

There is a growing body of evidence that all autoimmune disorders share to a varying degree their genetic susceptibility loci (9). It is clear that lack of statistical power due to limited sample size is a barrier to completely defining the genetic contribution to autoimmunity. As we increase our sample size, we can identify additional genes shared by some diseases but not shared by others. From the viewpoint that all autoimmune disorders are heterogeneous entities whose specific manifestations depend on the presence of several susceptibility genetic variants and

environmental triggers, the combined analysis of different autoimmune disorders will greatly increase our statistical power to detect modest genetic effects shared among them. This approach has already been successfully used to detect the shared genetic susceptibility component of RA, celiac disease, psoriasis and Crohn's disease (10–12). In the present study, we meta-analyze the GWASs of two autoimmune disorders that have been demonstrated to share a relatively large portion of their genetic component and clinical features: systemic sclerosis (SSc) and systemic lupus erythematosus (SLE) (13, 14), identifying new loci for each disease. We would suggest a similar approach for other autoimmune diseases in the future.

RESULTS

The overall process followed during this study is illustrated in Figure 1 and described in detail in the Materials and methods section. Briefly, from the pan-meta-analysis of the SSc and SLE GWAS (15,16), we selected all SNPs which presented a GC-corrected P -value of $<5 \times 10^{-6}$. From these, we further selected for replication all SNPs which were associated with both diseases independently at nominal level. For a more detailed description of the selection criteria, see the Materials and methods section.

According to these selection criteria, we investigated 19 SNPs whose statistics are shown in Table 1 and Figure 2. According to the significance criteria established (see Materials and methods section), we could confirm one new association shared by SSc (more specifically in the limited cutaneous subtype) and SLE: rs2275247 in *KIAA0319L* (combined P -value = 1.40×10^{-10} , OR = 1.51; SSc P -value = 3.92×10^{-6} ; SLE P -value = 4.44×10^{-6}) (Table 1). Furthermore, we were able to determine two new genetic susceptibility regions at genome-wide level of significance in the combined analysis of SSc which had been previously described only in SLE: *PXK* (rs2176082, combined P -value = 5.24×10^{-11} , OR = 1.21; SSc P -value = 4.44×10^{-7} ; SLE P -value = 1.12×10^{-5}) and *JAZF1* (combined P -value = 1.60×10^{-8} , OR = 1.14; SSc P -value = 3.63×10^{-5} ; SLE P -value = 1.19×10^{-4}) but with a modest association particularly

for the SSc analysis alone. This suggests a minor role of these genetic variants in SSc compared with SLE (Table 1). It is noteworthy that the associations present in the *KIAA0319L* and *PXK* regions were still significant at GWAS level after including as covariates the first three principal components in a logistic regression model, while the association in the *JAZF1* was slightly reduced below GWAS level after this (the P -value not corrected for principal components was 1.60×10^{-8} and after principal component correction was 5.52×10^{-8}), which suggests minor population stratification for this SNP (Table 1).

In whole blood gene expression data, we observed that *KIAA0319L* was significantly overexpressed in SLE patients compared with unaffected controls [$P = 5.36 \times 10^{-5}$, false discovery rate (FDR) = 2.94×10^{-3}] with a fold change of 1.9, while in SSc, there was a trend towards overexpression compared with healthy individuals ($P = 9.05 \times 10^{-3}$, FDR = 0.14) with a fold change of 1.3. No significant expression differences were observed for *PXK* and *JAZF1* in either SSc or SLE compared with controls (Table 2).

Finally, we wanted to address how the biological processes are differentially distributed and shared between SSc and SLE. For this, we assigned each SNP with a P -value of <0.05 and 5×10^{-6} to a gene (allowing a 20 kb window surrounding it). With this we were able to determine whether any particular biological process was overrepresented among these two sets of GWAS association ($P < 0.05$ and $P < 5 \times 10^{-6}$). As observed in Table 3, the biological processes most significantly overrepresented among the SSc and SLE associated genes were those belonging to the immune system (27 processes for SSc, being the best P -value = 2.05×10^{-13} ; and 8 processes for SLE being the best $P = 2.14 \times 10^{-8}$). Among these immune system processes involved in SSc and SLE, the topmost one was cellular response to interferon-gamma (GO term 0071346). Another set of processes apparently shared by SSc and SLE were those related to the nervous system, while signaling, molting and cell adhesion were exclusively overrepresented among the SSc associated genes (Table 3).

DISCUSSION

Utilizing a GWAS pan-meta-analysis strategy, we were able to identify a total of three new autoimmunity susceptibility loci, all of which are new for SSc and one of which is new for SLE. These new genetic susceptibility loci were undetected in the previous SSc and SLE GWASs due to the lack of statistical power.

The clearest example of this is rs2275247 in *KIAA0319L*, with a minor allele frequency of 3.43% in our GWAS cohorts, which would have gone undetected as Supplementary Material, Table S1, predicts. However, due to the combined analysis, we were able to capture this association in our meta-analysis. *KIAA0319L* was previously associated with learning and cognition disabilities (17). Interestingly, the protein contains among its evolutionary conserved domains a polycystic kidney disease domain, which is an immunoglobulin family-like domain of unclear function (18). Furthermore, *KIAA0319L* was significantly overexpressed in peripheral blood cells from SLE patients compared with those of healthy controls, also showed that a similar trend was observed in SSc (Table 2). Another possible lead pointing to the role of *KIAA0319L* in autoimmunity can be found in

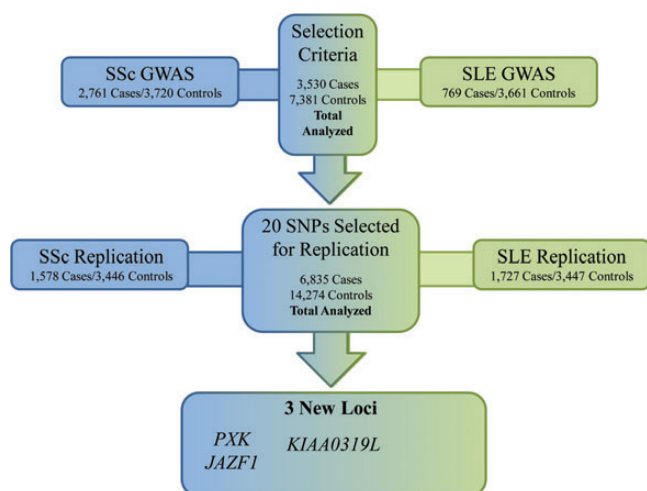


Figure 1. Workflow diagram showing the overall process followed during the present work.

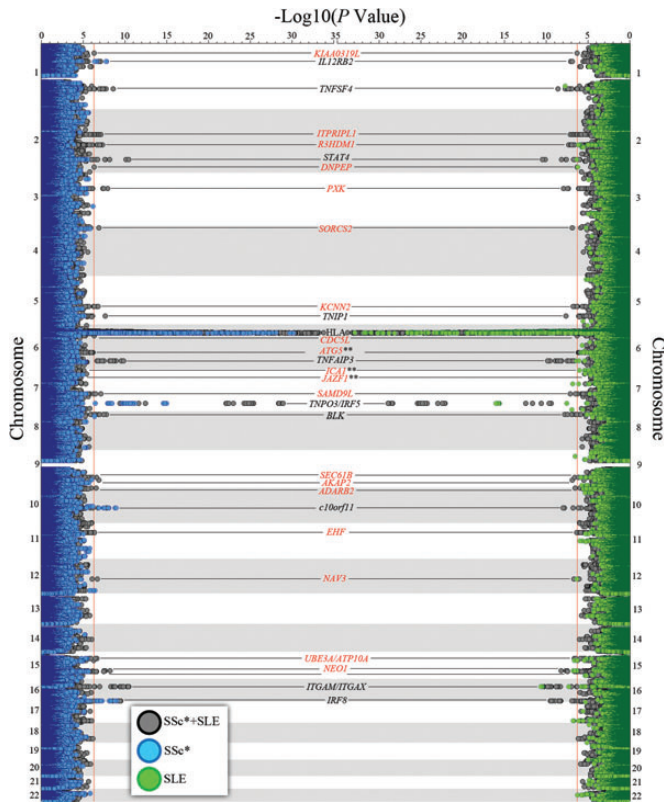


Figure 2. Twin Manhattan plot representing the results of the SSc (left side, blue), SLE (right side, green) and combined GWAS analysis (both sides, grey). Selected loci for replication in independent cohorts are marked in red. * SSc plotting represents either the total disease or any of its considered subphenotypes, i.e. ACA positive, ATA positive, lcSSc and dcSSc. ** *ICA1* and *JAZF1* SNPs were selected according to selection criteria three (see Materials and methods), and not because of reaching the significance threshold in the GWAS stage.

the expression profile of this gene in the bioGPS public database (<http://biogps.org/#goto=welcome>); although this gene is expressed ubiquitously, it has a particularly high expression in immune cells such as macrophages, natural killer cells and other hematopoietic cells in the mouse (Supplementary Material, Fig. S1) and CD33+ myeloid cells and CD14+ monocytes in humans (Supplementary Material, Fig. S2). Meanwhile, *PXX* and *JAZF1* were previously described as SLE genetic risk factors (15,19). The role of these three genes in SLE pathogenesis remains largely unknown, it is also unclear whether their pathogenic mechanism is the same or different in SSc. However, *PXX* has been described to be involved in the trafficking of the epidermal growth factor receptor, which poses an attractive role in both SSc and SLE (20). Furthermore, *PXX* has been described to be associated only with autoantibody production in SLE in a small cohort (21), analogously to what we have observed in SSc in this study (Table 1). As of *JAZF1*, it has been associated with skeletal frame size and height, relating it to bone morphogenesis and collagen deposit (22,23), a key process in disorders affecting skin such as SSc and SLE.

In GWAS data, we could classify the genetic associations into three groups: Tier 1 for the GWAS-level associations ($P < 5 \times 10^{-8}$), Tier 2 for suggestive associations ($P < 5 \times 10^{-6}$) and Tier 3 (all the remaining associations until 0.05). In the analysis

of biological processes performed in this study (Table 3, Supplementary Material, Table S2), we used two sets of genes: first the Tier 1 and Tier 2 associations (stronger associations in general with better P -values and ORs) and the Tier 3 associations, in which, should there be any true genetic association, it would be weaker. When using all SNPs (translated to nearby genes) from Tiers 1 and 2 in this analysis, we observed that many biological processes gathered under the immune system label are important for both SSc and SLE, which is to be expected. Still in the Tier 1 and 2 associations, we see that signaling processes (cytokine mediated) and molting (to process by which the external layers of complex living beings is renewed) are also important to SSc. In fact, we have already seen this in SSc genetic studies which have described genes involved in morphogenesis and fibrosis to play a role in this disease (2,16,24). Conversely, when we perform the biological process analysis including the Tier 3 genes, we observe that genes involved in the nervous system are important to both SSc and SLE (although different specific biological processes, Supplementary Material, Table S2) and cell adhesion is important to SSc. These biological processes being involved in SSc and SLE make sense based on the current knowledge (25–33), but so far we have failed to find the specific genetic markers which could induce the neurological complications of SLE or SSc or cell adhesion in SSc.

Often, the association of a genetic variant with a trait is referred to by some gene in the same region, picked by its proximity or functional attractiveness considering the studied trait. Nevertheless, we also wanted to assess whether any of the described associations with SSc or SLE were in the same region and represented the same signal attributed by different genes. For this, we performed a GRAIL analysis (34), taking into account all described genetic associations with a P -value of $< 5 \times 10^{-4}$ in the bibliography with any of the diseases. As seen in Figure 3, all of the associated regions are represented by the previously described genes in those regions, with one notable exception: in the genetic region represented by the association of rs12540874, the association was attributed to *GRB10* in SSc (2) and to *IKZF1* in SLE (15). Indeed, the signals were different in the studies by Gorlova *et al.* and Harley *et al.*, but according to the GRAIL analysis the most plausible associated gene based on the current knowledge is *IKZF1* (having connections with 12 other genes in the analysis) rather than *GRB10* (0 connections according to the literature on the other genes).

Due to the increased statistical power derived from the merging of two GWASs in SSc and SLE and the addition of large replication cohorts, we were able to establish one new functionally attractive susceptibility locus for SSc and SLE (*KIAA0319L*) and two new susceptibility loci for SSc (*PXX* and *JAZF1*) at GWAS level. This study, together with several others, adds evidence for the genetic continuum that underlies SSc, SLE and most autoimmune disorders.

MATERIALS AND METHODS

Study cohorts

The GWAS cohorts analyzed in this study were composed of a total of 3530 cases and 7381 healthy controls. Of these, 2761 cases and 3720 healthy controls belonged to a previous SSc GWAS conducted on cohorts from Spain, Germany, The

Table 2. Expression data of the successfully replicated loci from Table 1, whether at GWAS level or suggestive level of association

Chr.	Gene	SSc versus controls			SLE versus controls		
		Parametric <i>P</i>	FDR	Fold change	Parametric <i>P</i>	FDR	Fold change
1	<i>KIAA0319L</i>	9.05E-03	1.43E-01	1.328	5.36E-05	2.94E-03	1.937
3	<i>PXK</i>	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a	ND ^a
7	<i>JAZF1</i>	9.04E-01	9.60E-01	0.990	3.90E-01	6.17E-01	0.870

Significantly overexpressed genes are marked in bold.

^a*PXK* was not sufficiently expressed in whole blood samples and did not pass the filtering criteria.

Table 3. Significantly overrepresented groups of biological processes among the genes associated with either SSc or SLE and whether they are shared between them or not

Disease	SNP selection ^a	Group	Top biological process	Shared	Significant processes	GO term	Best <i>P</i> -value ^b
SSc	5.00E-06	Immune system	Cellular response to interferon-gamma	Yes	27	GO:0071346	2.05E-13
SSc	5.00E-06	Signaling	Cytokine-mediated signaling pathway	No	8	GO:0019221	4.57E-11
SSc	0.05	Nervous system	Transmission of nerve impulse	No	3	GO:0019226	9.71E-10
SSc	5.00E-06	Molting	Exogen	No	7	GO:0042638	2.61E-09
SSc	0.05	Cell adhesion	Cell adhesion	No	4	GO:0007155	3.29E-09
SLE	5.00E-06	Immune system	Cellular response to interferon-gamma	Yes	8	GO:0071346	2.14E-08
SLE	0.05	Nervous system	Nervous system development	No	4	GO:0007399	1.34E-07

Significantly overrepresented biological processes are marked in bold.

^aThreshold *P*-value for SNP selection from the GWAS data for the biological process analyses.

^bThe *P*-value for the FDR-corrected hypergeometric test for the most associated biological process for this group (see the Materials and methods section for more details).

Netherlands and USA (16). The rest of the cases and controls belonged to a previously published SLE GWAS conducted in the USA, SLEGEN cohort composed of 769 SLE patients and 3661 healthy controls (15) (Supplementary Material, Tables S3 and S4).

We selected independent SSc and SLE replication cohorts in order to confirm the results observed in the previous meta-GWAS stage. The SSc replication cohort was composed of 432 cases and 380 controls from Spain, 691 cases and 241 controls from Italy and 455 cases and 2826 controls from the UK, while the SLE replication cohort was composed of 375 cases and 380 controls from Spain, 335 cases and 240 controls from Italy and 1017 cases and 2826 controls from the UK (Supplementary Material, Table S3). Supplementary Material, Table S4, shows key features of the SSc cohorts analyzed. All the samples in the replication cohorts were recruited from hospitals and clinics of each country after approval by the corresponding ethics committees. Genotype data from the UK replication controls were obtained from the WTCCC data repositories, for which we were granted access.

All cases either met the American College of Rheumatology preliminary criteria for the classification of SSc (35) or had at least three of the five CREST (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasias) features and were classified according to their skin involvement and their auto-antibody production status (36,37). All SLE patients in the current study fulfilled the revised criteria for classification of SLE from the American College of Rheumatology (38). All individuals enrolled in the present study provided written informed consents.

Data quality control

GWAS data were filtered as previously described (16) using as a limits a 90% success call rate per SNP and individual, a deviation from the Hardy-Weinberg equilibrium of a *P*-value <0.0001 and a minor allele frequency <1%. The first 10 principal components were estimated and individuals who deviated more than three standard deviations from the centroid of their population in the first two principal components were excluded as outliers. The replication cohorts were filtered in the same way. The data for PCA adjustment in the logistic regression were obtained from Immunochip and GWAS data of our replication cohorts, except for the 455 SSc UK cases.

Genotyping

The GWAS genotyping of the SSc cases and controls was performed as follows: the Spanish SSc cases and controls together with Dutch and German SSc cases were genotyped at the Department of Medical Genetics of the University Medical Center Utrecht (The Netherlands) using the commercial release Illumina Human CNV370K BeadChip. Genotype data for Dutch and German controls were obtained from the Illumina Human 550KBeadChip available from a previous study. The SSc case group from the USA was genotyped at the Boas Center for Genomics and Human Genetics, Feinstein Institute for Medical Research, North Shore Long Island Jewish Health System using the Illumina Human 610-Quad BeadChip. CGEMS and Illumina iControl DBcontrols were genotyped on the Illumina Hap550KBeadChip. SNPs selected for the replication phase were genotyped in the

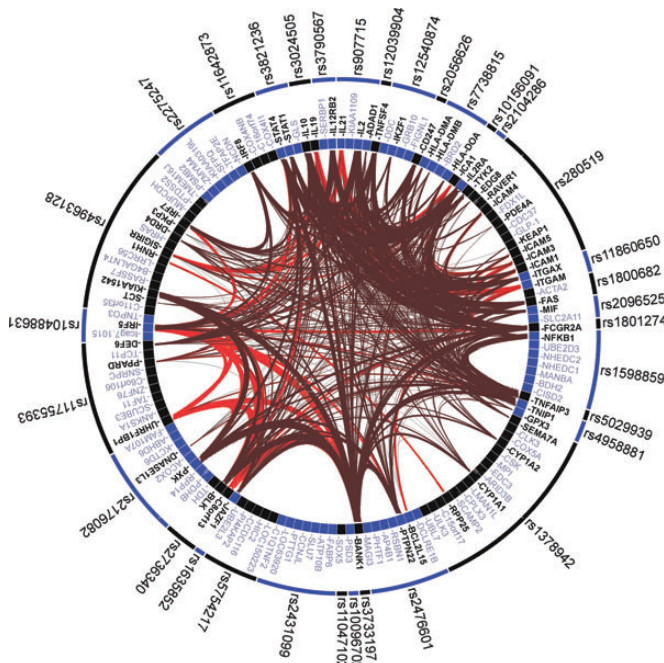


Figure 3. GRAIL analysis of all the described susceptibility loci for SSc and/or SLE with $P < 5 \times 10^{-4}$. The release 18 of the human genome and the PubMed text as of 2012 were used for the GRAIL analysis.

replication cohorts using Applied Biosystems' TaqMan SNP assays on ABI Prism 7900 HT real-time thermocyclers.

Study design

Considering the size of our cohort for both study phases, simultaneously we calculated the statistical power for the different scenarios according to Skol *et al.* (39). Supplementary Material, Table S1, shows the statistical power to detect different effect sizes. On average, we had 86% statistical power to detect an OR of 1.20 with a minor allele frequency of 0.20.

Taking this into consideration, for the GWAS analysis three different criteria were followed to select SNPs for replication and maximization of our success in signal detection:

- (1) In order to detect common signals for SSc and SLE which caused either risk or protection in both diseases, we selected SNPs that showed a P -value of the SSc and SLE meta-analysis $< 5 \times 10^{-7}$ and showed a nominally significant association with both diseases at a P -value of < 0.05 , as well as no significant heterogeneity in the SSc cohorts meta-analysis ($Q > 0.05$) (Supplementary Material, Table S5).
- (2) To detect common signals for SSc and SLE which caused either risk or protection in one of the diseases and the opposite effect in the other, we selected SNPs that showed a P -value of the SSc and SLE meta-analysis (using for SLE in this case $1/\text{OR}$ instead of the OR) $< 5 \times 10^{-7}$ and were associated with both a P -value of < 0.05 in both diseases, and without significant heterogeneity in the SSc cohorts meta-analysis ($Q > 0.05$) (Supplementary Material, Table S6).
- (3) To further detect any susceptibility variants previously reported for one of the diseases but not reported for the other, we selected any SNP with an overall P -value of

$< 5 \times 10^{-5}$ and associated separately in SSc and SLE (at a P -value of < 0.05) which had been previously described as a genetic risk factor for either disease (Supplementary Material, Table S7).

In any of these cases when performing the pan-meta-analysis of SSc and SLE, we also considered the most frequent SSc subphenotypes (classified as stated above): anti-centromere antibody (ACA) positive subgroup, anti-topoisomerase I antibody (ATA) positive subgroup, limited cutaneous subtype (lcSSc) and diffuse cutaneous subtype (dcSSc) (36,37). Supplementary Material, Figure S3, shows from which subphenotype of SSc each of the selected SNPs was derived from for the replication step.

After the replication stage was completed, we considered a signal to be statistically significant if the combined (SSc and SLE, GWAS and replication cohorts) meta-analysis P -value was $< 5 \times 10^{-8}$, and if the meta-analysis P -value (GWAS and replication cohorts) for each disease and stage was also significant.

Statistical analysis

Meta-analysis of the SSc GWAS data was performed with the Cochran–Mantel–Haenszel test correcting for the genomic inflation factor lambda (GC correction) (Supplementary Material, Fig. S4). Analysis of the SLE GWAS data was performed by a simple $\chi^2 2 \times 2$ test correcting the P -values for the genomic inflation factor lambda (GC correction) (Supplementary Material, Fig. S4). The pan-meta-analysis of the SSc and SLE GWAS data was performed using the inverse variance method calculating the resulting ORs. When the replication cohorts were included in the analysis, they were also analyzed using the inverse variance method. The analysis of the GWAS and replication cohorts together was also performed by means of logistic regression analysis using the first three principal components as covariates in order to rule out any population stratification effect. When different associations were found in the same locus, the underlying association hit was determined by means of conditional logistic regression analysis. All statistical analyses were performed using Plink version 1.07 (40) (<http://pngu.mgh.harvard.edu/~purcell/plink/>) and HelixTree SNP Variation Suite 7 (<http://www.goldenhelix.com/>).

Biological process analysis

We performed the biological process analysis of the SSc and SLE GWAS data according to Zhang *et al.* (41). For this, we performed two analyses on each disease, one selecting all SNPs with a P -value < 0.05 and the other selecting all SNPs with a P -value $< 5 \times 10^{-6}$. In both the cases, the test to determine whether a biological process (GO term) was overrepresented among GWAS associations was performed by means of a hypergeometric test with a cutoff P -value of 5×10^{-6} . All P -values in these analyses were multiple testing corrected according to Hochberg and Benjamini (42).

Gene expression Data

We had access to whole blood gene expression data from 74 SSc patients, 17 SLE patients and 21 healthy controls (43); which

were not included in the GWAS cohorts. None of the patients were treated with immunosuppressive agents (exception prednisone ≤ 5 mg or hydroxychloroquine). Blood samples for transcript studies were drawn directly into PAXgene tubes (PreAnalytiX, Franklin Lakes, NJ, USA). Total RNA was isolated according to the manufacturer's protocol using the PAXgene RNA kit (PreAnalytiX). The RNA quality and yield were assessed using a 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA) and an ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). Two hundred nanograms of total RNA were amplified and purified using the Illumina TotalPrep RNA Amplification kit (Applied Biosystems/Ambion, Austin, TX, USA) in accordance with the manufacturer's instructions. The amplified complementary RNA was hybridized on Illumina Human Ref-8 BeadChips, and the data were extracted with the Illumina Beadstudio software suite (Illumina, San Diego, CA, USA). A transcript was defined as differentially expressed when the significance level for comparison was $P \leq 0.05$ and the FDR was ≤ 0.10 using a random-variance *t*-test.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Sofia Vargas, Sonia Garcia and Gema Robledo for their excellent technical assistance, and all the patients and healthy controls for kindly accepting their essential collaboration. We also thank the following organizations: the EULAR Scleroderma Trials and Research (EUSTAR), the German Network of Systemic Sclerosis and Banco Nacional de ADN (University of Salamanca, Spain).

FUNDING

This work was supported by the following grants: J.M. was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009-11110 and SAF2012-34435 from the Ministerio de Economía y Competitividad, CTS-4977 from Junta de Andalucía (Spain), Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII, Spain), Fondo Europeo de Desarrollo Regional (FEDER) and the grant NIH NIAID 1U01AI09090-01. T.R.D.J.R. was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds) and is an ERC starting grant Laureate 2011. J.M. and T.R.D.J.R. were sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). B.P.C.K. is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). S.A. and M.D.M. were supported by NIH/NIAMS Scleroderma Family Registry and DNA Repository (N01-AR-0-2251), NIH/NIAMS-RO1- AR055258, and NIH/NIAMS Center of Research Translation in Scleroderma (1P50AR054144), the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111),

and K23-AR-061436. M.E.A.R. was funded by the Instituto de Salud Carlos III partially through FEDER funds of the European Union (PS09/00129), la Consejería de Salud de Andalucía (PI0012), la Fundación Ramón Areces and the Swedish Research Council. M.E.A.R. is the Chairman of the BIOLUPUS network funded by the European Science Foundation. T.W. is funded by the grant KFO 250, TP03, WI 1031/6-1.

REFERENCES

1. Terao, C., Ohmura, K., Ikari, K., Kochi, Y., Maruya, E., Katayama, M., Yurugi, K., Shimada, K., Murasawa, A., Honjo, S. *et al.* (2012) ACPA-negative RA consists of two genetically distinct subsets based on RF positivity in Japanese. *PLoS One*, **7**, e40067.
2. Gorlova, O., Martin, J.E., Rueda, B., Koeleman, B.P., Ying, J., Teruel, M., Diaz-Gallo, L.M., Broen, J.C., Vonk, M.C., Simeon, C.P. *et al.* (2011) Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet.*, **7**, e1002178.
3. Lundberg, K., Bengtsson, C., Kharlamova, N., Reed, E., Jiang, X., Kallberg, H., Pollak-Dorocic, I., Israelsson, L., Kessel, C., Padyukov, L. *et al.* (2012) Genetic and environmental determinants for disease risk in subsets of rheumatoid arthritis defined by the anticitrullinated protein/peptide antibody fine specificity profile. *Ann. Rheum. Dis.*, **72**, 652–658.
4. Mackie, S.L., Taylor, J.C., Martin, S.G., Wordsworth, P., Steer, S., Wilson, A.G., Worthington, J., Emery, P., Barrett, J.H. and Morgan, A.W. (2012) A spectrum of susceptibility to rheumatoid arthritis within HLA-DRB1: stratification by autoantibody status in a large UK population. *Genes Immun.*, **13**, 120–128.
5. de Vries, R.R., van der Woude, D., Houwing, J.J. and Toes, R.E. (2011) Genetics of ACPA-positive rheumatoid arthritis: the beginning of the end? *Ann. Rheum. Dis.*, **70**, i51–54.
6. Padyukov, L., Seielstad, M., Ong, R.T., Ding, B., Ronnelid, J., Seddighzadeh, M., Alfredsson, L. and Klareskog, L. (2011) A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann. Rheum. Dis.*, **70**, 259–265.
7. Sanchez, E., Nadig, A., Richardson, B.C., Freedman, B.I., Kaufman, K.M., Kelly, J.A., Niewold, T.B., Kamen, D.L., Gilkeson, G.S., Ziegler, J.T. *et al.* (2011) Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus. *Ann. Rheum. Dis.*, **70**, 1752–1757.
8. Chung, S.A., Taylor, K.E., Graham, R.R., Nititham, J., Lee, A.T., Ortmann, W.A., Jacob, C.O., Alarcon-Riquelme, M.E., Tsao, B.P., Harley, J.B. *et al.* (2011) Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production. *PLoS Genet.*, **7**, e1001323.
9. Cho, J.H. and Gregersen, P.K. (2011) Genomics and the multifactorial nature of human autoimmune disease. *N. Engl. J. Med.*, **365**, 1612–1623.
10. Festen, E.A., Goyette, P., Green, T., Boucher, G., Beauchamp, C., Trynka, G., Dubois, P.C., Lagace, C., Stokkers, P.C., Hommes, D.W. *et al.* (2011) 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.*, **7**, e1001283.
11. Zhernakova, A., Stahl, E.A., Trynka, G., Raychaudhuri, S., Festen, E.A., Franke, L., Westra, H.J., Fehrmann, R.S., Kurzeeman, F.A., Thomson, B. *et al.* (2011) Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. *PLoS Genet.*, **7**, e1002004.
12. Ellinghaus, D., Ellinghaus, E., Nair, R.P., Stuart, P.E., Esko, T., Metspalu, A., Debrus, S., Raelson, J.V., Tejasvi, T., Belouchi, M. *et al.* (2012) Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. *Am. J. Hum. Genet.*, **90**, 636–647.
13. Martin, J.E., Bossini-Castillo, L. and Martin, J. (2012) Unraveling the genetic component of systemic sclerosis. *Hum. Genet.*, **131**, 1023–1037.
14. Delgado-Vega, A., Sanchez, E., Lofgren, S., Castillejo-Lopez, C. and Alarcon-Riquelme, M.E. (2010) Recent findings on genetics of systemic autoimmune diseases. *Curr. Opin. Immunol.*, **22**, 698–705.
15. Harley, J.B., Alarcon-Riquelme, M.E., Criswell, L.A., Jacob, C.O., Kimberly, R.P., Moser, K.L., Tsao, B.P., Vyse, T.J., Langefeld, C.D., Nath, S.K. *et al.* (2008) Genome-wide association scan in women with systemic

- lupus erythematosus identifies susceptibility variants in ITGAM, PXX, KIAA1542 and other loci. *Nat. Genet.*, **40**, 204–210.
16. Radstake, T.R., Gorlova, O., Rueda, B., Martin, J.E., Alizadeh, B.Z., Palomino-Morales, R., Coenen, M.J., Vonk, M.C., Voskuyl, A.E., Schuerwegh, A.J. *et al.* (2010) Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet.*, **42**, 426–429.
 17. Poelmans, G., Buitelaar, J.K., Pauls, D.L. and Franke, B. (2011) A theoretical molecular network for dyslexia: integrating available genetic findings. *Mol Psychiatry*, **16**, 365–382.
 18. Bycroft, M., Bateman, A., Clarke, J., Hamill, S.J., Sandford, R., Thomas, R.L. and Chothia, C. (1999) The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *EMBO J.*, **18**, 297–305.
 19. Gateva, V., Sandling, J.K., Hom, G., Taylor, K.E., Chung, S.A., Sun, X., Ortmann, W., Kosoy, R., Ferreira, R.C., Nordmark, G. *et al.* (2009) A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat. Genet.*, **41**, 1228–1233.
 20. Takeuchi, H., Takeuchi, T., Gao, J., Cantley, L.C. and Hirata, M. (2010) Characterization of PXX as a protein involved in epidermal growth factor receptor trafficking. *Mol. Cell. Biol.*, **30**, 1689–1702.
 21. Yu, B., Wu, Q., Chen, Y., Li, P., Shao, Y., Zhang, J., Zhong, Q., Peng, X., Yang, H., Hu, X. *et al.* (2011) Polymorphisms of PXX are associated with autoantibody production, but not disease risk, of systemic lupus erythematosus in Chinese mainland population. *Lupus*, **20**, 23–27.
 22. Soranzo, N., Rivadeneira, F., Chinapen-Horsley, U., Malkina, I., Richards, J.B., Hammond, N., Stolk, L., Nica, A., Inouye, M., Hofman, A. *et al.* (2009) Meta-analysis of genome-wide scans for human adult stature identifies novel loci and associations with measures of skeletal frame size. *PLoS Genet.*, **5**, e1000445.
 23. Johansson, A., Marroni, F., Hayward, C., Franklin, C.S., Kirichenko, A.V., Jonasson, I., Hicks, A.A., Vitart, V., Isaacs, A., Axenovich, T. *et al.* (2009) Common variants in the JAZF1 gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.*, **18**, 373–380.
 24. Martin, J.E., Broen, J.C., Carmona, F.D., Teruel, M., Simeon, C.P., Vonk, M.C., van 't Slot, R., Rodriguez-Rodriguez, L., Vicente, E., Fonollosa, V. *et al.* (2012) Identification of CSK as a systemic sclerosis genetic risk factor through genome wide association study follow-up. *Hum. Mol. Genet.*, **21**, 2825–2835.
 25. Skhirtladze, C., Distler, O., Dees, C., Akhmetshina, A., Busch, N., Venalis, P., Zwerina, J., Spriewald, B., Pileckyte, M., Schett, G. *et al.* (2008) Src kinases in systemic sclerosis: central roles in fibroblast activation and in skin fibrosis. *Arthritis Rheum.*, **58**, 1475–1484.
 26. Fietta, P. and Delsante, G. (2011) Psychiatric and neuropsychological manifestations of systemic lupus erythematosus. *Acta Biomed.*, **82**, 97–114.
 27. Bertsias, G.K., Ioannidis, J.P., Aringer, M., Bollen, E., Bombardieri, S., Bruce, I.N., Cervera, R., Dalakas, M., Doria, A., Hanly, J.G. *et al.* (2010) EULAR Recommendations for the management of systemic lupus erythematosus with neuropsychiatric manifestations: report of a task force of the EULAR standing committee for clinical affairs. *Ann. Rheum. Dis.*, **69**, 2074–2082.
 28. Baubet, T., Ranque, B., Taieb, O., Berezne, A., Bricou, O., Mehallel, S., Moroni, C., Belin, C., Pagnoux, C., Moro, M.R. *et al.* (2011) Mood and anxiety disorders in systemic sclerosis patients. *Presse Med.*, **40**, e111–e119.
 29. Efthimiou, P. and Blanco, M. (2009) Pathogenesis of neuropsychiatric systemic lupus erythematosus and potential biomarkers. *Mod. Rheumatol.*, **19**, 457–468.
 30. Sofat, N., Malik, O. and Higgins, C.S. (2006) Neurological involvement in patients with rheumatic disease. *QJM*, **99**, 69–79.
 31. Milia, A.F., Conforti, M.L., Ibba-Manneschi, L. and Matucci-Cerinic, M. (2013) Differential expression of junctional adhesion molecules in different stages of systemic sclerosis. *Arthritis Rheum.*, **65**, 247–257.
 32. Yoshizaki, A., Yanaba, K., Iwata, Y., Komura, K., Ogawa, A., Akiyama, Y., Muroi, E., Hara, T., Ogawa, F., Takenaka, M. *et al.* (2010) Cell adhesion molecules regulate fibrotic process via Th1/Th2/Th17 cell balance in a bleomycin-induced scleroderma model. *J. Immunol.*, **185**, 2502–2515.
 33. Rabquer, B.J., Hou, Y., Del Galdo, F., Kenneth Haines, G. 3rd., Gerber, M.L., Jimenez, S.A., Seibold, J.R. and Koch, A.E. (2009) The proadhesive phenotype of systemic sclerosis skin promotes myeloid cell adhesion via ICAM-1 and VCAM-1. *Rheumatology (Oxford)*, **48**, 734–740.
 34. Raychaudhuri, S., Plenge, R.M., Rossin, E.J., Ng, A.C., Purcell, S.M., Sklar, P., Scolnick, E.M., Xavier, R.J., Altshuler, D. and Daly, M.J. (2009) Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.*, **5**, e1000534.
 35. 1980) 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.*, **23**, 581–590.
 36. LeRoy, E.C., Black, C., Fleischmajer, R., Jablonska, S., Krieg, T., Medsger, T.A. Jr., Rowell, N. and Wollheim, F. (1988) Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J. Rheumatol.*, **15**, 202–205.
 37. LeRoy, E.C. and Medsger, T.A. Jr. (2001) Criteria for the classification of early systemic sclerosis. *J. Rheumatol.*, **28**, 1573–1576.
 38. Hochberg, M.C. (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. *Arthritis Rheum.*, **40**, 1725.
 39. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–213.
 40. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 41. Zhang, B., Kirov, S. and Snoddy, J. (2005) Webgestalt: an integrated system for exploring gene sets in various biological contexts. *Nucleic Acids Res.*, **33**, W741–W748.
 42. Hochberg, Y. and Benjamini, Y. (1990) More powerful procedures for multiple significance testing. *Stat. Med.*, **9**, 811–818.
 43. Assassi, S., Mayes, M.D., Arnett, F.C., Gourh, P., Agarwal, S.K., McNearney, T.A., Chaussabel, D., Oommen, N., Fischbach, M., Shah, K.R. *et al.* (2010) Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum.*, **62**, 589–598.