CSIC ipbln
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS
INSTITUTO DE PARASITOLOGÍA Y BIOMEDICINA "LÓPEZ-NEYRA"

MINISTERIO
DE ECONOMÍA
Y COMPETITIVIDAD

Genetics of Systemic Sclerosis: a roadmap to a complex disease

Doctoral Thesis

Lara Mª Bossini Castillo

ugr | Universidad de Granada

Editor: Editorial de la Universidad de Granada
Autor: Lara Mª Bossini Castillo
D.L.: GR 1990-2014
ISBN: 978-84-9083-190-8

Tesis Doctoral

Genetics of Systemic Sclerosis: a roadmap to a complex disease

Memoria presentada por la licenciada en biología Lara María Bossini Castillo para optar al grado de Doctor Internacional por la Universidad de Granada.

Director: Javier Martín Ibáñez, Profesor de Investigación del CSIC.

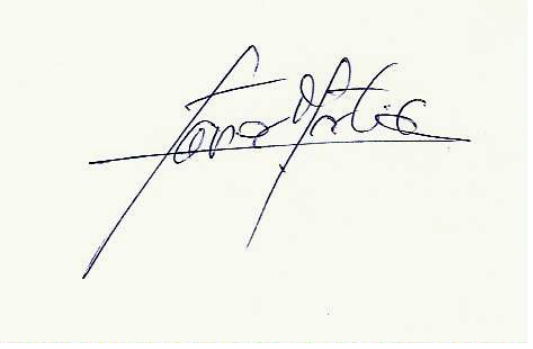
Instituto de Parasitología y Biomedicina López-Neyra, CSIC.
Granada, mayo de 2014.

La doctoranda Lara María Bossini Castillo y el director de la tesis Javier Martín Ibáñez garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

Así mismo, la doctoranda Lara María Bossini Castillo y el director de la tesis Javier Martín Ibáñez garantizamos, al firmar esta tesis doctoral, que las cinco publicaciones presentadas no se han utilizado en la defensa de ninguna otra tesis en España u otro país y que no serán utilizadas con tal propósito.

Granada, 21 mayo de 2014

Director de la Tesis



Fdo.: Javier Martín Ibáñez

Doctoranda

Fdo.: Lara María Bossini Castillo

INDEX

1. **Abbreviations** ……....1

2. **Summary** ……..……..9

3. Introduction

3.1. Systemic sclerosis, scleroderma ……...17

3.2. Environmental factors ……..34

3.3. Genetic component ……...39

4. **Objectives …..**……..58

5. Publications

5.1. GWAS follow-up and replication studies

5.1.1. *A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations ……..……..60*

5.1.2. *Confirmation of TNIP1 but not RHOB and PSORS1C1 as systemic sclerosis risk factors in a large independent replication study* ……68

5.2. Pulmonary involvement genetic markers

5.2.1. *A multicenter study confirms CD226 gene association with systemic sclerosis-related pulmonary fibrosis* ……..……...74

5.2.2. *KCNA5 gene is not confirmed as a systemic sclerosis-related pulmonary arterial hypertension genetic susceptibility factor* ……..81

5.3. Immunochip study

5.3.1. *Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis* ……...87

6. Discussion

6.1. Novel genetic findings ……103

6.2. An in-depth look into the HLA region ……....104

6.3. GWAS data mining and replication ……....114

6.4. Fine-mapping of novel and known associated regions ...117

6.5. Rare variant interrogation ……..……..…121

6.6. Genetic connections and biological relevance of the novel loci ……...……129

6.7. The causal variant quest …….……146

6.8. Pleiotropy with other phenotypes and ADs …....……150

6.9. Systemic sclerosis heritability ……154

7. **Conclusions** ...……...…….164

8. **Bibliography** ……..……168

ABBREVIATIONS

ABHD6: abhydrolase domain containing 6

ACA: anti-centromere antibodies

ACE: angiotensin-converting-enzyme

AD: autoimmune disease

ANA: antinuclear antibodies

ARA: anti-RNA polymerase III antibodies

AS: ankylosing spondylitis

ATA: anti-topoisomerase antibodies

ATD: autoimmune thyroid disease

ATG5: autophagy related 5

B3GALT4. UDP-Gal:betaGlcNAc beta 1,3-galactosyltransferase, polypeptide 4

BANK1: B-cell scaffold protein with ankyrin repeats 1

BLK: B lymphoid tyrosine kinase

CD226: CD226 molecule

CD247: CD247 molecule

CeD : celiac disease

ChIP: chromatin immunoprecipitation

CREST: calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia

CSK: c-src tyrosine kinase

dcSSc: diffuse cutaneous systemic sclerosis

DDX6: DEAD (Asp-Glu-Ala-Asp) box helicase 6

DEPAP: 1,2-di-oleyl ester

DNA: deoxyribonucleic acid

DNASE1L3: deoxyribonuclease I-like 3

EC: endothelial cells

ECM: extracellular matrix

eQTL: expression quantitative trait loci

ESYT1: extended synaptotagmin-like protein 1

FTO: fat mass and obesity associated

GRB10: growth factor receptor-bound protein 10

GWAS: genome wide association study

HCMV: human cytomegalovirus

HLA: human leukocyte antigen

HRCT: high resolution computed tomography

HSD17B8: hydroxysteroid (17-beta) dehydrogenase 8

HSD17B8: hydroxysteroid (17-beta) dehydrogenase 8

IFN: interferon

IKZF1: IKAROS family zinc finger 1

IL12A: interleukin 12A

IL12RB2: interleukin 12 receptor, beta 2

IL23R: interleukin 23 receptor

ILD: interstitial lung disease

IRF5: interferon regulatory factor 5

IRF7: interferon regulatory factor 7

IRF8: interferon regulatory factor 5

IRX3: iroquois homeobox 3

ITGA7: integrin, alpha 7

ITGAM: integrin, alpha M

JAZF1: JAZF zinc finger 1

JIA: juvenile idiopathic arthritis

KBAC: the kernel-based adaptive cluster test

KCNA5: potassium voltage-gated channel, shaker-related subfamily, member 5

KIAA0319L: KIAA0319-like

KIR: killer-cell immunoglobulin-like receptors

lcSSc: limited cutaneous systemic sclerosis

LD: linkage disequilibrium

LRT: the likelihood ratio test

MAF: minor allele frequency

NFAT: nuclear factor of activated T cells

NFKB1: nuclear factor of kappa light polypeptide gene enhancer in B-cells 1

NGS: next generation sequencing

NK: natural killer

PAH: pulmonary arterial hypertension

PBC: primary biliary cirrhosis

PCA: principal component analysis

POLR3A: polymerase (RNA) III (DNA directed) polypeptide A

PPARG: peroxisome proliferator-activated receptor gamma

PRDM2: PR domain containing 2, with ZNF domain

PS: psoriasis

PSC: primary sclerosing cholangitis

PSMB9: proteasome (prosome, macropain) subunit, beta type, 9

PSORS1C1: psoriasis susceptibility 1 candidate 1

PXK: PX domain containing serine/threonine kinase

RA: rheumatoid arthritis

RHOB: ras homolog family member B

RNA: ribonucleic acid

RPL41: ribosomal protein L41

RWAS: rare variant weighted aggregate statistic

SCHIP1: schwannomin interacting protein 1

SKAT: SNP-set (sequence) kernel association test

SLE: systemic lupus erythematosus

SNP: single nucleotide polymorphism

SOX5: SRY (sex determining region Y)-box 5

SRC: SSc-related renal crisis

SSc: systemic sclerosis

STAT4: signal transducer and activator of transcription 4

TAP: transporter ATP-binding cassette

TF: transcription factor

TGF-β: transforming growth factor β

TNF- α: tumor necrosis factor α

TNFAIP3: tumor necrosis factor α-induced protein 3

TNFSF4: tumor necrosis factor (ligand) superfamily, member 4

TNIP1: TNFAIP3 interacting protein 1

Treg: Regulatory T-cells

TREH: trehalase

VT: the variable threshold approach

WES: whole-exome sequencing

WGS: whole-genome sequencing

WSS: weighted-sum statistic test

ZNF512: zinc finger protein 512

α-SMA: α-smooth muscle actin

SUMMARY

Systemic sclerosis (SSc) is an autoimmune disorder that affects the connective tissue. SSc is characterized by fibrotic events in the skin and internal organs. Endothelial damage, increased extracellular matrix deposition and autoimmune imbalance are the main patho-physiological mechanisms that underlie SSc onset and progression. Moreover, SSc is classified as a complex disease since both genetic predisposition and environmental triggers contribute to the SSc development. SSc shows a wide range of phenotypical manifestations and heterogeneous clinical characteristics. Nevertheless, SSc patients can be classically classified according to the extent of the skin fibrosis (into the limited cutaneous and the diffuse cutaneous forms of the disease) and using their serological characteristics (based on the presence of two major auto-antibodies: anticentromere antibodies and anti-topoisomerase antibodies).

The present doctoral thesis is focused in the study of the genetic component of SSc. Significantly, the first Immunochip-based dense fine-mapping in SSc patients was included among the presented reports. Our Immunochip study led to the identification of

three novel SSc risk factors: *DNASE1L3*, *SCHIP1/ IL12A* and *ATG5*. Interestingly, the *DNASE1L3 locus* is the most significant association with SSc outside the HLA-region. The immune-directed Immunochip approach also allowed us to fine-map previously reported SSc-risk factors and identify *DDX6* as a suggestive SSc genetic marker. Furthermore, we performed the most comprehensive HLA-region analysis to date in SSc patients and identified a six amino acid and seven single nucleotide polymorphism (SNP) model that explained the observed association in the region.

In addition, this thesis comprises a genome-wide association study (GWAS) follow-up effort that identified for the first time the *IL12RB2 locus* as an SSc susceptibility factor. Our findings in the *IL12A* and *IL12RB2 loci*, together with previous evidences, point out the increasing relevance of the IL12 pathway in SSc.

Furthermore, we also include in this work compendium a replication study of the second GWAS carried out in SSc patients with European ancestry. Our data revealed that of the three proposed candidates, only *TNIP1* was a firm SSc risk factor, while

we discarded the *RHOB* and *PSORS1C1* variants as players in the SSc factor network.

All the above mentioned results increased the number of genome-wide level associated SSc genetic susceptibility factors to 15. Moreover, we have contributed to the better understanding of the SSc genetic component and the relation of SSc with other autoimmune diseases.

Pulmonary involvement accounts for a 60% of SSc-related deaths. Due to this great impact on the survival rates of SSc patients, we analyzed the association of two previously reported *loci*, *CD226* and *KCNA5*, with pulmonary fibrosis and pulmonary arterial hypertension in SSc patients, respectively. These studies showed that *CD226* is indeed associated with higher pulmonary fibrosis risk in SSc cases; but, we were not able to replicate previous finding in the *KCNA5 locus*.

Taking advantage of our large SSc cohorts, the previously knowledge about SSc genetic risk factors and our novel findings and using publically available databases and the latest *in silico* approaches and software, we have addressed the proposal of the

most likely causal variants for the known SSc genetic associations. Furthermore, we have performed a rare variant analysis in our Immunochip-genotyped cohorts, interrogated the connections and pathways that link the different known SSc-related *loci* and performed an estimation of SSc heritability.

RESUMEN

La esclerosis sistémica (SSc) es un desorden autoinmune que afecta al tejido conectivo. La SSc se caracteriza por evento fibróticos en la piel y los órganos internos. El daño endotelial, el depósito excesivo de matriz extracelular y el desequilibrio inmunológico son los principales mecanismos pato-fisiológicos que subyacen la aparición y la progresión de la SSc. Además, la SSc se clasifica como una enfermedad compleja pues tanto la predisposición genética como factores ambientales contribuyen al desarrollo de la SSc. La SSc muestra un amplio rango de manifestaciones fenotípicas y características clínicas heterogéneas. En cualquier caso, los pacientes de SSc se pueden clasificar de manera clásica según la extensión de la fibrosis en la piel (en la forma cutánea limitada y cutánea difusa de la enfermedad) y atendiendo a sus características serológicas (basándonos en la presencia de dos tipos mayoritarios de auto-anticuerpos: anticuerpos anti-centrómero y anticuerpos anti-topoisomerasa).

La presente tesis doctoral se centra en el estudio del componente genético de la SSc. Significativamente, el primer

estudio de mapeo fino usando la plataforma Immunochip que se ha realizado en pacientes de SSc se incluye entre los trabajos presentados. Nuestro estudio usando el Immunochip nos llevó a la identificación de tres nuevos factores de riesgo para la SSc: *DNASE1L3*, *SCHIP1*/*IL12A* y *ATG5*. Cabe resaltar que el *locus* del gen *DNASE1L3* constituye la asociación más significativa con SSc fuera de la región del HLA. La aproximación enfocada al sistema inmunológico que proporciona el Immunochip nos permitió realizar un mapeo fino de factores de riesgo a SSc ya conocidos e identificar al gen *DDX6* como un marcador de riesgo a SSc sugerente. También, llevamos a cabo el análisis más exhaustivo de la región del HLA en pacientes de SSc hasta la fecha e identificamos un modelo de seis aminoácidos y siete polimorfismos de un solo nucleótido (SNP) que explica todas las asociaciones observadas en la región.

Adicionalmente, esta tesis incluye un estudio de seguimiento de un estudio de asociación del genoma completo (GWAS) que identificó por primera vez al *locus IL12RB2* como un factor de susceptibilidad a SSc. Nuestros hallazgos en los *loci* de

IL12RB2 y de *IL12A*, junto con evidencias previas, resaltan la creciente relevancia de la ruta de la IL12 en la SSc.

Además, también hemos incluido en este compendio de trabajos un estudio de replicación del segundo GWAS realizado en pacientes de SSc de origen étnico europeo. Nuestros datos revelaron que el gen *TNIP1* es un factor de riesgo a SSc firme, mientras que descartaron a las variantes localizadas en *RHOB* y *PSORS1C1* como participantes en la red de factores de la SSc.

Todos los resultados mencionados incrementaron el número de factores genéticos de susceptibilidad a SSc hasta 15. Aún más, hemos contribuido al mejor conocimiento del componente genético de la SSc y la relación de ésta con otras enfermedades autoinmunes.

La afectación pulmonar es responsable de un 60% de las muertes debidas a la SSc. Debido a este importante impacto en la supervivencia de los pacientes de SSc, analizamos la asociación de dos *loci* previamente descritos, *CD226* y *KCNA5*, con fibrosis pulmonar e hipertensión arterial pulmonar, respectivamente. Estos estudios mostraron que *CD226* está realmente asociado con un

mayor riesgo de fibrosis pulmonar en los casos de SSc, pero no replicamos los hallazgos previos en el *locus* de *KCNA5*.

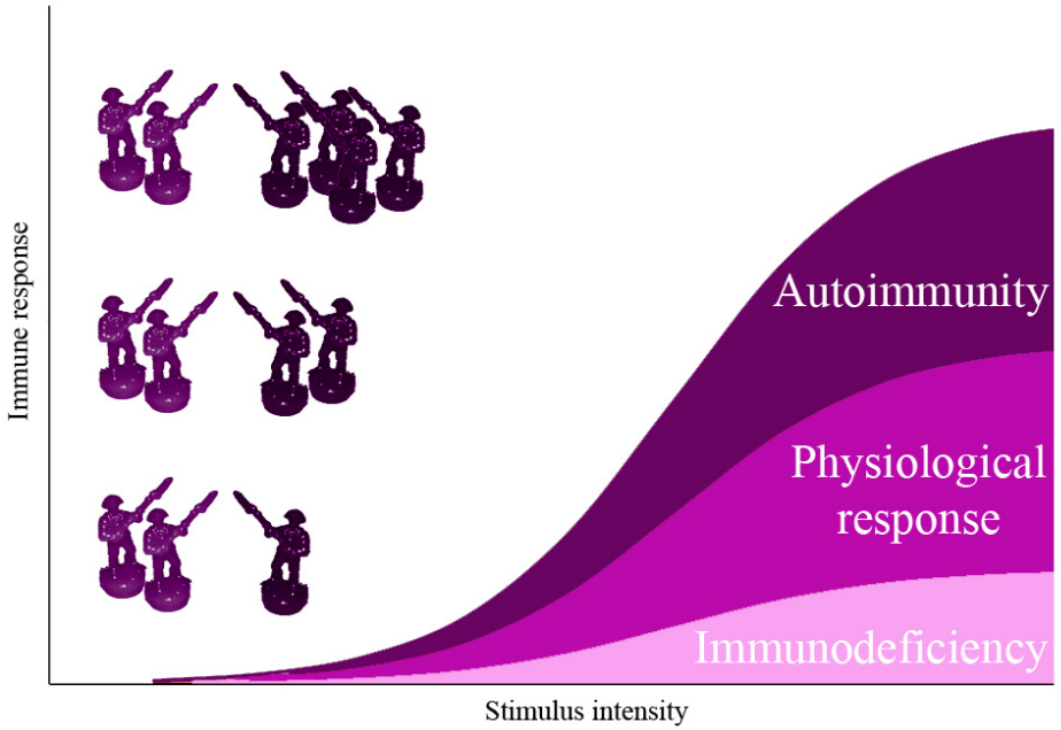
Aprovechando nuestras amplias cohortes de SSc, el conocimiento previo sobre los factores genéticos de riesgo a SSc y nuestros novedosos hallazgos y usando bases de datos públicas y las últimas estrategias y programas informáticos, abordamos la elaboración de una propuesta de identificación de las variantes causales de las asociaciones con SSc conocidas. Además, hemos realizado análisis de variantes raras, interrogado las conexiones y rutas que unen los *loci* relacionados con la SSc y hemos realizado una estimación de la heredabilidad de la SSc.

INTRODUCTION

Systemic sclerosis, scleroderma

Autoimmunity is a pathogenic process characterized by a loss of self-tolerance that leads to immune responses against self-antigens. The immune response is a continuum phenotype that is set relatively to the intensity of a stimulus (normally infection or tissue damage) [1]. In this range of gradual responses, autoimmune events are characterized by an excessive reaction against auto-antigens that are normally recognized and tolerated under physiological conditions (*Figure 1*). Autoimmune diseases (ADs) encompass a spectrum of disorders of unknown etiology. Unfortunately, most of them are chronic, incurable and debilitating. Although the prevalence of the individual ADs is variable among populations and often sex-biased, ADs collectively affect at least 6% of the Western population [2]. Therefore, these disorders cause both physical and psychological burden for the patients and important medical costs to society [3].

Figure 1. Immune response as a continuum phenotype determined by the stimulus intensity under both physiological and pathological conditions.



ADs, as non-infectious diseases, can be classified according to their etiology into: environmental diseases, which are caused by environmental factors and are not genetically transmitted to the offspring; monogenic or polygenic diseases, which arise from one or more mutations in the genome and will be inheritable if they affect the germ line; and complex diseases, which result of the combination of both genetic and environmental factors [4] (*Figure 2*). Complex diseases account for the vast majority of common human diseases [4]. However, the identification of the factors playing an important role in complex disorders is a difficult undertaking because of the complexity of the studied phenotypes and the confounding relations between factors [5].

Systemic sclerosis or scleroderma (SSc) is a complex chronic autoimmune disorder that affects the connective tissue [6]. Deregulation at the vascular, immune and extracellular matrix (ECM) deposition levels lead to the fibrotic phenomena that are characteristic of the disease (*Figure 3*) [7].

Several authors have postulated that the initial pathogenic process that takes places in the disease onset is a change in the function of endothelial cells (EC) and the structure of the

Figure 2. Classification of autoimmune diseases into Mendelian, environmental or complex according to their etiology.

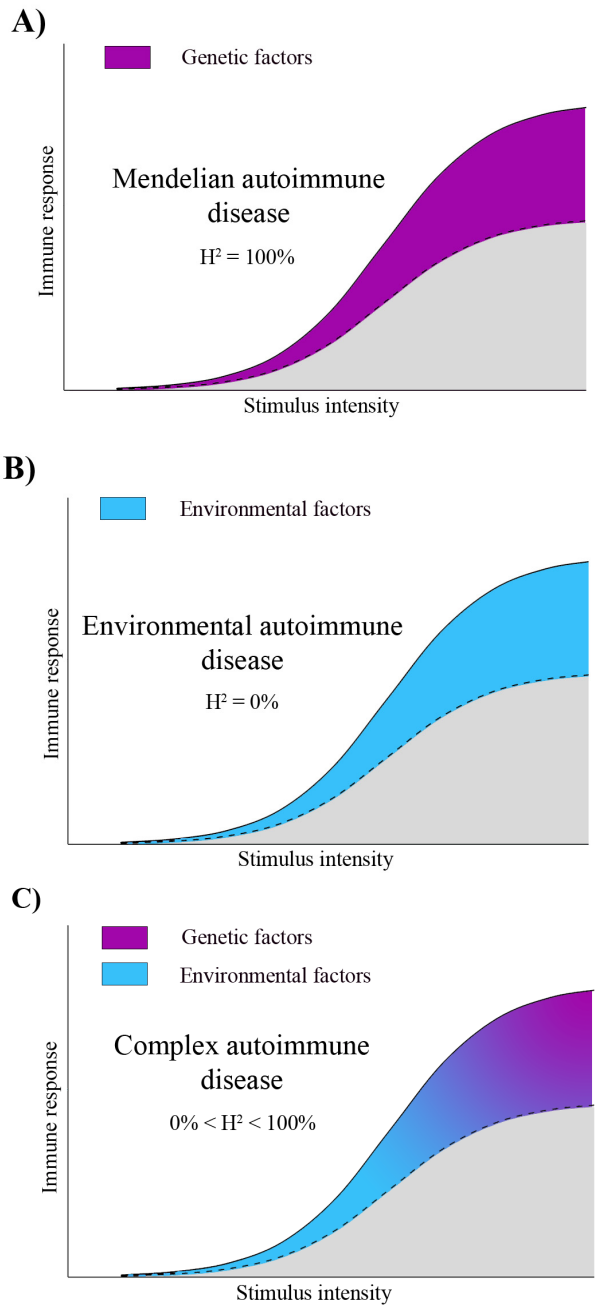
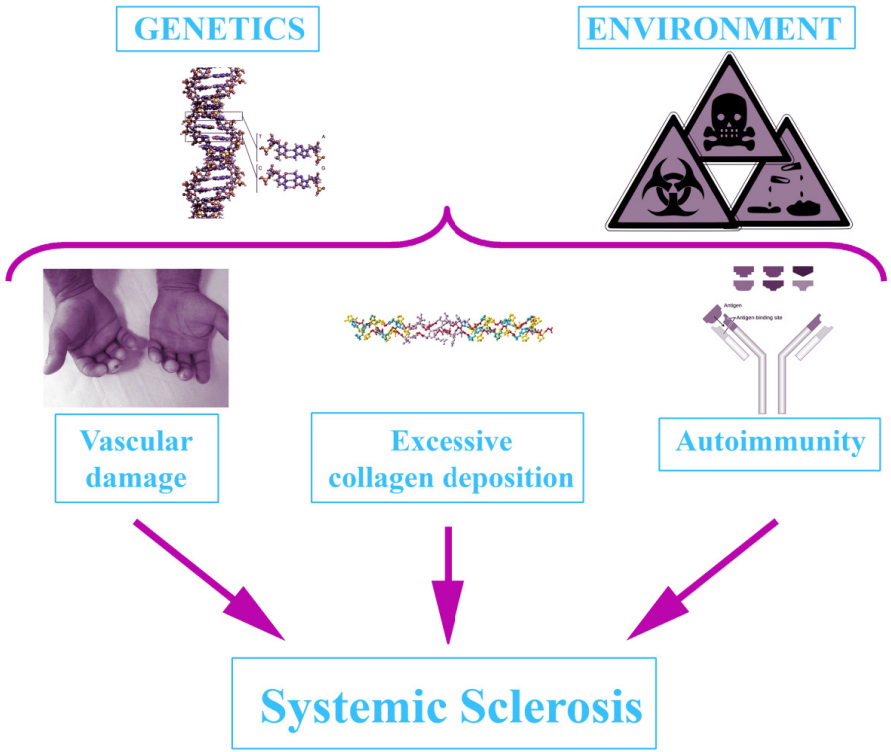


Figure 3. Systemic sclerosis etiological factors and most relevant characteristics.



microvasculature [6, 8-10]. Microvasculature modifications include malformed capillaries, Raynaud's phenomenon, and digital ulcers [11]. Later, vasculopathy can become life-threatening producing SSc-related renal crisis (SRC) and pulmonary arterial hypertension (PAH) [11]. Moreover, later stages of the disease are characterized by avascular areas, despite a general increase in many potent angiogenic factors [11]. These facts point towards a deeply impaired angiogenic response in patients with SSc [11]. EC damage is followed by an inflammatory phase characterized by inflammatory cell infiltration, progressive endothelial involvement, intimal fibrosis of small arteries and, finally dermal fibrosis arises [8-10]. Infiltration is more relevant in early SSc but seems to decrease when fibrotic events become more pronounced [9]. In the light of these findings, it can be hypothesized that the presence of big amounts inflammatory cells plays an important role in the induction of fibrosis [9]. Important alterations in several cell types from SSc patients have been reported, these abnormalities may be the basis that promotes the development of SSc pathogenic mechanisms. For example, SSc fibroblasts show a pronounced increase of procollagen mRNA expression, which allows the development of several autocrine cytokine loops in SSc fibroblasts

that could be present only partially in fibroblasts from healthy subjects [9]. Interestingly, SSc fibroblasts are known as myofibroblasts because they express α-smooth muscle actin (α-SMA) and exhibit a persistently activated phenotype with excessive production of collagen and other ECM proteins [12]. Moreover, myofibroblasts have an altered expression of and responses to cytokines, an increased proliferation, and a decreased apoptosis [12]. It has also been proposed that ECs may undergo a process called endothelial–mesenchymal transition, which allows them to gain myofibroblastic features [12]. Taking in account the relevance of the autoimmune events in SSc pathogenesis, it is not surprising that immune cells in SSc patients also show unique traits. T-cells in SSc patients are at the site of fibrosis, have an activated phenotype and show abnormal numbers and frequency in peripheral blood [13]. Furthermore, T-cells are a major component of the skin infiltrate (and also the lung infiltrate) of SSc patients during the inflammatory stage of the disease, with a high CD4+/CD8+ ratio [13]. These infiltrating T-cells in SSc skin lesions express activation markers and could be essential for the induction of the hyperactive SSc fibroblast phenotype [13]. Thus, although the recapitulation the disease onset is still controversial, all the three mentioned

processes: endothelial damage, increased extracellular matrix deposition and autoimmune imbalance interact to produce SSc.

SSc patients show a T-cell imbalance affecting different T-cell populations [9, 13]. For example, Th2 polarized T-cells have been found in SSc skin [9]. Therefore, soluble Th2 profibrotic mediators (such as IL-4, IL-6, IL-13) may interact with SSc altered fibroblasts and induce fibrosis [13]. In addition, NK T-cell subsets have reduced relative and absolute numbers and impaired function in SSc patients [9]. Recently, an important role for B-cells has also been elucidated. B-cells produce IL-6 that along with TGF-β, may promote matrix synthesis and reduce collagen degradation [14]. Additionally, SSc patients show chronic B-cell activation and present increased numbers of naive B-cells but reduced numbers of memory B-cells [14]. Moreover, memory B-cells seem to be abnormally activated in SSc [14]. Cytokines are immune system modulators; thus, as it could be expected, altered levels of different cytokines have been described in SSc patients, especially an imbalance between Th1 and Th2 cytokines and between Th17 and T_{reg} cytokines [15]. IL-4, IL-13, IL-5, IL-6, and IL-10, which are common in a Th2 response, lead to tissue fibrosis, while Th1 (IFN-

γ, TNF-α, IL-1, IL-2, and IL-12) and Th17 (IL-17, IL-21, IL-22) cytokines are proinflammatory in SSc patients [15]. Interestingly, plasma cytokine levels have been reported to be different among the ACA+, ATA+ and ARA+ SSc patients and disease duration and internal organ involvement seem to have an impact on the cytokine profiles [16]. Re-establishing the Th1/Th2 and Th17/T_{reg} balances might be a therapeutic target in SSc [15].

SSc is considered a rare disease, but disease prevalence rates range from 7/million to 700/ million showing noteworthy inter-study discrepancies [17]. Remarkable differences among populations have been identified, for example: black populations show an increased prevalence compared to whites and Asians and more severe outcomes are more frequent in Hispanics and Native Americans than in populations with European ancestry [18, 19]. A Native American tribe, the Choctaws, has the highest SSc prevalence described to date (660/million) [20]. In addition, an SSc North-to-South gradient has been described [17]. Thus, ethnical origin and ancestry factors greatly impact the onset and prognosis of the disease.

It is also worth mentioning that SSc, as other related ADs, has a marked sex-bias and is more frequent in females, especially middle-aged, than men (in a female-to-male mean ratio ranging from 3:1 to 14:1) [7, 21, 22]. As it can be observed in *Table 1*, the complete set of patients included in the different manuscripts

Table 1. Population characteristics considering all the individuals included in all the studies presented in this thesis.

Phenotype		**%**
TOTAL N	**7,962**	
Sex		
N=7,597	Female	86.52
	Male	13.48
Clinical subtype		
N=6,886	lcSSc	67.15
	dcSSc	32.85
Anticentromere antibodies		
N=7,244	Positive	43.94
	Negative	56.06
Antitopoisomerase antibodies		
N=7,220	Positive	31.69
	Negative	68.31
Interstitial Lung Disease		
N=4,363	Positive	32.78
	Negative	67.22
Pulmonary Arterial Hypertension		
N=3,691	Positive	12.38
	Negative	87.62

contained in the present doctoral thesis showed a 9:1 female-to-male ratio [23-27]. The origin of such a conspicuous female preponderance among ADs remains uncovered. However, different plausible causes have been proposed: sex-specific genetic traits (*i.e.* X-chromosome dosage effects, mutations in specific X-linked genes, skewed X-chromosome inactivation, loss of mosaicism, fetal microchimerism), and the levels and effects of sex hormones [21, 22, 28].

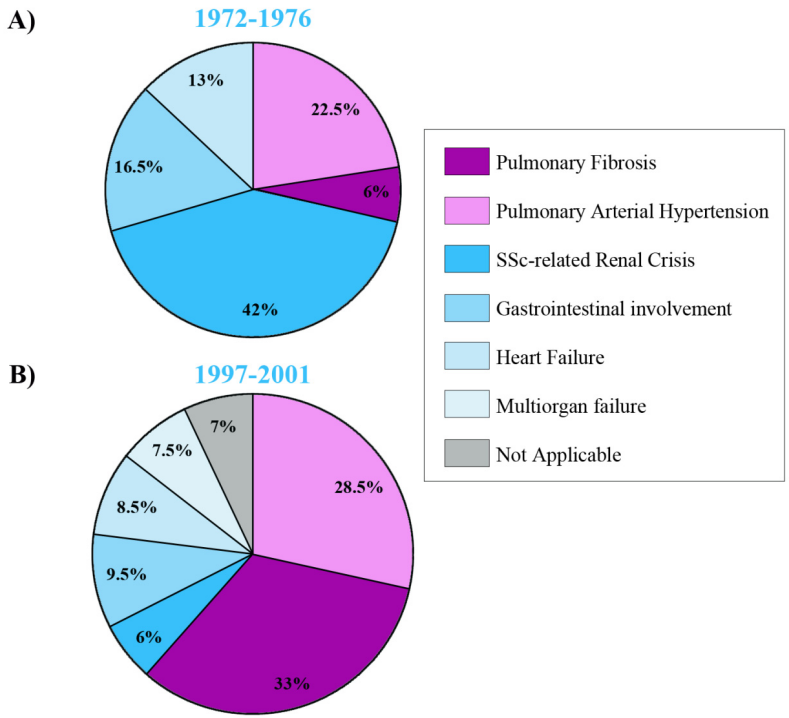
SSc is a phenotypically heterogeneous disorder and consequently diagnosis is complex and the definition of sensible and specific classification criteria is a major issue in clinical practice [29]. All the patients included in the present piece of work fulfill the 1980 preliminary classification criteria of the American Rheumatism Association (now the American College of Rheumatology) or the early systemic sclerosis criteria proposed by Leroy *et al*. or the presence of at least three out of five CREST features typical of SSc (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasia) [30, 31]. The additions of the two later additional criteria were used to increase the limited sensitivity of the classical SSc classification criteria[31].

SSc patients are clinically stratified into two main subtypes: patients with a limited cutaneous form of the disease (lcSSc) and those with diffuse cutaneous SSc (dcSSc) [32]. Both groups are defined based on the magnitude of skin tightening: limited to hands, face, feet, and forearms in lcSSc cases and extensive in dcSSc [32]. Moreover, the more frequent lcSSc is characterized by a slower disease progression with Raynaud's phenomenon for several years, dilated nailfold capillary loops [32]. Nevertheless, lcSSc patients have a high incidence of pulmonary hypertension (with or without interstitial lung disease), calcinosis and telangiectasias [32]. In the case of dcSSc, the disease course is more aggressive with an early onset of Raynaud's phenomenon, development tendon friction rubs, nailfold capillary dilation and capillary destruction [32]. Additionally, severe internal organ involvement appears early, *i.e*. interstitial lung disease, renal failure, gastrointestinal disease, and myocardial involvement [32].

Severe dysfunction and failure of visceral organs (kidneys, esophagus, heart, and lungs) is an essential determinant of patient prognosis [7]. Steen *et al*. reported that while in the early 1970s the main factor behind the death of SSc patients was SRC, in the late

1990s pulmonary involvement (comprising both PAH, and pulmonary fibrosis or interstitial lung disease, ILD) accounted for 60% of SSc-linked deaths (*Figure 4*) [33]. This decrease in the relevance of SRC became possible thanks to the use of angiotensin-converting-enzyme (ACE) inhibitors, which regulates blood pressure inhibiting angiotensin-induced vasoconstriction via the renin-angiotensin system) [33]. Nevertheless, due to the severe complications of pulmonary involvement, SSc is still a highly life-threatening disease with a 10-year survival of only 63% [33]. Moreover, the overall pooled standardized mortality ratio of patients with SSc reaches 3.53, which is markedly higher than in other ADs [34, 35]. Therefore, a great interest in deciphering the factors leading pulmonary dysfunction has arisen in the medical and scientific community. Since we shared these concerns with our clinical collaborators, two out of five studies included in this doctoral thesis were focused in SSc-related pulmonary involvement [24, 26]. The diagnosis of pulmonary involvement in SSc is a changing area with constant definition changes and fluctuating prevalence rates. Using current gold standards and following previous reports and clinical counseling, we considered that the patients in the analyzed cohorts had confirmed PAH using right

Figure 4. Main systemic sclerosis-related death causes before the use of angiotensin-converting-enzyme inhibitors (A) and after (B). Data extracted from Steen *et al*.2007.



heart catheterization (mean resting pulmonary artery pressure > 25 mmHg and a pulmonary capillary wedge pressure ≤ 15 mmHg) [36-38]. Furthermore, ILD was determined by the visualization of interstitial abnormalities in High Resolution Computed Tomography (HRCT) [36, 39, 40]. The frequencies of these clinical complications in the complete set of SSc patients included in our studies are reported in *Table 1*.

The presence of antinuclear antibodies is also a major SSc hallmark derived from its characteristic autoimmune events. From a serological point of view SSc patients can be encompassed in three highly SSc-specific main subgroups: anticentromere antibody positive (ACA+), anticentromere antibody positive (ATA+) and antiRNA-polimerase III antibody positive (ARA+) SSc patients [7, 41, 42]. Despite important inter-population differences, the approximate ACA+, ATA+ and ARA+ frequencies are 16-39%, 9-39% and 4-25% respectively [7, 41, 42]. The ACA+, ATA+ and ARA+ groups comprise around half of the patients and the presence of each previously mentioned auto-antibody is generally exclusive from the others [42]. Nevertheless, a small fraction of patients are positive for both ACA+ and ATA+ [42]. The remaining patients

either have rarer auto-antibodies or are negative for the presence of antinuclear autoantibodies (3-11%) [42]. Usually, the antibody subset of a patient is maintained during time and it is widely accepted that autoantibodies are better biomarkers for organ involvement and survival than clinical subtypes (lcSSc or dcSSc) [41, 42]. However, serological and clinical subsets are clearly overlapping. Indeed, a 70-80% of lcSSc patients are ACA+, although a small proportion of ACA+ cases (5–7%) develop dcSSc [42]. Moreover, around 60% of ATA+ and the majority of ARA+ (67–93%) SSc patients are included in the dcSSc subgroup [42]. Regarding the relation between organ involvement and autoantibodies, the ACA+ SSc patients show a higher incidence of PAH without ILD than the ATA+ subgroup [42]. In addition, primary biliary cirrhosis (PBC) is more frequent in the ACA+ cases [42]. On the contrary, the presence of ACA seems to be protective for digital ulcers, ILD and SRC [42]. It is worth mentioning that, ILD is greatly linked with ATA+, which causes the highest mortality rates among all the autoantibody-based subsets [42]. Furthermore, ATA+ patients are more prone to developing digital ulcers and more aggressive skin involvement and disease activity [42]. Additionally, SRC appears in up to 43% of ARA+ SSc cases

[42]. In the pre-ACE inhibitors era, SRC caused a dramatic mortality of these patients [42]. However, as previously commented, nowadays the survival of this subgroup has greatly improved [42].

Clinical characteristics and prognosis are greatly variable in SSc patients. Therefore, grouping patients based on disease subsets or autoantibody profiles is a useful tool both for clinical practice and genetic studies. Of note, all the individuals included in our cohorts were of European ancestry, self-reported and checked using Principal Component Analysis (PCA) when possible (this is, if high-throughput genetic data were available). Hence, as illustrated in *Table 1*, clinical subtype, pulmonary involvement and autoantibody profile based stratification were applied when necessary in the presented studies [23-27]. However, the subgroups in this artificial classification should not be considered as isolated entities but as part of quantitative traits, continuous phenotypes.

As it was stated previously, both environmental factors and genetic susceptibility interact in the disease onset and maintenance. Therefore, the current knowledge on both aspects will be addressed in this introduction paragraph.

ENVIRONMENTAL FACTORS

Occupational exposition to several chemical compounds has been found to increase SSc susceptibility:

1) *Silica:* crystalline silica or quartz is released typically in mining but also in other construction works. Silica exposure has been associated with a 3.2 higher risk of suffering SSc [3]. In spite of its long known association with SSc no clear mechanism for silica-related immune dysfunction has been elucidated yet. It has been suggested that it may interact with alveolar macrophages, activate T and B lymphocytes, trigger autoimmunity-related apoptosis and increase fibroblast proliferation [43]. It should be noted that Rocha-Parise *et al*. reported that silica-exposed individuals present increased serum soluble IL-2 receptor levels, decreased production of IL-2 and increased levels of pro-inflammatory mediators (IFN-γ, IL-1α, TNF-α, IL-6) and anti-inflammatory cytokines (IL-10 and TGF-β) [44]. Moreover, these individuals were more likely to show enhanced lymphoproliferation [44]. Furthermore, lymphocytes stimulated with silica express high levels of Fas receptor (CD95) and undergo apoptosis [45]. Then, altered proteins are released and autoimmune events may be

initiated [45]. Thus, silica might act as both a haptene (modifying proteins and triggering the immune response) and a direct immune system activator [45].

2) *Organic solvents and solvent-like chemicals*: chlorinated solvents, aromatic solvents, white spirit, trichloroethylene, ketones and other molecules sharing structural characteristics (such as vinyl chloride, epoxy resins, perchloroethylene, or mixed solvents) have been reported to cause SSc-like syndrome. The estimated SSc-risk due to solvent exposure has been estimated in 2.4 and it is greater in men than in women and increases in highly exposed individuals [3, 43]. It has been proposed that organic solvents may link with nucleic acids and proteins and disrupt the immune response [43]. Solvents initiate cellular and humoral autoimmunity and stimulate fibrogenic responses and they can also enhance the immunogenicity of intracellular molecules [45].

3) Welding fumes: are a complex mixture of metallic oxides, silicates and fluorides, which are formed when a metal is heated above its boiling point and then the vapours condense. Welding fumes have also been linked with increased susceptibility to SSc (with an OR reaching 3.54) [3, 43, 45].

Finally, we would like to mention the toxic oil syndrome (TOS) as an example of SSc-like affection caused by chemicals. In the 1980s, 356 people died and thousands were diagnosed with TOS in Spain [46]. This disease was caused by consumption of rapeseed oil that was contaminated with 1,2-di-oleyl ester (DEPAP) and oleic anilide [46]. The main effect of this chemicals was a non-necrotizing vasculitis of several organs [46]. It has been postulated that the intermediates of the metabolism of the parent compounds could act as haptenes and activate autoreactive T-cells, disrupt signal transduction, or induce apoptosis and necrosis [46]. Curiously, disease intensity was linked with certain HLA-DR2 alleles and polymorphisms in metabolism and immune response genes [46].

On the contrary, no association was found between SSc and other non-occupational chemical factors such as smoking, the use of drugs (anorexigens, pentazocine, bromocriptine, L-tryptophan), implants (prosthesis, silicone implants, and contact lenses) and dyeing hair [43].

Infectious agents were suggested as possible causes that break self-tolerance and may trigger autoimmunity via molecular

mimicry, endothelial cell damage, super-antigens, and microchimerism processes [47, 48]. In the case of SSc, multiple bacterial and viral infections have been studied:

1) Parvovirus B19: this virus is related with other ADs. Moreover, in SSc B19 has been found in bone marrow biopsies from patients, it has been shown that SSc fibroblast can be persistently infected by B19 and EC damage in SSc may reflect a combination of direct viral toxicity and humoral immunity towards this virus [47, 48].

2) The herpesvirus family: Human cytomegalovirus (HCMV) infects both ECs and monocytes/macrophages and produces immune dysregulation (with the development of ATA, for example) and promotes fibrogenic cytokines [47, 48]. Latent HCMV in SSc patients may represent an allotypic stimulus to T-cells and increase allograft rejection (microchimerism) [48]. Of note, Epstein-Barr virus (EBV) is a common risk factor for many ADs, which causes a cross reaction between its molecules and self-antigens [48]. Remarkably, SSc IgG recognised the HCMV late protein UL94 and the endothelial cell surface integrin–NAG-2 protein complex, which induced endothelial cell apoptosis [49].

3) Endogenous retroviruses: can affect the immune response due to their encoded proteins or their insertion in the genome. Shockingly, a common protein sequence between DNA topoisomerase I and p30gag retroviral proteins has been found (another example of molecular mimicry) [48].

4) *Helicobacter pylori*: may be involved in endothelial damage and vascular changes [48].

Pregnancy, if considered as an environmental factor, has also been reported to induce higher SSc rates [7]. Particularly, SSc female patients have higher levels of fetal microchimerism, this is fetal DNA is found in high proportions in maternal blood [50]. Additionally, HLA-DR alleles which were indistinguishable in mother and child correlated with an increased risk of SSc onset in the mother [50].

Supplementary physical or life-style factors such as ionizing radiation, dietary habits and food contaminants have been reported to contribute to the development of other ADs [3]. However, there are no confirmed data on SSc that can prove these hypotheses.

Unfortunately, SSc manifestations are very heterogeneous and often it is hard to identify and quantify the environmental exposures. As a consequence, the knowledge about the link between environmental triggers and SSc pathogenesis is still scarce. Thus, environmental factors and gene-environment interactions could not be analyzed in this work.

GENETIC COMPONENT

SSc is a complex disease with a well-established genetic susceptibility background. Indeed, high autoantibody concordance has been described between twins and in SSc multi-case families [51, 52]. As it could be expected, familial history of SSc is the major risk factor, showing relative risks 13-fold higher than in the general population in first degree relatives and 15-fold higher in siblings [53]. As mentioned above, ancestry has also a relevant role in SSc susceptibility [17-20].

Nevertheless, it should be noted that the number of disease-associated *loci* that are involved in a complex disease is usually unknown [54, 55]. Moreover, despite that familial aggregation is common, the inheritance of complex disorders does

not fit Mendelian patterns [4, 56]. As a consequence, familial studies in complex diseases have not been as successful as they are in Mendelian traits. Additionally, genetic risk factors individually have normally a very modest impact on complex disease susceptibility, *i.e*. penetrance of individual variants is low and they cannot be identified within pedigrees [1, 57].

Case-control studies are observational epidemiological studies that comprise non-related individuals, and in which cases (affected individuals or individuals with a certain trait) and controls (healthy individuals or individuals without the trait) are compared on the basis of an attribute (the presence of a certain allele in a particular polymorphic position). Genetic case-control studies can be designed based on previous knowledge (candidate gene studies) or in a hypothesis-free fashion (genome-wide association studies, GWASs). Candidate gene studies begin with the selection of the analyzed locus and polymorphisms based on an *a priori* functional or positional possible implication in the disease. The functional role of the selected *loci* for candidate gene studies is normally based on the possible or known biological implication of the encoded protein in the disease. Moreover, those variants that are located in coding or

regulatory regions are often preferentially studied. In the case of the positional clues to select a *locus*, they are usually due to the association of a region through pedigree analyses and then the different *loci* that map within that region or the most likely (considering the previously described functional evidence) are tested. Thus, candidate gene studies depend on previous evidences and are directed by the researchers. Nevertheless, these studies are a powerful approach to analyze the contribution of specific *loci*, especially when large cohorts are analyzed.

The initial genetic interrogation of ADs was originally focused on candidate genes and comprised small cohorts; thus, this pioner studies resulted in the identification of few firm genetic susceptibility factors outside the HLA [58]. To date, the advances in genotyping technologies and the gathering of wide patient cohorts have made it possible to genotype large numbers of common variants in large collections [58]. Thus, recent well-powered candidate gene studies and especially Genome-Wide Association Studies (GWAS) have resulted in the identification of multiple common genetic polymorphisms related to the variety of ADs [57, 59].

GWASs test markers for association all through the genome. This strategy takes advantage of the known linkage disequilibrium patterns in the genome to reduce the number of variants to be included in the analysis. Moreover, it has been possible thanks to the availability of high-throughput genotyping platforms. These studies deal with restrictive multiple testing correction thresholds due to the large number of comparisons. The standard significance threshold for GWASs (also known as genome-wide significance level) is established at *p*-value < $5*10^{-8}$, which corresponds to a Bonferroni correction assuming all tests as independent and based on an estimation of the number of independent SNPs in the genome. Moreover, large cohorts are necessary to reach a favorable statistical power to identify significant association signals [60]. It is worth mentioning that GWASs tend to show inflated effects of the associated variants, this effect is known as the 'winner's curse'. As a consequence, the replication of the *loci* identified in GWAS in independent populations is mandatory. Our group was involved in the first GWAS in SSc in white populations [61]. In this study *CD247* was reported as a novel SSc risk factor, and the previously reported associations in the HLA, *STAT4* and *IRF5 loci* were confirmed at

the GWAS level [61]. Our novel findings in *CD247* were later independently replicated by Dieudé *et al*. [62]. A second SSc GWAS that comprised an independent French population was published during the period of this thesis [63]. Therefore, we used our large SSc cohort to perform an independent replication of the reported results in Allanore *et al.* [63]. The results in our study, which was consequently included in the present thesis, confirmed the previously described signals in the *TNIP1 locus* with SSc but discarded *RHOB* and *PSORS1C1* as SSc genetic risk factors [25].

In the case of SSc, as in most of ADs, the HLA region is the major genetic association described to date [64, 65]. A number of classical alleles have showed an increased frequency in SSc patients, being the most important of them the *HLA-DRB1*11:04/HLA-DQA1*0501/HLA-DQB1*0301* haplotype [66-71]. Besides, firm SSc protective HLA haplotypes have been described as the *HLA-DRB1*0701/HLA-DQA1*0201/HLA-DQB1*0202* [66]. Additionally, different alleles have been reported to increase SSc susceptibility (for example, *HLA-DRB1*01* and *HLA-DQB1*0501* with the ACA+ subgroup and *HLA-DPB1*1301* with the ATA+) but these markers vary with subtypes

Table 2. Non-HLA *loci* that were identified prior to the publication of the first GWAS in white systemic sclerosis (SSc) patients and that have reached genome-wide level (*p*-value $<5x10^{-8}$) associations with SSc.

Locus	Locus Coordinates	Locus Function	Associated Genetic polymorphism/s	Risk/ Protection	Reference
STAT4	2:191,029,576-191,172,671	T cell maduration and signalling, Th1 and Th17 differentiation	rs7574865, rs11889341, rs8179673, rs10181656, rs6752770, rs3821236	Risk	Rueda 2009, Tsuchiya 2009, Gourh 2009, Dieude 2009; Radstake 2010, Allanore 2011, Mayes 2014
IRF5	7:128,937,940-128,950,035	Virus-mediated activation of interferon, modulation of cell growth, differentiation, apoptosis and immune system activity	rs2004640, rs10954213, rs2280714, rs10488631, rs12537284, rs4728142, rs3757385	Risk	Dieude 2009,2010 Ito 2009 Radstake 2010, Allanore 2011, Mayes 2014

Table 3. Non-HLA *loci* that were identified after the publication of the first GWAS in white systemic sclerosis (SSc) patients was published and that have reached genome-wide level (p-value $<5x10^{-8}$) associations with SSc.

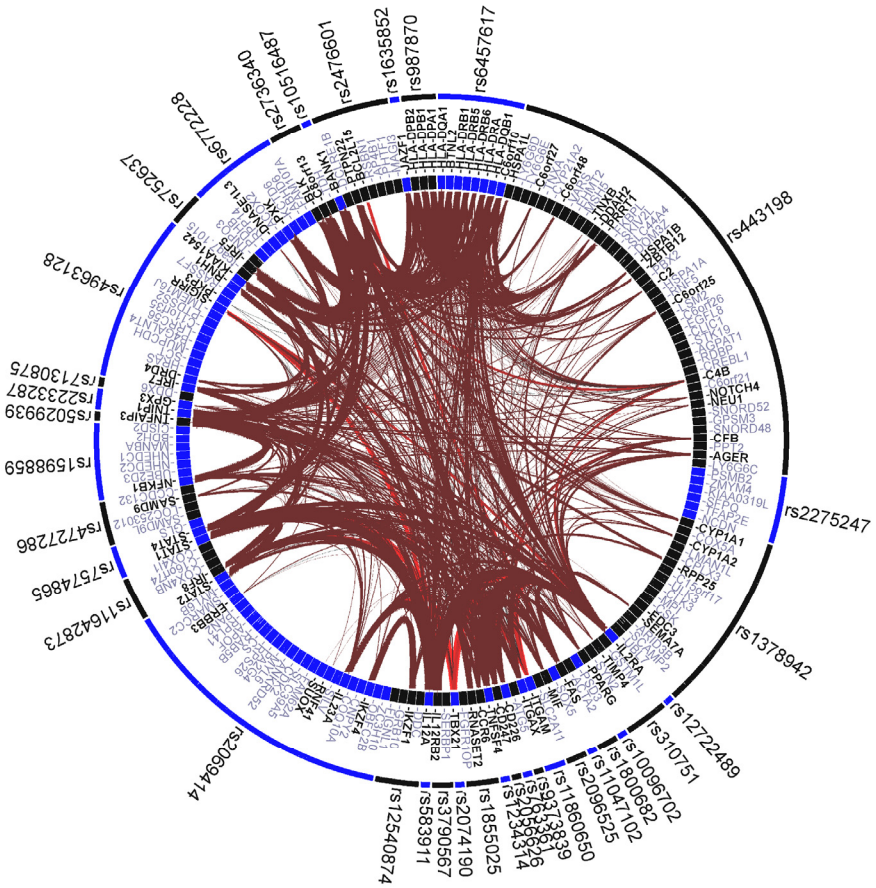
Locus	Locus Coordinates	Locus Function	Associated genetic polymorphism/s	Risk /Protection	Reference
CD247	1:167,430,640-167,518,610	T-cell receptor-CD3complex	rs2056626	Protection	Radstake 2010, Dieude 2010 Allanore 2011,
TNIP1	5:151,029,943-151,087,660	Regulation of NFKB	rs2233287,rs4958881, rs3792783	Risk	Allanore 2011 Bossini-Castillo 2013
CSK	15:74,782,084-74,803,198	Collagen deposition regulation	rs1378942	Risk	Martin 2012
IRF8	16:85,899,168-85,922,606	Lineage commitment and myeloid cell maturation	rs11642873	Protection	Gorlova 2011
KIAA0319L	1:35,433,490-35,557,670	Unknown	rs2275247	Risk	Martin 2013
JAZF1	7:27,830,573-28,180,818	Transcriptional repressor	rs1635852	Risk	Martin 2013
NOTCH4	6:32,194,843-32,224,067	Vascular, renal and hepatic development	rs443198,rs9296015	Protection	Gorlova 2011
TNFAIP3	6:137,867,188-137,883,314	Inhibits NF-kappa B activation and TNF-mediated apoptosis	rs5029939	Risk	Dieude 2010
ITGAM	16:31,259,967-31,332,892	Adherence of neutrophils and monocytes and phagocytosis	rs1143679	Risk	Carmona 2011 Coustet 2011

and ethnical origin [66, 71-76]. Remarkably, it has been recently suggested that the HLA association may be confined to the autoantibody subsets (ACA+ and ATA+) [72]. Currently, a number of non-HLA *loci* have been firmly associated with SSc [65, 77, 78]. *Table 2* illustrates the non-HLA genetic l*oci* known to affect SSc or any of its major subphenotypes susceptibility to date, excluding those that were identified in the articles included in the present thesis. All SNP associations reaching the GWAS significance level were included in *Tables 2-3*. We consider these genome-wide level associated *loci* as SSc confirmed genetic susceptibility factors. However, in order to maximize the scope of the *in silico* analyses in this thesis we also included in the analyzed gene sets those *loci* that have reached second tier level associations (*p*-value < $5x10^{-5}$) or that have been replicated in different populations (*Figures 5-6*). The polymorphism/*locus* correspondence was established as implemented in Gene Relationships Across Implicated Loci, GRAIL software, [79] using the release 18 of the Human Genome and the PubMed text as of 2012 (*Figure 5*). According to GRAIL predictions the association of the rs12540874 variant, attributed to the *GRB10 locus* in previous reports [72] will be referred as likely corresponding to the *IKZF1 locus* (*Figure 5*). Following the same

criteria, we will simplify the nomenclature of the suggestive *TREH/DDX6 locus* in Mayes *et al*. as *DDX6* (*Figure 5*) [23]. However, due to the confusing GRAIL results for the previously reported SSc-associated variants rs1378942 and rs11171747 we maintained their initial nomenclature as *CSK* and *RPL41*/*ESYT1* (*Figure 5*). We have done the same in the case of rs443198 and rs9296015 that will be considered variants related to the *NOTCH4* gene (*Figure 5*). A pathway analysis based on the previously described *loci* was also conducted by the means of the Panther Classification System [80, 81]. The results of this pathway analysis are shown in *Figure 6*. The known SSc genetic risk factors play a role in both adaptative and innate immune response and several of them are involved in cytokine pathways (Figure 6). Furthermore, the function of some of them is related to angiogenic processes, apoptosis and cellular commitment (*Figure 6*). The involvement of a great number of different compartments and pathways shows the tangled network that leads to SSc pathogenesis.

However, despite the great advance allowed by the GWASs carried out in SSc and other complex diseases, several concerns about GWAS results have been pointed out by various authors, such

Figure 5. Gene Relationships Across Implicated Loci (GRAIL) results for the *loci* comprising the firmest systemic sclerosis genetic associations (release 18 of the human genome and the PubMed text as of 2012 was used).



as (reviewed in Stranger *et al.* and Visscher *et al*. [57, 82]):

1) Multiple small effects may contribute significantly to heritability but might have remained undetected because of the needed sample size to identify them.

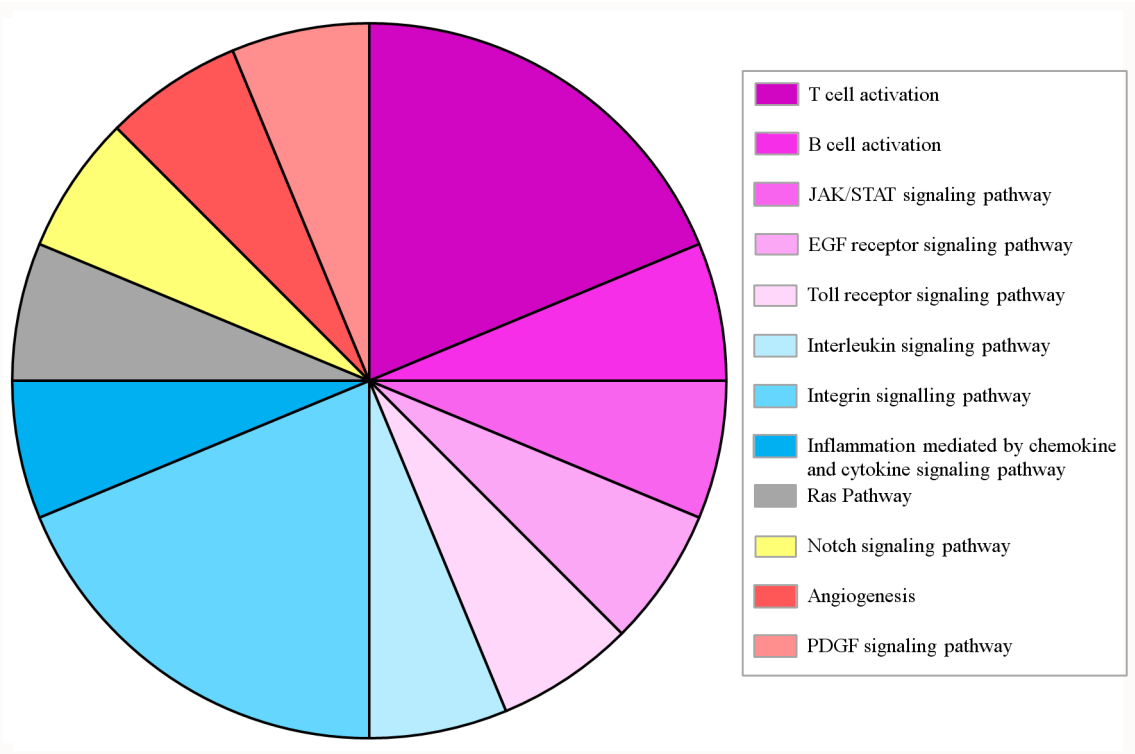
2) Incomplete LD or interpopulation LD differences between the associated variants and the causal variant/s that may cause an underestimation of the magnitude of the real effect of the susceptibility *locus*.

3) Low-frequency polymorphisms (MAF=0.01-0.05) and rare variants (MAF < 0.01) may play an important role in the unknown genetic variability. It should be taken into account that the common SNPs that are included in GWASs normally have a minor allele frequency (MAF) higher than 5%. Therefore, the effect of these low-frequency variants may have been overlooked.

4) Most of causal variants remain unknown and biological functions are unclear.

Although the first GWAS limitation could be overtaken by increasing the sample size and using meta-analysis techniques, this is not always possible. Consequently, it is now evident that GWAS

Figure 6. Pathway analysis as implemented in the Panther Classification System for the firmest systemic sclerosis genetic associations excluding the *loci* identified in the present thesis.



findings need to be complemented with additional approaches. Firstly, it is essential to data mine the GWAS genotype information. The GWAS significance thresholds are very strict to prevent the identification of false positive variants [83]. However, some suggestive second tier associations might be true but need to be explored in additional cohorts [83, 84]. Following this line, three novel SSc susceptibility *loci* have been identified via GWAS follow-up studies: *IL12RB2* (which is included among the presented studies), *CSK*, *PPARG* [27, 85, 86].

Except in a few cases in which the GWAS identified variant is assumed to be the causal variant for the association, the lead variant identified through GWAS in a *locus* is presumed to be in LD with the causal, functional variant [82, 87]. Thus, fine-mapping of an associated *locus*, either including all the known variants or using a tag-SNP strategy might help defining statistically independent association signals that could reveal the existence of a single associated haplotype or could suggest multiple causal variants [82]. These fine-mapping strategies (and sequencing approaches) will help to overcome the second and the third previously described GWAS handicaps. Then, it is possible to redefine the linked variants and prioritize the putatively functional polymorphisms

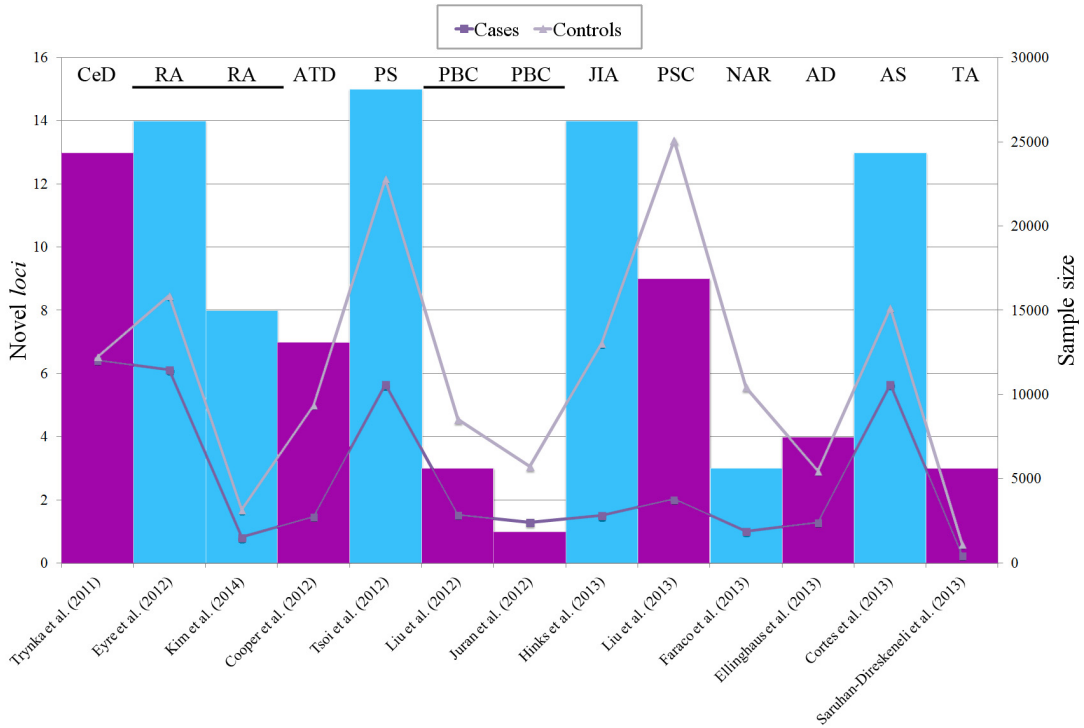
[82]. Finally, functional analysis of the variants will be needed to discern the real effect of the associated *loci* and solve the fourth of the mentioned GWAS restrains.

Especially directed towards ADs and taking advantage of the well-known autoimmunity shared genetic component, a novel fine mapping approach, known as the Immunochip (HumanImmuno Beadchip by Illumina), has reported very successful results recently. This platform is a custom SNP genotyping array that includes 196,524 variants (SNPs and small insertions-deletions) that comprise a fine-mapping of 186 known autoimmunity susceptibility *loci* [88]. For the custom array design all the described variants for white (European ancestry) populations were considered (SNPdb, 1000 Genomes Project February 2010 release, and additional sequencing projects) [88]. Remarkably, rare variants and those variants that are predicted to have a putative functional role were also included in the SNP panel, together with a dense set of HLA, KIR (Killer-cell immunoglobulin-like receptors) and ancestry markers [88]. Several Immunochip-based dense mapping studies have been published lately a variety of ADs: celiac disease (CeD), rheumatoid arthritis (RA), autoimmune thyroid disease (ATD),

psoriasis (PS), PBC, juvenile idiopathic arthritis (JIA), primary sclerosing cholangitis (PSC), narcolepsy, ankylosing spondylitis (AS), atopic dermatitis, and Takayasu arteritis [89-101]. Each one of these reports has led to the identification of variable number of common AD susceptibility factors probably, but not only, depending on their sample size, as it is graphically illustrated in *Figure 7*. These findings have greatly contributed to the better knowledge of the shared genetic background of autoimmunity [102]. It is widely accepted that SSc and systemic lupus erythematosus (SLE), and other related ADs as rheumatoid arthritis (RA) have an especially overlapping genetic background [65, 103, 104]. Therefore, we followed the Immunochip strategy in a large SSc case and control set [23]. The results of this study, recently published in the *American Journal of Human Genetics*, are part of the present doctoral thesis [23].

As stated above, the HLA region is the most significantly associated *locus* with ADs [64]. However, little is known about the peptides that are presented by the associated HLA alleles or the functional causes behind this outstanding peak of association with

Figure 7. Number of novel *loci* identified in all the Immunochip studies carried out to date. The number of cases and samples included in each study is shown in the secondary Y axis.



ADs. Recently, a new computing procedure has achieved an accurate imputation of the classical HLA alleles at four digit resolution and at the amino acid level resolution with high imputation accuracy [59]. This method starts from a large reference panel composed of two European ancestry reference panels (HapMap-CEPH pedigrees and the Type 1 Diabetes Genetics Consortium) and has successfully identified a five-amino acid model that explains the observed association in the HLA region with RA [59, 105-107]. Moreover, it has been recently extended to non-European populations [108]. Due to its dense coverage of the HLA region, we were able to follow perform this analysis in our SSc Immunochip study and identified a six polymorphic amino acid position and seven SNP model that explained the observed HLA associations [23].

Finally, bioinformatics and publically available functional features databases currently offer a wide variety of information that can be used to prioritize variants for experimental studies to test putative functional variants for a real effect [82]. The most straightforward analysis is to interpret coding or transcribed variants with tools such as SIFT and PolyPhen-2 [109, 110]. However, most

of the associated GWAS polymorphisms map in non-transcribed DNA and it is likely that their underlying mechanism is regulatory [87]. Expression quantitative trait *loci* (eQTL) are polymorphisms that influence gene expression, either closely located genes (cis-eQTLs) or distant genes (trans-eQTLs). Intriguingly, GWAS associated variants have been reported to be significantly enriched for eQTLs [111-115].

The most ambitious project for integration of functional information is without doubt the Encyclopedia of DNA Elements (ENCODE) which aims to classify all functional elements in the human genome, including identifying and quantifying RNA species in whole cells and in sub-cellular compartments, mapping protein-coding regions, delineating chromatin and DNA accessibility and structure with nucleases and chemical probes, mapping of histone modifications and transcription factor (TF) binding sites by chromatin immunoprecipitation (ChIP), and measurement of DNA methylation and multiple smaller-scale efforts [116].

Taking all the above findings into account, together with the public access available for most of them, we consider that we are in

the position to integrate these sources of information with our results and we will address this point in the discussion section.

OBJECTIVES

The main aim of this doctoral thesis was to further investigate the genetic component of systemic sclerosis (SSc), we specifically aimed:

1. To identify novel loci or validate already proposed genetic markers associated with SSc susceptibility using Genome-Wide Association Study follow-up or independent replication approaches.

2. To independently replicate newly identified SSc-related pulmonary involvement genetic susceptibility markers.

3. To uncover novel SSc genetic susceptibility markers and to fine-map the genetic regions that have been associated with SSc, by the means the immune-focused Immunochip genotyping platform.

4. To perform a comprehensive analysis of the Human Leukocyte Antigen (HLA) region association with SSc using the dense set of HLA polymorphisms included in the Immunochip genotyping platform and taking advance of novel HLA imputation methods.

GWAS follow-up and replication studies

HMG Advance Access published November 29, 2011

Human Molecular Genetics, 2011 **1–8**
doi:10.1093/hmg/ddr522

A GWAS follow-up study reveals the association of the *IL12RB2* gene with systemic sclerosis in Caucasian populations

Lara Bossini-Castillo[1,*], Jose-Ezequiel Martin[1], Jasper Broen[2], Olga Gorlova[3], Carmen P. Simeón[4], Lorenzo Beretta[5], Madelon C. Vonk[2], Jose Luis Callejas[6], Ivan Castellví[7], Patricia Carreira[8], Francisco José García-Hernández[9], Mónica Fernández Castro[10], the Spanish Scleroderma Group[†], Marieke J.H. Coenen[11], Gabriela Riemekasten[12], Torsten Witte[13], Nicolas Hunzelmann[14], Alexander Kreuter[15], Jörg H.W. Distler[16], Bobby P. Koeleman[17], Alexandre E. Voskuyl[18], Annemie J. Schuerwegh[19], Øyvind Palm[20], Roger Hesselstrand[21], Annika Nordin[22], Paolo Airó[23], Claudio Lunardi[24], Raffaella Scorza[5], Paul Shiels[25], Jacob M. van Laar[26], Ariane Herrick[27], Jane Worthington[27], Christopher Denton[28], Filemon K. Tan[29], Frank C. Arnett[29], Sandeep K. Agarwal[29], Shervin Assassi[29], Carmen Fonseca[28], Maureen D. Mayes[29], Timothy R.D.J. Radstake[2] and Javier Martin[1]

[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Granada 18100, Spain [2]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen 6525, The Netherlands [3]Department of Epidemiology, M.D. Anderson Cancer Center, Houston, TX 77030, USA [4]Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona 08035, Spain [5]IRCCS Fondazione Policlinico-Mangiagalli-Regina Elena & University of Milan, Allergy, Clinical Immunology and Rheumatology, Milan 20122, Italy [6]Unidad de Enfermedades Sistémicas Autoinmunes, Servicio de Medicina Interna, Hospital Clínico Universitario San Cecilio, Granada 18012, Spain [7]Servicio de Reumatología, Hospital Sant Pau, Barcelona 08025, Spain [8]Servicio de Reumatología, Hospital 12 de Octubre, Madrid 28041, Spain [9]Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla 41013, Spain [10]Servicio de Reumatología, Hospital Universitario Puerta del Hierro Majadahonda, Madrid 28222, Spain [11]Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen 6525, The Netherlands [12]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin 12203, Germany [13]Hannover Medical School, Hannover 30625, Germany [14]Department of Dermatology, University of Cologne, Cologne 50923, Germany [15]Ruhr University of Bochum, Bochum 44801, Germany [16]Department of Internal Medicine 3, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen 91054, Germany [17]Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3508AB, The Netherlands [18]Department of Rheumatology, VU University Medical Center, Amsterdam 1081, The Netherlands [19]Department of Rheumatology, Leiden University Medical Center, Leiden 2333, The Netherlands [20]Department of Rheumatology, Rikshospitalet, Oslo University Hospital, Oslo 0027, Norway [21]Department of Rheumatology, Lund University, Lund 222 41, Sweden [22]Karolinska Institute, Stockholm 171 77, Sweden [23]Servizio di Reumatologia ed Immunologia Clinica Spedali Civili, Brescia 25018, Italy [24]Department of Medicine, Università degli Studi di Verona, Verona 37129, Italy [25]University of Glasgow, Glasgow G12 8QQ, UK [26]Institute of Cellular Medicine, Newcastle University, Newcastle NE2 4HH, UK [27]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9PT, UK [28]Centre for Rheumatology, Royal Free and University College Medical School, London NW3 2QG, UK [29]The University of Texas Health Science Center–Houston, Houston, TX 77030, USA

Received June 1, 2011; Revised October 28, 2011; Accepted November 7, 2011

[*]To whom correspondence should be addressed at: Instituto de Parasitología y Biomedicina López-Neyra (IPBLN-Ctable SIC), Consejo Superior de Investigaciones Científicas, Parque Tecnológico Ciencias de la Salud. Avenida del Conocimiento s/n 18100-Armilla, Granada, Spain. Tel: +34 958181621; Fax: +34 958181632; Email: larabc@ipb.csic.es
[†]The authors are listed in Appendix.

A single-nucleotide polymorphism (SNP) at the *IL12RB2* locus showed a suggestive association signal in a previously published genome-wide association study (GWAS) in systemic sclerosis (SSc). Aiming to reveal the possible implication of the *IL12RB2* gene in SSc, we conducted a follow-up study of this *locus* in different Caucasian cohorts. We analyzed 10 GWAS-genotyped SNPs in the *IL12RB2* region (2309 SSc patients and 5161 controls). We then selected three SNPs (rs3790567, rs3790566 and rs924080) based on their significance level in the GWAS, for follow-up in an independent European cohort comprising 3344 SSc and 3848 controls. The most-associated SNP (rs3790567) was further tested in an independent cohort comprising 597 SSc patients and 1139 controls from the USA. After conditional logistic regression analysis of the GWAS data, we selected rs3790567 [$P_{MH} = 1.92 \times 10^{-5}$ odds ratio (OR) = 1.19] as the genetic variant with the firmest independent association observed in the analyzed GWAS peak of association. After the first follow-up phase, only the association of rs3790567 was consistent ($P_{MH} = 4.84 \times 10^{-3}$ OR = 1.12). The second follow-up phase confirmed this finding ($P_{\chi 2} = 2.82 \times 10^{-4}$ OR = 1.34). After performing overall pooled-analysis of all the cohorts included in the present study, the association found for the rs3790567 SNP in the *IL12RB2* gene region reached GWAS-level significant association ($P_{MH} = 2.82 \times 10^{-9}$ OR = 1.17). Our data clearly support the *IL12RB2* genetic association with SSc, and suggest a relevant role of the interleukin 12 signaling pathway in SSc pathogenesis.

INTRODUCTION

Systemic sclerosis or scleroderma (SSc) is a rare complex connective tissue disorder characterized by extensive fibrosis of multiple organs produced by vascular damage and autoimmune dysfunction (1,2). Patients are commonly classified into two major subgroups: the limited cutaneous SSc (lcSSc) and the diffuse cutaneous (dcSSc) form of the disease (3). Positive autoantibody titers are a main feature of this disabling condition, especially anticentromere autoantibodies (ACA) and antitopoisomerase autoantibodies (ATA) (1,2). To date, a number of genes have been implicated in an increased susceptibility to SSc, confirming the genetic component of this complex disease (4,5). Some of these genes are shared with other related autoimmune diseases, supporting the idea of common pathogenic pathways underlying autoimmune imbalance (6,7).

Recently, our group published the first genome-wide association study (GWAS) conducted in Caucasian SSc patients (5). GWASs are often followed by follow-up studies focused on the regions where association peaks are observed, not only in the associations which reached the GWAS significance level, but also those which are below the GWAS level but might result in true association with the disease. In this line, a single-nucleotide polymorphism (SNP) at the *IL12RB2* locus showed a suggestive association signal in the previously mentioned GWAS [$P_{MH} = 1.92 \times 10^{-5}$ odds ratio (OR) = 1.19 (1.10–1.29)] (5).

Noteworthy, interleukin 12 (IL-12) binding to its receptor powerfully induces IFNγ production and promotes T helper differentiation in Th1 cells (8). In addition, several experimental and clinical studies have implicated IL-12 and IFNγ in the development of autoimmune inflammation (8,9). The IL-12 receptor (IL-12R) comprises two subunits, IL-12R β1 subunit (IL-12Rβ1) and IL-12R β2 subunit (IL-12Rβ2), which are both homologous to gp130 (a shared component of the receptors for several type I cytokines) (10).

IL12RB2 encodes IL-12Rβ2, which constitutes the transducing component of the receptor heterodimer and recruits different tyrosine kinases, signal transducers and activators of transcription (11–13). Interestingly, animal models lacking IL12Rβ2 signaling develop autoimmune events (14). Moreover, polymorphisms in the *IL12RB2* gene region and upstream this locus have been related to several human autoimmune disorders, such as psoriasis (PS) (15), primary biliary cirrhosis (PBC) (16), Behçet disease (17,18) and giant cell arteritis (GCA) (19).

Hence, with the aim of investigating the possible role of the *IL12RB2* gene in SSc, we conducted a GWAS follow-up study in different European and US Caucasian cohorts.

RESULTS

IL12RB2 region analysis in the GWAS set

Ten SNPs in the *IL12RB2* region were included in the initial GWAS analysis set, six of them were found to be significantly associated with SSc, but only four remained significant after GC correction (Table 1). However, conditioned logistic regression revealed that among the initially observed associations, only the rs3790567 association was independent from the others (Table 1). HapMap linkage disequilibrium patterns defined rs3790566 (not included in the GWAS phase) as the unique tag-SNP for rs3790567. Hence, both the most-associated SNP (rs3790567) and this tag-SNP (rs3790566) were selected for replication.

Despite the loss of the observed association after correction for multiple testing (Table 1), we also included rs924080 in the first follow-up phase. This genetic variant was located in the intergenic region between *IL12RB2* and *IL23R*, and it was the last GWAS SNP contained in the *IL23R* haplotype block (Fig. 1). In addition, this polymorphism mapped in a recombination hotspot identified in the HapMap Project (Phase II, Caucasian and Asian populations; http://www.hapmap.org) and previous reports (17).

European follow-up phase

Table 2 shows the pooled analysis of seven independent white European cohorts of the three SNPs analyzed in the first follow-up phase. No evidence of association was observed

Table 1. Pooled logistic regression of *IL12RB2* genetic variants in the GWAS cohort (2309 SSc patients and 5161 controls)

SNP	Chr: 1 position (bp)	Minor allele	P_{log}	OR	P_{GC}	*P*-value: add to rs3790567	OR and rs3790567	*P*-value rs3790567 add to SNP	OR rs3790567 and to SNP	r^2 with rs3790567
rs924080	67,532,728	G	2.93×10^{-2}	1.08	3.91×10^{-2}	0.12	1.06	2.16×10^{-5}	1.19	0.02
rs12131065	67,541,594	A	0.16	0.94	0.18	0.20	0.95	7.72×10^{-6}	1.19	0.001
rs3790558	67,549,609	C	0.31	1.04	0.34	0.19	0.95	4.74×10^{-6}	1.23	0.23
rs10489627	67,552,264	G	4.88×10^{-2}	1.08	0.06	0.83	0.99	4.98×10^{-5}	1.20	0.23
rs2066445	67,554,563	A	0.09	0.93	0.11	0.11	0.93	7.62×10^{-6}	1.20	0.0005
rs3790567	67,594,965	A	6.36×10^{-6}	1.20	1.92×10^{-5}	NA	NA	NA	NA	NA
rs3828069	67,612,161	G	4.24×10^{-2}	0.91	0.05	0.44	0.96	4.28×10^{-5}	1.19	0.08
rs4297265	67,624,923	G	1.71×10^{-2}	1.09	2.41×10^{-2}	0.39	0.96	9.86×10^{-5}	1.23	0.44
rs2270614	67,628,609	A	1.66×10^{-2}	1.09	2.34×10^{-2}	0.41	0.96	1.04×10^{-4}	1.23	0.44
rs7555183	67,633,215	A	0.24	1.05	0.27	0.92	1.00	1.31×10^{-5}	1.20	0.08

Chr, chromosome; P_{log}, logistic regression *P*-value; OR, odds ratio; P_{GC}, GC corrected *P*-value. Last columns, single locus test *P*-value when SNP added to rs3790567, single locus test OR when SNP added to rs3790567, single locus test *P*-value when rs3790567 added to SNP in logistic regression analyses, single locus test OR when rs3790567 added to SNP in logistic regression analyses and pairwise r^2 of SNP with rs3790567. NA, not applicable.

for rs924080. Despite an initial association of rs3790566 and rs3790567, after performing Bonferroni multiple test correction only the association of rs3790567 remained significant (Table 2). The pooled analysis of this genetic variant in the GWAS cohort and the independent follow-up set reached a notable statistically significant association [$P_{MH} = 5.19 \times 10^{-7}$ OR = 1.16 (1.09–1.22), Table 3].

The subgroup and autoantibody titer stratified pooled analyses comprising the GWAS and the European follow-up cohorts showed firm statistically significant risk association signals in all the subgroups of the disease considered (Supplementary Material, Tables S1–S2).

US follow-up phase

In order to confirm the rs3790567 signal, an independent US cohort was included (597 SSc and 1139 controls). Case–control frequency analysis revealed a strong association [$P_{\chi 2} = 2.82 \times 10^{-4}$ OR = 1.34 (1.14–1.57), Table 3]. After stratification, only lcSSc subgroup reached statistical significance, probably due to a lack of power since the other subgroups are relatively smaller (Supplementary Material, Tables S1–S2).

The overall pooled analysis of rs3790567 comprising the GWAS set and both the European and the US follow-up sets reached GWAS-level statistically significant association in the whole set of SSc patients [$P_{MH} = 2.82 \times 10^{-9}$ OR = 1.17 (1.11–1.24)] and remained significant after stratification in all the subgroups (Table 3 and Supplementary Material, Tables S1–S2). Hence, we suggest that the association found in rs3790567 most likely belonged to the whole SSc set of patients rather than any of its subgroups. The rs3790567 individual population allele distributions and association tests are shown in Supplementary Material, Tables S3–S5.

IL23R locus dependence analysis

Aiming to further confirm the independence of the reported *IL12RB2* signal from the *IL23R* locus, we analyzed the association of the SNPs in the *IL23R* region which were included in the GWAS initial phase and their effect on the *IL12RB2* rs3790567 association. The *IL23R* region comprised 27 SNPs and only 4 of them showed some marginal association with SSc, considering uncorrected *P*-values (Supplementary Material, Table S7). Nevertheless, the association observed in rs3790567 was found independent of these weak signals (Supplementary Material, Table S7).

DISCUSSION

Our data clearly support an association of *IL12RB2* rs3790567 with SSc. The risk effect of the *IL12RB2* rs3790567 minor allele is consistent in all the analyzed cohorts with the exception of the Italian population. In contrast to other cohorts, the minor allele rs3790567*A is over-represented in controls compared with SSc patients in the Italian sample set. The Italian control group showed the highest minor allele frequency among all the included populations, and the linkage disequilibrium between rs3790567 and rs3790566 in the Italian cohort was considerably lower ($r^2 = 0.70$) than in the other European populations ($r^2 > 0.90$). In addition, this over-representation of the rs3790567*A minor allele is also observed in the TSI (Tuscan in Italy) population in the HapMap Project (Phase III) ($MAF_{TSI} = 0.30$) when compared with the CEU population ($MAF_{CEU} = 0.26$). However, the linkage disequilibrium observed between rs3790567 and rs3790566 in the Hapmap TSI population compared with the CEU population decreased very slightly ($r^2_{TSI} = 0.97$, $r^2_{CEU} = 1$). Hence, it is likely that the observed discrepancies in the Italian set were due to ethnic differences in linkage disequilibrium patterns. Supporting this notion, BD test revealed significant heterogeneity in the lcSSc overall pooled analysis caused by the Italian patients (P_{BD} with the Italian population = 0.04; P_{BD} without the Italian population = 0.45). Although cases and controls were geographically matched, the potential effect of population substructure in the replication cohorts could not be controlled by deriving principal components on a population-specific basis, as it was performed for the GWAS cohorts, due to the lack of high-throughput genotype information for these individuals. Considering the reported heterogeneous genetic background for Italian populations (20), the influence of this factor on the deviation observed in our Italian subset cannot be ignored.

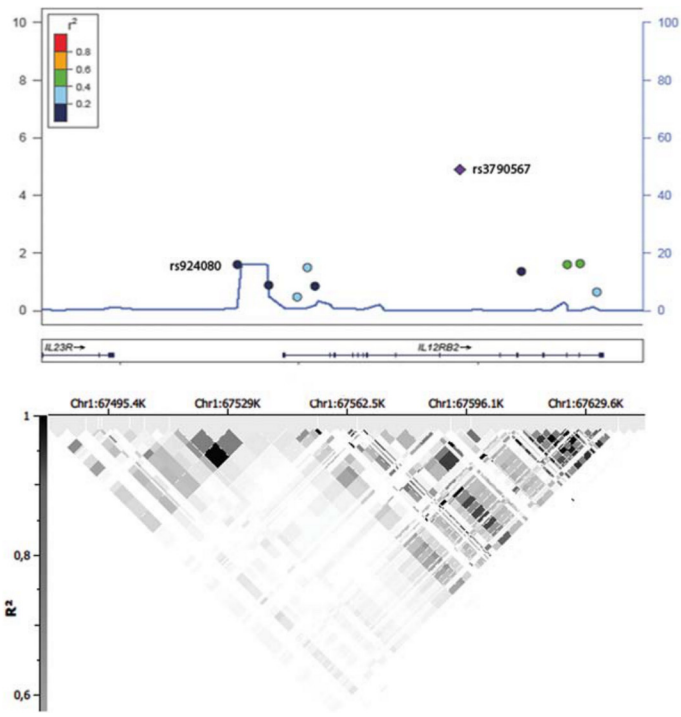


Figure 1. GWAS phase of the *IL12RB2* region. Regional association plot, recombination rate, linkage disequilibrium pattern and pairwise r^2 of the SNPs with rs3790567.

Table 2. Genotype and allele distribution of *IL12RB2* genetic variants in the European SSc patients and controls follow-up study (3344 SSc/3848 controls)

SNP	1/2	CTRL				SSc								
		1/1 (*n*)	1/2 (n)	2/2 (*n*)	MAF	1/1 (n)	1/2 (*n*)	2/2 (*n*)	MAF	P_{MH}	OR	95% CI	P_{Bonf}	P_{BD}
rs924080	C/T	0.22 (827)	0.48 (1807)	0.29 (1094)	0.46	0.22 (687)	0.49 (1545)	0.30 (934)	0.46	0.96	1.00	0.93–1.07	1	NS
rs3790566	T/C	0.08 (280)	0.37 (1334)	0.55 (1978)	0.26	0.09 (273)	0.39 (1196)	0.52 (1617)	0.28	3.35×10^{-2}	1.09	1.01–1.18	0.10	NS
rs3790567	A/G	0.08 (241)	0.37 (1169)	0.56 (1773)	0.26	0.09 (282)	0.38 (1187)	0.52 (1616)	0.28	4.84×10^{-3}	1.12	1.04–1.22	0.01	NS

SSc, systemic sclerosis patients; CTRL, healthy controls; 1/2, minor allele/major allele; MAF, minor allele frequency; P_{MH}, allelic Mantel–Haenszel fixed effects model *P*-value; OR, odds ratio; 95% CI, 95% confidence interval; P_{BD}, Breslow–Day test *P*-value; NS, not statistically significant.

As stated above, different *IL12RB2* genetics variants have been associated with multiple autoimmune disorders (15–19). However, the fact that the same *IL12RB2* variant, rs3790567, has been associated with increased susceptibility to both PBC and GCA (16,19), together with the lack of association in our data of a nearby highly linked variant (rs3790566), suggest that rs3790567 intronic SNP may be tagging a functional variant or even has a yet unknown functional implication itself.

The *IL12RB2* gene maps close to the IL-23R coding gene (*IL23R*), which are located <50 kb from each other. IL-23R binds IL-12Rβ1 chain constituting the heterodimeric receptor for IL-23 (21). Although *IL23R* polymorphisms have been associated with different autoimmune diseases (22–28), its implication in SSc is not clear (29–31). In this report, conditional regression analyses showed that the association of *IL12RB2* rs3790567 with SSc is independent from all the studied *IL23R* genetic polymorphisms, even from *IL23R* rs11209026 (Arg281Gln) missense variant. Hence, we suggest that the reported association of the *IL12RB2* gene with SSc susceptibility does not rely on the *IL23R* locus. Nevertheless, further studies will be necessary to investigate the possible effect of *IL12RB2* genetic variants on *IL23R* gene expression.

IL-12 levels are increased in the serum of SSc patients as well as in the alveolar lavage fluid (BAL-f) from patients with SSc-associated interstitial lung disease (ILD) (32,33). Although IL-12 classical implication in immune imbalance

Table 3. Genotype and allele distribution of *IL12RB2* rs3790567 genetic variant in SSc patients and controls in a three-step association study

Population (CTRL/SSc)	CTRL AA (*n*)	AG (*n*)	GG (*n*)	MAF	SSc AA (*n*)	AG (*n*)	GG (*n*)	MAF	P_{MH}	OR	95% CI	P_{BD}
GWAS cohort (5161/2309)	0.06 (332)	0.37 (1911)	0.57 (2918)	0.25	0.08 (196)	0.40 (919)	0.52 (1194)	0.28	1.92×10^{-5}	1.19	1.10–1.29	NS
European follow-up (3183/3085)	0.08 (241)	0.37 (1169)	0.56 (1773)	0.26	0.09 (282)	0.38 (1187)	0.52 (1161)	0.28	4.84×10^{-3}	1.12	1.04–1.22	NS
GWAS + European follow-up (8344/5394)	0.07 (573)	0.37 (3080)	0.56 (4691)	0.25	0.09 (478)	0.39 (2106)	0.52 (2810)	0.28	5.19×10^{-7}	1.16	1.09–1.22	NS
US follow-up (1139/597)	0.05 (60)	0.37 (417)	0.58 (662)	0.24	0.10 (59)	0.39 (231)	0.51 (307)	0.29	[a]2.82×10^{-4}	1.34	1.14–1.57	NA
GWAS + European + US follow-up (9483/5991)	0.07 (633)	0.37 (3497)	0.56 (5353)	0.25	0.09 (537)	0.39 (2337)	0.52 (3117)	0.28	2.82×10^{-9}	1.17	1.11–1.24	NS

Controls are used as reference for all comparisons. CTRL, healthy controls; SSc, systemic sclerosis; MAF, minor allele (A) frequency; P_{MH}, allelic Mantel–Haenszel fixed effects model *P*-value; [a]allelic Chi-square uncorrected *P*-value; OR, odds ratio; 95% CI, 95% confidence interval; P_{BD}, Breslow–Day test *P*-value; NS, not statistically significant; NA, not applicable.

has been mainly related to a pro-inflammatory cell-mediated immunity and Th1 response (9) and increased levels of IL-12 correlate with renal vascular damage (32), the role of IL-12 in SSc pathogenesis should be considered cautiously. Indeed, SSc patients and especially those with ILD have a Th2-polarized response (34). Additionally, it has been suggested that IL-12 drives a drift from a Th2 to Th1 response which improves skin score in SSc patients (35). Moreover, IL-12 is known to have anti-fibrotic effects in fibroblasts (36), and the administration of IL-12 coding plasmid to the tight skin SSc mouse model prevents collagen accumulation in the skin (37). On the other hand, the implication IL-12Rβ2 in autoimmune events seems to be complex as well. For instance, IL12rb2 knock-out mice do not display IL-12-mediated NK cytotoxicity (38) and the IL-12/IL-12Rβ2 axis is known to be critical for the generation of Th1 autoreactive cells (39), but, despite this, these mice develop spontaneous autoimmune pathology (immune-complex glomerulonephritis) and B-cell tumors by a strong IL-6 up-regulation (14,40). In addition, IL-12R signals predominantly through the STAT pathway, especially STAT4 (37,40). In this regard, it should be noted that polymorphisms in the *STAT4* gene are well-established risk factors for SSc (4). Hence, it is likely that genetically predisposed individuals may present subtle differences in IL-12 signaling pathway regulation that could influence the prognosis of SSc.

To date, only a few SSc-related loci have reached a GWAS-level significance (i.e. *P*-value $< 5.00 \times 10^{-8}$), both in the previously mentioned GWAS and recent studies: the HLA region, *STAT4*, *TNPO3-IRF5*, *CD247*, *PSORS1C1*, *TNIP1* and *IRF8* (5,41,42). Hence, we consider that the reported GWAS-level significant association may firmly contribute to the genetic knowledge of the disease.

In conclusion, we report for the first time the association of an *IL12RB2* genetic variant with SSc. Our data together with previous reports identify *IL12RB2* as a common genetic risk factor for autoimmunity.

MATERIALS AND METHODS

Subjects

The GWAS cohort was comprised of 2309 SSc patients and 5161 controls of Caucasian ancestry from Spain, Germany, The Netherlands and USA from a previously published study (5). The first follow-up phase consisted of 3085 SSc patients and 3183 controls from seven European Caucasian cohorts (Spain, Germany, The Netherlands, Italy, Sweden, UK and Norway). The second follow-up step comprised 1736 additional USA Caucasian individuals (597 SSc and 1139 controls). All the patients fulfilled the 1980 American College of Rheumatology (ACR) classification criteria for SSc (43) or the criteria proposed for early-SSc (44). In addition, patients were classified as having lcSSc or dcSSc as described in LeRoy *et al.* (3).

The following clinical data were collected for the ascertainment of the clinical phenotype of SSc patients: age, gender and presence of SSc-specific autoantibodies (Ab) ATA and ACA (Supplementary Material, Table S6). The control population consisted of unrelated healthy individuals recruited in the same geographical regions as SSc patients and matched by age, sex and ethnicity with the SSc patients groups.

The study was approved by local ethical committees from all the participating centers. Both patients and controls were included in the study after written informed consent. DNA from patients and controls were obtained using standard methods.

SNP selection

In the screening GWAS phase, we included a 116 kb region spanning the *IL12RB2* region and ~13 kb upstream and downstream from this locus, from base pair 67 530 000 to 67 646 000 in chromosome 1, in the GWAS cohorts. After QC filtering as described in Radstake *et al.* (5), genotyping data for 10 SNPs over this region on chromosome 1 were available. The same procedure was applied for the analysis of the *IL23R* region, which comprised 163 kb and 27 SNPs.

TaqMan SNP genotyping of the follow-up cohorts was performed in a 7900HT Real-Time Polymerase Chain Reaction (PCR) System from Applied Biosystems following the manufacturer's suggestions (Foster City, CA, USA).

Statistical analysis

Significance was calculated using 2×2 contingency tables and Fisher's exact test or χ^2 when necessary, to obtain *P*-values, OR and 95% confidence intervals using PLINK

(v1.07) software (http://pngu.mgh.harvard.edu/purcell/plink/). *P*-values below 0.05 were considered statistically significant. Bonferroni correction and GC as described in Radstake *et al*. (5) were applied. The Hardy–Weinberg equilibrium (HWE) was tested for all the SNPs comparing the observed genotype distribution in controls with the expected genotype distribution under HWE by means of Fisher's exact test or χ^2 when necessary as described in Radstake *et al*. (5). The logistic regression and conditioned logistic regression analyses (considering the different cohorts as covariables) were performed using PLINK software. Linkage disequilibrium patterns across the region in the HapMap Project Phase I and II (CEU population) defined the haplotype-tagging SNPs using Haploview (v.4.2) software (http://www.broadinstitute.org/haploview/haploview). The SNPs included in the GWAS phase were forced-included in the list of SNPs. Over this region on chromosome 1, the recombination rate was estimated from HapMap public database using LocusZoom (v.1.1) software (http://csg.sph.umich.edu/locuszoom/) (45). SNP & Variation Suite Version 7.5.1 (Golden Helix Inc.) and LocusZoom software were used for the composition of Figure 1. Cochran–Mantel–Haenszel meta-analysis was performed to control for the differences among populations as implemented in PLINK software. In addition, the Breslow–Day test (BD test) was performed as implemented in PLINK in each meta-analysis to assess the homogeneity of the association among populations. The power of the whole set of SSc patients and controls reached 100%. Power was calculated using the software Power Calculator for Genetic Studies 2006 (46) and assuming an additive model (*P*-value = 0.01 OR = 1.20). The genotyping success call rate in the GWAS cohort was of 99.8%, while in the replication set was over 95%.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We thank Sofia Vargas, Sonia García and Gema Robledo for their excellent technical assistance and all the patients and control donors for their essential collaboration. We thank Banco Nacional de ADN (University of Salamanca, Spain) and the Norwegian Bone Marrow Donor Registry who supplied part of the control DNA samples. We are also thankful to EUSTAR (The EULAR Scleroderma Trials and Research group) and the German Network of Systemic Sclerosis for the facilitation of this project.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by the following grants: J.M. was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009-11110 from the Spanish Ministry of Science, CTS-4977 from Junta de Andalucía, Spain, in part by Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain and by Fondo Europeo de Desarrollo Regional (FEDER). T.R.D.J.R. was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). 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). T.W. was granted by DFG WI 1031/6.1 and DFG KFO 250 TP03. N.O. was funded by PI-0590-2010, Consejería de Salud, Junta de Andalucía, Spain. The USA studies were supported by NIH/NIAMS Scleroderma Registry and DNA Repository (N01-AR-0-2251), NIH/NIAMS-RO1-AR055258 and NIH/NIAMS Center of Research Translation in Scleroderma (1P50AR054144) and the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111).

APPENDIX

Spanish Scleroderma Group: Norberto Ortego-Centeno and Raquel Ríos, Unidad de Enfermedades Sistémicas Autoinmunes, Servicio de Medicina Interna, Hospital Clínico Universitario San Cecilio, Granada; Nuria Navarrete, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Rosa García Portales, Servicio de Reumatología, Hospital Virgen de la Victoria, Málaga; María Teresa Camps, Servicio de Medicina Interna, Hospital Carlos Haya, Málaga; Antonio Fernández-Nebro, Servicio de Reumatología, Hospital Carlos Haya, Málaga; María F. González-Escribano, Servicio de Inmunología, Hospital Virgen del Rocío, Sevilla; Julio Sánchez-Román and Mª Jesús Castillo, Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Servicio de Reumatología, Hospital Reina Sofía, Córdoba; Benjamín Fernández-Gutiérrez and Luis Rodríguez-Rodríguez, Servicio de Reumatología, Hospital Clínico San Carlos, Madrid; Esther Vicente, Servicio de Reumatología, Hospital La Princesa, Madrid; José Luis Andreu, Servicio de Reumatología, Hospital Puerta del Hierro, Madrid; Paloma García de la Peña, Servicio de Reumatología, Hospital Madrid Norte Sanchinarro, Madrid; Francisco Javier López-Longo and Lina Martínez, Servicio de Reumatología, Hospital General Universitario Gregorio Marañón, Madrid; Vicente Fonollosa, Servicio de Medicina Interna, Hospital Valle de Hebrón, Barcelona; Gerard Espinosa, Servicio de Medicina Interna, Hospital Clinic, Barcelona; Carlos Tolosa, Servicio de Medicina Interna, Hospital Parc Tauli, Sabadell; Anna Pros, Servicio de Reumatología, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Servicio de Medicina Interna, Hospital Universitari Mútua Terrasa, Barcelona; Francisco Javier Narváez, Servicio de Reumatología, Hospital Universitari de Bellvitge, Barcelona; Miguel Ángel González-Gay, Servicio de Reumatología, Hospital Universitario Marqués de Valdecilla, Santander; Bernardino Díaz, Luis Trapiella and María Gallego, Servicio de Medicina Interna, Hospital Central de Asturias, Oviedo; María del Carmen Freire and Inés Vaqueiro, Unidad de Trombosis y Vasculitis, Servicio de Medicina Interna, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; María Victoria Egurbide, Servicio de Medicina Interna, Hospital de Cruces, Barakaldo; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital

Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Servicio de Reumatología, Hospital Universitario de Canarias, Tenerife; Emma Beltrán and Servicio de Reumatología, Hospital del Doctor Peset Aleixandre, Valencia; José Andrés Román-Ivorra, Servicio de Reumatología, Hospital Universitari i Politecnic La Fe, Valencia.

REFERENCES

1. Gabrielli, A., Avvedimento, E.V. and Krieg, T. (2009) Scleroderma. *N. Engl. J. Med.*, **360**, 1989–2003.
2. Katsumoto, T.R., Whitfield, M.L. and Connolly, M.K. (2011) The pathogenesis of systemic sclerosis. *Annu. Rev. Pathol.*, **6**, 509–537.
3. 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.
4. Martin, J. and Fonseca, C. (2011) The genetics of scleroderma. *Curr. Rheumatol. Rep.*, **13**, 13–20.
5. 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.
6. Zhernakova, A., van Diemen, C.C. and Wijmenga, C. (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat. Rev. Genet.*, **10**, 43–55.
7. Gregersen, P.K. and Olsson, L.M. (2009) Recent advances in the genetics of autoimmune disease. *Annu. Rev. Immunol.*, **27**, 363–391.
8. Paunovic, V., Carroll, H.P., Vandenbroeck, K. and Gadina, M. (2008) Signalling, inflammation and arthritis: crossed signals: the role of interleukin (IL)-12, -17, -23 and -27 in autoimmunity. *Rheumatology (Oxford)*, **47**, 771–776.
9. Caspi, R.R. (1998) IL-12 in autoimmunity. *Clin. Immunol. Immunopathol.*, **88**, 4–13.
10. Hunter, C.A. (2005) New IL-12-family members: IL-23 and IL-27, cytokines with divergent functions. *Nat. Rev. Immunol.*, **5**, 521–531.
11. Bacon, C.M., McVicar, D.W., Ortaldo, J.R., Rees, R.C., O'Shea, J.J. and Johnston, J.A. (1995) Interleukin 12 (IL-12) induces tyrosine phosphorylation of JAK2 and TYK2: differential use of Janus family tyrosine kinases by IL-2 and IL-12. *J. Exp. Med.*, **181**, 399–404.
12. Jacobson, N.G., Szabo, S.J., Weber-Nordt, R.M., Zhong, Z., Schreiber, R.D., Darnell, J.E. Jr and Murphy, K.M. (1995) Interleukin 12 signaling in T helper type 1 (Th1) cells involves tyrosine phosphorylation of signal transducer and activator of transcription (Stat)3 and Stat4. *J. Exp. Med.*, **181**, 1755–1762.
13. Ahn, H.J., Tomura, M., Yu, W.G., Iwasaki, M., Park, W.R., Hamaoka, T. and Fujiwara, H. (1998) Requirement for distinct Janus kinases and STAT proteins in T cell proliferation versus IFN-gamma production following IL-12 stimulation. *J. Immunol.*, **161**, 5893–5900.
14. Airoldi, I., Di Carlo, E., Cocco, C., Sorrentino, C., Fais, F., Cilli, M., D'Antuono, T., Colombo, M.P. and Pistoia, V. (2005) Lack of Il12rb2 signaling predisposes to spontaneous autoimmunity and malignancy. *Blood*, **106**, 3846–3853.
15. Liu, Y., Helms, C., Liao, W., Zaba, L.C., Duan, S., Gardner, J., Wise, C., Miner, A., Malloy, M.J., Pullinger, C.R. *et al.* (2008) A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. *PLoS Genet.*, **4**, e1000041.
16. Hirschfield, G.M., Liu, X., Xu, C., Lu, Y., Xie, G., Gu, X., Walker, E.J., Jing, K., Juran, B.D., Mason, A.L. *et al.* (2009) Primary biliary cirrhosis associated with HLA, IL12A, and IL12RB2 variants. *N. Engl. J. Med.*, **360**, 2544–2555.
17. Remmers, E.F., Cosan, F., Kirino, Y., Ombrello, M.J., Abaci, N., Satorius, C., Le, J.M., Yang, B., Korman, B.D., Cakiris, A. *et al.* (2010) Genome-wide association study identifies variants in the MHC class I, IL10, and IL23R-IL12RB2 regions associated with Behcet's disease. *Nat. Genet.*, **42**, 698–702.
18. Mizuki, N., Meguro, A., Ota, M., Ohno, S., Shiota, T., Kawagoe, T., Ito, N., Kera, J., Okada, E., Yatsu, K. *et al.* (2010) Genome-wide association studies identify IL23R-IL12RB2 and IL10 as Behcet's disease susceptibility loci. *Nat. Genet.*, **42**, 703–706.
19. Rodriguez-Rodriguez, L., Carmona, F.D., Castaneda, S., Miranda-Filloy, J.A., Morado, I.C., Narvaez, J., Mari-Alfonso, B., Gomez-Vaquero, C., Amigo-Diaz, E., Rios-Fernandez, R. *et al.* (2011) Role of rs1343151 IL23R and rs3790567 IL12RB2 polymorphisms in biopsy-proven giant cell arteritis. *J. Rheumatol.*, **38**, 889–892.
20. Nelis, M., Esko, T., Magi, R., Zimprich, F., Zimprich, A., Toncheva, D., Karachanak, S., Piskackova, T., Balascak, I., Peltonen, L. *et al.* (2009) Genetic structure of Europeans: a view from the North-East. *PLoS ONE*, **4**, e5472.
21. Trinchieri, G., Pflanz, S. and Kastelein, R.A. (2003) The IL-12 family of heterodimeric cytokines: new players in the regulation of T cell responses. *Immunity*, **19**, 641–644.
22. Duerr, R.H., Taylor, K.D., Brant, S.R., Rioux, J.D., Silverberg, M.S., Daly, M.J., Steinhart, A.H., Abraham, C., Regueiro, M., Griffiths, A. *et al.* (2006) A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*, **314**, 1461–1463.
23. Oliver, J., Rueda, B., Lopez-Nevot, M.A., Gomez-Garcia, M. and Martin, J. (2007) Replication of an association between IL23R gene polymorphism with inflammatory bowel disease. *Clin. Gastroenterol. Hepatol.*, **5**, 977–981.
24. Cargill, M., Schrodi, S.J., Chang, M., Garcia, V.E., Brandon, R., Callis, K.P., Matsunami, N., Ardlie, K.G., Civello, D., Catanese, J.J. *et al.* (2007) A large-scale genetic association study confirms IL12B and leads to the identification of IL23R as psoriasis-risk genes. *Am. J. Hum. Genet.*, **80**, 273–290.
25. Burton, P.R., Clayton, D.G., Cardon, L.R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D.P., McCarthy, M.I., Ouwehand, W.H., Samani, N.J. *et al.* (2007) Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat. Genet.*, **39**, 1329–1337.
26. Rueda, B., Orozco, G., Raya, E., Fernandez-Sueiro, J.L., Mulero, J., Blanco, F.J., Vilches, C., Gonzalez-Gay, M.A. and Martin, J. (2008) The IL23R Arg381Gln non-synonymous polymorphism confers susceptibility to ankylosing spondylitis. *Ann. Rheum. Dis.*, **67**, 1451–1454.
27. Nunez, C., Dema, B., Cenit, M.C., Polanco, I., Maluenda, C., Arroyo, R., de las Heras, V., Bartolome, M., de la Concha, E.G., Urcelay, E. *et al.* (2008) IL23R: a susceptibility locus for celiac disease and multiple sclerosis? *Genes Immun.*, **9**, 289–293.
28. Jiang, Z., Yang, P., Hou, S., Du, L., Xie, L., Zhou, H. and Kijlstra, A. (2010) IL-23R gene confers susceptibility to Behcet's disease in a Chinese Han population. *Ann. Rheum. Dis.*, **69**, 1325–1328.
29. Rueda, B., Broen, J., Torres, O., Simeon, C., Ortego-Centeno, N., Schrijvenaars, M.M., Vonk, M.C., Fonollosa, V., van den Hoogen, F.H., Coenen, M.J. *et al.* (2009) The interleukin 23 receptor gene does not confer risk to systemic sclerosis and is not associated with systemic sclerosis disease phenotype. *Ann. Rheum. Dis.*, **68**, 253–256.
30. Agarwal, S.K., Gourh, P., Shete, S., Paz, G., Divecha, D., Reveille, J.D., Assassi, S., Tan, F.K., Mayes, M.D. and Arnett, F.C. (2009) Association of interleukin 23 receptor polymorphisms with anti-topoisomerase-I positivity and pulmonary hypertension in systemic sclerosis. *J. Rheumatol.*, **36**, 2715–2723.
31. Farago, B., Magyari, L., Safrany, E., Csongei, V., Jaromi, L., Horvatovich, K., Sipeky, C., Maasz, A., Radics, J., Gyetvai, A. *et al.* (2008) Functional variants of interleukin-23 receptor gene confer risk for rheumatoid arthritis but not for systemic sclerosis. *Ann. Rheum. Dis.*, **67**, 248–250.
32. Sato, S., Hanakawa, H., Hasegawa, M., Nagaoka, T., Hamaguchi, Y., Nishijima, C., Komatsu, K., Hirata, A. and Takehara, K. (2000) Levels of interleukin 12, a cytokine of type 1 helper T cells, are elevated in sera from patients with systemic sclerosis. *J. Rheumatol.*, **27**, 2838–2842.
33. Meloni, F., Caporali, R., Marone Bianco, A., Paschetto, E., Morosini, M., Fietta, A.M., Patrizio, V., Bobbio-Pallavicini, F., Pozzi, E. and Montecucco, C. (2004) BAL cytokine profile in different interstitial lung diseases: a focus on systemic sclerosis. *Sarcoidosis Vasc. Diffuse Lung Dis.*, **21**, 111–118.
34. Boin, F., De Fanis, U., Bartlett, S.J., Wigley, F.M., Rosen, A. and Casolaro, V. (2008) T cell polarization identifies distinct clinical phenotypes in scleroderma lung disease. *Arthritis Rheum.*, **58**, 1165–1174.
35. Matsushita, T., Hasegawa, M., Hamaguchi, Y., Takehara, K. and Sato, S. (2006) Longitudinal analysis of serum cytokine concentrations in systemic sclerosis: association of interleukin 12 elevation with spontaneous regression of skin sclerosis. *J. Rheumatol.*, **33**, 275–284.

36. Banning, U., Krutmann, J. and Korholz, D. (2006) The role of IL-4 and IL-12 in the regulation of collagen synthesis by fibroblasts. *Immunol. Invest.*, **35**, 199–207.
37. Tsuji-Yamada, J., Nakazawa, M., Takahashi, K., Iijima, K., Hattori, S., Okuda, K., Minami, M., Ikezawa, Z. and Sasaki, T. (2001) Effect of IL-12 encoding plasmid administration on tight-skin mouse. *Biochem. Biophys. Res. Commun.*, **280**, 707–712.
38. Wu, C., Wang, X., Gadina, M., O'Shea, J.J., Presky, D.H. and Magram, J. (2000) IL-12 receptor beta 2 (IL-12R beta 2)-deficient mice are defective in IL-12-mediated signaling despite the presence of high affinity IL-12 binding sites. *J. Immunol.*, **165**, 6221–6228.
39. Shevach, E.M., Chang, J.T. and Segal, B.M. (1999) The critical role of IL-12 and the IL-12R beta 2 subunit in the generation of pathogenic autoreactive Th1 cells. *Springer Semin. Immunopathol.*, **21**, 249–262.
40. Pistoia, V., Cocco, C. and Airoldi, I. (2009) Interleukin-12 receptor beta2: from cytokine receptor to gatekeeper gene in human B-cell malignancies. *J. Clin. Oncol.*, **27**, 4809–4816.
41. Allanore, Y., Saad, M., Dieude, P., Avouac, J., Distler, J.H., Amouyel, P., Matucci-Cerinic, M., Riemekasten, G., Airo, P., Melchers, I. *et al.* (2011) Genome-Wide Scan Identifies TNIP1, PSORS1C1, and RHOB as Novel Risk Loci for Systemic Sclerosis. *PLoS Genet.*, **7**, e1002091.
42. 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.
43. Preliminary Criteria for the Classification of Systemic Sclerosis (Scleroderma). (1980) Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. *Arthritis Rheum.*, **23**, 581–590.
44. LeRoy, E.C. and Medsger, T.A. Jr (2001) Criteria for the classification of early systemic sclerosis. *J. Rheumatol.*, **28**, 1573–1576.
45. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
46. 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.

EXTENDED REPORT

Confirmation of *TNIP1* but not *RHOB* and *PSORS1C1* as systemic sclerosis risk factors in a large independent replication study

Lara Bossini-Castillo,[1] Jose Ezequiel Martin,[1] Jasper Broen,[2] Carmen P Simeon,[3] Lorenzo Beretta,[4] Olga Y Gorlova,[5] Madelon C Vonk,[2] Norberto Ortego-Centeno,[6] Gerard Espinosa,[7] Patricia Carreira,[8] Paloma García de la Peña,[9] Natividad Oreiro,[10] José Andrés Román-Ivorra,[11] María Jesús Castillo,[12] Miguel A González-Gay,[13] Luis Sáez-Comet,[14] Ivan Castellví,[15] Annemie J Schuerwegh,[16] Alexandre E Voskuyl,[17] Anna-Maria Hoffmann-Vold,[18] Roger Hesselstrand,[19] Annika Nordin,[20] Claudio Lunardi,[21] Raffaella Scorza,[4] Jacob M van Laar,[22] Paul G Shiels,[23] Ariane Herrick,[24] Jane Worthington,[24] Carmen Fonseca,[25] Christopher Denton,[25] Filemon K Tan,[26] Frank C Arnett,[26] Shervin Assassi,[26] Bobby P Koeleman,[27] Maureen D Mayes,[26] Timothy R D J Radstake,[2,28] Javier Martin,[1] the Spanish Scleroderma Group*

► Additional supplementary data are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2012-201888.bmj.com).

For numbered affiliations see end of article.

Correspondence to
Dr Lara Bossini-Castillo, Department of Immunology, Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Consejo Superior de Investigaciones Científicas, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18100-Armilla, Granada, Spain; larabc@ipb.csic.es

Received 18 April 2012
Accepted 20 July 2012
Published Online First
15 August 2012

ABSTRACT

Introduction A recent genome-wide association study in European systemic sclerosis (SSc) patients identified three loci (*PSORS1C1*, *TNIP1* and *RHOB*) as novel genetic risk factors for the disease. The aim of this study was to replicate the previously mentioned findings in a large multicentre independent SSc cohort of Caucasian ancestry.

Methods 4389 SSc patients and 7611 healthy controls from different European countries and the USA were included in the study. Six single nucleotide polymorphisms (SNP): rs342070, rs13021401 (*RHOB*), rs2233287, rs4958881, rs3792783 (*TNIP1*) and rs3130573 (*PSORS1C1*) were analysed. Overall significance was calculated by pooled analysis of all the cohorts. Haplotype analyses and conditional logistic regression analyses were carried out to explore further the genetic structure of the tested loci.

Results Pooled analyses of all the analysed SNPs in *TNIP1* revealed significant association with the whole disease (rs2233287 $p_{MH}=1.94\times10^{-4}$, OR 1.19; rs4958881 $p_{MH}=3.26\times10^{-5}$, OR 1.19; rs3792783 $p_{MH}=2.16\times10^{-4}$, OR 1.19). These associations were maintained in all the subgroups considered. *PSORS1C1* comparison showed association with the complete set of patients and all the subsets except for the anti-centromere-positive patients. However, the association was dependent on different HLA class II alleles. The variants in the *RHOB* gene were not associated with SSc or any of its subsets.

Conclusions These data confirmed the influence of *TNIP1* on an increased susceptibility to SSc and reinforced this locus as a common autoimmunity risk factor.

INTRODUCTION

Systemic sclerosis or scleroderma (SSc) is a complex autoimmune disorder that affects the connective tissue causing fibrosis in the skin and different internal organs.[1] The contribution of different genetic factors to the development and prognosis of the disease is now widely accepted.[2] Over the past few years, genome-wide association studies (GWAS) have been a useful tool in the genetic dissection of autoimmune pathologies and other complex diseases.[3] Radstake *et al*[4] performed the first SSc GWAS in Caucasian populations, which represented the first large-scale GWAS in SSc. This work reinforced the association within the HLA region, especially with the *HLA-DQB1* gene, which was also reported in a comprehensive multiethnic SSc HLA study.[5] It also confirmed the associations found in *STAT4* and *IRF5* and identified *CD247* as a new SSc risk locus. It is worth mentioning that the role of *CD247* in SSc has recently been independently replicated.[6] This GWAS has led to three follow-up studies, which have described several novel SSc susceptibility factors, ie, *IRF8*, *GRB10*, *SOX5*, *NOTCH4*, *IL12RB2*, *CSK*, *PSD3* and *NFKB1*.[7–9] Interestingly, *SOX5* and *NOTCH4* are directly related to the fibrotic process, which is a main hallmark of SSc.

A GWAS has recently been performed in a French Caucasian SSc discovery cohort.[10] In this GWAS, 17 single-nucleotide polymorphisms (SNP) showing tier two associations were selected for follow-up in independent cohorts. Three of the selected SNP were located within the HLA region corresponding to the *HLA-DQB1* and *PSORS1C1* genes; and the remaining SNP were located in six independent non-HLA loci. After the replication step, the associations of *HLA-DQB1*, *CD247*, *STAT4* and *IRF5* were confirmed, and six SNP located in three loci (*TNIP1*, *RHOB*, *PSORS1C1*) were proposed as novel SSc risk factors.

It has been observed that associations identified from a single GWAS, even passing the established statistical significance thresholds, tend to have

inflated effect sizes.[11] This effect size is called the winner's curse, and it also affects the predictive ability of the discovered associations and the estimate of the risk variance explained by the associations.[11] Replication in independent comparable populations is thus essential for firmly establishing a genotype–phenotype association.[11 12] Therefore, we aimed to perform a large-scale replication study of the novel SSc genetic risk factors identified by GWAS strategy in an independent white European and US SSc population.

PATIENTS AND METHODS

Subjects

4389 SSc patients and 7611 controls of Caucasian ancestry (Spain, The Netherlands, USA, Italy, Sweden, UK and Norway) were included in this study. Patients were classified as having limited or diffuse SSc, as defined by LeRoy *et al*.[13] The following clinical data were collected for ascertainment of the clinical phenotype of the patients with SSc: age, gender, disease duration and presence of SSc-associated autoantibodies, anti-topoisomerase (ATA) and anti-centromere (ACA). Supplementary table S1 (available online only) shows the cohort-specific SSc patient data. The control population consisted of unrelated healthy individuals recruited in the same geographical regions as SSc patients and matched by age, sex and ethnicity with the SSc patient groups. Local ethics committees from all the participating centres approved the study. Both patients and controls were included in the study after written informed consent.

In the meta-analysis with previously published data by Allanore *et al*,[10] which includes 2246 SSc patients and 5702 healthy controls from France, Italy, Germany and Eastern Europe, the total cohort size reached 6635 patients and 13 313 controls (except for rs13021401 and rs3792783, which were not available for The Netherlands and US GWAS cohorts, respectively).

Genotyping

Genotype data of six SNP (rs342070, rs13021401 (*RHOB*), rs2233287, rs4958881, rs3792783 (*TNIP1*) and rs3130573 (*PSORS1C1*)) was obtained from both available GWAS genotyping platforms and SNP genotyping assays. When possible, genotypes from the Spanish, Dutch and US cohorts from Radstake *et al*[4] were used (Spain I, The Netherlands I and US I cohorts). In addition, additional Spanish SSc patients and controls were genotyped using the Illumina HumanCytoSNP-12 DNA Analysis BeadChip and Illumina Human1M-Duo DNA Analysis BeadChip ((Illumina Inc., San Diego, CA, USA), respectively, this information was thus included in the Spain I set when available. The remaining European cohorts (Spain II, The Netherlands II, Italy, Sweden, UK and Norway) were analysed using TaqMan SNP genotyping assays in a 7900HT Real-Time PCR System from Applied Biosystems following the manufacturer's suggestions (Applied Biosystems, Foster City, California, USA). The differences in the number of samples included in the analyses of each polymorphism correspond to the availability of the genotype data in each platform (see supplementary table S2, available online only). Stringent quality control filters and principal component analysis were applied to the GWAS and the HLA imputed data as described in Radstake *et al*.[4] The genotyping call-rate for the individuals genotyped using TaqMan assays reached: rs342070 93.77%, rs13021401 94.98%, rs2233287 96.08%, rs4958881 93.75%, rs3792783 93.29% and rs3130573 95.19%.

Statistical analysis

PLINK (V.1.07) software (http://pngu.mgh.harvard.edu/purcell/plink/) was used for individual population association tests (significance was calculated by 2×2 contingency tables and Fisher's exact test or χ^2 when necessary, and in the case of the haplotypes each haplotype was tested against all others), logistic regression and conditional logistic regression analyses. The different cohorts were considered covariables in the logistic regression analyses. OR and their 95% CI were reported. *TNIP1* haplotypes were constructed using PLINK (V.1.07) and HaploView 4.2 (http://www.broadinstitute.org/haploview/haploview) only with those individuals successfully genotyped for the three included variants (2432 SSc patients and 3496 healthy controls). The Breslow–Day test was performed as implemented in PLINK and StatsDirect to assess the homogeneity of the association among populations. Pooled analyses and meta-analyses were carried out using a Mantel–Haenszel test under a fixed effects by PLINK (V.1.07), METAL[14] (http://www.sph.umich.edu/csg/abecasis/metal/) and StatsDirect (V.2.6.6 StatsDirect Ltd) in the case of haplotypes. Significant heterogeneity among populations was found in the meta-analysis of *RHOB* locus polymorphisms; consequently, in this case a random effects model was applied using StatsDirect. Genotypic frequency distributions for the meta-analysis were kindly provided by Allanore *et al*[10] for the meta-analysis by personal communication. All cohorts were in Hardy–Weinberg equilibrium at a significance level of 0.01 for all the included SNP. Power was calculated using the software Power Calculator for Genetic Studies 2006 and assuming an additive model at the 5% significance level and previously reported OR (rs342070 minor allele (A) frequency (MAF) 0.226, OR 1.20; rs13021401 MAF 0.225, OR 1.21; rs2233287 MAF 0.096, OR 1.31; rs4958881 MAF 0.115, OR 1.29; rs3792783 MAF 0.152, OR 1.29; rs3130573 MAF 0.321, OR 1.25).[15]

RESULTS

Non-HLA loci analysis

In this study we analysed five SNP located in two non-HLA loci, *TNIP1* and *RHOB*. Regarding the *TNIP1* locus, we replicated the previously described associations, with rs4958881 showing the most significant relationship, $p_{MH}=3.26\times10^{-5}$, OR 1.19, 95% CI 1.09 to 1.29 (table 1). We also observed that the associations were consistent through the different clinical and serological subsets. The Breslow–Day test showed no evidence of interpopulation heterogeneity either in the whole disease analyses or in the stratified groups (supplementary table S3, available online only, shows individual cohort analyses). Statistical power was over 99% for the three SNP. Moreover, all the *TNIP1* genetic variants showed significant association at GWAS level in the meta-analysis with the initial report (rs2233287 $p_{MH}=1.7\times10^{-9}$, OR 1.23 95%, CI 1.15 to 1.32; rs4958881 $p_{MH}=2.88\times10^{-11}$, OR 1.24, 95% CI 1.16 to 1.32; rs3792783 $p_{MH}=9.11\times10^{-16}$, OR 1.31, 95% CI 1.23 to 1.40; figure 1).

As previously described, the three SNP examined belong to the same haplotype block.[10] As reported by Allanore *et al*,[10] the polymorphisms studied in the *TNIP1* region showed moderate to high linkage disequilibrium (see supplementary figures S1 and S2, available online only). Therefore, haplotype analysis was performed. The Breslow–Day test show homogeneity in the association of the haplotypes among populations. Haplotype block analysis revealed the association of two haplotypes with the disease. Haplotypes CTT and TCC (SNP order

Table 1 Pooled analysis of the novel SSc non-HLA susceptibility loci

CHR	BP	SNP	Locus	1/2	Subgroup (N)	Genotype, N (%)				Allele test		
						1/1	1/2	2/2	MAF N (%)	p_{MH}*	OR (95% CI)	p_{BD}
2	20548952	rs342070	*RHOB*	C/T	Controls (n=7193)	448 (6.23)	2713 (37.72)	4032 (56.05)	3609 (25.09)			
					SSc (n=4249)	292 (6.87)	1515 (35.66)	2442 (57.47)	2099 (24.70)	0.76	1.01 (0.95 to 1.08)	0.08
					lcSSc (n=2649)	184 (6.95)	953 (35.98)	1512 (57.08)	1321 (24.93)	0.75	1.01 (0.94 to 1.09)	0.24
					dcSSc (n=1227)	80 (6.52)	444 (36.19)	703 (57.29)	604 (24.61)	0.95	1.00 (0.90 to 1.10)	0.21
					ACA+ (n=1492)	92 (6.17)	529 (35.46)	871 (58.38)	713 (23.89)	0.52	0.97 (0.88 to 1.07)	0.52
					ATA+ (n=864)	71 (8.22)	303 (35.07)	490 (56.71)	445 (25.75)	0.22	1.08 (0.96 to 1.21)	0.04
2	20552000	rs13021401	*RHOB*	T/C	Controls (n=6557)	405 (6.18)	2416 (36.85)	3736 (56.98)	3226 (24.60)			
					SSc (n=4259)	280 (6.57)	1524 (35.78)	2455 (57.64)	2084 (24.47)	0.61	1.02 (0.95 to 1.09)	0.06
					lcSSc (n=2644)	174 (6.58)	966 (36.54)	1504 (56.88)	1314 (24.85)	0.49	1.03 (0.95 to 1.11)	0.11
					dcSSc (n=1239)	79 (6.38)	444 (35.84)	716 (57.79)	602 (24.29)	0.96	1.00 (0.90 to 1.10)	0.38
					ACA+ (n=1487)	91 (6.12)	530 (35.64)	866 (58.24)	712 (23.94)	0.91	0.99 (0.90 to 1.09)	0.39
					ATA+ (n=872)	65 (7.45)	316 (36.24)	491 (56.31)	446 (25.57)	0.17	1.09 (0.97 to 1.22)	0.11
5	150420290	rs2233287	*TNIP1*	T/C	Controls (n=7164)	93 (1.30)	1226 (17.11)	5845 (81.59)	1412 (9.85)			
					SSc (n=4152)	52 (1.25)	859 (20.69)	3241 (78.06)	963 (11.60)	1.94E-04	1.19 (1.08 to 1.30)	0.72
					lcSSc (n=2586)	33 (1.28)	528 (20.42)	2025 (78.31)	594 (11.48)	1.69E-03	1.18 (1.07 to 1.31)	0.70
					dcSSc (n=1194)	15 (1.26)	253 (21.19)	926 (77.55)	283 (11.85)	2.81E-03	1.23 (1.07 to 1.41)	0.76
					ACA+ (n=1442)	19 (1.32)	309 (21.43)	1114 (77.25)	347 (12.03)	2.27E-03	1.22 (1.07 to 1.39)	0.86
					ATA+ (n=849)	14 (1.65)	186 (21.91)	649 (76.44)	214 (12.60)	4.94E-03	1.26 (1.07 to 1.47)	0.73
5	150430429	rs4958881	*TNIP1*	C/T	Controls (n=7182)	131 (1.82)	1512 (21.05)	5539 (77.12)	1774 (12.35)			
					SSc (n=4226)	98 (2.32)	1035 (24.49)	3093 (73.19)	1231 (14.56)	3.26E-05	1.19 (1.09 to 1.29)	0.33
					lcSSc (n=2637)	57 (2.16)	637 (24.16)	1943 (73.68)	751 (14.24)	1.23E-03	1.17 (1.06 to 1.28)	0.58
					dcSSc (n=1213)	31 (2.56)	302 (24.90)	880 (72.55)	364 (15.00)	5.57E-04	1.24 (1.10 to 1.41)	0.29
					ACA+ (n=1490)	31 (2.08)	378 (25.37)	1081 (72.55)	440 (14.77)	3.29E-03	1.19 (1.06 to 1.34)	0.14
					ATA+ (n=868)	29 (3.34)	232 (26.73)	607 (69.93)	290 (16.71)	1.31E-05	1.36 (1.19 to 1.57)	0.57
5	150435925	rs3792783	*TNIP1*	C/T	Controls (n=3704)	113 (3.05)	995 (26.86)	2596 (70.09)	1221 (16.48)			
					SSc (n=2704)	108 (3.99)	829 (30.66)	1767 (65.35)	1045 (19.32)	2.16E-04	1.19 (1.09 to 1.31)	0.70
					lcSSc (n=1708)	65 (3.81)	518 (30.33)	1125 (65.87)	648 (18.97)	4.44E-03	1.17 (1.05 to 1.30)	0.56
					dcSSc (n=715)	30 (4.20)	222 (31.05)	463 (64.76)	282 (19.72)	5.41E-03	1.23 (1.06 to 1.42)	0.59
					ACA+ (n=1041)	40 (3.84)	326 (31.32)	675 (64.84)	406 (19.50)	6.09E-03	1.20 (1.05 to 1.36)	0.18
					ATA+ (n=623)	24 (3.85)	208 (33.39)	391 (62.76)	256 (20.55)	3.36E-03	1.26 (1.08 to 1.47)	0.70

Controls are used as reference for all comparisons.
*All p values have been calculated for the allelic model.
ACA+, anti-centromere autoantibody-positive patients; ATA+, anti-topoisomerase autoantibody-positive patients; BP, base pair; CHR, chromosome; CTRL, healthy controls; dcSSc, diffuse cutaneous systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis; MAF, minor allele (A) frequency; p_{MH}, Mantel–Haenszel test under fixed effect; p_{BD}, homogeneity Breslow–Day test; 1/2, minor allele/major allele; SNP, single nucleotide polymorphism.

rs2233287–rs4958881–rs3792783) represent the combinations of the major and minor alleles of each SNP, respectively, and subsequently show a protective or a susceptibility role that is concordant with the individual SNP associations, ie, major alleles are protective while minor alleles are risk variants (see supplementary table S4, available online only). However, haplotype block analysis did not show more significant p values than individual SNP analyses, and no additive or multiplicative effect of the SNP was observed. With the aim of clarifying possible underlying dependence among the SNP, we performed conditional logistic regression analysis. Nevertheless, due to the linkage disequilibrium between the analysed SNP, this approach did not enable us to identify an independent association signal (see supplementary table S5, available online only).

As shown in table 1, none of the tested polymorphisms in *RHOB* showed significant associations with SSc, or any of the examined subgroups. Only weak association signals could be detected in the Italian cohort. The power for the analyses of the rs342070 and rs13021401 *RHOB* genetic variants in the overall cohort was of 99% in both cases. Meta-analyses with the previous report showed significant OR heterogeneity in the Breslow–Day tests and no significant association under a random effects model for both polymorphisms (rs342070 P_{random}=0.19; rs13021401 P_{random}=0.13).

PSORS1C1 analysis

The study of the *PSORS1C1* reported variant, rs3130573, showed a suggestive but heterogeneous association of this polymorphism with increased SSc susceptibility (table 2). Moreover, the association was maintained in all the subgroups (including the ACA-negative and ATA-negative subsets) except for the ACA-positive patients. Considering that the association of the HLA region with SSc is influenced primary by the autoantibody profile of the patients,[5 7] we aimed to test for an uncovered influence of the HLA genes. We thus carried out a step-wise logistic regression conditional analysis of the analysed *PSORS1C1* variant with all the independent signals from the most significant to the lowest observed p values in the HLA region. The considered conditions included SNPs, imputed aminoacidic positions and imputed classic HLA-alleles (as described in Raychaudhuri *et al*)[16] in the analysed GWAS cohorts (Spain I, The Netherlands I and USA I) (unpublished data). Standard logistic regression analyses in the GWAS cohorts showed evidence of association only in the whole disease versus controls and in the ACA-negative patients versus controls comparisons (P_{log}=0.034, OR 1.09; P_{log}=0.01, OR 1.12, respectively). However, the previously mentioned association with the whole set of SSc patients lost its significance when it was conditioned to the HLA-DPB1*1301 allele (p value

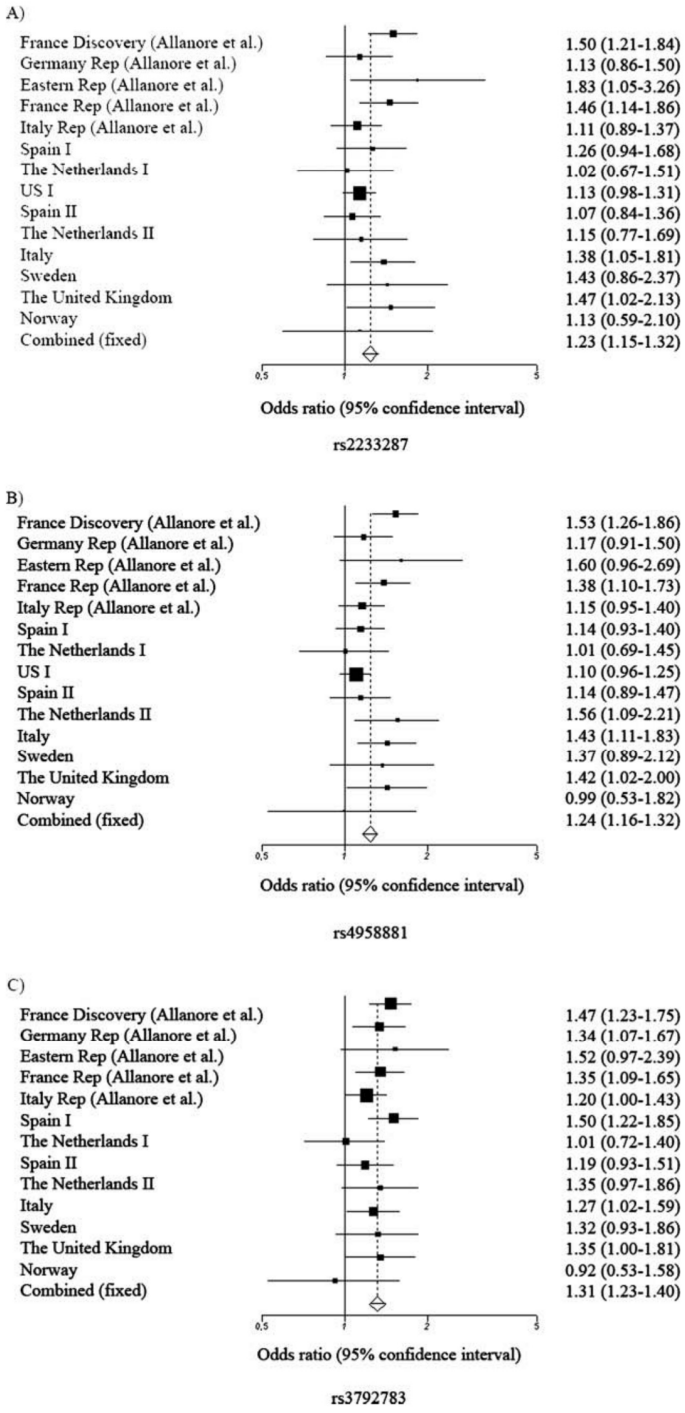


Figure 1 (A) Forest plot for the meta-analysis of the rs2233287 *TNIP1* genetic variant. (B) Forest plot for the meta-analysis of the rs4958881 *TNIP1* genetic variant. (C) Forest plot for the meta-analysis of the rs3792783 *TNIP1* genetic variant.

Table 2 Pooled-analysis of the rs3130573 *PSORS1C1* HLA-region genetic variant

CHR	BP	SNP	1/2	Subgroup (N)	Genotype, N (%)			MAF N (%)	Allele test		
					1/1	1/2	2/2		p_{MH}*	OR (95% CI)	p_{BD}
6	31214247	rs3130573	G/A	Controls (n=7139)	892 (12.49)	3220 (45.10)	3027 (42.40)	5004 (35.05)			
				SSc (n=4130)	574 (13.90)	1953 (47.29)	1603 (38.81)	3101 (37.54)	1.17E-05	1.14 (1.07 to 1.21)	0.02
				lcSSc (n=2575)	350 (13.59)	1221 (47.42)	1004 (38.99)	1921 (37.30)	1.16E-03	1.12 (1.05 to 1.20)	0.07
				dcSSc (n=1187)	177 (14.91)	561 (47.26)	449 (37.83)	915 (38.54)	3.09E-04	1.18 (1.08 to 1.29)	0.07
				ACA+ (n=1446)	181 (12.52)	666 (46.06)	599 (41.42)	1028 (35.55)	0.28	1.05 (0.96 to 1.14)	0.12
				ACA− (n=2511)	356 (15.07)	1143 (48.37)	864 (36.56)	1855 (39.25)	1.01E-07	1.21 (1.13 to 1.29)	0.05
				ATA+ (n=845)	123 (14.56)	411 (48.64)	311 (36.80)	657 (38.88)	6.48E-03	1.16 (1.04 to 1.29)	0.51
				ATA− (n=3147)	413 (13.92)	1395 (47.00)	1160 (39.08)	2221 (37.42)	1.03E-04	1.13 (1.07 to 1.21)	0.03

Controls are used as reference for all comparisons.
*All p values have been calculated for the allelic model.
ACA±, anti-centromere autoantibody-positive/negative patients; ATA±, anti-topoisomerase autoantibody-positive/negative patients; BP, base pair; CHR, chromosome; CTRL, healthy controls; dcSSc, diffuse cutaneous systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis; MAF, minor allele (A) frequency; p_{MH}, Mantel–Haenszel test under fixed effect; p_{BD}, homogeneity Breslow–Day test; 1/2, minor allele/major allele; SNP, single nucleotide polymorphism.

conditioning on DPB1*13 : 01=0.06; OR conditioning on DPB1*13 : 01=1.08), HLA-DRB1*11 : 04 (p value conditioning on DRB1*11 : 04=0.08; OR conditioning on DRB1*11 : 04=1.07) and the HLA-DQA1*05 : 01 (p value conditioning on DQA1*05 : 01=0.36; OR conditioning on DQA1*05 : 01=1.04) alleles. Moreover, the association with the ACA-negative subset of patients was also shown to be dependent on the HLA-DQA1*05 : 01 allele (p value conditioning on DQA1*05 : 01=0.27; OR add to DQA1*05 : 01=1.05). Therefore, our data suggest that the association of *PSORS1C1* with SSc is not independent from the HLA region.

DISCUSSION

In this study we conducted a large multicentre replication of the novel SSc risk variants identified by Allanore *et al*,[10] and we confirmed the association of the *TNIP1* locus with SSc. However, the associations observed in the *RHOB* gene and the independence of *PSORS1C1* from the HLA region were not supported by our data.

Due to the lack of association observed in the *RHOB* locus, we suggest that the initial association reported in this gene might have been a false positive finding. It is worth mentioning that *RHOB* has never been robustly associated with an autoimmune disease, and the previously reported association with SSc in this gene did not reach the GWAS significance level even after replication.[10]

PSORS1C1 was proposed as an HLA-independent SSc risk factor.[10] The authors performed a dependence analysis controlling for the described association in the *HLA-DQB1* gene and reported the independence of both loci. Nevertheless, *HLA-DQB1* has been specifically related with the ACA-positive subset of patients,[5 7] and both in our data and in the previous study no association of *PSORS1C1* with ACA positivity has been shown.[10] Therefore, a deeper analysis of this locus was needed, and we performed for the first time conditional logistic regression including all the independent signals in the HLA region. In our initial approach we found a signal in the *PSORS1C1* gene that was comparable to the one described in the previous work;[10] however, a comprehensive analysis showed that the *PSORS1C1* association is dependent from the HLA-DPB1*13 : 01, HLA-DQA1*05 : 01 and HLA-DRB1*11 : 04 alleles (especially the HLA-DQA1*0501). These HLA loci have previously been described as ATA positivity risk factors,[5 7] which is consistent with the lack of association in the ACA-positive subgroup. Altogether, our data do not confirm *PSORS1C1* as an independent player in the SSc genetic susceptibility network.

Regarding the association of TNFAIP3 interacting protein 1 (*TNIP1*), our data clearly support *TNIP1* as a SSc risk factor. Our replication study confirmed that the association of the three SNPs tested is maintained in all the subsets, indicating that this association peak corresponds to the whole disease. Therefore, *TNIP1* might be implicated in the development of the disease but may not act as a disease modifier. Remarkably, *TNIP1* is involved in HIV replication, acts as a negative regulator of the nuclear factor κB pathway (a key regulator of the immune response, which has also recently been associated with SSc),[9] and also represses agonist-bound retinoic acid receptors and peroxisome proliferator-activated receptors.[17–19] Furthermore, recent studies focused on the control of *TNIP1* transcription have reported a complex mechanism behind *TNIP1* expression that combines constitutive transcription factors and inducible factors (nuclear factor κB and peroxisome proliferator-activated receptors).[20] Interestingly, Allanore *et al*[10] showed that the transcription and expression of *TNIP1* is decreased both in the skin of SSc patients and SSc cultured fibroblasts, thus the anti-inflammatory effect of this molecule may be reduced in SSc patients, providing evidence for a relevant role of *TNIP1* in the disease. *TNIP1* is also a well-established risk factor for different autoimmune diseases, such as psoriasis, psoriatic arthritis and systemic lupus erythematosus.[21–23] Furthermore, *TNIP1* association with psoriasis has also been reported in Asian populations,[24] suggesting that the role of *TNIP1* in autoimmune diseases is consistent through different ethnicities. Therefore, this locus can be considered a common autoimmune disease risk factor that can be used as a new therapeutic target.

To conclude, our replication study has reinforced the influence of *TNIP1* in an increased susceptibility to SSc and its role as a new player in the autoimmunity genetic background. Future research will identify the causal variant for the *TNIP1* association and its implication in SSc pathophysiology.

Author affiliations
[1]Department of Immunology, Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Granada, Spain
[2]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[3]Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona, Spain
[4]Department of Clinical Immunology and Rheumatology, Allergy, IRCCS Fondazione Policlinico-Mangiagalli-Regina Elena and University of Milan, Milan, Italy

[5]Department of Epidemiology, M.D. Anderson Cancer Center, The University of Texas, Houston, Texas, USA
[6]Servicio de Medicina Interna, Hospital Clínico Universitario, Granada, Spain
[7]Servicio de Enfermedades Autoinmunes, Hospital Clinic, Barcelona, Spain
[8]Servicio de Reumatología, Hospital Universitario 12 de Octubre, Madrid, Spain
[9]Servicio de Reumatología, Hospital Madrid Norte Sanchinarro, Madrid, Spain
[10]Servicio de Reumatología, INIBIC-Hospital Universitario A Coruña, La Coruña, Spain
[11]Servicio de Reumatología, Hospital Universitari i Politecnic La Fe, Valencia, Spain
[12]Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla, Spain
[13]Servicio de Reumatología, Hospital Universitario Marqués de Valdecilla, IFIMAV, Santander, Spain
[14]Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital Universitario Miguel Servet, Zaragoza, Spain
[15]Servicio de Reumatología, Hospital Sant Pau, Barcelona, Spain
[16]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[17]Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands
[18]Department of Rheumatology, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[19]Department of Rheumatology, Lund University, Lund, Sweden
[20]Karolinska Institute, Stockholm, Sweden
[21]Department of Medicine, Università degli Studi di Verona, Verona, Italy
[22]Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle, UK
[23]University of Glasgow, Glasgow, UK
[24]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[25]Centre for Rheumatology, Royal Free and University College Medical School, London, UK
[26]Department of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center-Houston, Houston, Texas, USA
[27]Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
[28]Department of Rheumatology and Clinical Immunology, Utrecht Medical Center, Utrech, The Netherlands
*For members of the Spanish Scleroderma group see supplementary appendix, available online only).

Acknowledgements The authors would like to thank Sofia Vargas, Sonia García and Gema Robledo for excellent technical assistance and all the patients and control donors for their essential collaboration. They thank Banco Nacional de ADN (University of Salamanca, Spain) and the Norwegian Bone Marrow Donor Registry who supplied part of the control DNA samples. They are also grateful to EUSTAR (the EULAR Scleroderma Trials and Research group) for the facilitation of this project.

Contributors LBC and JEM contributed to the analysis and interpretation of data and in drafting the article. JB, CPS, LB, OG, MCV, NOC, GE, PC, PGdlP, NO, JARI, MJC, MAGG, LSC, IC, AJS, AEV, AMHV, RH, AN, CL, RS, JMvL, PS, AH, JW, CF, CD, FKT, FCA and SA participated in analysis and interpretation of data and critically revised the manuscript draft. BPK, MDM, TRDJR and JM were involved in the conception and design of the study and critically revised the submitted version of the manuscript. All authors approved the final version to be published.

Funding This work was supported by the following grants: JM was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009–11110 from the Spanish Ministry of Science, CTS-4977 from Junta de Andalucía, Spain, in part by Redes Temáticas de Investigación Cooperativa Sanitaria Programme, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain and by Fondo Europeo de Desarrollo Regional (FEDER). TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). JM and TRDJR were sponsored by the Orphan Disease Programme grant from the European League Against Rheumatism (EULAR). BPCK is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). TW was granted by DFG WI 1031/6.1. This study was also funded by PI-0590-2010, Consejería de Salud, Junta de Andalucía, Spain. The USA studies were supported by NIH/NIAMS Scleroderma Registry and DNA Repository (N01-AR-0-2251), NIH/NIAMS-RO1-AR055258 and NIH/NIAMS Center of Research Translation in Scleroderma (1P50AR054144), and the Department of Defense Congressionally Directed Medical Research Programmes (W81XWH-07-01-0111).

Competing interests None.

Patient consent Obtained.

Ethics approval Local ethics committees from all the participating centres approved the study.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **Gabrielli A,** Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009;**360**:1989–2003.
2. **Martin JE,** Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. *Hum Genet* 2012;**131**:1023–37.
3. **Visscher PM,** Brown MA, McCarthy MI, *et al*. Five years of GWAS discovery. *Am J Hum Genet* 2012;**90**:7–24.
4. **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**:426–9.
5. **Arnett FC,** Gourh P, Shete S, *et al*. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann Rheum Dis* 2010;**69**:822–7.
6. **Dieude P,** Boileau C, Guedj M, *et al*. Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. *Ann Rheum Dis* 2011;**70**:1695–6.
7. **Gorlova O,** Martin JE, Rueda B, *et al*. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet* 2011;**7**:e1002178.
8. **Bossini-Castillo L,** Martin JE, Broen J, *et al*. A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. *Hum Mol Genet* 2012;**21**:926–33.
9. **Martin JE,** Broen JC, Carmona FD, *et al*. Identification of CSK as a systemic sclerosis genetic risk factor through genome wide association study follow-up. *Hum Mol Genet* 2012;**21**:2825–35.
10. **Allanore Y,** Saad M, Dieude P, *et al*. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS Genet* 2011;**7**:e1002091.
11. **Ioannidis JP,** Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. *Nat Rev Genet* 2009;**10**:318–29.
12. **Chanock SJ,** Manolio T, Boehnke M, *et al*. Replicating genotype–phenotype associations. *Nature* 2007;**447**:655–60.
13. **LeRoy EC,** Black C, Fleischmajer R, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;**15**:202–5.
14. **Willer CJ,** Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;**26**:2190–1.
15. **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**:209–13.
16. **Raychaudhuri S,** Sandor C, Stahl EA, *et al*. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;**44**:291–6.
17. **Gupta K,** Ott D, Hope TJ, *et al*. A human nuclear shuttling protein that interacts with human immunodeficiency virus type 1 matrix is packaged into virions. *J Virol* 2000;**74**:11811–24.
18. **Mauro C,** Pacifico F, Lavorgna A, *et al*. ABIN-1 binds to NEMO/IKKgamma and co-operates with A20 in inhibiting NF-kappaB. *J Biol Chem* 2006;**281**:18482–8.
19. **Gurevich I,** Aneskievich BJ. Liganded RARalpha and RARgamma interact with but are repressed by TNIP1. *Biochem Biophys Res Commun* 2009;**389**:409–14.
20. **Gurevich I,** Zhang C, Encarnacao PC, *et al*. PPARgamma and NF-kappaB regulate the gene promoter activity of their shared repressor, TNIP1. *Biochim Biophys Acta* 2012;**1819**:1–15.
21. **Nair RP,** Duffin KC, Helms C, *et al*. Genome-wide scan reveals association of psoriasis with IL-23 and NF-kappaB pathways. *Nat Genet* 2009;**41**:199–204.
22. **Bowes J,** Orozco G, Flynn E, *et al*. Confirmation of TNIP1 and IL23A as susceptibility loci for psoriatic arthritis. *Ann Rheum Dis* 2011;**70**:1641–4.
23. **Gateva V,** Sandling JK, Hom G, *et al*. A large-scale replication study identifies TNIP1, PRDM1, JAZF1, UHRF1BP1 and IL10 as risk loci for systemic lupus erythematosus. *Nat Genet* 2009;**41**:1228–33.
24. **Sun LD,** Cheng H, Wang ZX, *et al*. Association analyses identify six new psoriasis susceptibility loci in the Chinese population. *Nat Genet* 2010;**42**:1005–9.

Pulmonary involvement genetic markers

Bossini-Castillo et al. Arthritis Research & Therapy 2012, 14:R85
http://arthritis-research.com/content/14/2/R85

arthritis research&therapy

RESEARCH ARTICLE Open Access

A multicenter study confirms *CD226* gene association with systemic sclerosis-related pulmonary fibrosis

Lara Bossini-Castillo[1*], Carmen P Simeon[2], Lorenzo Beretta[3], Jasper C Broen[4], Madelon C Vonk[4], Raquel Ríos-Fernández[5], Gerard Espinosa[6], Patricia Carreira[7], María T Camps[8], Maria J Castillo[9], Miguel A González-Gay[10], Emma Beltrán[11], María del Carmen Freire[12], Javier Narváez[13], Carlos Tolosa[14], Torsten Witte[15], Alexander Kreuter[16], Annemie J Schuerwegh[17], Anna-Maria Hoffmann-Vold[18], Roger Hesselstrand[19], Claudio Lunardi[20], Jacob M van Laar[21], Meng May Chee[22], Ariane Herrick[23], Bobby PC Koeleman[24], Christopher P Denton[25], Carmen Fonseca[25], Timothy RDJ Radstake[26†] and Javier Martin[1†], for the Spanish Scleroderma Group

Abstract

Introduction: *CD226* genetic variants have been associated with a number of autoimmune diseases and recently with systemic sclerosis (SSc). The aim of this study was to test the influence of *CD226* loci in SSc susceptibility, clinical phenotypes and autoantibody status in a large multicenter European population.

Methods: A total of seven European populations of Caucasian ancestry were included, comprising 2,131 patients with SSc and 3,966 healthy controls. Three *CD226* single nucleotide polymorphisms (SNPs), rs763361, rs3479968 and rs727088, were genotyped using Taqman 5'allelic discrimination assays.

Results: Pooled analyses showed no evidence of association of the three SNPs, neither with the global disease nor with the analyzed subphenotypes. However, haplotype block analysis revealed a significant association for the TCG haplotype (SNP order: rs763361, rs34794968, rs727088) with lung fibrosis positive patients (P_{Bonf} = 3.18E-02 OR 1.27 (1.05 to 1.54)).

Conclusion: Our data suggest that the tested genetic variants do not individually influence SSc susceptibility but a *CD226* three-variant haplotype is related with genetic predisposition to SSc-related pulmonary fibrosis.

Introduction

Systemic sclerosis (SSc) is a connective tissue disorder in which fibrotic collagen deposition, vascular damage, autoimmunity and autoantibody production (especially anticentromere (ACA), and antitopoisomerase, (ATA) antibodies) are the main hallmarks [1]. SSc patients can be classified classically in two major subgroups, those suffering from limited cutaneous SSc (lcSSc) and those with the diffuse cutaneous form of the disease (dcSSc) [2].

The genetic component of SSc has been recently reinforced and several genes involved in immune regulation have been proposed as risk factors for the development of SSc [3]. A number of loci such as *IRF5* [4], *STAT4* [5,6], *BANK1* [7,8], *C8orf13-BLK* [9,10], *CD247* [11,12] and *TNFSF4* [13,14], have been associated with genetic predisposition to SSc in Caucasian populations. Some of these loci are shared with other autoimmune diseases, such as rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE), reinforcing the theory of a common genetic background in autoimmune diseases [15].

CD226 (cluster of differentiation 226)/PTA1 (platelet and T-cell activation antigen 1)/DNAM-1 (DNAX accessory molecule 1) is a member of the immunoglobulin superfamily and it plays an important role in the co-

* Correspondence: larabc@ipb.csic.es
† Contributed equally
[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Avenida del Conocimiento s/n, Granada, 18100, Spain
Full list of author information is available at the end of the article

stimulation pathways of natural killer (NK) cells and activated T cells [16]. Furthermore, CD226 is constitutively expressed on NK cells, CD4+ and CD8+ T cells, monocytes, platelets and certain B cells playing a pleiotropic role in the immune system [16,17], thus subtle changes in CD226 expression could be involved in SSc immune imbalance.

CD226 gene polymorphisms have been correlated with an increasing number of autoimmune pathologies. Thus, the *CD226* rs763361/Gly307Ser non-synonymous polymorphism was first correlated to type 1 diabetes susceptibility [18,19], later to multiple autoimmune diseases [19,20] and recently, to SSc [21]. Interestingly, the minor allele rs763361*T encodes a non-synonymous mutation (Gly307Ser) in the cytoplasmic tail of CD226 protein (exon 7). In addition, the rs763361 glycine to serine substitution could interfere in the phosphorylation of CD226 at 322Tyr and 329Ser residues, and the downstream signaling pathway may be modulated by these posttranslational modifications [16,22].

In a recent study performed in SLE, a three-variant haplotype in *CD226* gene, comprising *CD226* rs763361-rs34794968-rs727088, was found to be associated with SLE and the authors proposed that rs727088 may be the single nucleotide polymorphism (SNP) with a functional influence on *CD226* transcription levels [23].

The aim of this study was to test in a large European population the previously reported association of *CD226* gene rs763361/Gly307Ser with SSc, and to analyze, for the first time, the role of two additional polymorphisms, rs34794968 and rs727088, and the effect of *CD226* rs763361-rs34794968-rs727088 haplotype in SSc.

Materials and methods

Subjects

A total of 2,131 SSc cases and 3,966 controls from seven European Caucasian cohorts (Spain, Germany, the Netherlands, Italy, Sweden, the United Kingdom and Norway) were included in this study. Patients were diagnosed as having SSc using the criteria proposed by the 1980 ACR and/or LeRoy and Medsger criteria [24,25]. In addition, patients were classified as having limited or diffuse SSc as defined by LeRoy *et al.* [26]. The following clinical data was collected for ascertainment of clinical phenotype of all the patients with SSc: age, gender, disease duration and presence of SSc-associated autoantibodies, ATA and ACA. Pulmonary fibrosis was diagnosed by High Resolution Computed Tomography (HRCT). Considering the previously reported subphenotype associations, the subtype, autoantibody status and pulmonary fibrosis data were available for all the patients included in this report (Table S1 in Additional file 1 shows these clinical data). The control population consisted of unrelated healthy individuals recruited in the same geographical regions as SSc patients and matched by age, sex and ethnicity with the SSc patients groups. Local ethical committees from all the participating centers approved the study. Both patients and controls were included in the study after written informed consent.

CD226 genotyping and statistical analysis

Three SNPs, rs763361 and rs3479968 located in exon 7, and rs727088 in the 3'UTR region were selected as genetic markers. The SNPs were analyzed by Taqman SNP genotyping assays in a 7900HT real-time polymerase chain reaction (PCR) System following the manufacturer's suggestions (Applied Biosystems, Foster City, CA, USA).

All cohorts were in Hardy-Weinberg equilibrium (HWE) at significance level = 0.01 for all the included SNPs. PLINK (v1.07) software [27] was used for individual population association tests (significance was calculated by 2 × 2 contingency tables and Fisher's exact test or χ^2 when necessary) and pooled analysis. Bonferroni and Benjamini and Hochberg false discovery rate method correction (FDR) were applied for multiple test correction [28]. In addition, the Breslow-Day test (BD test) was performed as implemented in PLINK to assess the homogeneity of the association among populations. Haplotype pooled analysis was performed by Haploview (Cambridge, MA, USA) and Statsdirect (Altrincham, Cheshire, UK) software. Power was calculated using the software Power Calculator for Genetic Studies 2006 and assuming an additive model [29].

Results

CD226 rs763361/Gly307Ser, rs34794968 and rs727088 analysis

Table 1 shows the genotype and allelic frequencies and pooled analyses of the three *CD226* SNPs included in this report in the global disease and the considered subgroups. Tables S2 to S4 in Additional file 1 show the genotype and allele distribution of each of the tested variants in the seven analyzed European cohorts. The BD test revealed no statistically significant differences between the seven cohorts included, hence we performed a pooled analysis using the Mantel-Haenszel test under fixed effects for each of the tested polymorphisms. As opposed to Dieudé *et al.* [21], the pooled analysis of rs763361 showed no evidence of association with the whole set of SSc patients. Then, we interrogated the major SSc subphenotypes as defined by LeRoy *et al.* [26], and in addition by autoantibody status and by lung fibrosis as described in Dieudé *et al.*[21]. Allele frequencies in each subgroup were compared to control frequencies and no evidence of association was found at any of the considered subgroups. In addition, we

Table 1 Genotype and allele distribution of *CD226* rs763361 (chr:18, 65,682,622 bp), rs34794968 (chr:18; 65,682,006 bp), rs727088 (chr:18, 65,681,419 bp) genetic variants and pooled analysis.

		Genotype, N (%)				Allele test			
SNP	Subgroup (N)	1/1	1/2	2/2	MAF (%)	P_{MH}	P_{FDR_BH}	OR [CI 95%]	P_{BD}
rs763361	Controls (n = 3811)	906 (23.77)	1841 (48.31)	1064 (27.92)	47.93				
	SSc (n = 2023)	480 (23.73)	990 (48.94)	553 (27.34)	48.2	0.56	0.73	1.02 [0.95-1.10]	0.56
	lcSSc (n = 1397)	332 (23.77)	681 (48.75)	384 (27.49)	48.14	0.64	0.95	1.02 [0.94-1.11]	0.8
	dcSSc (n = 626)	148 (23.64)	309 (49.36)	169 (27.00)	48.32	0.6	0.94	1.03 [0.92-1.17]	0.14
	ACA+ (n = 797)	176 (22.08)	396 (49.69)	225 (28.23)	46.93	0.68	0.68	0.98 [0.88-1.09]	0.85
	ATA+ (n = 503)	133 (26.44)	239 (47.51)	131 (26.04)	50.2	0.22	0.28	1.09 [0.95-1.24]	0.63
	Fib+ (n = 729)	176 (24.14)	359 (49.25)	194 (26.61)	48.77	0.48	0.58	1.04 [0.93-1.17]	0.5
rs34794968	Controls (n = 3858)	669 (17.34)	1842 (47.74)	1347 (34.91)	41.21				
	SSc (n = 2060)	348 (16.89)	978 (47.48)	734 (35.63)	40.63	0.73	0.73	0.99 [0.91-1.07]	0.37
	lcSSc (n = 1422)	234 (16.46)	685 (48.17)	503 (35.37)	40.54	0.74	0.95	0.99 [0.90-1.08]	0.71
	dcSSc (n = 638)	114 (17.87)	293 (45.92)	231 (36.21)	40.83	0.94	0.94	1.00 [0.88-1.12]	0.05
	ACA+ (n = 816)	129 (15.81)	390 (47.79)	297 (36.40)	39.71	0.35	0.68	0.95 [0.85-1.06]	0.6
	ATA+ (n = 518)	100 (19.31)	249 (48.07)	169 (32.63)	43.34	0.28	0.28	1.08 [0.94-1.23]	0.38
	Fib+ (n = 755)	122 (16.16)	362 (47.95)	271 (35.89)	40.13	0.58	0.58	0.97 [0.87-1.09]	0.44
rs727088	Controls (n = 3815)	917 (24.04)	1869 (48.99)	1029 (26.97)	48.53				
	SSc (n = 2042)	489 (23.95)	1014 (49.66)	539 (26.40)	48.78	0.59	0.73	1.02 [0.95-1.10]	0.86
	lcSSc (n = 1409)	336 (23.85)	702 (49.82)	371 (26.33)	48.76	0.63	0.95	1.02 [0.94-1.11]	0.66
	dcSSc (n = 633)	153 (24.17)	312 (49.29)	168 (26.54)	48.82	0.71	0.94	1.02 [0.91-1.15]	0.26
	ACA+ (n = 814)	178 (21.87)	414 (50.86)	222 (27.27)	47.3	0.54	0.68	0.97 [0.87-1.08]	0.56
	ATA+ (n = 514)	138 (26.85)	250 (48.64)	126 (24.51)	51.17	0.14	0.28	1.10 [0.97-1.26]	0.61
	Fib+ (n = 739)	185 (25.03)	363 (49.12)	191 (25.85)	49.59	0.4	0.58	1.05 [0.94-1.17]	0.86

Controls are used as reference for all comparisons. ACA, anti-centromere antibodies; ATA, anti-topoisomerase antibodies; 95% CI, 95% confidence interval; dcSSc, diffuse cutaneous systemic sclerosis; Fib+, lung fibrosis positive SSc patients; lcSSc, limited cutaneous systemic sclerosis; MAF, minor allele frequency; NS, not significant; OR, odds ratio; P_{BD}, Breslow-Day homogeneity test *p*-value; P_{FDR_BH}, corrected *P*-value using Benjamini and Hochberg false discovery rate; P_{MH}, allelic Mantel-Haenszel fixed effects model *P*-value; SSc, systemic sclerosis.

observed no significant association in the pooled analyses of the *CD226* rs34794968 and rs727088 genetic variants with neither the global disease nor with the considered phenotypic subgroups.

CD226 haplotype block analysis

Considering the *CD226* haplotype block association described in SLE [23], we analyzed the possible effect of this haplotype block in SSc patients. In this analysis, considering only haplotypes with frequency > 5% in the pooled analysis, we reported no significant association with SSc susceptibility. However, stratification revealed that the TCG haplotype (SNP order: rs763361-rs34794968- rs727088) was over-represented in the lung fibrosis positive subgroup of patients. This haplotype, composed of the minor alleles of rs763361 and rs727088 polymorphisms and the major allele of rs34794968, was not equivalent to that associated with SLE susceptibility, but showed a modest risk effect in the lung fibrosis positive group of patients. This association with lung fibrosis remained significant even after performing Bonferroni multiple test correction, (P_{Bonf} = 3.18E-02 OR 1.27 (1.05 to1.54), Table 2). Remarkably, a trend of association for the TCG haplotype was observed in the comparison between the lung fibrosis positive subset of patients and the fibrosis negative group of patients (Table 2). We consider that the lack of statistical significance may be possibly due to a reduced statistical power, as the lung fibrosis negative subgroup (n = 1,572) is smaller than the control group (n = 3,966).

Discussion

We carried out a well-powered case-control study aiming to test the contribution of three *CD226* genetic variants (rs763361, rs34794968 and rs727088) to SSc susceptibility and clinical phenotypes. We report the association of a *CD226* three-variant haplotype with SSc-related pulmonary fibrosis. However, we observed a lack of individual association of these three *CD226* polymorphisms with SSc or with its serological and clinical manifestations.

The allele and genotype frequencies observed in our control group were similar to those described in the HapMap public database phase 3 (CEPH: Utah residents with ancestry from northern and western Europe; abbreviation: CEU), and considerably comparable with those in Dieudé *et al.* report [21]. Although the present study did not confirm the previously reported association of

Table 2 Association of *CD226* haplotype block in the Fib+ subset of systemic sclerosis patients.

rs763361	rs34794968	rs727088	Subgroup	MAF (%)	P_{MH}	P_{Bonf}	OR [CI 95%]	P_{BD}
			Fib+	0.40				
T	A	G	CTRL	0.41	0.59	NS	0.97 [0.86-1.08]	0.45
			Fib-	0.41	0.76	NS	0.98 [0.86-1.11]	0.77
			Fib+	0.10				
T	C	G	CTRL	0.08	**1.59E-02**	**3.18E-02**	1.27 [1.05-1.54]	0.48
			Fib-	0.08	0.069	NS	1.24 [0.99-1.56]	0.51
			Fib+	0.50				
C	C	A	CTRL	0.51	0.44	NS	0.96 [0.86-1.07]	0.71
			Fib-	0.51	0.50	NS	0.95 [0.84-1.09]	0.44

For the haplotype-specific analyses, the odds ratio (OR) with 95% confidence interval (95% CI) was determined for each allele variant in the haplotype tested against all of the others pooled together using the Mantel-Haenszel test under fixed effects model, considering no single reference haplotype. CTRL, healthy controls; Fib-, lung fibrosis negative SSc patients; Fib+, lung fibrosis positive SSc patients; MAF, minor allele frequency; P_{Bonf}, corrected *P*-value using Bonferroni multiple test correction.

rs763361 with SSc, we found that one of the haplotypes containing the rs763361*T allele showed evidence of association with the lung fibrosis positive group of patients [21]. We consider that the lack of association for rs763361 individually could have arisen due to a type II error (false negative), because our sample size is similar to the study by Dieudé *et al.* and reaches 99% power to detect the previously reported effect of the polymorphism. It is worth mentioning that Dieudé *et al.* analyzed this polymorphism both in an Italian and German replication cohorts, the latest being the most associated with the FA+ group. Consequently our FA+ German cohort, in spite of being smaller than the one in Dieudé *et al.*, showed the highest trend of association observed in this group (Table S1 in Additional file 1). Given the high power of our pooled analysis to detect similar associations to those reported in the previous SSc study, our data suggest that the previously reported effect for the association between *CD226* rs763361*T allele and SSc susceptibility may be influenced by ethnical factors. Remarkably, Dieudé *et al.* studied the implication of the amino acid change encoded by rs763361*T in CD226 expression, but no significant differences were observed [21]. Our data together with SLE reports [23], suggest that haplotypic allele combinations might be considered for further functional studies.

Tao *et al.* described an increased susceptibility of NKT cells to apoptosis via CD95-CD95L or TCR-CD3 in NKT cells isolated from active SLE patients, and linked it with a deficient expression of CD226 [30]. Later, genetic studies revealed that rs727088 minor allele was part of a haplotype associated with SLE susceptibility and was responsible for a decrease in *CD226* gene expression [23]. Despite experimental evidence of the role of rs727088 in *CD226* transcription modulation in SLE, we detected no association signal of rs727088 or the SLE-risk haplotype with SSc susceptibility. These findings suggest that although *CD226* is a common risk factor for SLE and SSc, the genetic variants in this gene may have potentially divergent roles in both diseases or the causal variant might be disease specific. Noteworthy, it is well-established that SSc and SLE share a number of associated loci, but the associated variants are not necessarily the same, for example in the human leukocyte antigen (HLA) region [31]. Considering that fibrosis is a main hallmark of SSc, the specific association of *CD226* with SSc fibrosis-positive patients might reflect the influence of this locus in diverse pathways. Interestingly, *NLRP1* a common autoimmune disease risk factor has been reported to be associated with the ATA and lung fibrosis-positive subgroups of SSc patients [32]. Moreover, *NLRP1* was reported to contribute to SLE in families suffering SLE and vitiligo or other autoimmune or autoinflammatory diseases [33] and it has been recently found to be associated with SLE in a Brazilian population [34]. Nevertheless, rs2670660 and an rs12150220-rs2670660 haplotype were the associated variants in SLE, but only rs8182352 was associated with SSc-related pulmonary fibrosis (although both rs12150220 and rs2670660 were analyzed) [32,34]. Hence, although the population size in the SLE study is limited, this differential association in SLE of a novel SSc fibrosis risk factor is analogous to the reported *CD226* results.

It is well-established that SSc patients exhibit reduced numbers and impaired function of NKT cells [35-37] and the highest CD226 expression levels in healthy donors are found in NKT cells [21,23]. Thus, the impact of the associated *CD226* haplotype reported in this study on the functions of SSc patients' NK and NKT cells should be further explored. Furthermore, new therapeutic approaches based on anti-CD226 mAb treatment have already been tested in autoimmunity animal models [38], and CD226 has been recently implied in novel T cell activation pathways [39] and NK-driven tissue injury in SLE patients [40]. Hence, the implication

of CD226 in cell-mediated cytotoxicity should be considered. In addition, the possibility of lung fibrosis being a marker of generalized internal organ damage, which was not analyzed in this study and represented a limitation of our findings, should be interrogated.

To date just a few loci have been reported to be associated with SSc-related pulmonary involvement, such as *IRF5* [6], *STAT4* [6], *TNFAIP3* [41], *KCNA5* [42], *NLRP1* [32] and *HGF* [43]. Hence, we consider that the confirmation of *CD226* as a pulmonary involvement marker might be valuable in the deciphering of the mechanisms that underlie the lung fibrosis process in SSc patients.

Conclusions

Our data suggest that previously autoimmune-associated *CD226* gene polymorphisms play a role in the SSc pulmonary fibrosis events in European Caucasian populations, and confirm *CD226* as an important shared autoimmune factor in SSc.

Additional material

Additional file 1: Population specific characteristics and genotype and allelic distribution of the three analyzed variants in each population. This file contains: Table S1 showing the population specific composition of the complete SSc set of patients for the analyzed features of the disease; Tables S2 to 4 showing the genotype and allele distributions of *CD226* rs763361, rs34794968 and rs727088 genetic variants in seven European cohorts.

Abbreviations

ACA: anti-centromere autoantibodies; ATA: anti-topoisomerase antibodies; BD test: Breslow-Day test; CD226: cluster of differentiation 226; CEPH/CEU: Utah residents with ancestry from northern and western Europe; dcSSc: diffuse cutaneous systemic sclerosis; DNAM-1: DNAX accessory molecule 1; FDR: false discovery rate method correction; HLA: human leukocyte antigen; HRCT: high resolution computed tomography; HWE: Hardy-Weinberg equilibrium; lcSSc: limited cutaneous systemic sclerosis; PCR: polymerase chain reaction; PTA1: platelet and T-cell activation antigen 1; RA: rheumatoid arthritis; SLE: systemic lupus erythematosus; SNP: single nucleotide polymorphism; SSc: systemic sclerosis.

Acknowledgements

We thank Sofia Vargas, Sonia García and Gema Robledo for their excellent technical assistance and all the patients and control donors for their essential collaboration. We thank Banco Nacional de ADN (University of Salamanca, Spain) and the Norwegian Bone Marrow Donor Registry, who supplied part of the control DNA samples. We are also thankful to EUSTAR (the EULAR Scleroderma Trials and Research group) and the German Network of Systemic Sclerosis for the facilitation of this project.
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 from Junta de Andalucía, Spain, in part by Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain and by Fondo Europeo de Desarrollo Regional (FEDER). TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). JM and TRDJR were sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). BPCK is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). TW was granted by DFG WI 1031/6.1. This study was also funded by PI-0590-2010, Consejería de Salud, Junta de Andalucía, Spain.
The Spanish Scleroderma Group: Norberto Ortego-Centeno and Jose Luis Callejas, Unidad de Enfermedades Sistémicas Autoinmunes, Servicio de Medicina Interna, Hospital Clínico Universitario San Cecilio, Granada; Nuria Navarrete, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Rosa García Portales, Servicio de Reumatología, Hospital Virgen de la Victoria, Málaga; Antonio Fernández-Nebro, Servicio de Reumatología, Hospital Carlos Haya, Málaga; María F. González-Escribano, Servicio de Inmunología, Hospital Virgen del Rocío, Sevilla; Julio Sánchez-Román and Francisco José García-Hernández, Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Servicio de Reumatología, Hospital Reina Sofía, Córdoba; Benjamín Fernández-Gutiérrez and Luis Rodríguez-Rodríguez, Servicio de Reumatología, Hospital Clínico San Carlos, Madrid; José Luis Andreu and Mónica Fernández de Castro, Servicio de Reumatología, Hospital Puerta del Hierro, Madrid; Paloma García de la Peña, Servicio de Reumatología, Hospital Madrid Norte Sanchinarro, Madrid; Francisco Javier López-Longo and Lina Martínez, Servicio de Reumatología, Hospital General Universitario Gregorio Marañón, Madrid; Vicente Fonollosa, Servicio de Medicina Interna, Hospital Valle de Hebrón, Barcelona; Iván Castellví, Servicio de Reumatología, Hospital Sant Pau, Barcelona; Anna Pros, Servicio de Reumatología, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Servicio de Medicina Interna, Hospital Universitari Mútua Terrasa, Barcelona; Bernardino Díaz, Luis Trapiella and María Gallego, Servicio de Medicina Interna, Hospital Central de Asturias, Oviedo; Inés Vaqueiro, Unidad de Trombosis y Vasculitis, Servicio de Medicina Interna, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; María Victoria Egurbide, Servicio de Medicina Interna, Hospital de Cruces, Barakaldo; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Servicio de Reumatología, Hospital Universitario de Canarias, Tenerife; José Andrés Román-Ivorra, Servicio de Reumatología, Hospital Universitari i Politecnic La Fe, Valencia.

Author details

[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Avenida del Conocimiento s/n, Granada, 18100, Spain. [2]Servicio de Medicina Interna, Hospital Valle de Hebron, Passeig de la Vall d'Hebron, 119-129, Barcelona, 08035, Spain. [3]IRCCS Fondazione Policlinico-Mangiagalli-Regina Elena and University of Milan, Allergy, Clinical Immunology and Rheumatology, Via Francesco Sforza 28, Milan, 20122, Italy. [4]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Geert Grooteplein Zuid 10, Nijmegen, 6525 GA, The Netherlands. [5]Servicio de Medicina Interna, Hospital Clínico Universitario, Avenida Doctor Olóriz 16, Granada, 18012, Spain. [6]Servicio de Enfermedades Autoinmunes, Hospital Clinic, C/Villarroel, 170, Barcelona, 08036, Spain. [7]Servicio de Reumatología, Hospital 12 de Octubre, Avenida de Córdoba, s/n, Madrid, 28041, Spain. [8]Servicio de Medicina Interna, Hospital Carlos Haya, Avenida Carlos Haya s/n, Málaga, 29010, Spain. [9]Servicio de Medicina Interna, Hospital Virgen del Rocío, Avenida Manuel Siurot s/n, Sevilla, 41013, Spain. [10]Servicio de Reumatología, Hospital Universitario Marqués de Valdecilla, IFIMAV, Avenida Valdecilla 25, Santander, 39008, Spain. [11]Servicio de Reumatología, Hospital del Doctor Peset Aleixandre, Avenida Gaspar Aguilar 90, Valencia, 46017, Spain. [12]Unidad Trombosis y Vasculitis, Servicio de Medicina Interna, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Rua Pizarro 22, Vigo, 36204, Spain. [13]Servicio de Reumatología, Hospital Universitario de Bellvitge, C/Feixa Llarga s/n, 08907, Barcelona, Spain. [14]Servicio de Medicina Interna, Hospital Parc Taulí, Parc del Taulí s/n, Sabadell, 08208, Spain. [15]Hannover Medical School, Carl-Neuberg-Straße 1, Hannover, 30625, Germany. [16]Ruhr University of Bochum, Universitätsstraße 150, Bochum, 44801, Germany. [17]Department of Rheumatology, Leiden University Medical Center, Albinusdreef 2, Leiden, 2333 ZA, The Netherlands. [18]Department of Rheumatology, Rikshospitalet, Oslo University Hospital, Rikshospitalet-Radiumhospitalet Medical Center, Oslo, 0027, Norway. [19]Department of Rheumatology, Lund University, Paradisgatan 2, Lund, SE-221 00, Sweden. [20]Department of Medicine, Università degli Studi di Verona, Via dell'Artigliere, 19, Verona 37129, Italy. [21]Institute of Cellular Medicine, Newcastle University, Framlington Place, Newcastle upon Tyne, Newcastle, NE2 4HH, UK. [22]Centre for Rheumatic Diseases, Glasgow Royal Infirmary, 84 Castle Street, Glasgow G4 0SF, UK.

[23]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Stopford Building, Oxford Road, Manchester, M13 9PT, UK. [24]Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, Utrecht, 3584 CG, The Netherlands. [25]Centre for Rheumatology, Royal Free and University College Medical School, University College London, Royal Free Campus, Rowland Hill Street, London, NW3 PF, UK. [26]Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Heidelberglaan 100, Utrecht, 3584 CX, The Netherlands.

Authors' contributions

LBC has contributed to the analysis and interpretation of data and to the drafting the manuscript. CPS, LB and JCB have participated in the acquisition of data and the drafting of the manuscript. CF, TRDJR and JM contributed to the conception and design of the study and have critically revised the manuscript. MCV, RRF, GE, PC, MTC, MJC, MAGG, EB, MCF, JN, CT, TW, AK, AJS, AMHV, RH, CL, JMvL, MMC, AH, CPD, BPK have been involved in the acquisition of data and the revision of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 9 January 2012 Revised: 27 March 2012
Accepted: 24 April 2012 Published: 24 April 2012

References

1. Gabrielli A, Avvedimento EV, Krieg T: **Scleroderma.** *N Engl J Med* 2009, **360**:1989-2003.
2. Steen VD: **The many faces of scleroderma.** *Rheum Dis Clin North Am* 2008, **34**:1-15, v.
3. Martin J, Fonseca C: **The genetics of scleroderma.** *Curr Rheumatol Rep* 2011, **13**:13-20.
4. Dieude P, Guedj M, Wipff J, Avouac J, Fajardy I, Diot E, Granel B, Sibilia J, Cabane J, Mouthon L, Cracowski JL, Carpentier PH, Hachulla E, Meyer O, Kahan A, Boileau C, Allanore Y: **Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis.** *Arthritis Rheum* 2009, **60**:225-233.
5. Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suarez H, Ortego-Centeno N, Riemekasten G, Fonollosa V, Vonk MC, van den Hoogen FH, Sanchez-Roman J, Aguirre-Zamorano MA, Garcia-Portales R, Pros A, Camps MT, Gonzalez-Gay MA, Coenen MJ, Airo P, Beretta L, Scorza R, van Laar J, Gonzalez-Escribano MF, Nelson JL, Radstake TR, Martin J: **The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype.** *Hum Mol Genet* 2009, **18**:2071-2077.
6. Dieude P, Guedj M, Wipff J, Ruiz B, Hachulla E, Diot E, Granel B, Sibilia J, Tiev K, Mouthon L, Cracowski JL, Carpentier PH, Amoura Z, Fajardy I, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis.** *Arthritis Rheum* 2009, **60**:2472-2479.
7. Dieude P, Wipff J, Guedj M, Ruiz B, Melchers I, Hachulla E, Riemekasten G, Diot E, Hunzelmann N, Sibilia J, Tiev K, Mouthon L, Cracowski JL, Carpentier PH, Distler J, Amoura Z, Tarner I, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4.** *Arthritis Rheum* 2009, **60**:3447-3454.
8. Rueda B, Gourh P, Broen J, Agarwal SK, Simeon C, Ortego-Centeno N, Vonk MC, Coenen M, Riemekasten G, Hunzelmann N, Hesselstrand R, Tan FK, Reveille JD, Assassi S, Garcia-Hernandez FJ, Carreira P, Camps M, Fernandez-Nebro A, Garcia de la Pena P, Nearney T, Hilda D, Gonzalez-Gay MA, Airo P, Beretta L, Scorza R, Radstake TR, Mayes MD, Arnett FC, Martin J: **BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians.** *Ann Rheum Dis* 2010, **69**:700-705.
9. Gourh P, Agarwal SK, Martin E, Divecha D, Rueda B, Bunting H, Assassi S, Paz G, Shete S, McNearney T, Draeger H, Reveille JD, Radstake TR, Simeon CP, Rodriguez L, Vicente E, Gonzalez-Gay MA, Mayes MD, Tan FK, Martin J, Arnett FC: **Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations.** *J Autoimmun* 2010, **34**:155-162.
10. Coustet B, Dieude P, Guedj M, Bouaziz M, Avouac J, Ruiz B, Hachulla E, Diot E, Kracowski JL, Tiev K, Sibilia J, Mouthon L, Frances C, Amoura Z, Carpentier P, Cosnes A, Meyer O, Kahan A, Boileau C, Chiocchia G, Allanore Y: **C8orf 13/BLK is a genetic risk locus for systemic sclerosis and has additive effect with BANK1: Results from a large French cohort and meta-analysis.** *Arthritis Rheum* 2011, **63**:2091-2096.
11. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, Broen JC, van Riel PL, van 't Slot R, Italiaander A, Ophoff RA, Riemekasten G, Hunzelmann N, Simeon CP, Ortego-Centeno N, Gonzalez-Gay MA, Gonzalez-Escribano MF, Airo P, van Laar J, Herrick A, Worthington J, Hesselstrand R, Smith V, de Keyser F, Houssiau F, Chee MM, *et al*: **Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus.** *Nat Genet* 2010, **42**:426-429.
12. Dieude P, Boileau C, Guedj M, Avouac J, Ruiz B, Hachulla E, Diot E, Cracowski JL, Tiev K, Sibilia J, Mouthon L, Frances C, Amoura Z, Carpentier P, Cosnes A, Meyer O, Kahan A, Chiocchia G, Allanore Y: **Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor.** *Ann Rheum Dis* 2011, **70**:1695-1696.
13. Gourh P, Arnett FC, Tan FK, Assassi S, Divecha D, Paz G, McNearney T, Draeger H, Reveille JD, Mayes MD, Agarwal SK: **Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis.** *Ann Rheum Dis* 2010, **69**:550-555.
14. Bossini-Castillo L, Broen JC, Simeon CP, Beretta L, Vonk MC, Ortego-Centeno N, Espinosa G, Carreira P, Camps MT, Navarrete N, Gonzalez-Escribano MF, Vicente-Rabaneda E, Rodriguez L, Tolosa C, Roman-Ivorra JA, Gomez-Gracia I, Garcia-Hernandez FJ, Castellvi I, Gallego M, Fernandez-Nebro A, Garcia-Portales R, Egurbide MV, Fonollosa V, de la Pena PG, Pros A, Gonzalez-Gay MA, Hesselstrand R, Riemekasten G, Witte T, Coenen MJ, *et al*: **A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort.** *Ann Rheum Dis* 2011, **70**:638-641.
15. Zhernakova A, van Diemen CC, Wijmenga C: **Detecting shared pathogenesis from the shared genetics of immune-related diseases.** *Nat Rev Genet* 2009, **10**:43-55.
16. Xu Z, Jin B: **A novel interface consisting of homologous immunoglobulin superfamily members with multiple functions.** *Cell Mol Immunol* 2010, **7**:11-19.
17. Maier LM, Hafler DA: **Autoimmunity risk alleles in costimulation pathways.** *Immunol Rev* 2009, **229**:322-336.
18. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, *et al*: **Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.** *Nat Genet* 2007, **39**:857-864.
19. Hafler JP, Maier LM, Cooper JD, Plagnol V, Hinks A, Simmonds MJ, Stevens HE, Walker NM, Healy B, Howson JM, Maisuria M, Duley S, Coleman G, Gough SC, Worthington J, Kuchroo VK, Wicker LS, Todd JA: **CD226 Gly307Ser association with multiple autoimmune diseases.** *Genes Immun* 2009, **10**:5-10.
20. Maiti AK, Kim-Howard X, Viswanathan P, Guillen L, Qian X, Rojas-Villarraga A, Sun C, Canas C, Tobon GJ, Matsuda K, Shen N, Chernavsky AC, Anaya JM, Nath SK: **Non-synonymous variant (Gly307Ser) in CD226 is associated with susceptibility to multiple autoimmune diseases.** *Rheumatology (Oxford)* 2010, **49**:1239-1244.
21. Dieude P, Guedj M, Truchetet ME, Wipff J, Revillod L, Riemekasten G, Matucci-Cerinic M, Melchers I, Hachulla E, Airo P, Diot E, Hunzelmann N, Mouthon L, Cabane J, Cracowski JL, Riccieri V, Distler J, Amoura Z, Valentini G, Camaraschi P, Tarner I, Frances C, Carpentier P, Brembilla NC, Meyer O, Kahan A, Chizzolini C, Boileau C, Allanore Y: **Association of the CD226 307Ser variant with systemic sclerosis: Evidence for a contribution of co-stimulation pathways in SSc pathogenesis.** *Arthritis Rheum* 2011, **63**:1097-1105.
22. Shibuya K, Lanier LL, Phillips JH, Ochs HD, Shimizu K, Nakayama E, Nakauchi H, Shibuya A: **Physical and functional association of LFA-1 with DNAM-1 adhesion molecule.** *Immunity* 1999, **11**:615-623.
23. Lofgren SE, Delgado-Vega AM, Gallant CJ, Sanchez E, Frostegard J, Truedsson L, de Ramon Garrido E, Sabio JM, Gonzalez-Escribano MF, Pons-Estel BA, D'Alfonso S, Witte T, Lauwerys BR, Endreffy E, Kovacs L,

Vasconcelos C, Martins da Silva B, Martin J, Alarcon-Riquelme ME, Kozyrev SV: **A 3'-untranslated region variant is associated with impaired expression of CD226 in T and natural killer T cells and is associated with susceptibility to systemic lupus erythematosus.** *Arthritis Rheum* 2010, **62**:3404-3414.

24. **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.
25. LeRoy EC, Medsger TA Jr: **Criteria for the classification of early systemic sclerosis.** *J Rheumatol* 2001, **28**:1573-1576.
26. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, Rowell N, Wollheim F: **Scleroderma (systemic sclerosis): classification, subsets and pathogenesis.** *J Rheumatol* 1988, **15**:202-205.
27. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559-575.
28. 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.
29. Skol AD, Scott LJ, Abecasis GR, Boehnke M: **Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.** *Nat Genet* 2006, **38**:209-213.
30. Tao D, Shangwu L, Qun W, Yan L, Wei J, Junyan L, Feili G, Boquan J, Jinquan T: **CD226 expression deficiency causes high sensitivity to apoptosis in NK T cells from patients with systemic lupus erythematosus.** *J Immunol* 2005, **174**:1281-1290.
31. Martin JE, Bossini-Castillo L, Martin J: **Unraveling the genetic component of systemic sclerosis.** *Hum Genet* 2012.
32. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Airo P, Melchers I, Hachulla E, Cerinic MM, Diot E, Hunzelmann N, Caramaschi P, Sibilia J, Tiev K, Mouthon L, Riccieri V, Cracowski JL, Carpentier PH, Distler J, Amoura Z, Tarner I, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis.** *Ann Rheum Dis* 2011, **70**:668-674.
33. Jin Y, Mailloux CM, Gowan K, Riccardi SL, LaBerge G, Bennett DC, Fain PR, Spritz RA: **NALP1 in vitiligo-associated multiple autoimmune disease.** *N Engl J Med* 2007, **356**:1216-1225.
34. Pontillo A, Girardelli M, Kamada AJ, Pancotto JA, Donadi EA, Crovella S, Sandrin-Garcia P: **Polimorphisms in Inflammasome Genes Are Involved in the Predisposition to Systemic Lupus Erythematosus.** *Autoimmunity* 2012.
35. Riccieri V, Parisi G, Spadaro A, Scrivo R, Barone F, Moretti T, Bernardini G, Strom R, Taccari E, Valesini G: **Reduced circulating natural killer T cells and gamma/delta T cells in patients with systemic sclerosis.** *J Rheumatol* 2005, **32**:283-286.
36. Kojo S, Adachi Y, Keino H, Taniguchi M, Sumida T: **Dysfunction of T cell receptor AV24AJ18+, BV11+ double-negative regulatory natural killer T cells in autoimmune diseases.** *Arthritis Rheum* 2001, **44**:1127-1138.
37. Mitsuo A, Morimoto S, Nakiri Y, Suzuki J, Kaneko H, Tokano Y, Tsuda H, Takasaki Y, Hashimoto H: **Decreased CD161+CD8+ T cells in the peripheral blood of patients suffering from rheumatic diseases.** *Rheumatology (Oxford)* 2006, **45**:1477-1484.
38. Dardalhon V, Schubart AS, Reddy J, Meyers JH, Monney L, Sabatos CA, Ahuja R, Nguyen K, Freeman GJ, Greenfield EA, Sobel RA, Kuchroo VK: **CD226 is specifically expressed on the surface of Th1 cells and regulates their expansion and effector functions.** *J Immunol* 2005, **175**:1558-1565.
39. Joller N, Hafler JP, Brynedal B, Kassam N, Spoerl S, Levin SD, Sharpe AH, Kuchroo VK: **Cutting edge: TIGIT has T cell-intrinsic inhibitory functions.** *J Immunol* 2011, **186**:1338-1342.
40. Huang Z, Fu B, Zheng SG, Li X, Sun R, Tian Z, Wei H: **Involvement of CD226+ NK cells in immunopathogenesis of systemic lupus erythematosus.** *J Immunol* 2011, **186**:3421-3431.
41. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M, Melchers I, Hachulla E, Airo P, Diot E, Hunzelmann N, Cabane J, Mouthon L, Cracowski JL, Riccieri V, Distler J, Meyer O, Kahan A, Boileau C, Allanore Y: **Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population.** *Ann Rheum Dis* 2010, **69**:1958-1964.
42. Wipff J, Dieude P, Guedj M, Ruiz B, Riemekasten G, Cracowski JL, Matucci-Cerinic M, Melchers I, Humbert M, Hachulla E, Airo P, Diot E, Hunzelmann N, Caramaschi P, Sibilia J, Valentini G, Tiev K, Girerd B, Mouthon L, Riccieri V, Carpentier PH, Distler J, Amoura Z, Tarner I, Degano B, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **Association of a KCNA5 gene polymorphism with systemic sclerosis-associated pulmonary arterial hypertension in the European Caucasian population.** *Arthritis Rheum* 2010, **62**:3093-3100.
43. Hoshino K, Satoh T, Kawaguchi Y, Kuwana M: **Association of hepatocyte growth factor promoter polymorphism with severity of interstitial lung disease in Japanese patients with systemic sclerosis.** *Arthritis Rheum* 2011, **63**:2465-2472.

doi:10.1186/ar3809
Cite this article as: Bossini-Castillo *et al.*: A multicenter study confirms *CD226* gene association with systemic sclerosis-related pulmonary fibrosis. *Arthritis Research & Therapy* 2012 **14**:R85.

Bossini-Castillo et al. Arthritis Research & Therapy 2012, 14:R273
http://arthritis-research.com/content/14/6/R273

arthritis research&therapy

RESEARCH ARTICLE Open Access

KCNA5 gene is not confirmed as a systemic sclerosis-related pulmonary arterial hypertension genetic susceptibility factor

Lara Bossini-Castillo[1*], Carmen P Simeon[2], Lorenzo Beretta[3], Jasper Broen[4,5], Madelon C Vonk[6], José Luis Callejas[7], Patricia Carreira[8], Luis Rodríguez-Rodríguez[9], Rosa García-Portales[10], Miguel A González-Gay[11], Ivan Castellvi[12], María Teresa Camps[13], Carlos Tolosa[14], Esther Vicente-Rabaneda[15] and María Victoria Egurbide[16], for the Spanish Scleroderma Group, Annemie J Schuerwegh[17], Roger Hesselstrand[18], Claudio Lunardi[19], Jacob M van Laar[20], Paul Shiels[21], Ariane Herrick[22], Jane Worthington[22], Christopher Denton[23], Timothy RDJ Radstake[4,5], Carmen Fonseca[23†] and Javier Martin[1†]

Abstract

Introduction: Potassium voltage-gated channel shaker-related subfamily member 5 (KCNA5) is implicated in vascular tone regulation, and its inhibition during hypoxia produces pulmonary vasoconstriction. Recently, a protective association of the *KCNA5 locus* with systemic sclerosis (SSc) patients with pulmonary arterial hypertension (PAH) was reported. Hence, the aim of this study was to replicate these findings in an independent multicenter Caucasian SSc cohort.

Methods: The 2,343 SSc cases (179 PAH positive, confirmed by right-heart catheterization) and 2,690 matched healthy controls from five European countries were included in this study. Rs10744676 single-nucleotide polymorphism (SNP) was genotyped by using a TaqMan SNP genotyping assay.

Results: Individual population analyses of the selected *KCNA5* genetic variant did not show significant association with SSc or any of the defined subsets (for example, limited cutaneous SSc, diffuse cutaneous SSc, anti-centromere autoantibody positive and anti-topoisomerase autoantibody positive). Furthermore, pooled analyses revealed no significant evidence of association with the disease or any of the subsets, not even the PAH-positive group. The comparison of PAH-positive patients with PAH-negative patients showed no significant differences among patients.

Conclusions: Our data do not support an important role of *KCNA5* as an SSc-susceptibility factor or as a PAH-development genetic marker for SSc patients.

Introduction

Systemic sclerosis (SSc) is a life-threatening fibrotic connective tissue disorder that affects the skin and different internal organs [1]. The 10-year survival of SSc patients reaches only 63%, with pulmonary involvement the leading cause of death [2]. SSc is a complex disorder in which the environmental triggers and genetic susceptibility factors co-act in the development and maintenance of the disease [3]. As a clearer picture of the genetic component of this disease is being revealed, interest in genetic markers for specific clinical features, especially lung involvement, is increasing. An SSc phenotype-restricted genome-wide analysis was carried out recently [4]. Remarkably, two of the four new SSc genetic-susceptibility markers identified in the previously mentioned study might play a relevant role in the SSc-related fibrotic process, *SOX5* and *NOTCH4* [4]. However, no such strategy has yet been considered for pulmonary involvement, and only some studies have reported significant genetic association with SSc-related lung involvement [5-9]. Furthermore, only the association of *CD226* with pulmonary fibrosis has been

* Correspondence: larabc@ipb.csic.es
† Contributed equally
[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Avenida del Conocimiento s/n, Granada, 18100, Spain
Full list of author information is available at the end of the article

independently replicated [10]. Wipff *et al.* [11] recently described the association of potassium voltage-gated channel shaker-related subfamily member 5 (*KCNA5*) with SSc-related pulmonary arterial hypertension (PAH).

Potassium voltage-gated channels (Kv channels) are homo- or heterotetramers of structural α-subunits and regulatory β-subunits, which control potassium (K) flux. K balance is known to regulate the apoptotic cell shrinkage, a main event in the apoptotic process [12]. Specifically, KCNA5 participates in pulmonary artery smooth muscle cell (PASMC) apoptosis control [13]. It has been reported that overexpression of the *KCNA5* gene induces accelerated K efflux and increases caspase-3 proteolytic activity, promoting apoptosis [13].

KCNA5 is also involved in membrane-potential maintenance and vascular-tone regulation [14]. Hypoxic conditions produce a specific inhibition of KCNA5 in the PASMCs, causing pulmonary vasoconstriction [15-17]. Moreover, it has been reported that primary pulmonary hypertension patients have an intrinsic reduced level of *KCNA5* mRNA in their PASMCs, and this characteristic might play an important role in the pathogenesis of the disease [18].

The role of *KCNA5* genetic polymorphisms in pulmonary disease also was explored. Remillard *et al.* [19] analyzed the influence of *KCNA5* genetic variants in IPAH, and showed that different single-nucleotide polymorphisms (SNPs) were related to abnormal function or drug responsiveness. Recently, the importance of *KCNA5* variants in SSc-related PAH was analyzed [11]. In the previously mentioned study, the association of *KCNA5* with SSc and specifically with PAH-positive (PAH^+) patients was described [11]. Moreover, the rs10744676 polymorphism was reported as the variant underlying the observed association [11].

In light of the previous evidence, the main goal of this report is to replicate the association of *KCNA5* rs10744676 polymorphism with SSc and SSc-related PAH in an independent European population of Caucasian ancestry.

Materials and methods

Subjects

Our study comprised 2,343 SSc cases and 2,690 controls from Spain, The Netherlands, Italy, Sweden, and The United Kingdom. All the individuals included in this report were of Caucasian ancestry. Patients were classified as having limited (lcSSc) or diffuse SSc (dcSSc), according to their skin involvement, as defined by LeRoy *et al.* [20]. The serologic subgroup stratification of the patients was based on the presence of SSc-associated autoantibodies (anti-centromere antibodies (ACAs) and anti-topoisomerase (ATA)). Pulmonary fibrosis was diagnosed by the presence of interstitial abnormalities in high-resolution computed tomography (HRCT). Patients defined as PAH^+ showed a mean resting pulmonary artery pressure > 25 mm Hg and a pulmonary capillary wedge pressure ≤15 mm Hg, at the time of a right-heart catheterization [11]. The control population consisted of unrelated healthy individuals recruited in the same geographic regions as the SSc patients and matched by age, sex, and ethnicity with the SSc-patient groups. Approval of local ethical committees was obtained from all the participating centers (Comité de Bioética del Consejo Superior de Investigaciones Científicas, Comitato Etico Azienda Ospedaliera Universitaria Integrata di Verona, Local Ethics Committee of the Radboud University Nijmegen Medical Centre, Medische Ethische Commissie Leids Universitair Medisch Centrum, The regional Ethical Review Board in Lund, Local Research Ethics Committee at Glasgow Royal Infirmary, U.O. Comitato di Etica e Sperimentazione Farmaci Fondazione IRCCS Ca' Granda-Ospedale Maggiore Policlinico di Milano, Local Research Ethics Committee at Glasgow Royal Infirmary, Royal Free Hospital and Medical School Research Ethics Committee, Manchester University Research Ethics Committee). Written informed consent was required for both patients and controls to be included in the study.

Genotyping and statistical analysis

The rs10744676 *KCNA5* biallelic variant was analyzed with TaqMan SNP genotyping assay in a 7900HT Real-Time polymerase chain reaction (PCR) system from Applied Biosystems by following the manufacturer's suggestions (Foster City, CA, USA).

None of the included cohorts showed significant deviation from Hardy-Weinberg equilibrium (HWE) (*P* value significance threshold, 0.05). Significance for the allelic model in the individual cohort analyses was calculated by 2×2 contingency tables and the Fisher Exact test or χ^2 when necessary. Odds ratios (ORs) were calculated according to the Woolf method. A Breslow-Day test (BD test) was performed to assess the homogeneity of the association among populations, and pooled analysis under a Cochran-Mantel-Haenszel fixed-effect model was used to analyze jointly all the included cohorts. Statistical analyses were performed as implemented in PLINK (v1. 07) software package [21]. Power was calculated by using the software Power Calculator for Genetic Studies 2006 and assuming an additive model and previously reported minor allele frequency (MAF) and ORs [22].

Results

The power of our replication study in each stratified analysis is summarized in Table 1. We emphasize that the size of our pooled PAH^+ patient subgroup represents the largest SSc-related PAH^+ cohort analyzed to date (n = 179).

Table 1 Overall statistical power of the study for *KCNA5* rs10744676 SNP in each analyzed disease subtype at the 5% significance level assuming an additive effect model and a minor allele (rs10744676*C) frequency = 0.10 (MAF_{CEU})

	Phenotype						
Statistical power (%)	SSc	lcSSc	dcSSc	ACA^+	ATA^+	Fib^+	PAH^+
OR = 0.62	100	100	99	100	96	99	95[a]
OR = 0.48	100	100	100	100	100	100	100[b]
OR = 0.79	96	91	67	77	59	71	45[c]

[a]Reference OR, 0.62, except for the PAH^+ group, in which OR = 0.47. [b]Reference OR (lower previously reported 95% CI) = 0.48, except for the PAH^+ group, in which OR = 0.32. [c]Reference OR (upper previously reported 95% CI) = 0.79, except for the PAH^+ group, in which OR = 0.71. SSc, systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis; dcSSc, diffuse cutaneous systemic sclerosis; ACA^+, anti-centromere autoantibody-positive patients; ATA^+, anti-topoisomerase autoantibody-positive patients; Fib^+, lung fibrosis-positive patients (HRCT); PAH^+, pulmonary arterial hypertension-positive patients (right-heart catheterization).

The power of the study of this clinical feature in our population to detect an association equivalent to that previously reported by Wipff *et al.* is 95%, at the 5% significance level.

Table 2 shows the results of the comparison of the complete set of SSc and each of the previously defined subgroups with the healthy control group. As observed in the table, the analyzed polymorphism showed no significant association with either the disease or any of the subsets, not even the PAH-positive group. Furthermore, no significant association was observed in the individual population analyses, either in the whole disease comparison or in the different subphenotypes (see Additional file 1, Table S1). The different cohorts showed a high inter-population combinability, as can be observed in the Breslow-Day test results (Table 2).

The minor allele in the Spanish and Italian control populations included in this study was less frequent than that in the pooled population described by the Wipff *et al.* (rs10744676*C $Frequency_{Spain}$ = 10.48%; rs10744676*C $Frequency_{Italy}$ = 10.00%; rs10744676*C $Frequency_{pooled\_Wipff}$ = 14.7%). However, these frequencies are in concordance with those reported for the HapMap CEU population (MAF_{HapMap_CEU} = 10.00%). Moreover, MAF differences have been reported in different European populations ($MAF_{1000Genomes\_CEU}$ = 17.64%, $MAF_{1000Genomes\_GBR}$ = 15.73%, $MAF_{1000Genomes\_FIN}$ = 13.98%, $MAF_{1000Genomes\_TSI}$ = 8.16%).

In addition, the analysis of PAH^+ versus PAH^- patients revealed no phenotype-restricted association in this subgroup (P_{MH} = 0.59; OR, 0.91; 95% CI, 0.64 to 1.28).

Discussion

Despite the previous findings of an association of the rs10744676 *KCNA5* genetic variant with PAH^+ SSc patients, our data do not corroborate the reported protective effect of the minor allele of this polymorphism on SSc-related PAH.

It is worth mentioning that the prevalence of right-heart catheterization-confirmed PAH in our pooled population (7.64%) is consistent with previous reports [11]. However, SSc is a chronic progressive disease, and some patients classified as PAH^- might develop PAH in the future. Therefore, we consider this issue a limitation of our study and similar reports. In this context, we point out that the mean disease duration for the patient group included in this study (mean disease duration, 13.9 ± 14.4 years, as reported in Bossini-Castillo *et al.* [23]) was higher than that in Wipff *et al.* [11] (11.9 ± 9.4 years in the discovery

Table 2 Pooled analysis and stratified analyses of SSc patients and healthy controls for rs10744676 genetic variant, located in the *KCNA5* gene

	Genotype, *n* (%)				Allele test		
Subgroup (*n*)	C/C	C/T	T/T	MAF (%)	P_{MH}	OR (95% CI)	P_{BD}
Controls (*n* = 2,690)	37 (1.38)	597 (22.19)	2,056 (76.43)	12.47			
SSc (*n* = 2,343)	43 (1.84)	471 (20.10)	1,829 (78.06)	11.89	0.49	0.96 (0.85-1.08)	0.53
lcSSc (*n* = 1,642)	31 (1.89)	327 (19.91)	1,284 (78.20)	11.85	0.48	0.95 (0.83-1.09)	0.63
dcSSc (*n* = 701)	12 (1.71)	144 (20.54)	545 (77.75)	11.98	0.80	0.98 (0.81-1.17)	0.59
ACA^+ (*n* = 931)	18 (1.93)	197 (21.16)	716 (76.91)	12.51	0.49	1.06 (0.90-1.24)	0.96
ATA+ (*n* = 568)	8 (1.41)	117 (20.60)	443 (77.99)	11.71	0.85	0.98 (0.80-1.20)	0.94
Fib^+ (*n* = 771)	14 (1.82)	154 (19.97)	603 (78.21)	11.80	0.34	0.92 (0.77-1.09)	0.92
PAH^+ (*n* = 179)	2 (1.12)	37 (20.67)	140 (78.21)	11.45	0.44	0.88 (0.63-1.23)	0.70

MAF, minor allele frequency; P_{MH}, Mantel-Haenszel test under fixed-effect *P* values. Controls are used as reference for all comparisons, and *P* values have been calculated for the allelic model. OR, odds ratio for the minor allele; 95% CI, 95% confidence interval; SSc, systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis; dcSSc, diffuse cutaneous systemic sclerosis; ACA^+, anti-centromere autoantibody-positive patient; ATA^+, anti-topoisomerase autoantibody-positive patient; Fib^+, lung fibrosis-positive patient (HRCT); PAH^+, pulmonary arterial hypertension-positive patients (right-heart catheterization).

cohort and 13.9 ± 14.4 years in the replication), which might influence the observed differences.

The PAH^+ cohort studied in this report reached a 95% estimated statistical power to detect an association equivalent to that observed by Wipff *et al.* within this subgroup (OR, 0.47, in the comparison between PAH^+ SSc patients and controls of the discovery phase and two replication steps [11]). Hence, we consider that the lack of replication observed in our study is unlikely to be caused by a type II error (false negative) because of a reduced statistical power. However, autoimmune-associated variants usually show modest degrees of risk, especially in the case of non-HLA loci [24]. Wipff *et al.* reported an OR for *KCNA5* rs10744676 polymorphism that was remarkably more protective than the SSc genetic-association standards for non-HLA loci (that is, modest ORs between 0.70 and 1). Thus, we reflect that if the influence of rs10744676 in the SSc-patient genetic predisposition to PAH is modest, the statistical power to detect a possible association in the PAH^+ cohort analyzed in the present article might be insufficient, and a possible modest effect of *KCNA5* rs10744676 might be overlooked (Table 1).

In addition, it is established that PAH is more frequent in ACA^+ patients [1], and despite the suggestive association with PAH^+ patients, no significant association with the ACA^+ subphenotype was identified in the previous SSc study [11] or in this report. This fact is also consistent with a lack of association of the selected polymorphism with PAH development.

Although rs10744676 might have a functional role in *KCNA5* expression because of its location in its putative promoter, no evidence confirms a functional role for this variant. Therefore, we speculate that the previously mentioned association with SSc could be the consequence of a tagged causal variant yet to be discovered. Moreover, because PAH onset time has not been considered for the analyses, it might act as a confounding factor in the discrepant results.

Conclusions

In summary, our data do not support an important role of rs10744676 as a PAH genetic marker in SSc patients.

Additional material

Additional file 1: Genotype and minor allele frequencies of *KCNA5* rs10744676 genetic variant in five European cohorts. This file contains Table S1, showing the genotype and allele distributions of *KCNA5* rs10744676 genetic variant in five European cohorts (2,343 SSc cases and 2,690 controls).

Abbreviations

ACA: anti-centromere autoantibodies; ATA: anti-topoisomerase antibodies; BD test: Breslow-Day test; *CD226*: cluster of differentiation 226; CEU: Utah residents with ancestry from northern and western Europe; dcSSc: diffuse cutaneous systemic sclerosis; hRCT: High-resolution computed tomography; HWE: Hardy-Weinberg equilibrium; K: potassium; *KCNA5*: potassium voltage-gated channel shaker-related subfamily member 5; Kv channels: potassium voltage-gated channels; lcSSc: limited cutaneous systemic sclerosis; MAF: minor allele frequency; *NOTCH4*: Notch (*Drosophila*) homologue 4; OR: odds ratio; PAH: pulmonary arterial hypertension; PASMC: pulmonary artery smooth muscle cell; PCR: polymerase chain reaction; SNP: single-nucleotide polymorphism; *SOX5*: SRY (sex-determining region Y)-box 5; SSc: systemic sclerosis.

Authors' contributions

LBC contributed to the analysis and interpretation of data and to the drafting of the manuscript. CPS, LB, and JB participated in the acquisition of data and the drafting of the manuscript. CF, TRDJR, and JM contributed to the conception and design of the study and critically revised the manuscript. MCV, JLC, PC, LRR, RGP, MAGG, IC, MTC, CT, EVR, MVE, AJS, RH, CL, JMvL, PS, AH, JW, CD, and the Spanish Scleroderma Group were involved in the acquisition of data and the revision of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Sofia Vargas, Sonia García, and Gema Robledo for excellent technical assistance and all the patients and control donors for their essential collaboration. We thank Banco Nacional de ADN (University of Salamanca, Spain), who supplied part of the control DNA samples. We are also thankful to EUSTAR (The EULAR Scleroderma Trials and Research group for the facilitation of this project. We thank the following Spanish Scleroderma Group members for their contribution in this study: Norberto Ortego-Centeno and Raquel Ríos, Unidad de Enfermedades Sistémicas Autoinmunes, Servicio de Medicina Interna, Hospital Clínico Universitario San Cecilio, Granada; Nuria Navarrete, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Antonio Fernández-Nebro, Servicio de Medicina Reumatología, Hospital Carlos Haya, Málaga; María F. González-Escribano, Servicio de Inmunología, Hospital Virgen del Rocío, Sevilla; Julio Sánchez-Román, Mª Jesús Castillo and Francisco José García-Hernández, Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Servicio de Reumatología, Hospital Reina Sofía, Córdoba; Benjamín Fernández-Gutiérrez, Servicio de Reumatología, Hospital Clínico San Carlos, Madrid; José Luis Andreu and Mónica Fernández de Castro, Servicio de Reumatología, Hospital Puerta del Hierro, Madrid; Paloma García de la Peña, Servicio de Reumatología, Hospital Madrid Norte Sanchinarro, Madrid; Francisco Javier López-Longo and Lina Martínez, Servicio de Reumatología, Hospital General Universitario Gregorio Marañón, Madrid; Vicente Fonollosa, Servicio de Medicina Interna, Hospital Valle de Hebrón, Barcelona; Gerard Espinosa, Servicio de Medicina Interna, Hospital Clinic, Barcelona; Anna Pros, Servicio de Reumatología, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Servicio de Medicina Interna, Hospital Universitari Mútua Terrasa, Barcelona; Francisco Javier Narváez, Servicio de Reumatología, Hospital Universitari de Bellvitge, Barcelona; Bernardino Díaz, Luis Trapiella and María Gallego, Servicio de Medicina Interna, Hospital Central de Asturias, Oviedo; María del Carmen Freire and Inés Vaqueiro, Unidad de Trombosis y Vasculitis, Servicio de Medicina Interna, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Servicio de Reumatología, Hospital Universitario de Canarias, Tenerife; Juan José Alegre, Servicio de Reumatología, Hospital del Doctor Peset Aleixandre, Valencia; José Andrés Román-Ivorra, Servicio de Reumatología, Hospital Universitari i Politecnic La Fe, Valencia. Francisco J. Blanco-García and Natividad Oreiro, Servicio de Reumatología, INIBIC-Hospital Universitario A Coruña, La Coruña.

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 from Junta de Andalucía, Spain, in part by Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain, and by Fondo Europeo de Desarrollo Regional (FEDER). TRDJR was funded by the VIDI

laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). JM and TRDJR were sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). BPCK is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). This study was also funded by PI-0590-2010, Consejería de Salud, Junta de Andalucía, Spain.

Author details
[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Avenida del Conocimiento s/n, Granada, 18100, Spain. [2]Servicio de Medicina Interna, Hospital Valle de Hebron, Passeig de la Vall d'Hebron, 119-129, Barcelona, 08035, Spain. [3]Referral Center for Systemic Autoimmune Diseases Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, via Francesco Sforza 28, Milan, 20122, Italy. [4]Department of Rheumatology & Clinical Immunology, University Medical Center, Heidelberglaan 100, 3485 CX Utrecht, The Netherlands. [5]Laboratory of Translational Immunology, University Medical Center, Heidelberglaan 100, 3485 CX Utrecht, The Netherlands. [6]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Geert Grooteplein-Zuid 10, Nijmegen, 6525 GA, The Netherlands. [7]Servicio de Medicina Interna, Hospital Clínico Universitario, Avenida Doctor Olóriz 16, Granada, 18012, Spain. [8]Servicio de Reumatología, Hospital Universitario 12 de Octubre, Avenida de Córdoba, s/n, Madrid, 28041, Spain. [9]Servicio de Reumatología, Hospital Clínico San Carlos, C/ Profesor Martín Lagos s/n, Madrid, 28040, Spain. [10]Servicio de Reumatología, Hospital Universitario Virgen de la Victoria, Campus Universitario Teatinos s/n, Málaga, 29010, Spain. [11]Servicio de Reumatología, Hospital Universitario Marqués de Valdecilla, IFIMAV, Avenida Valdecilla 25, Santander, 39008, Spain. [12]Servicio de Reumatología, Hospital de la Santa Creu i Sant Pau, C/Sant Antoni Maria Claret 167, Barcelona, 08025, Spain. [13]Servicio de Medicina Interna, Hospital Carlos Haya, Avenida Carlos Haya s/n, Málaga, 29010, Spain. [14]Servicio de Medicina Interna, Hospital Parc Tauli, Parc del Taulí 1, Sabadell, 08208, Spain. [15]Servicio de Reumatología, Hospital Universitario de la Princesa, C Diego de León 62, Madrid, 28006, Spain. [16]Servicio de Medicina Interna, Hospital de Cruces, Plaza de Cruces 2, Barakaldo, 48903, Spain. [17]Department of Rheumatology, Leiden University Medical Center, Albinusdreef 2, Leiden, 2333 ZA, The Netherlands. [18]Department of Rheumatology, Lund University, Paradisgatan 2, Lund, SE-221 00, Sweden. [19]Department of Medicine, Università degli Studi di Verona, Via dell'Artigliere 19, Verona, 37129, Italy. [20]Institute of Cellular Medicine, Newcastle University, Framlington Place, Newcastle upon Tyne, Newcastle, NE2 4HH, UK. [21]Centre for Rheumatic Diseases, Glasgow Royal Infirmary, 84 Castle Street, Glasgow G4 0SF, UK. [22]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Stopford Building, Oxford Road, Manchester, M13 9PT, UK. [23]Centre for Rheumatology, Royal Free and University College Medical School, University College London, Royal Free Campus, Rowland Hill Street, London, NW3 PF, UK.

Received: 16 July 2012 Revised: 6 December 2012
Accepted: 20 December 2012 Published: 27 December 2012

References

1. Gabrielli A, Avvedimento EV, Krieg T: **Scleroderma.** *N Engl J Med* 2009, **360**:1989-2003.
2. Steen VD, Medsger TA: **Changes in causes of death in systemic sclerosis, 1972-2002.** *Ann Rheum Dis* 2007, **66**:940-944.
3. Martin JE, Bossini-Castillo L, Martin J: **Unraveling the genetic component of systemic sclerosis.** *Hum Genet* 2012, **131**:1023-1037.
4. Gorlova O, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, Diaz-Gallo LM, Broen JC, Vonk MC, Simeon CP, Alizadeh BZ, Coenen MJ, Voskuyl AE, Schuerwegh AJ, van Riel PL, Vanthuyne M, van 't Slot R, Italiaander A, Ophoff RA, Hunzelmann N, Fonollosa V, Ortego-Centeno N, Gonzalez-Gay MA, Garcia-Hernandez FJ, Gonzalez-Escribano MF, Airo P, van Laar J, Worthington J, Hesselstrand R, Smith V, *et al*: **Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy.** *PLoS Genet* 2011, **7**:e1002178.
5. Dieude P, Guedj M, Wipff J, Ruiz B, Hachulla E, Diot E, Granel B, Sibilia J, Tiev K, Mouthon L, Cracowski JL, Carpentier PH, Amoura Z, Fajardy I, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis.** *Arthritis Rheum* 2009, **60**:2472-2479.
6. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Airo P, Melchers I, Hachulla E, Cerinic MM, Diot E, Hunzelmann N, Caramaschi P, Sibilia J, Tiev K, Mouthon L, Riccieri V, Cracowski JL, Carpentier PH, Distler J, Amoura Z, Tarner I, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis.** *Ann Rheum Dis* 2011, **70**:668-674.
7. Dieude P, Guedj M, Wipff J, Ruiz B, Riemekasten G, Matucci-Cerinic M, Melchers I, Hachulla E, Airo P, Diot E, Hunzelmann N, Cabane J, Mouthon L, Cracowski JL, Riccieri V, Distler J, Meyer O, Kahan A, Boileau C, Allanore Y: **Association of the TNFAIP3 rs5029939 variant with systemic sclerosis in the European Caucasian population.** *Ann Rheum Dis* 2010, **69**:1958-1964.
8. Hoshino K, Satoh T, Kawaguchi Y, Kuwana M: **Association of hepatocyte growth factor promoter polymorphism with severity of interstitial lung disease in Japanese patients with systemic sclerosis.** *Arthritis Rheum* 2011, **63**:2465-2472.
9. Dieude P, Guedj M, Truchetet ME, Wipff J, Revillod L, Riemekasten G, Matucci-Cerinic M, Melchers I, Hachulla E, Airo P, Diot E, Hunzelmann N, Mouthon L, Cabane J, Cracowski JL, Riccieri V, Distler J, Amoura Z, Valentini G, Camaraschi P, Tarner I, Frances C, Carpentier P, Brembilla NC, Meyer O, Kahan A, Chizzolini C, Boileau C, Allanore Y: **Association of the CD226 Ser(307) variant with systemic sclerosis: evidence of a contribution of costimulation pathways in systemic sclerosis pathogenesis.** *Arthritis Rheum* 2011, **63**:1097-1105.
10. Bossini-Castillo L, Simeon CP, Beretta L, Broen JC, Vonk MC, Rios-Fernandez R, Espinosa G, Carreira P, Camps MT, Castillo MJ, Gonzalez-Gay MA, Beltran E, Carmen Freire MD, Narvaez J, Tolosa C, Witte T, Kreuter A, Schuerwegh AJ, Hoffmann-Vold AM, Hesselstrand R, Lunardi C, van Laar JM, Chee MM, Herrick A, Koeleman BP, Denton CP, Fonseca C, Radstake TR, Martin J: **A multicenter study confirms CD226 gene association with systemic sclerosis-related pulmonary fibrosis.** *Arthritis Res Ther* 2012, **14**:R85.
11. Wipff J, Dieude P, Guedj M, Ruiz B, Riemekasten G, Cracowski JL, Matucci-Cerinic M, Melchers I, Humbert M, Hachulla E, Airo P, Diot E, Hunzelmann N, Caramaschi P, Sibilia J, Valentini G, Tiev K, Girerd B, Mouthon L, Riccieri V, Carpentier PH, Distler J, Amoura Z, Tarner I, Degano B, Avouac J, Meyer O, Kahan A, Boileau C, Allanore Y: **Association of a KCNA5 gene polymorphism with systemic sclerosis-associated pulmonary arterial hypertension in the European Caucasian population.** *Arthritis Rheum* 2010, **62**:3093-3100.
12. Bortner CD, Hughes FM Jr, Cidlowski JA: **A primary role for K^+ and Na^+ efflux in the activation of apoptosis.** *J Biol Chem* 1997, **272**:32436-32442.
13. Brevnova EE, Platoshyn O, Zhang S, Yuan JX: **Overexpression of human KCNA5 increases IK V and enhances apoptosis.** *Am J Physiol Cell Physiol* 2004, **287**:C715-722.
14. Yuan XJ: **Voltage-gated K^+ currents regulate resting membrane potential and $[Ca^{2+}]_i$ in pulmonary arterial myocytes.** *Circ Res* 1995, **77**:370-378.
15. Archer SL, Souil E, Dinh-Xuan AT, Schremmer B, Mercier JC, El Yaagoubi A, Nguyen-Huu L, Reeve HL, Hampl V: **Molecular identification of the role of voltage-gated K^+ channels, Kv1.5 and Kv2.1, in hypoxic pulmonary vasoconstriction and control of resting membrane potential in rat pulmonary artery myocytes.** *J Clin Invest* 1998, **101**:2319-2330.
16. Platoshyn O, Brevnova EE, Burg ED, Yu Y, Remillard CV, Yuan JX: **Acute hypoxia selectively inhibits KCNA5 channels in pulmonary artery smooth muscle cells.** *Am J Physiol Cell Physiol* 2006, **290**:C907-C916.
17. Coppock EA, Martens JR, Tamkun MM: **Molecular basis of hypoxia-induced pulmonary vasoconstriction: role of voltage-gated K^+ channels.** *Am J Physiol Lung Cell Mol Physiol* 2001, **281**:L1-L12.
18. Yuan XJ, Wang J, Juhaszova M, Gaine SP, Rubin LJ: **Attenuated K^+ channel gene transcription in primary pulmonary hypertension.** *Lancet* 1998, **351**:726-727.
19. Remillard CV, Tigno DD, Platoshyn O, Burg ED, Brevnova EE, Conger D, Nicholson A, Rana BK, Channick RN, Rubin LJ, O'Connor DT, Yuan JX: **Function of Kv1.5 channels and genetic variations of KCNA5 in patients with idiopathic pulmonary arterial hypertension.** *Am J Physiol Cell Physiol* 2007, **292**:C1837-C1853.
20. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, Rowell N, Wollheim F: **Scleroderma (systemic sclerosis): classification, subsets and pathogenesis.** *J Rheumatol* 1988, **15**:202-205.

21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *Am J Hum Genet* 2007, **81**:559-575.
22. Skol AD, Scott LJ, Abecasis GR, Boehnke M: **Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies.** *Nat Genet* 2006, **38**:209-213.
23. Bossini-Castillo L, Simeon CP, Beretta L, Vonk MC, Callejas-Rubio JL, Espinosa G, Carreira P, Camps MT, Rodriguez-Rodriguez L, Rodriguez-Carballeira M, Garcia-Hernandez FJ, Lopez-Longo FJ, Hernandez-Hernandez V, Saez-Comet L, Egurbide MV, Hesselstrand R, Nordin A, Hoffmann-Vold AM, Vanthuyne M, Smith V, De Langhe E, Kreuter A, Riemekasten G, Witte T, Hunzelmann N, Voskuyl AE, Schuerwegh AJ, Lunardi C, Airo P, Scorza R, *et al*: **Confirmation of association of the macrophage migration inhibitory factor gene with systemic sclerosis in a large European population.** *Rheumatology (Oxford)* 2011, **50**:1976-1981.
24. Cho JH, Gregersen PK: **Genomics and the multifactorial nature of human autoimmune disease.** *N Engl J Med* 2011, **365**:1612-1623.

doi:10.1186/ar4124
Cite this article as: Bossini-Castillo *et al.*: *KCNA5* gene is not confirmed as a systemic sclerosis-related pulmonary arterial hypertension genetic susceptibility factor. *Arthritis Research & Therapy* 2012 **14**:R273.

Immunochip study

Immunochip Analysis Identifies Multiple Susceptibility Loci for Systemic Sclerosis

Maureen D. Mayes,[1,52,*] Lara Bossini-Castillo,[2,52,*] Olga Gorlova,[3,52] José Ezequiel Martin,[2,52] Xiaodong Zhou,[1] Wei V. Chen,[3] Shervin Assassi,[1] Jun Ying,[3] Filemon K. Tan,[1] Frank C. Arnett,[1] John D. Reveille,[1] Sandra Guerra,[4] María Teruel,[2] Francisco David Carmona,[2] Peter K. Gregersen,[5] Annette T. Lee,[5] Elena López-Isac,[2] Eguzkine Ochoa,[2] Patricia Carreira,[6] Carmen Pilar Simeón,[7] Iván Castellví,[8] Miguel Ángel González-Gay,[9] the Spanish Scleroderma Group, Alexandra Zhernakova,[10] Leonid Padyukov,[11] Marta Alarcón-Riquelme,[12,13] Cisca Wijmenga,[10] Matthew Brown,[14] Lorenzo Beretta,[15] Gabriela Riemekasten,[16] Torsten Witte,[17] Nicolas Hunzelmann,[18] Alexander Kreuter,[19] Jörg H.W. Distler,[20] Alexandre E. Voskuyl,[21] Annemie J. Schuerwegh,[22] Roger Hesselstrand,[23] Annika Nordin,[11] Paolo Airó,[24] Claudio Lunardi,[25] Paul Shiels,[26] Jacob M. van Laar,[27] Ariane Herrick,[28] Jane Worthington,[28] Christopher Denton,[4] Fredrick M. Wigley,[29] Laura K. Hummers,[29] John Varga,[30] Monique E. Hinchcliff,[30] Murray Baron,[31] Marie Hudson,[31] Janet E. Pope,[32] Daniel E. Furst,[33] Dinesh Khanna,[34] Kristin Phillips,[34] Elena Schiopu,[34] Barbara M. Segal,[35] Jerry A. Molitor,[36] Richard M. Silver,[37] Virginia D. Steen,[38] Robert W. Simms,[39] Robert A. Lafyatis,[39] Barri J. Fessler,[40] Tracy M. Frech,[41] Firas AlKassab,[42] Peter Docherty,[43] Elzbieta Kaminska,[44] Nader Khalidi,[45] Henry Niall Jones,[46] Janet Markland,[47] David Robinson,[48] Jasper Broen,[49,50] Timothy R.D.J. Radstake,[49,50,52] Carmen Fonseca,[4,52] Bobby P. Koeleman,[51,52] and Javier Martin[2,52]

In this study, 1,833 systemic sclerosis (SSc) cases and 3,466 controls were genotyped with the Immunochip array. Classical alleles, amino acid residues, and SNPs across the human leukocyte antigen (HLA) region were imputed and tested. These analyses resulted in a model composed of six polymorphic amino acid positions and seven SNPs that explained the observed significant associations in the region. In

[1]The University of Texas Health Science Center at Houston, Houston, TX 77030, USA; [2]Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas, Granada 18016, Spain; [3]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA; [4]Centre for Rheumatology, University College London Medical School, London WC1E 6BT, UK; [5]The Feinstein Institute for Medical Research, North Shore – Long Island Jewish Health System, Manhasset, NY 11030, USA; [6]Department of Rheumatology, Hospital Universitario 12 de Octubre, Madrid 28041, Spain; [7]Department of Internal Medicine, Hospital Valle de Hebrón, Barcelona 08035, Spain; [8]Department of Rheumatology, Hospital de la Santa Creu i Sant Pau, Barcelona 08025, Spain; [9]Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, Instituto de Formación e Investigación Marqués de Valdecilla, Santander 39008, Spain; [10]Department of Genetics, University Medical Center Groningen, Groningen 9700, the Netherlands; [11]Rheumatology Unit, Department of Medicine, Karolinska University Hospital, Karolinska Institute, 171 76 Stockholm, Sweden; [12]Área de Variabilidad del ADN Humano, Centro Pfizer – Universidad de Granada – Junta de Andalucía de Genómica e Investigación Oncológica, Granada 18016, Spain; [13]Arthritis and Clinical Immunology Program, Oklahoma Medical Research Foundation, Oklahoma City, OK 73104, USA; [14]University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, QLD 4072, Australia; [15]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan 20122, Italy; [16]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin 10117, Germany; [17]Department of Clinical Immunology, Hannover Medical School, Hannover 30625, Germany; [18]Department of Dermatology, University of Cologne, Cologne 50924, Germany; [19]Department of Dermatology, Venereology, and Allergology, HELIOS St. Elisabeth Hospital, Oberhausen 46045, Germany; [20]Department of Internal Medicine, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen 91054, Germany; [21]Department of Rheumatology, VU University Medical Center, Amsterdam 1081 HV, the Netherlands; [22]Department of Rheumatology and Pathology and Central Medical Immunology Laboratory, Leiden University Medical Center, Leiden 2300 RC, the Netherlands; [23]Department of Rheumatology, Lund University, Lund 221 85, Sweden; [24]Servizio di Reumatologia ed Immunologia Clinica Spedali Civili, Brescia 25123, Italy; [25]Department of Medicine, Universita degli Studi di Verona, Verona 37134, Italy; [26]Glasgow Biomedical Research Centre, University of Glasgow, Glasgow G61 1BD, UK; [27]Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne NE2 4HH, UK; [28]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester M13 9PT, UK; [29]Division of Rheumatology, Johns Hopkins University, Baltimore, MD 21224, USA; [30]Division of Rheumatology, Northwestern University, Chicago, IL 60611, USA; [31]Department of Rheumatology, Jewish General Hospital, Montreal, QC H3T 1E2, Canada; [32]Division of Rheumatology, University of Western Ontario, London, ON N6A 4V2, Canada; [33]Division of Rheumatology, University of California, Los Angeles, Los Angeles, CA 90095-1670, USA; [34]Division of Rheumatology, University of Michigan, Ann Arbor, MI 48109, USA; [35]Division of Rheumatology, University of Minnesota, Minneapolis, MN 55455, USA; [36]Division of Rheumatic and Autoimmune Diseases, University of Minnesota, Minneapolis, MN 55455, USA; [37]Division of Rheumatology, Medical University of South Carolina, Charleston, SC 29425, USA; [38]Division of Rheumatology, Georgetown University Medical Center, Washington, DC 20007, USA; [39]Division of Rheumatology, Boston University, Boston, MA 02118, USA; [40]Division of Rheumatology, University of Alabama Birmingham, Birmingham, AL 35294-3408, USA; [41]Division of Rheumatology, University of Utah, Salt Lake City, UT 84132-2101, USA; [42]Department of Rheumatology, Carolinas Healthcare System, Charlotte, NC 28211, USA; [43]Department of Rheumatology, The Moncton Hospital, Moncton, NB E1C 6Z8, Canada; [44]Rheumatology Section, Department of Medicine, Alberta Health Services, Calgary, AB T2N 2T9, Canada; [45]Division of Rheumatology, Department of Medicine, McMaster University, Hamilton, ON L8N 1Y2, Canada; [46]Department of Rheumatology, University of Edmonton, Edmonton, AB T5M 0H4, Canada; [47]Department of Rheumatology, University of Saskatchewan, Saskatoon, SK S7K 0H6, Canada; [48]Department of Rheumatology, Arthritis Center, University of Manitoba, Winnipeg, MB R3A 1M4, Canada; [49]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen 6500 HC, the Netherlands; [50]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht 3584 CX, the Netherlands; [51]Complex Genetics Section, Department of Medical Genetics, University Medical Center Utrecht, Utrecht 3584 CX, the Netherlands
[52]These authors contributed equally to this work
*Correspondence: maureen.d.mayes@uth.tmc.edu (M.D.M.), larabossc@gmail.com (L.B.-C.)
http://dx.doi.org/10.1016/j.ajhg.2013.12.002.

addition, a replication step comprising 4,017 SSc cases and 5,935 controls was carried out for several selected non-HLA variants, reaching a total of 5,850 cases and 9,401 controls of European ancestry. Following this strategy, we identified and validated three SSc risk loci, including *DNASE1L3* at 3p14, the *SCHIP1-IL12A* locus at 3q25, and *ATG5* at 6q21, as well as a suggested association of the *TREH-DDX6* locus at 11q23. The associations of several previously reported SSc risk loci were validated and further refined, and the observed peak of association in *PXK* was related to *DNASE1L3*. Our study has increased the number of known genetic associations with SSc, provided further insight into the pleiotropic effects of shared autoimmune risk factors, and highlighted the power of dense mapping for detecting previously overlooked susceptibility loci.

Introduction

Systemic sclerosis (SSc, scleroderma [MIM 181750]) is an autoimmune disease characterized by three main features: (1) fibrosis of the skin and internal organs, (2) a noninflammatory vasculopathy, and (3) autoantibody production.[1] It is a multiorgan system disease with considerable phenotypic heterogeneity, resulting in a broad spectrum of disease severity. From a clinical point of view, SSc is divided into limited cutaneous systemic sclerosis (lcSSc) or diffuse cutaneous systemic sclerosis (dcSSc).[2] From an immunological point of view, SSc is typically classified according to mutually exclusive and disease-specific autoantibodies (anticentromere antibodies [ACAs] and antitopoisomerase antibodies [ATAs]), which are found in approximately 50% of SSc cases.[3]

Several studies have established that SSc is a complex genetic disease with contributions from multiple genetic loci.[4,5] The evidence supporting a genetic predisposition for the disease has revealed a major contribution from the major histocompatibility complex (MHC), as well as a number of other gene regions.[4–6] In fact, genome-wide association studies (GWASs) have confirmed the MHC class II region as the most significant genetic region associated with SSc.[7,8] Interestingly, a recent report restricted the MHC associations to the different autoantibody subsets.[9] In addition, multiple non-MHC loci, such as *IRF5* (MIM 607218), *STAT4* (MIM 600558), *CD247* (MIM 186780), *TNIP1* (MIM 607714), *IRF8* (MIM 601565), *IL12RB2* (MIM 601642), *CSK* (MIM 124095), *KIAA0319L* (MIM 613535), *PXK* (MIM 611450), *JAZF1* (MIM 606246), *BLK* (MIM 191305), *ITGAM* (MIM 120980), and *TNFAIP3* (MIM 191163), have been associated with this condition at the genome-wide significance level ($p < 5 \times 10^{-8}$).[7–13] Both GWASs and candidate-gene strategies have clearly identified SSc susceptibility factors involved in different components of the immune system (innate immune response, adaptive immune response, cytokines, cytokine receptors, etc.), as well as genes involved in pathways that might play a role in vascular damage and fibrotic processes.[4–6] Despite the success of these approaches, innovative strategies are needed for dealing with the remaining unmapped SSc heritability.

Hence, the next step in the genetic dissection of complex autoimmune diseases (AIDs) is to identify the causal variants for the already established susceptibility loci, as well as to discover variants with lower penetrance with the use of larger cohorts. Furthermore, these goals need to be achieved with a cost-efficient strategy.[14,15] To this end, an international group of collaborators formed the Immunochip Consortium to develop and implement the Immunochip array, a custom SNP genotyping array that provides high-density mapping of AID-associated loci for large cohorts at reduced costs. The Immunochip is based on an Illumina Infinium array platform containing 196,524 variants across 186 known autoimmunity risk loci.[14] The variants included in the Immunochip encompass all variants that have been previously described for white European populations (SNPdb, 1000 Genomes Project February 2010 release, and other available sequencing projects). Also, the Immunochip design includes several rare variants considered to have significant functional effects and that might have been previously overlooked.[14,15]

This approach has recently shown encouraging results in different AIDs, such as celiac disease (CeD [MIM 212750]), rheumatoid arthritis (RA [MIM 180300]), autoimmune thyroid disease (ATD [MIM 275000 and 140300]), psoriasis (PS [MIM 177900]), primary biliary cirrhosis (PBC [MIM 109720]), juvenile idiopathic arthritis (JIA [MIM 604302]), primary sclerosing cholangitis (PSC [MIM 613806]), narcolepsy (MIM 161400), ankylosing spondylitis (MIM 106300), atopic dermatitis (MIM 603165), and Takayasu arteritis (MIM 207600).[16–27] These studies resulted in the identification of susceptibility genes and the narrowing of the associations in previously reported risk loci. Moreover, these reports increased the number of shared genetic markers between the different disorders, further supporting the common genetic component of autoimmunity.

Taking the above into consideration, the goal of this study was to explore SSc risk loci shared with other autoimmune diseases included on the Immunochip and to fine map these areas, which comprised many, but not all, previously associated SSc loci.

Subjects and Methods

Case Definition

SSc cases were defined on the basis of the 1980 preliminary classification criteria of the American Rheumatism Association (now the American College of Rheumatology)[28] or the presence of at least three out of five CREST (*c*alcinosis, *R*aynaud's phenomenon, *e*sophageal dysmotility, *s*clerodactyly, *t*elangiectasia) features typical of SSc. The designation of lcSSc or dcSSc was determined according to the method of Leroy.[2] The SSc-specific autoantibodies, ACA and ATA, were determined by standard means as previously described.[29]

Table 1. Main Clinical Features of the Studied Cohorts

		Status and Gender				SSc Subgroups			
Sample Population	Total After QC	Cases	% Female Cases	Controls	% Female Controls	lcSSc	dcSSc	ACA^+	ATA^+
US discovery	3,697	956	89%	2,741	65%	594	327	281	155
Spain discovery	1,602	877	89%	725	66%	533	247	390	192
US and Canada replication	1,845	927	86%	918	53%	584	339	279	163
Spain replication	1,237	449	89%	788	56%	238	108	160	85
Germany	1,110	694	83%	416	54%	345	240	237	179
Netherlands	1,470	385	71%	1,085	42%	250	117	90	99
Italy	1,918	648	92%	1,270	81%	436	156	301	213
Sweden	587	229	77%	358	75%	163	66	67	39
UK	1,785	685	84%	1,100	87%	491	191	253	118
Meta-analysis	15,251	5,850	86%	9,401	65%	3,634	1,791	2,058	1,243

Note: 62.17% of cases were lcSSc, and 30.53% were dcSSc (7.3% were not designated); 35.18% of cases were ACA^+, and 58.85% were ACA^- (5.97% were not designated); 21.26% of cases were ATA^+, and 72.15% were ATA^- (6.59% were not designated); and 36.70% of cases were ACA^-ATA^- (6.86% were not designated). Abbreviations are as follows: ACA^+, anticentromere-antibody-positive cases; ATA^+, antitopoisomerase-antibody-positive cases; dcSSc, diffuse cutaneous SSc; lcSSc, limited cutaneous SSc; and SSc, systemic sclerosis cases.

Sample Population

After quality-control (QC) measures were applied, the combined data set consisted of 5,850 SSc cases and 9,401 unrelated healthy controls. Table 1 shows the sample populations by sex, origin, and SSc subtype. The discovery cohort consisted of 1,833 SSc cases and 3,466 controls of white European ancestry from the US and Spain. The validation cohort was drawn from an independent group of cases and controls of similar ancestry from North America (US and Canada), Spain, Germany, the Netherlands, Italy, Sweden, and the UK (4,017 SSc cases and 5,935 controls). The populations included in this study partially overlapped with those in previously published SSc GWASs.[7,9] This study was approved by the local ethics committees of all the centers that recruited the participating individuals, and all participating individuals gave written informed consent.

The discovery population's power to detect an association with an odds ratio (OR) = 1.5 under an additive model at the 5×10^{-6} significance level was 100% (minor allele frequency [MAF] = 0.2), 96% (MAF = 0.1), 56% (MAF = 0.05), and 1% (MAF = 0.01). The overall meta-analysis had a statistical power of 100% (MAF = 0.05) and 18% (MAF = 0.01) with an OR = 1.5 but reached a power of 99% with an OR = 2.00 (the remaining parameters were maintained as previously mentioned). Power calculations were performed as implemented in Power Calculator for Genome-wide Analyses.[30]

Genotyping

The genotyping for the Immunochip custom array of the discovery cohorts was performed in accordance with Illumina protocols in two centers: the Feinstein Institute for Medical Research (Manhasset) and the Translational Research Institute at the University of Queensland Diamantina Institute (Brisbane). The US and Spanish control cohorts have already been included in previous Immunochip studies.[17,26]

The genotyping of the replication cohorts was performed with either (1) the TaqMan SNP genotyping technology in the Applied Biosystems 7900HT Fast Real-Time PCR System according to the manufacturer's suggestions or (2) the Immunochip platform (the controls sets overlapped with those from previous Immunochip reports).[16–20,22–24,26]

Genotype Calling and QC

Genotype calling on the discovery cohort was performed with the Immunochip platform according to the manufacturer's instructions with the use of the Illumina iScan System and the Genotyping Module (v.1.8.4) of the GenomeStudio Data Analysis software. Stringent QC parameters were applied in all data sets: (1) individuals who generated genotypes at <90% loci were not considered in the analysis, (2) markers with call rates ≤ 90% were excluded, and (3) markers with allele distributions strongly deviating from Hardy-Weinberg equilibrium (HEW) in controls ($p < 1 \times 10^{-5}$) were discarded. After QC measures were applied, 126,270 markers passing QC (101,692 of them showed a MAF > 0.001) were included in the analysis of the discovery cohort, resulting in a genotyping rate of 99.8%. A total of 1,886 SSc cases (895 Spanish and 981 US cases) and 4,629 unaffected controls (781 Spanish and 3,848 US individuals) passed these genotyping QCs.

Individuals in the discovery phase were also excluded on the basis of inferred ethnicity by principal-component analysis (PCA) as implemented in SNP & Variation Suite v.7 (Golden Helix). The first three principal components for each individual were plotted, and those individuals deviating from the cluster centroid were discarded for further analysis. Considering the different reported ancestries in the US control set, we included HapMap Phase 2 samples as reference populations in an initial PCA. Then, we discarded those individuals who deviated more than 6 SDs from the European ancestry cluster centroid (Figure S1, available online). In addition, individuals who deviated more than 4 SDs from the centroid of their cohort were excluded from the analysis. Moreover, to identify duplicate pairs or highly related individuals among data sets, we performed pairwise comparisons by using the genome function in PLINK[31] and HapMap2 populations as

references (which required Pi-HAT of ≥ 0.5), and we removed one sample from each pair. By these means, 43 SSc cases (18 Spanish individuals and 25 US individuals) and 1,163 controls (56 Spanish and 1,107 US individuals) were not included in the analyzed discovery cohort.

To calculate the genomic-inflation factor (λ) while overcoming the skewed nature of the SNP selection process in the Immunochip design, we used a set of 3,120 "null" SNPs not associated with autoimmune diseases (originated by J.C. Barrett and provided by John Bowes, personal communications).[19,22,23,25,26] The SNPs that did not pass QC or showed an association p value above the 90th percentile in the SSc meta-analysis were discarded. The λ for the discovery sets was estimated at 1.07 (US) and 1.085 (Spain). It has been demonstrated that λ increases with sample size.[32] Therefore, it is informative to calculate a rescaled λ for an equivalent study of 1,000 cases and 1,000 controls ($\lambda_{1,000}$).[32] In this report, $\lambda_{1,000}$ was estimated at 1.049 (US) and 1.098 (Spain). In the case of the discovery-phase meta-analysis, $\lambda = 1.065$ and $\lambda_{1,000} = 1.02$ (including the associations in the 90th percentile, $\lambda = 1.209$ and $\lambda_{1,000} = 1.064$). All loci showing one or more independent genome-wide-significant associations in the discovery phase remained significantly associated after correction for λ. However, all the identified loci (showing either genome-wide or suggestive significance, Table S1) underwent a replication step.

Imputation

We performed SNP genotype imputation of the identified SSc susceptibility loci (5 Mb regions centered in the lead SNP were considered). We included only discovery-phase genotyped SNPs that passed QC in the imputation process. Genotypes were imputed as implemented in IMPUTE2 with the use of the 1000 Genomes Phase 3 integrated reference panel.[33,34] We assessed imputed SNP quality by establishing a probability threshold for merging genotypes at 0.9. Moreover, after imputation, stringent QC was applied: (1) individuals who generated genotypes at <90% loci were not considered in the analysis, (2) markers with call rates $\leq$ 95% were excluded, (3) markers with allele distributions deviated from HWE in controls ($p < 1 \times 10^{-3}$) were discarded, and (4) variants with MAF < 0.01 were not included in the analyses. A total of 390 additional variants were imputed in *ATG5* (autophagy-related 5 [MIM 604261]), 541 in the *SCHIP1* (schwannomin-interacting protein 1[MIM 612008])-*IL12A* (interleukin 12A [MIM 161560]) locus, and 628 in *DNASE1L3* (deoxyribonuclease 1-like 3 [MIM 602244]).

Imputation of the HLA Region

The 5,955 SNPs obtained in the MHC after all QCs were used for the imputation process. We used as a reference panel for the imputation (1) 2,767 European-descent individuals[35] with four-digit typing of human leukocyte antigen (HLA) class I and II molecules and (2) the genotypes of more than 7,500 common SNPs and indel polymorphisms across the extended MHC (xMHC).[36] The imputation process was performed with Beagle software[37] according to a previously described method.[38] Imputed data in the xMHC for SNPs, amino acids, or the HLA four-digit code were filtered as follows: (1) variants with a success call rate below 95%, (2) variants with a MAF below 1%, and (3) all individuals with a SNP success call rate below 95% were excluded. After these filters, a total of 7,261 SNPs remained. Also, through the imputation process, a total of 449 polymorphic amino acid positions were obtained. At last, the alleles of the class I MHC (*HLA-A* [MIM 142800], *HLA-B* [MIM 142830], and *HLA-C* [MIM 142840]) and class II MHC (*HLA-DPA1* [MIM 142880], *HLA-DPB1* [MIM 142858], *HLA-DQA1* [MIM 146880], *HLA-DQB1* [MIM 604305], and *HLA-DRB1* [MIM 142857]) were obtained. In order to assess the accuracy of the imputation, we used partial data from previously genotyped HLA alleles from 490 individuals from the US and 466 individuals from Spain for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1*, and *HLA-DRB1* at four digits, which were partially included in previous reports.[39,40] This resulted in an accuracy of 93.51% in the US cohort and 91.70% in the Spanish cohort in the four-digit comparison (Table S2).

Statistical Analysis

The association statistics for the discovery cohorts were calculated separately in each data set via logistic regression including sex as a covariate. Then, meta-analysis combining ORs and SEs of both cohorts was performed via the inverse-variance method under the assumption of a fixed effect as implemented in PLINK v.1.07.[31] Heterogeneity across the data sets was assessed with Cochran's Q test, and those loci showing significant heterogeneity ($Q < 0.05$) were not considered for the validation step.

The validation SNPs were chosen for the validation phase if they (1) had a meta-analysis $p < 5 \times 10^{-6}$, (2) had a nominal association in both cohorts at $p < 0.05$, (3) had consistent OR directions in both cohorts, (4) were not located in a gene previously identified as SSc associated, (5) and mapped outside the HLA region (considered the loci located in chr6: 20,000,000–40,000,000 bp).

The possibility of different independent signals in the same locus was evaluated with a stepwise logistic regression as described in Eyre et al.[17] In brief, the most associated variant in the region was considered a covariate and the association analyses were calculated for the remaining variants. Independent signals should have met the following criteria: (1) $p < 5 \times 10^{-6}$, (2) no high correlation with the lead SNP ($r^2 < 0.6$), and (3) no substantial difference between the conditioned and the unconditioned p values ($p < 5 \times 10^{-4}$).

The combined analysis of the discovery populations with the replication cohorts was performed via the inverse-variance method under a fixed-effects model on the basis of population-specific logistic regression results. Only variants showing a genome-wide-significant association ($p < 5 \times 10^{-8}$), either in the whole disease or in any of the subgroup analyses, were considered confirmed associations with SSc susceptibility.

For the analyses of the HLA region, we used logistic regression to test for allelic association between the imputed variant and disease status. Individual amino acid positions were tested with a model including all the possible amino acid residues. Statistical significance was established by comparison of the deviance model to the null model.[38] For the conditional analyses, we assumed that the null model comprised the previously associated variants in a stepwise manner until no genome-wide-significant associations remained in the HLA region.

Functional Prioritization of Noncoding Variants

In addition to analyzing coding variants (functional prediction was based on PolyPhen-2),[41] we propose most likely functional variants that might explain the associations on the basis of noncoding element data. We used the Regulome, Genevar, and Blood eQTL browser resources to explore the evidence of transcription factor binding sites, DNase hypersensitivity sites, and expression quantitative trait loci (eQTLs) in the publicly available

ENCODE database and additional eQTL databases.[42–44] Molecular graphics and analyses of DNASE1L3, HLA-DPAβ1, HLA-DQα1, and HLA-DRβ1 structure were performed with the UCSF Chimera package.[45]

Results

The HLA region showed the most significant associations in the discovery phase, and three already reported non-HLA SSc risk factors (*STAT4*, *PXK*, and the *TNPO3* [MIM 610032]-*IRF5* locus)[7,13,46–54] showed genome-wide-significant associations with the disease (Table S3). No independent signals as previously defined were observed in these loci. Additionally, we calculated blocks of high linkage disequilibrium (LD) for the leading signals (defined as those SNPs showing $r^2 > 0.9$ with the lead variant according to the 1000 Genomes Project CEU [Utah residents with ancestry from northern and western Europe from the CEPH collection] population) in these loci.[34] By these means, we established an association region comprising 35 kb for the leading variant in *STAT4*, 126 kb in the case of the *TNPO3-IRF5* locus, and 91 kb in *PXK*. Because of their proximity, we tested for possible dependence between the signals in *PXK* and *DNASE1L3*. Interestingly, both the genome-wide association observed in our data in *PXK* (rs4681851) and the reported association with SSc in *PXK* (rs2176082)[13] were dependent on the association located in *DNASE1L3* (p_{cond} rs4681851 = 0.19, p_{cond} rs35677470 = 3.71×10^{-8}; p_{cond} rs2176082 = 0.12, p_{cond} rs35677470 = 6.41×10^{-11}). In addition, several previously described SSc susceptibility factors showed significant associations with SSc or the different subphenotypes of the disease and were confirmed in our analysis (Table S3). In this study, we performed a comprehensive analysis of the associations in the HLA region and focused on the associations previously undescribed in non-HLA loci.

Associations between the HLA Region and Systemic Sclerosis

We also conducted the imputation of the HLA region (chr6: 28,000,000–34,000,000 bp) in our data by using a large reference panel as previously described. Resulting from the imputation, we obtained 7,261 SNPs, 449 polymorphic amino acid positions, and 298 four-digit HLA alleles for both class I and class II molecules. Compared to HLA typing, the imputation had an overall accuracy of 93.08%.

Taking into account the serologically restricted HLA association in previous reports,[9] we conducted different analyses to compare total SSc, ACA^+, and ATA^+ between cases and controls in this locus. The overall results from the analysis specific to the HLA region can be observed in Figure 1 and Table 2. After stepwise conditional multiple logistic regression analysis in the aforementioned groups, we obtained a model composed of six polymorphic amino acid positions (two and four of which were associated with ACA and ATA, respectively, Table 2) and seven SNPs (five and two of which were associated with SSc and ACA, respectively, Table 2), which successfully conditioned all observed associations in the HLA region in either SSc or any of its serological subphenotypes (Figure 1). Consequently, the identified models conditioned the known SSc-related HLA four-digit alleles (Table S4).[39]

Loci Revealed in the Immunochip Analyses

Eight SNPs were selected as putative SSc risk factors (Table S1). We genotyped these variants in a large multicenter replication cohort of European ancestry in order to confirm the initial associations in the discovery phase. One SNP failed to genotype in the North American replication cohort, as indicated in Table S5. As illustrated in Figure 2 and Table 3, the meta-analysis of the European and North American discovery and replication cohorts identified three non-HLA loci to be associated with overall SSc ($p < 5 \times 10^{-8}$, see Table S5 for results of all selected SNPs in all groups). The associated variants included a missense SNP (rs35677470) in *DNASE1L3* at 3p14 (resulting in a cysteine substitution for arginine in the gene product, Figure S2), a SNP (rs77583790) in the intergenic region between *SCHIP1* and *IL12A* at 3q25, and a SNP (rs9373839) intronic to *ATG5* at 6q21. In addition, a SNP (rs7130875) in the intergenic region between *TREH* (trehalase [MIM 275360]) and *DDX6* (DEAD-box RNA helicase 6 [MIM 600326]) at 11q23 showed suggestive association ($p = 1.29 \times 10^{-7}$, OR = 1.17).[17]

The association signals for *DNASE1L3* and *SCHIP1-IL12A* were also significant in the lcSSc and ACA^+ subgroups. Moreover, in the case of *DNASE1L3*, the signal was strongly related to ACA^+ SSc ($p = 4.25 \times 10^{-31}$, OR = 2.03), and the most significant association in the *SCHIP1-IL12A* region was found with lcSSc ($p = 1.53 \times 10^{-11}$, OR = 2.81) (Table 3). Of note, there was no previously unreported genome-wide-significant association with either the dcSSc or the ATA^+ subset of disease in this cohort. It is worth mentioning that the risk effects for the replicated polymorphisms were maintained in all the analyzed populations (Figure 2 and Figures S3 and S4).

In order to further dissect the associated loci, we imputed the associated regions by using the 1000 Genomes Project reference panel, as described previously. Imputation resulted in a significant increase in the number of analyzed variants, but only slight differences between the top genotyped SNPs and the top imputed SNPs were observed. Considering the dense coverage of the fine-mapped loci in the Immunochip, this low gain of information from imputation was concordant with previous Immunochip reports.[22,26] As expected, the linked variants in the regions surrounding the association peaks showed concordant associations (Figure 3). However, no additional independent signals, neither SNPs nor haplotypes, showed up after conditioning on the lead variant (Figure 3).

In the case of *DNASE1L3*, the lead variant (rs35677470) encodes a probably damaging nonsynonymous change, as mentioned above. No additional variant in high LD

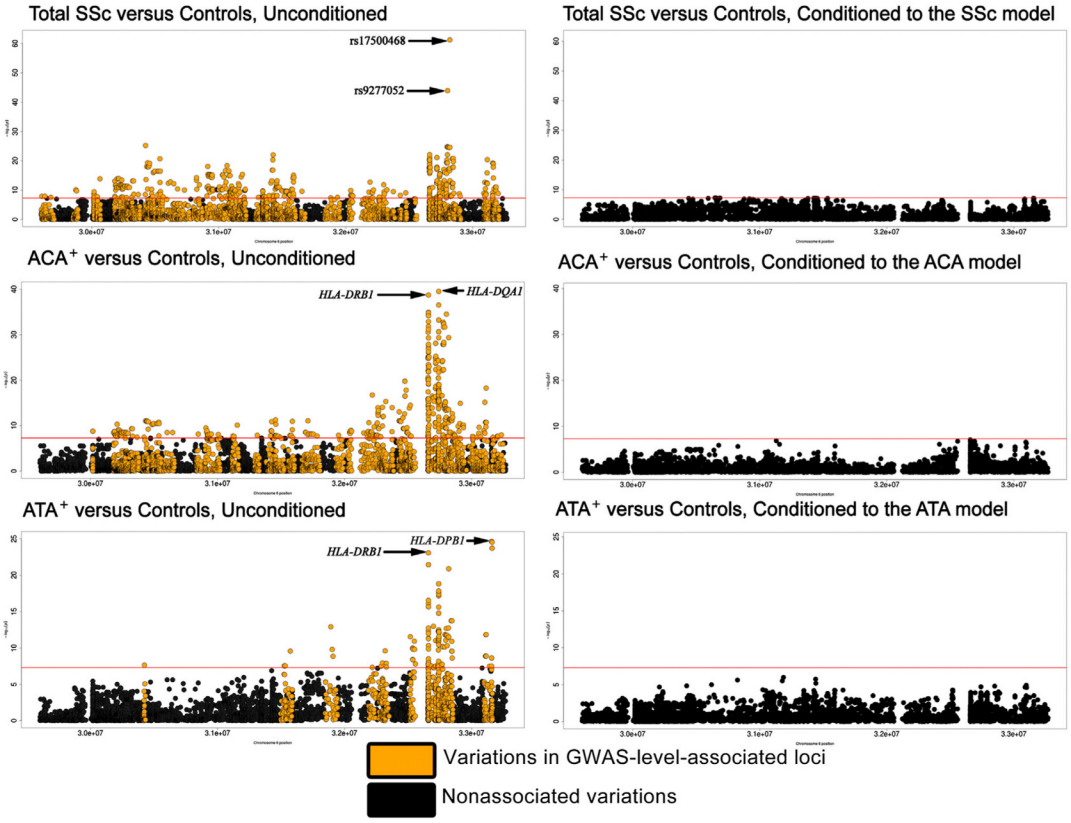


Figure 1. HLA Region: 28,000,000–36,000,000 bp
Manhattan plots in which each dot represents the p value of a variant ($-\log_{10}$(p value)) on the vertical axis and the position in chromosome 6 on the horizontal axis. The ACA model comprised HLA-DRβ1 amino acid 13, HLA-DQα1 amino acid 69, and SNPs rs12528892 and rs6933319. The ATA model comprised HLA-DRβ1 amino acids 67 and 86 and HLA-DPβ1 amino acids 76 and 96. The SSc model included the previously described ACA and ATA models and SNPs rs17500468, rs9277052, rs2442719, and rs4713605.

with this SNP ($r^2 > 0.9$) was located in a coding region or has been associated with eQTLs. Therefore, we suggest that the lead variant itself might be the most suitable causal candidate.

No variants linked with the *ATG5* leading variant have protein-coding consequences or have been associated with eQTLs. Four variants located in the fifth *ATG5* intron (rs62422881, rs3804333, rs1885450, and rs698029) were located in transcription factor binding sites and DNase peaks (Regulome scores = 2a–4). This region showed suggestive evidence of binding of different proteins with roles in the immune system (e.g., PRDM1, EP300, or CEBPB). In addition, rs7763652, a variant in the *ATG5* promoter region, showed minimal evidence of affecting the binding of CTCF (Regulome score = 4). Further research might shed light on the functional relevance of this region.

Variants in the *SCHIP1-IL12A* locus showed only minimal evidence of binding. Nevertheless, this lack of functional information might be due to the low MAF.

Discussion

In the present study, we were able to further dissect the long-known association between the HLA region and SSc. We narrowed the genetic factors in this region down to a model of six amino acids and seven SNPs, all in the ACA^+ and ATA^+ subgroups (except for rs17500468, rs9277052, rs2442719, rs4713605, and rs443623, which presented their association in the overall disease). We further checked all associated SNPs for any functional explanation of their association. Among other methods, we checked whether any of the SNPs were in a putative eQTL according to Westra et al.[44] SNPs rs17500468 and rs2442719 (associated with SSc) were in an eQTL affecting the expression of the non-HLA gene *TAP2* (MIM 170261; $p = 4.90 \times 10^{-10}$ and 4.23×10^{-8}, respectively), rs9277052 (associated with SSc) was in an eQTL affecting the expression of *HLA-DPB1* ($p = 2.86 \times 10^{-69}$), SNP rs4713605 was in an eQTL affecting the expression of

Table 2. All Independent Associations Found between the HLA Region and SSc or Any of Its Subphenotypes

SSc Subgroup	Type	Gene	Variation	AA or Nucleotide	Frequency in Cases	Frequency in Controls	p Value[a]	OR[b]	95% CI[b]
ACA^+	AA	*HLA-DRB1*	AA 13[c]	Tyr	0.048	0.139	1.79×10^{-39}	0.295	0.22–0.38
				Phe	0.232	0.136		1.817	1.56–2.11
ACA^+	AA	*HLA-DQA1*	AA 69[c]	Leu	0.484	0.581	4.46×10^{-23}	0.657	0.58–0.75
				T	0.076	0.020		3.788	2.85–5.02
ACA^+	SNP	*TAP2*	rs12528892[d]	T	0.064	0.011	2.74×10^{-11}	9.137	6.49–12.85
ACA^+	SNP	*HLA-DOA*	rs6933319	T	0.030	0.056	7.13×10^{-16}	0.304	0.21–0.42
SSc	SNP	*TAP2*	rs17500468[d]	G	0.186	0.082	5.87×10^{-62}	2.868	2.52–3.25
SSc	SNP	*HLA-DPB1*	rs9277052[d]	C	0.275	0.182	4.08×10^{-21}	1.572	1.43–1.72
SSc	SNP	*HLA-B*	rs2442719[d]	G	0.491	0.410	9.55×10^{-23}	1.512	1.39–1.64
SSc	SNP	*HLA-DPB1*	rs4713605[d]	A	0.381	0.435	2.16×10^{-13}	0.726	0.66–0.79
SSc	SNP	*HLA-DOA*	rs443623	A	0.227	0.304	3.49×10^{-9}	0.770	0.70–0.84
ATA^+	AA	*HLA-DRB1*	AA 67	Leu	0.227	0.388	3.55×10^{-22}	0.461	0.38–0.55
				Phe	0.311	0.173		2.158	1.80–2.58
ATA^+	AA	*HLA-DPB1*	AA 96	no Lys or Arg[e]	0.037	0.001	3.21×10^{-23}	75.230	28.09–201.50
ATA^+	AA	*HLA-DRB1*	AA 86[c]	Val	0.573	0.476	1.77×10^{-5}	1.423	1.21–1.67
ATA^+	AA	*HLA-DPB1*	AA 76[c]	Ile	0.074	0.028	3.05×10^{-8}	2.577	1.85–3.59

Abbreviations are as follows: AA, amino acid; ACA^+, anticentromere-antibody-positive cases; ATA^+, antitopoisomerase-antibody-positive cases; CI, confidence interval; OR, odds ratio; and SSc, systemic sclerosis cases.
[a]Logistic regression omnibus test for variations with more than two alleles.
[b]ORs and CIs for each of the associated alleles.
[c]Amino acid part of the binding pocket of the molecule.
[d]Located in a putative eQTL with a p value of at least 5×10^{-8} according to Westra et al.
[e]The absence of a lysine or an arginine in this position is the associated variant, which is very infrequent in healthy individuals of European ancestry.

HLA-DPB1 ($p = 2.26 \times 10^{-81}$), and rs12528892 was also in an eQTL affecting the expression of *TAP2* ($p = 8.53 \times 10^{-68}$). Finally, SNPs rs443623 and rs6933319 (associated with SSc and ACA production, respectively) were not predicted to be in any eQTL or a likely transcription factor binding site; the closest gene was *HLA-DOA* (MIM 142930), whose product forms an heterodimer with HLA-DOβ to assemble the HLA-DO molecule, found in lysosomes in B cells regulating HLA-DM-mediated peptide loading on MHC class II molecules.[55] Regarding the other described genes affected by eQTL SNPs, *HLA-DPB1* encodes the β chain of the HLA-DP MHC receptor complex, whereas *TAP2* is involved in antigen presentation.[56]

Regarding the amino-acid-associated positions, we found that the associated amino acid at position 13 in HLA-DRβ1 was part of the binding pocket of this molecule with epitope,[57] which was also the case with amino acid 69 in HLA-DQα1;[58] both of these amino acids are associated with ACA production (Figure S5). Amino acids 86 and 76 in HLA-DRβ1 and HLA-DPβ1, respectively (both associated with ATA production), were also part of the binding pocket of their respective molecules.[57,59] Conversely, amino acids 67 and 96 in HLADRβ1 and HLA-DPβ1, respectively (associated with ATA production, Table 2), were not part of the binding pocket but could affect the 3D structure of the MHC. Hence, the complexity and heterogeneity of the associations between the HLA region and SSc reflect the complex and heterogeneous nature of this disorder.

Three non-HLA variants associated with SSc susceptibility were clearly identified in this report, increasing the number of loci that have been associated with SSc and providing additional insight into SSc pathogenesis. Moreover, the nonsynonymous risk variant in *DNASE1L3* showed a highly significant association with the ACA^+ group considering the known non-HLA genetic associations with SSc.[4–6]

DNASE1L3 codes for deoxyribonuclease 1-like 3, a member of the human DNase 1 family, which is expressed primarily in the liver and spleen and secreted by macrophages.[60] This protein plays a role in DNA fragmentation during apoptosis and in the generation of resected double-strand breaks in immunoglobulin-encoding genes.[61–63] Moreover, it has been implicated in susceptibility to SLE and RA.[17,64] The nonsynonymous rs35677470 SNP in *DNASE1L3* exon 8 results in a Arg-to-Cys substitution at amino acid position 206 of the protein (p.Arg206Cys). This amino acid change results in the loss of a hydrogen bond in the protein (Figure S2), and in vitro studies have indicated that the protein constructed with this substitution lacks DNase activity; this evidence suggests a potential role in autoimmunity for this SNP.[65] Moreover, the observed genome-wide associations in *PXK*

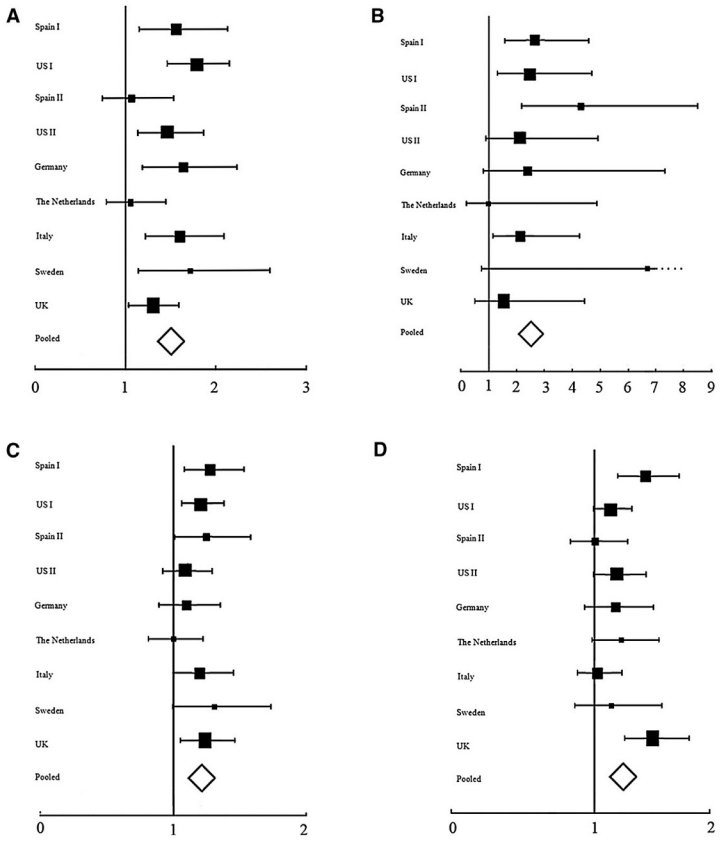


Figure 2. Forest Plots Showing the Population-Specific and Pooled Analyses of the Variants at Genome-wide or Suggestive Significance in the SSc versus Control Analysis
(A) rs35677470 in *DNASE1L3*.
(B) rs77583790 in the *SCHIP1-IL12A* locus.
(C) rs9373839 in *ATG5*.
(D) rs7130875 in the *TREH-DDX6* locus.

were in modest LD with and dependent on the *DNASE1L3* variant. Thus, we hypothesize that because of its functional implications, the rs35677470 polymorphism is likely to be the causal variant for the reported associations with SSc in the region. Considering that this variant is exclusive to populations with European ancestry and has a relatively low frequency and only two proxy SNPs ($r^2 > 0.8$), it is not surprising that it remained undiscovered and that fine mapping of the region was required for identifying this highly significant association. Interestingly, vascular injury and immune deregulation, which are both SSc hallmarks, are related to cell damage and deregulated apoptosis.[66] Therefore, the association between SSc (especially ACA^+ SSc) and the rs35677470 loss-of-function variant in *DNASE1L3* might provide a link between a defective apoptotic DNA breakdown and the production of ACA.

Utilizing the Immunochip fine-mapping platform, Eyre et al. confirmed the same SNP as the SNP most highly associated with RA in the region,[17] which supports the role of this gene and this specific variant as a common autoimmunity risk factor. Direct genotyping of this polymorphism in nonfamilial SLE might shed light on the relevance of this gene beyond the currently reported association in familial SLE cases.[63]

As noted in the Subjects and Methods, our study reached a high statistical power to detect associations with large effects despite low frequency of the minor allele. In fact, a rare variant in the intergenic region between *SCHIP1* and *IL12A* has been firmly associated with SSc, particularly the lcSSc subset. The same region was also identified in the Immunochip analysis in CeD.[16] Trynka et al. identified three different signals (rs76830965, rs1353248, and rs2561288) in this region.[16] However, these variants were not associated in our data ($p > 0.05$), and our lead SNP (rs77583790) was not linked to the previously mentioned polymorphisms. On the contrary, rs80014155, one of the four signals observed in PBC,[20] showed $r^2 = 0.97$ with our lead SNP. Therefore, both signals were equivalent. The location of the SSc signal in the region upstream of *IL12A* might point to an effect of this SNP (or a linked variant) in the regulation of *IL12A* expression. Moreover, different loci in the IL12 pathway have been associated with SSc (i.e., *IL12RB2* and *STAT4*), and altered levels of IL12 have been detected in SSc-affected individuals,[7,11,52,54,67] supporting a relevant role for this interleukin pathway in the disease. Nevertheless, because of the lack of functional data on this variant, its low frequency, and the high level of LD in this region, an unknown role of *SCHIP1* or other surrounding loci should not be discarded.

Regarding the association located in *ATG5*, the *ATG5*-encoded protein forms a complex with ATG12 and assists in autophagosomal elongation.[68] Remarkably, this signal at the genome-wide level links SSc with autophagy. Autophagy is a degradation pathway that mediates pathogen clearance and allows cells to degrade unwanted cytoplasmic material and to recycle nutrients.[68] It plays a key role in both innate and adaptive immune system development and responses and has been associated with a number of AIDs.[68] Along this line, several GWASs have found associations between the autophagy-related gene *ATG5* and SLE susceptibility, as well as risk of childhood and

Table 3. Non-HLA Loci Associated with SSc and Its Subsets and Identified through Immunochip Analysis

Locus	SNP	Chromosomal Region	Minor Allele	LD Region $r^2 > 0.9$	Region size (kb)	Comments	Phenotype	MAF Cases/Controls	p Value	OR
DNASE1L3	rs35677470	3p14	A	rs139325797, rs35677470	2	missense Arg > Cys	SSc	0.100/0.062	3.36×10^{-16}	1.47
							lcSSc	0.108/0.062	1.28×10^{-20}	1.62
							ACA^+	**0.133/0.062**	**4.25×10^{-31}**	**2.03**
SCHIP1-IL12A	rs77583790	3q25	A	rs183665802, rs186702062, rs184541632, rs185514275, rs80014155, rs77583790, rs114591798, rs73154557, rs187938609, rs200875363, rs73155915, rs73155982	1,641	intergenic	SSc	0.017/0.005	1.22×10^{-11}	2.57
							lcSSc	**0.016/0.005**	**1.53×10^{-11}**	**2.81**
							ACA^+	0.016/0.005	2.40×10^{-8}	2.76
ATG5	rs9373839	6q25	G	rs34599047, rs34582442, rs1322178, rs9386514, rs62422862, rs34936565, rs9398073, rs9399978, rs9373839, rs9372120, rs2299864, rs9386516, rs62422878, rs77791277, rs3827644, rs3804329, rs34843857, rs62422881, rs9398075, rs11752888, rs10484577, rs62422886, rs3804333, rs9373842, rs9372121, rs151290510, rs9373843, rs9398078, rs66469051, rs9373846, rs7763652	157	intronic	SSc	0.241/0.185	3.75×10^{-8}	1.19

The phenotypes with the most significant associations are shown in bold. Abbreviations are as follows: ACA^+, anticentromere-antibody-positive cases; ATA^+, antitopoisomerase-antibody-positive cases; dcSSc, diffuse cutaneous SSc; lcSSc, limited cutaneous SSc; LD, linkage disequilibrium; MAF, minor allele frequency; OR, odds ratio; p value, p value of the meta-analysis using the inverse variance method under a fixed effects model; and SSc, systemic sclerosis cases.

adult asthma and decline in lung function.[68] However, recently published fine-mapping studies in RA, CeD, PS, ATD, PBC, JIA, and narcolepsy did not find associations with this gene region.[16–24] Thus, this locus appears to be restricted to some, but not all, AIDs. In any case, the intronic location and the protein binding data of the variants in the *ATG5* region suggest that the functional meaning of these variants might be complex and involve distant genes (even *PRDM1* [MIM 603423], which maps downstream of *ATG5* and has been associated with different autoimmune diseases[69–71]). Moreover, there is evidence of a possible binding of CTCF in the rs7763652 region. This protein is involved in chromatin architecture and DNA-loop formation,[72] which brings up the possibility of a complex DNA structure in the region. Therefore, we consider that future research will be needed for determining the functional implications of this signal.

We also identified a suggestive association in the 11q23 intergenic region between *TREH* and *DDX6*, which has been shown to be a shared susceptibility region among several AIDs. Nevertheless, the different lead SNPs in *DDX6* in the Immunochip analysis of RA (rs4938573),[17] CeD (rs10892258),[16] and PBC (rs80065107)[20] were not linked with this SSc-suggestive variant. The gene product of *DDX6* (an RNA helicase that is important for efficient miRNA-induced gene silencing) has been shown to regulate vascular endothelial growth factor under hypoxic conditions, which might provide a clue to the vasculopathy and fibrosis that characterize SSc.[73]

The association of these loci provides genetic evidence of the possible role of defects in DNA elimination during apoptosis, introduces autophagy as a pathogenic mechanism, and reinforces the role of the IL12 pathway in the pathogenic processes that lead to SSc onset and disease progression. Therefore, we consider that the present study, together with previous knowledge about the genetic component of SSc, has contributed to the notion of this disorder as a complex condition in which several biological mechanisms, such as the innate immune response (Toll-like receptor and type I IFN pathways), the adaptive immune response (especially the IL12 pathway), tissue damage, fibrotic processes, and now DNA-clearance mechanisms during apoptosis, interact. However, neither these mechanisms nor the involved genetic loci should be considered independent compartments but rather pleiotropic players in a genetic and phenotypic continuum.

The analysis of genome-wide associations in already known loci resulted in the fine mapping of previously reported signals in *STAT4* and *IRF5*. *STAT4* is a shared susceptibility factor among RA, CeD, PBC, and JIA, but the most associated SNPs in the different conditions vary.[16,17,20–22] In the case of the *TNPO3-IRF5* locus, different variants have been associated with multiple autoimmune diseases and previous Immunochip analyses have reported different lead SNPs in RA and PBC.[17,20,21] In our study, the leading signals in these genes were highly linked with the RA-associated variants. The dense mapping of these established SSc

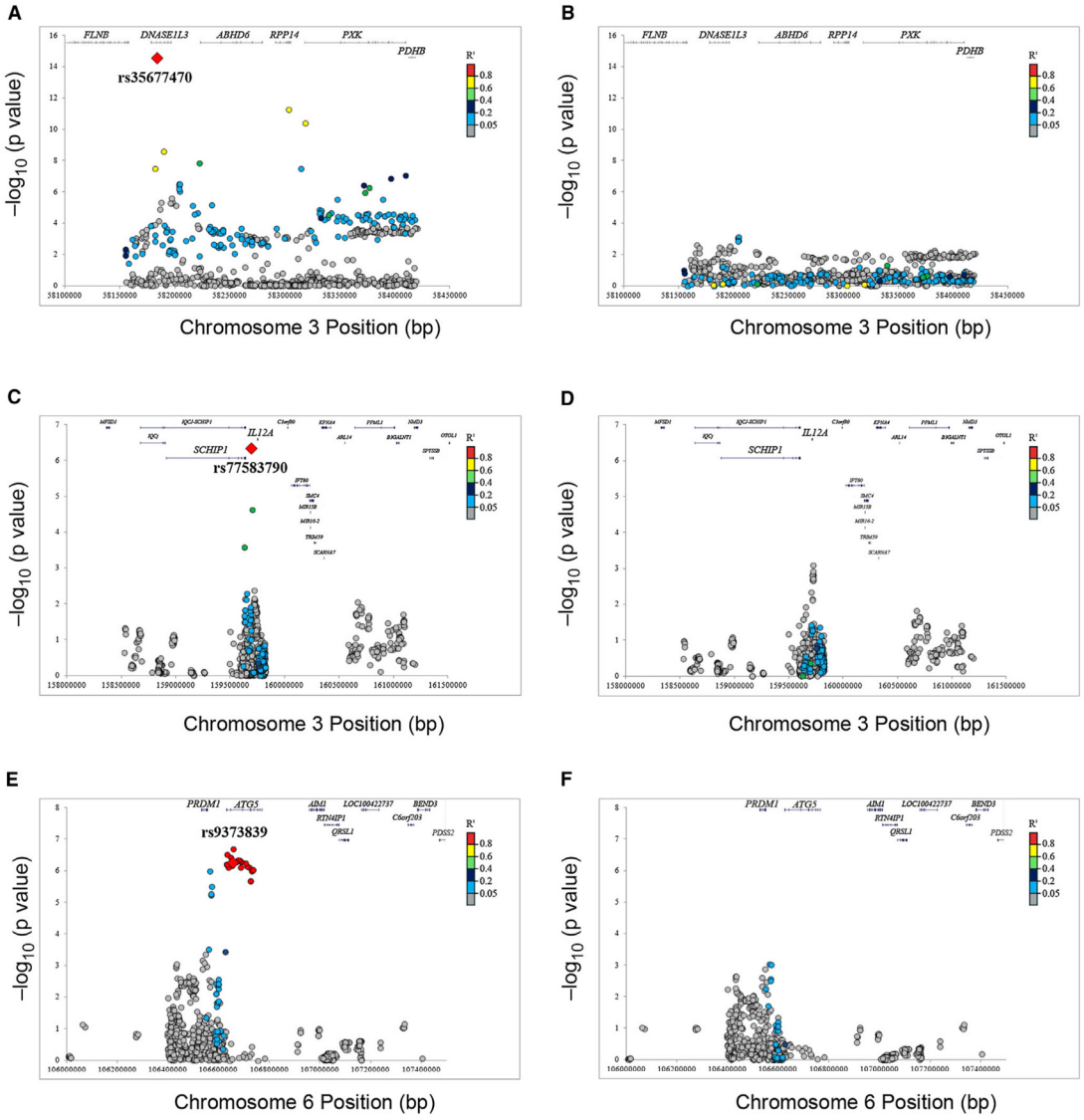


Figure 3. Regional Plots of the Associations Replicated at Genome-wide Significance in SSc Cases or Different Subgroups in the Overall Meta-analysis after Imputation

(A) *DNASE1L3* associations in the ACA^+ versus control comparison.
(B) *DNASE1L3* associations in the ACA^+ versus control comparison after conditioning on the lead variant (rs35677470).
(C) *SCHIP1-IL12A* locus associations in the lcSSc versus control comparison.
(D) *SCHIP1-IL12A* locus associations in the lcSSc versus control comparison after conditioning on the lead variant (rs77583790).
(E) *ATG5* associations in the dcSSc versus control comparison.
(F) *ATG5* associations in the dcSSc versus control comparison after conditioning on the lead variant (rs9373839).
p values correspond to the discovery phase.

susceptibility factors resulted in the confirmation of previous signals and narrowed association regions in these loci (35 kb in *STAT4* and 126 kb in *IRF5*, considering the SNPs showing $r^2 > 0.9$ with the lead variant in the locus according to the 1000 Genomes Project CEU population).[34]

In the case of *IL12RB2* and *TNIP1*, both loci showed suggestive associations but did not reach genome-wide significance. In previous reports, similar results were observed in the discovery phases in which these loci did not reach genome-wide significance and large replication cohorts

were needed for confirming these associations.[8,10,11] Therefore, we consider that our data reinforce the evidence of a role for these genes in SSc; however, larger cohorts should be fine mapped for narrowing down the observed signals.

As a result of design limitations, not all the known variants in *CD247* and *CSK* were included in the Immunochip (which was planned prior to the SSc GWAS), and these regions were therefore only covered by a tag SNP strategy. Unfortunately, after QC, the tagging SNPs for the reported associations in these regions were not maintained for further analysis. Thus, the initial signals were not covered, and this study could not address the confirmation of these previous findings.

It is worth mentioning that, as intended in the Immunochip design, our data showed an expected common genetic background between SSc and other AIDs. The genetic similarities between SSc and SLE are well known,[4,13] but our results reinforce the evidence of overlap with other AIDs, such as RA or PBC. The identification of common autoimmune risk factors is essential for the characterization of different pathogenic mechanisms that contribute to autoimmunity. Moreover, these common pathways might shed light on the origin of coautoimmunity and the preclinical stages of autoimmunity. In addition, relevant information can be inferred from the existence of different signals in the susceptibility regions or the lack of association between certain loci and specific diseases. Therefore, it has been suggested that platforms such as the Immunochip will help to dissect the pathogenic mechanisms underlying multiple disease states and lead to both more sensitive diagnostics and novel therapies.[74]

In summary, we have provided a comprehensive analysis of associations between the HLA region and SSc and its subphenotypes. Moreover, we report associations between *DNASE1L3*, the *SCHIP1-IL12A* locus, and *ATG5* and SSc and a suggestive association between the *TREH-DDX6* locus and SSc. The Immunochip-based interrogation of the analyzed cohorts revealed shared associations with other autoimmune diseases, which was the goal of the Immunochip Consortium, but also identified intriguing differences. Moreover, our data underline the need for direct genotyping of virtually all functional polymorphisms and rare variants in large cohorts for identifying variants that have strong effects in disease susceptibility but that might have been ignored thus far.

Supplemental Data

Supplemental Data include full affiliations for the Spanish Scleroderma Group, five figures, and five tables and can be found with this article online at http://www.cell.com/AJHG.

Consortia

The members of the Spanish Scleroderma Group are Norberto Ortego-Centeno, Raquel Ríos, José Luis Callejas, Nuria Navarrete, Rosa García Portales, María Teresa Camps, Antonio Fernández-Nebro, María F. González-Escribano, Julio Sánchez-Román, Francisco José García-Hernández, María Jesús Castillo, María Ángeles Aguirre, Inmaculada Gómez-Gracia, Benjamín Fernández-Gutiérrez, Luis Rodríguez-Rodríguez, Esther Vicente, José Luis Andreu, Mónica Fernández de Castro, Paloma García de la Peña, Francisco Javier López-Longo, Lina Martínez, Vicente Fonollosa, Gerard Espinosa, Carlos Tolosa, Anna Pros, Mónica Rodríguez Carballeira, Francisco Javier Narváez, Manel Rubio Rivas, Vera Ortiz Santamaría, Bernardino Díaz, Luis Trapiella, María del Carmen Freire, Adrián Sousa, María Victoria Egurbide, Patricia Fanlo Mateo, Luis Sáez-Comet, Federico Díaz, Vanesa Hernández, Emma Beltrán, José Andrés Román-Ivorra, Elena Grau, Juan José Alegre Sancho, Francisco J. Blanco García, Natividad Oreiro, and Luis Fernández Sueiro.

Acknowledgments

This study was supported by National Institutes of Health (NIH) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Centers of Research Translation (CORT) grant P50AR054144, NIH grant KL2RR024149-04, NIH NIAMS grant N01-AR02251, and NIH National Center for Research Resources grant 3UL1RR024148. This study was supported by GEN-FER from the Spanish Society of Rheumatology, SAF2009-11110 and SAF2012-34435 from the Spanish Ministry of Science, CTS-4977 from Junta de Andalucia, and RD08/0075 and RD12/0009/0004 from the Redes Temáticas de Investigación Cooperativa en Salud (Red de Investigación en Inflamación y Enfermedades Reumáticas). We are indebted to the Immunochip Consortium for provision of the control data. We are grateful to Julio Charles, Marilyn Perry, Suzanne S. Taillefer (National Study Coordinator for the Canadian Scleroderma Research Group), Sofia Vargas, Sonia García, and Gema Robledo. We also thank the Scleroderma Foundation, the Spanish Scleroderma Patient Group, the Canadian Scleroderma Patient Group, the Federation of European Scleroderma Associations, National DNA Bank Carlos III (University of Salamanca), the European League against Rheumatism Scleroderma Trials and Research group, the German Network of Systemic Sclerosis, and the affected individuals, friends, spouses, and others who generously provided the samples for these studies.

Received: July 19, 2013
Accepted: December 3, 2013
Published: January 2, 2014

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://www.1000genomes.org/
BEAGLE, http://faculty.washington.edu/browning/beagle/beagle.html
Blood eQTL Browser, http://genenetwork.nl/bloodeqtlbrowser/
ENCODE, http://genome.ucsc.edu/ENCODE/
Genevar (Gene Expression Variation), http://www.sanger.ac.uk/resources/software/genevar/
HLA Nomenclature, http://hla.alleles.org/
HUGO Gene Nomenclature Committee, http://www.genenames.org/
IMPUTE2, http://mathgen.stats.ox.ac.uk/impute/impute_v2.html
LocusZoom, http://csg.sph.umich.edu/locuszoom/

NCBI, http://www.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM), http://omim.org/
PLINK, http://pngu.mgh.harvard.edu/~purcell/plink/
PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/
RegulomeDB, http://RegulomeDB.org/
UCSF Chimera, http://www.cgl.ucsf.edu/chimera/

Accession Numbers

The dbGaP accession number for the genotype information of the US cohort presented in this paper is phs000357.v2.p2.

References

1. Bolster, M.B., and Silver, R.S. (2011). Clinical features of systemic sclerosis. In Rheumatology, *Fifth Edition*, M.C. Hochberg, A.J. Silman, J.S. Smolen, M.E. Weinblatt, and M.H. Weisman, eds. (Philadelphia: Mosby, Elsevier), pp. 1373–1386.
2. 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.
3. Steen, V.D. (2005). Autoantibodies in systemic sclerosis. Semin. Arthritis Rheum. *35*, 35–42.
4. Martín, J.E., Bossini-Castillo, L., and Martín, J. (2012). Unraveling the genetic component of systemic sclerosis. Hum. Genet. *131*, 1023–1037.
5. Mayes, M.D. (2012). The genetics of scleroderma: looking into the postgenomic era. Curr. Opin. Rheumatol. *24*, 677–684.
6. Assassi, S., Radstake, T.R., Mayes, M.D., and Martin, J. (2013). Genetics of scleroderma: implications for personalized medicine? BMC Med. *11*, 9.
7. 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.; Spanish Scleroderma Group (2010). Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat. Genet. *42*, 426–429.
8. Allanore, Y., Saad, M., Dieudé, P., Avouac, J., Distler, J.H., Amouyel, P., Matucci-Cerinic, M., Riemekasten, G., Airo, P., Melchers, I., et al. (2011). Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet. *7*, e1002091.
9. 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.; Spanish Scleroderma Group (2011). Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet. *7*, e1002178.
10. Bossini-Castillo, L., Martin, J.E., Broen, J., Simeon, C.P., Beretta, L., Gorlova, O.Y., Vonk, M.C., Ortego-Centeno, N., Espinosa, G., Carreira, P., et al.; Spanish Scleroderma Group (2013). Confirmation of TNIP1 but not RHOB and PSORS1C1 as systemic sclerosis risk factors in a large independent replication study. Ann. Rheum. Dis. *72*, 602–607.
11. Bossini-Castillo, L., Martin, J.E., Broen, J., Gorlova, O., Simeón, C.P., Beretta, L., Vonk, M.C., Callejas, J.L., Castellví, I., Carreira, P., et al.; Spanish Scleroderma Group (2012). A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. Hum. Mol. Genet. *21*, 926–933.
12. 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.; Spanish Scleroderma Group (2012). Identification of CSK as a systemic sclerosis genetic risk factor through Genome wide Association Study follow-up. Hum. Mol. Genet. *21*, 2825–2835.
13. Martin, J.E., Assassi, S., Diaz-Gallo, L.M., Broen, J.C., Simeon, C.P., Castellvi, I., Vicente-Rabaneda, E., Fonollosa, V., Ortego-Centeno, N., González-Gay, M.A., et al.; Spanish Scleroderma Group; SLEGEN consortium; U.S. Scleroderma GWAS group; BIOLUPUS (2013). A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum. Mol. Genet. *22*, 4021–4029.
14. Cortes, A., and Brown, M.A. (2011). Promise and pitfalls of the Immunochip. Arthritis Res. Ther. *13*, 101.
15. Polychronakos, C. (2011). Fine points in mapping autoimmunity. Nat. Genet. *43*, 1173–1174.
16. Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al.; Spanish Consortium on the Genetics of Coeliac Disease (CEGEC); PreventCD Study Group; Wellcome Trust Case Control Consortium (WTCCC) (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat. Genet. *43*, 1193–1201.
17. Eyre, S., Bowes, J., Diogo, D., Lee, A., Barton, A., Martin, P., Zhernakova, A., Stahl, E., Viatte, S., McAllister, K., et al.; Biologics in Rheumatoid Arthritis Genetics and Genomics Study Syndicate; Wellcome Trust Case Control Consortium (2012). High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. Nat. Genet. *44*, 1336–1340.
18. Cooper, J.D., Simmonds, M.J., Walker, N.M., Burren, O., Brand, O.J., Guo, H., Wallace, C., Stevens, H., Coleman, G., Franklyn, J.A., et al.; Wellcome Trust Case Control Consortium (2012). Seven newly identified loci for autoimmune thyroid disease. Hum. Mol. Genet. *21*, 5202–5208.
19. Tsoi, L.C., Spain, S.L., Knight, J., Ellinghaus, E., Stuart, P.E., Capon, F., Ding, J., Li, Y., Tejasvi, T., Gudjonsson, J.E., et al.; Collaborative Association Study of Psoriasis (CASP); Genetic Analysis of Psoriasis Consortium; Psoriasis Association Genetics Extension; Wellcome Trust Case Control Consortium 2 (2012). Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat. Genet. *44*, 1341–1348.
20. Liu, J.Z., Almarri, M.A., Gaffney, D.J., Mells, G.F., Jostins, L., Cordell, H.J., Ducker, S.J., Day, D.B., Heneghan, M.A., Neuberger, J.M., et al.; UK Primary Biliary Cirrhosis (PBC) Consortium; Wellcome Trust Case Control Consortium 3 (2012). Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. Nat. Genet. *44*, 1137–1141.
21. Juran, B.D., Hirschfield, G.M., Invernizzi, P., Atkinson, E.J., Li, Y., Xie, G., Kosoy, R., Ransom, M., Sun, Y., Bianchi, I., et al.; Italian PBC Genetics Study Group (2012). Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. Hum. Mol. Genet. *21*, 5209–5221.
22. Hinks, A., Cobb, J., Marion, M.C., Prahalad, S., Sudman, M., Bowes, J., Martin, P., Comeau, M.E., Sajuthi, S., Andrews, R.,

et al.; Boston Children's JIA Registry; British Society of Paediatric and Adolescent Rheumatology (BSPAR) Study Group; Childhood Arthritis Prospective Study (CAPS); Childhood Arthritis Response to Medication Study (CHARMS); German Society for Pediatric Rheumatology (GKJR); JIA Gene Expression Study; NIAMS JIA Genetic Registry; TREAT Study; United Kingdom Juvenile Idiopathic Arthritis Genetics Consortium (UKJIAGC) (2013). Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. Nat. Genet. *45*, 664–669.

23. Liu, J.Z., Hov, J.R., Folseraas, T., Ellinghaus, E., Rushbrook, S.M., Doncheva, N.T., Andreassen, O.A., Weersma, R.K., Weismüller, T.J., Eksteen, B., et al.; UK-PSCSC Consortium; International IBD Genetics Consortium; International PSC Study Group (2013). Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. Nat. Genet. *45*, 670–675.
24. Faraco, J., Lin, L., Kornum, B.R., Kenny, E.E., Trynka, G., Einen, M., Rico, T.J., Lichtner, P., Dauvilliers, Y., Arnulf, I., et al. (2013). ImmunoChip study implicates antigen presentation to T cells in narcolepsy. PLoS Genet. *9*, e1003270.
25. Ellinghaus, D., Baurecht, H., Esparza-Gordillo, J., Rodríguez, E., Matanovic, A., Marenholz, I., Hübner, N., Schaarschmidt, H., Novak, N., Michel, S., et al. (2013). High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. Nat. Genet. *45*, 808–812.
26. Cortes, A., Hadler, J., Pointon, J.P., Robinson, P.C., Karaderi, T., Leo, P., Cremin, K., Pryce, K., Harris, J., Lee, S., et al.; International Genetics of Ankylosing Spondylitis Consortium (IGAS); Australo-Anglo-American Spondyloarthritis Consortium (TASC); Groupe Française d'Etude Génétique des Spondylarthrites (GFEGS); Nord-Trøndelag Health Study (HUNT); Spondyloarthritis Research Consortium of Canada (SPARCC); Wellcome Trust Case Control Consortium 2 (WTCCC2) (2013). Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. Nat. Genet. *45*, 730–738.
27. Saruhan-Direskeneli, G., Hughes, T., Aksu, K., Keser, G., Coit, P., Aydin, S.Z., Alibaz-Oner, F., Kamalı, S., Inanc, M., Carette, S., et al. (2013). Identification of multiple genetic susceptibility loci in Takayasu arteritis. Am. J. Hum. Genet. *93*, 298–305.
28. (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.
29. Reveille, J.D., Fischbach, M., McNearney, T., Friedman, A.W., Aguilar, M.B., Lisse, J., Fritzler, M.J., Ahn, C., and Arnett, F.C.; GENISOS Study Group (2001). Systemic sclerosis in 3 US ethnic groups: a comparison of clinical, sociodemographic, serologic, and immunogenetic determinants. Semin. Arthritis Rheum. *30*, 332–346.
30. 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.
31. Purcell, S.N.B., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., and Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. *81*, 559–575.
32. de Bakker, P.I., Ferreira, M.A., Jia, X., Neale, B.M., Raychaudhuri, S., and Voight, B.F. (2008). Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. *17* (R2), R122–R128.
33. Howie, B.N., Donnelly, P., and Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. *5*, e1000529.
34. Abecasis, G.R., Altshuler, D., Auton, A., Brooks, L.D., Durbin, R.M., Gibbs, R.A., Hurles, M.E., and McVean, G.A.; 1000 Genomes Project Consortium (2010). A map of human genome variation from population-scale sequencing. Nature *467*, 1061–1073.
35. Brown, W.M., Pierce, J., Hilner, J.E., Perdue, L.H., Lohman, K., Li, L., Venkatesh, R.B., Hunt, S., Mychaleckyj, J.C., and Deloukas, P.; Type 1 Diabetes Genetics Consortium (2009). Overview of the MHC fine mapping data. Diabetes Obes. Metab. *11* (*Suppl 1*), 2–7.
36. de Bakker, P.I., McVean, G., Sabeti, P.C., Miretti, M.M., Green, T., Marchini, J., Ke, X., Monsuur, A.J., Whittaker, P., Delgado, M., et al. (2006). A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat. Genet. *38*, 1166–1172.
37. Browning, S.R., and Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. *81*, 1084–1097.
38. Jia, X., Han, B., Onengut-Gumuscu, S., Chen, W.M., Concannon, P.J., Rich, S.S., Raychaudhuri, S., and de Bakker, P.I. (2013). Imputing amino acid polymorphisms in human leukocyte antigens. PLoS ONE *8*, e64683.
39. Arnett, F.C., Gourh, P., Shete, S., Ahn, C.W., Honey, R.E., Agarwal, S.K., Tan, F.K., McNearney, T., Fischbach, M., Fritzler, M.J., et al. (2010). Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. Ann. Rheum. Dis. *69*, 822–827.
40. Karp, D.R., Marthandan, N., Marsh, S.G., Ahn, C., Arnett, F.C., Deluca, D.S., Diehl, A.D., Dunivin, R., Eilbeck, K., Feolo, M., et al. (2010). Novel sequence feature variant type analysis of the HLA genetic association in systemic sclerosis. Hum. Mol. Genet. *19*, 707–719.
41. Adzhubei, I., Jordan, D.M., and Sunyaev, S.R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. Curr. Protoc. Hum. Genet. *Chapter 7*, 20.
42. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.
43. Yang, T.P., Beazley, C., Montgomery, S.B., Dimas, A.S., Gutierrez-Arcelus, M., Stranger, B.E., Deloukas, P., and Dermitzakis, E.T. (2010). Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics *26*, 2474–2476.
44. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans

eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.

45. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.
46. Carmona, F.D., Martin, J.E., Beretta, L., Simeón, C.P., Carreira, P.E., Callejas, J.L., Fernández-Castro, M., Sáez-Comet, L., Beltrán, E., Camps, M.T., et al.; Spanish Scleroderma Group (2013). The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. PLoS ONE *8*, e54419.
47. Sharif, R., Mayes, M.D., Tan, F.K., Gorlova, O.Y., Hummers, L.K., Shah, A.A., Furst, D.E., Khanna, D., Martin, J., Bossini-Castillo, L., et al. (2012). IRF5 polymorphism predicts prognosis in patients with systemic sclerosis. Ann. Rheum. Dis. *71*, 1197–1202.
48. Dieude, P., Dawidowicz, K., Guedj, M., Legrain, Y., Wipff, J., Hachulla, E., Diot, E., Sibilia, J., Mouthon, L., Cabane, J., et al. (2010). Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. J. Rheumatol. *37*, 987–992.
49. Ito, I., Kawaguchi, Y., Kawasaki, A., Hasegawa, M., Ohashi, J., Hikami, K., Kawamoto, M., Fujimoto, M., Takehara, K., Sato, S., et al. (2009). Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. Arthritis Rheum. *60*, 1845–1850.
50. Dieudé, P., Guedj, M., Wipff, J., Avouac, J., Fajardy, I., Diot, E., Granel, B., Sibilia, J., Cabane, J., Mouthon, L., et al. (2009). Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. Arthritis Rheum. *60*, 225–233.
51. Gourh, P., Agarwal, S.K., Divecha, D., Assassi, S., Paz, G., Arora-Singh, R.K., Reveille, J.D., Shete, S., Mayes, M.D., Arnett, F.C., and Tan, F.K. (2009). Polymorphisms in TBX21 and STAT4 increase the risk of systemic sclerosis: evidence of possible gene-gene interaction and alterations in Th1/Th2 cytokines. Arthritis Rheum. *60*, 3794–3806.
52. Dieudé, P., Guedj, M., Wipff, J., Ruiz, B., Hachulla, E., Diot, E., Granel, B., Sibilia, J., Tiev, K., Mouthon, L., et al. (2009). STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis. Arthritis Rheum. *60*, 2472–2479.
53. Tsuchiya, N., Kawasaki, A., Hasegawa, M., Fujimoto, M., Takehara, K., Kawaguchi, Y., Kawamoto, M., Hara, M., and Sato, S. (2009). Association of STAT4 polymorphism with systemic sclerosis in a Japanese population. Ann. Rheum. Dis. *68*, 1375–1376.
54. Rueda, B., Broen, J., Simeon, C., Hesselstrand, R., Diaz, B., Suárez, H., Ortego-Centeno, N., Riemekasten, G., Fonollosa, V., Vonk, M.C., et al. (2009). The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum. Mol. Genet. *18*, 2071–2077.
55. Bellemare-Pelletier, A., Tremblay, J., Beaulieu, S., Boulassel, M.R., Routy, J.P., Massie, B., Lapointe, R., and Thibodeau, J. (2005). HLA-DO transduced in human monocyte-derived dendritic cells modulates MHC class II antigen processing. J. Leukoc. Biol. *78*, 95–105.
56. Bahram, S., Arnold, D., Bresnahan, M., Strominger, J.L., and Spies, T. (1991). Two putative subunits of a peptide pump encoded in the human major histocompatibility complex class II region. Proc. Natl. Acad. Sci. USA *88*, 10094–10098.
57. Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L., and Wiley, D.C. (1993). Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature *364*, 33–39.
58. Tollefsen, S., Hotta, K., Chen, X., Simonsen, B., Swaminathan, K., Mathews, I.I., Sollid, L.M., and Kim, C.Y. (2012). Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. J. Biol. Chem. *287*, 13611–13619.
59. Dai, S., Murphy, G.A., Crawford, F., Mack, D.G., Falta, M.T., Marrack, P., Kappler, J.W., and Fontenot, A.P. (2010). Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proc. Natl. Acad. Sci. USA *107*, 7425–7430.
60. Baron, W.F., Pan, C.Q., Spencer, S.A., Ryan, A.M., Lazarus, R.A., and Baker, K.P. (1998). Cloning and characterization of an actin-resistant DNase I-like endonuclease secreted by macrophages. Gene *215*, 291–301.
61. Shiokawa, D., and Tanuma, S. (2001). Characterization of human DNase I family endonucleases and activation of DNase gamma during apoptosis. Biochemistry *40*, 143–152.
62. Errami, Y., Naura, A.S., Kim, H., Ju, J., Suzuki, Y., El-Bahrawy, A.H., Ghonim, M.A., Hemeida, R.A., Mansy, M.S., Zhang, J., et al. (2013). Apoptotic DNA fragmentation may be a cooperative activity between caspase-activated deoxyribonuclease and the poly(ADP-ribose) polymerase-regulated DNAS1L3, an endoplasmic reticulum-localized endonuclease that translocates to the nucleus during apoptosis. J. Biol. Chem. *288*, 3460–3468.
63. Okamoto, M., Okamoto, N., Yashiro, H., Shiokawa, D., Sunaga, S., Yoshimori, A., Tanuma, S., and Kitamura, D. (2005). Involvement of DNase gamma in the resected double-strand DNA breaks in immunoglobulin genes. Biochem. Biophys. Res. Commun. *327*, 76–83.
64. Al-Mayouf, S.M., Sunker, A., Abdwani, R., Abrawi, S.A., Almurshedi, F., Alhashmi, N., Al Sonbul, A., Sewairi, W., Qari, A., Abdallah, E., et al. (2011). Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. Nat. Genet. *43*, 1186–1188.
65. Ueki, M., Takeshita, H., Fujihara, J., Iida, R., Yuasa, I., Kato, H., Panduro, A., Nakajima, T., Kominato, Y., and Yasuda, T. (2009). Caucasian-specific allele in non-synonymous single nucleotide polymorphisms of the gene encoding deoxyribonuclease I-like 3, potentially relevant to autoimmunity, produces an inactive enzyme. Clin. Chim. Acta *407*, 20–24.
66. Gu, Y.S., Kong, J., Cheema, G.S., Keen, C.L., Wick, G., and Gershwin, M.E. (2008). The immunobiology of systemic sclerosis. Semin. Arthritis Rheum. *38*, 132–160.
67. Sato, S., Hanakawa, H., Hasegawa, M., Nagaoka, T., Hamaguchi, Y., Nishijima, C., Komatsu, K., Hirata, A., and Takehara, K. (2000). Levels of interleukin 12, a cytokine of type 1 helper T cells, are elevated in sera from patients with systemic sclerosis. J. Rheumatol. *27*, 2838–2842.
68. Choi, A.M., Ryter, S.W., and Levine, B. (2013). Autophagy in human health and disease. N. Engl. J. Med. *368*, 1845–1846.
69. 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.
70. Raychaudhuri, S., Thomson, B.P., Remmers, E.F., Eyre, S., Hinks, A., Guiducci, C., Catanese, J.J., Xie, G., Stahl, E.A., Chen, R., et al.; BIRAC Consortium; YEAR Consortium (2009). Genetic

variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat. Genet. *41*, 1313–1318.
71. Ellinghaus, D., Zhang, H., Zeissig, S., Lipinski, S., Till, A., Jiang, T., Stade, B., Bromberg, Y., Ellinghaus, E., Keller, A., et al. (2013). Association between variants of PRDM1 and NDP52 and Crohn's disease, based on exome sequencing and functional studies. Gastroenterology *145*, 339–347.
72. Merkenschlager, M., and Odom, D.T. (2013). CTCF and cohesin: linking gene regulatory elements with their targets. Cell *152*, 1285–1297.
73. de Vries, S., Naarmann-de Vries, I.S., Urlaub, H., Lue, H., Bernhagen, J., Ostareck, D.H., and Ostareck-Lederer, A. (2013). Identification of DEAD-box RNA helicase 6 (DDX6) as a cellular modulator of vascular endothelial growth factor expression under hypoxia. J. Biol. Chem. *288*, 5815–5827.
74. Voight, B.F., and Cotsapas, C. (2012). Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity. Curr. Opin. Immunol. *24*, 552–557.

DISCUSSION

Novel genetic findings

The first Immunochip-based dense genotyping performed in SSc patients was included in this thesis [23]. Interestingly, three novel genome-wide level associations were identified in this study: *DNASE1L3*, *ATG5* and *SCHIP1/IL12A* [23]. Moreover, a novel HLA imputation method allowed us to describe a model composed of six polymorphic amino acid positions and seven SNPs that explained the associations in this region [23, 59, 105]. In addition, we confirmed at the GWAS significance level the two firmest non-HLA SSc risk factors, *STAT4* and *IRF5*, and assigned the previously reported signal in *PXK* to the functional polymorphism discovered in the *DNASE1L3 locus* [23, 103]. Likewise, a suggestively associated *locus*, *DDX6*, was proposed [23]. Furthermore, previously SSc-related *loci* showed suggestive second tier associations (p-value < $5x10^{-5}$) in our data, such as: *BLK*, *TNFSF4* and *ITGAM* [78, 103, 117-122]. Besides our Immunochip study, the article compendium that integrates this doctoral thesis led to the identification of *IL12RB2* and the confirmation of *TNIP1* as SSc

genetic markers of also at the $\alpha = 5x10^{-8}$ level [25, 27]. Likewise, the role of *CD226* in SSc-related ILD was strengthened and a risk haplotype for this devastating clinical outcome was described [26]. This thesis also contributed to the rejection of *RHOB* and *PSORS1C1* as important independent players in SSc pathogenesis and the discard of *KCNA5* as a PAH marker [24, 25].

Remarkably, the presented pieces of work, together with previous and contemporary studies, have identified a total of 15 SSc risk *loci* reaching genome-wide significance associations.

An in-depth look into the HLA region

This doctoral thesis includes the most comprehensive SSc HLA association dissection performed to date, which led to the identification of an explanatory model of the region reaching the amino acid level resolution in Mayes *et al*. [23]. In this line, we would like to emphasize that although the HLA association with SSc had been long known and the associated genes had been already reported, the causal reasons for this association remain obscure. Several HLA associations with SSc have been previously described (they are illustrated in *Table 4*). Ancestry has been proven

to influence the association of classical HLA alleles associated with SSc and its different subtypes [65, 66]. Therefore, due to the ethnical origin of all the individuals included the current studies, we will focus in the associations described to date with white populations as reported in *Table 4*. Briefly, the HLA-DRB1*1104/ DQA1*0501/ DQB1*0301 haplotype and those HLA-DQB1 alleles that encode a non-leucine residue at position 26 (HLA-DQB1 26 epi) have been shown to be more frequent in SSc patients [66]. Furthermore, the *HLA-DRB1*0701*/ *DQA1*0201*/ *DQB1*0202* and the *DRB1*1501* classical haplotypes have been firmly associated with protective OR to SSc [66, 74]. In addition, the most significant SNP association in the two published SSc GWAS corresponded to the *HLA-DQB1 locus* (rs6457617 in Radtake *et al*. and rs9275224 in Allanore *et al*.) [61, 63]. However, it is clear now that the different subsets defined by the patient antibody status show differential associations in the HLA genes [66, 72]. Thus, ACA+ patients have been reported to have increased frequency of the *HLA-DQA1*0401*, *HLA-DQA1*0101* alleles [66]. But also different variants in the *HLA-DQB1* (*HLA-DQB1*0501*, rs6457617, rs9275390) and *HLA-DRB1* (*HLA- DRB1*0101*, *HLA-DRB1*0801*)

Table 4. Previously described HLA region associated alleles and their relation with the amino acid model in Mayes *et al*.

Locus	Variation	Subgroup	Effect	$P_{Immunochip}$	Conditioned by model*
HLA-DRB1	04:04	SSc	Risk	0.85	NA
HLA-DRB1	07:01†	SSc	Protection	**4.58E-17**	0.74
HLA-DRB1	11:04‡	SSc	Risk	**2.65E-11**	1.43E-02
HLA-DRB1	15:01	SSc	Protection	0.88	NA
HLA-DQB1	02:02†	SSc	Protection	**1.37E-14**	0.97
HLA-DQB1	03:01‡	SSc	Risk	5.56E-07	0.21
HLA-DQB1	06:02	SSc	Protective	0.76	NA
HLA-DQB1	26 epi	SSc	Risk	Not included	NA
HLA-DQB1	rs6457617	SSc	Risk	**7.62E-13**	0.33
HLA-DQB1	rs9275224	SSc	Risk	**1.46E-12**	0.45
HLA-DQA1	01:02	SSc	Protective	0.61	NA
HLA-DQA1	02:01†	SSc	Protection	**6.58E-17**	*A*
HLA-DQA1	05:01‡	SSc	Risk	0.07	NA
HLA-DRB1	01:01	ACA	Risk	**6.54E-17**	1.21E-04
HLA-DRB1	04:01	ACA	Risk	1.11E-04	1.86E-05
HLA-DRB1	04:04	ACA	Risk	0.11	NA
HLA-DQB1	03:01	ACA	Risk	0.30	NA
HLA-DQB1	05:01	ACA	Risk	**4.30E-18**	1.43E-05
HLA-DQB1	26 epi	ACA	Risk	Not included	NA
HLA-DQB1	rs6457617	ACA	Risk	**1.59E-23**	3.91E-06
HLA-DQB1	rs9275390	ACA	Risk	**2.34E-28**	0.002451
HLA-DQA1	01:01	ACA	Risk	**1.15E-17**	2.09E-05
HLA-DQA1	04:01	ACA	Risk	**5.82E-20**	*a*

*P-values after conditioning by the model corresponding to the subgroup were the initial association was described are reported; a: Determined by the DQA1_69_T / DQA1_69_L residues; Not included: this variant was not included in the Immunochip; NA: not applicable since no association was found in the Immunochip analysis.

Table 4 (Cont.)

Locus	Variation	Subgroup	Effect	$P_{Immunochip}$	Conditioned by model*
HLA-DRB1	11:04‡	ATA	Risk	**1.98E-26**	1.09E-05
HLA-DQB1	03:01‡	ATA	Risk	**3.38E-12**	0.36
HLA-DQB1	05:01	ATA	Protective	1.16E-02	2.04E-02
HLA-DQA1	05:01‡	ATA	Risk	3.33E-03	0.22
HLA-DQA1/ DRB1	rs3129763	ATA	Risk	5.77E-04	0.36
HLA-DPB1	13:01	ATA	Risk	**1.23E-10**	0.01
HLA-DPA1/ DPB1	rs987870	ATA	Risk	6.16E-08	2.19E-02
HLA-DPA1/ DPB1	rs3135021	ATA	Risk	1.40E-03	0.29
HLA-DPA1/ DPB1	rs6901221	ATA	Risk	NA	NA
HLA-DRA	rs3129882	ATA	Risk	**7.03E-12**	0.58

*P-values after conditioning by the model corresponding to the subgroup were the initial association was described are reported; a: Determined by the DQA1_69_T / DQA1_69_L residues; Not included: this variant was not included in the Immunochip; NA: not applicable since no association was found in the Immunochip analysis.

genes have been associated with ACA+ SSc patients [66, 72, 74, 75, 123]. On the other hand, the most prominent association with ATA+ SSc patients corresponds to the *HLA-DPB1* locus: *HLA-DPB1*1301* and the combination of rs987870, rs3135021 and rs6901221 [66, 72, 75]. However, the *HLA-DRB1*1104/ DQA1*0501/ DQB1*0301* haplotype has been also reported to be significantly increased in this subset [66, 71, 72, 74, 75, 123].

The analysis of AD susceptibility caused by the HLA regions is a complex problem [124]. The linkage disequilibrium is extremely high in this region, creating long haplotypes. Furthermore, the region is greatly rich in immune-related genes and contains MHC class I and class II genes, *TNF* and complement genes, *TAP* and *HLA-DM* genes [124]. Thus, it is difficult to discern which gene/s is/are responsible for disease susceptibility [124]. Furthermore, the gene association in the HLA region is often not unique and multiple genes within the HLA *locus* can contribute to disease risk and modify the disease risk [124].

The HLA region is the most polymorphic region in the human genome, especially in the peptide binding groove region which allows the presentation of a wide range of peptides [124, 125]. Several, and probably simultaneous, mechanisms may connect AD susceptibility and HLA-mediated peptide presentation: presentation of key auto-antigens enabled by certain polymorphisms, poor presentation of essential epitopes during thymic lymphocyte maturation which results the escape of auto-reactive cells, certain polymorphisms create especial peptide-binding groove conformations that may amplify the immune response, protective

HLA haplotypes may influence the T_{reg} repertoire and keep inflammation under control, disease-associated HLA alleles present also viral and microbial peptides that expand self-reactive T-cell clones [124]. The core binding motif of both HLA class I and HLA class II molecules is approximately nine amino acids long [126]. But the ends of the HLA class II peptide binding groove are open and it can accommodate longer peptides [127, 128]. Therefore, the development of predictive algorithms for peptide binding in HLA class II is a difficult task since the peptides need to be correctly aligned before identifying the nine amino acid core-binding motif [129]. As it was reported in Mayes *et al*., our Immunochip-based approach allowed us to identify a six amino acid model, with serological subgroup specificity that explained the observed association in our data, including the association of classical HLA alleles [23]. Remarkably, both polymorphic amino acid positions associated with the ACA+ subset and two of the four associated with ATA+ SSc patients were located in the peptide binding groove of the corresponding HLA-DRB1, HLA-DQA1 and HLA-DPB1 proteins [23, 130-132]. As it is shown in *Figure 8* and in *Table 4*, the previously described classical HLA alleles were not only conditioned by our model but also, in the majority of them, the

presence of risk/protection amino acids as described in our model was concordant with the previously described associated subgroup. Consequently, those alleles associated with ACA+ risk have ACA+ risk amino acids or ATA+ protective amino acids from our model, and the opposite situation was found in the ATA+ subgroup. We consider that this overlap supports the consistency of our model (*Figure 8* and *Table 4*). However, the association of some alleles such as *HLA-DRB1*0404* and *HLA-DQA1*0401* are statistically conditioned but not "functionally" explained by our model, which suggests that additional amino acids might play an important role and will be discovered increasing the statistical power by using larger sample populations.

In addition, according to the data displayed in the HLA-DRB1 sequence logos (which are the most accurately predicted) accessible in the MHC Motif Viewer, it is clear that the amino acid preferences of the binding anchor in ACA+ and ATA+ alleles are different (*Figure 8*) [133]. The predictions used in the MHC Motif Viewer web tool were generated using the NetMHCIIpan peptide-HLA binding prediction method [134]. As it can be observed in

Figure 8. Previously reported classical four digit HLA alleles associated with either anti-centromere antibody positive (ACA+) or anti-topoisomerase positive (ATA+) systemic sclerosis patients and their relation with the novel HLA amino acid based model described in Mayes *et al*. 2014 (A). HLA-DRB1 sequence logos for the associated alleles are also shown (B).

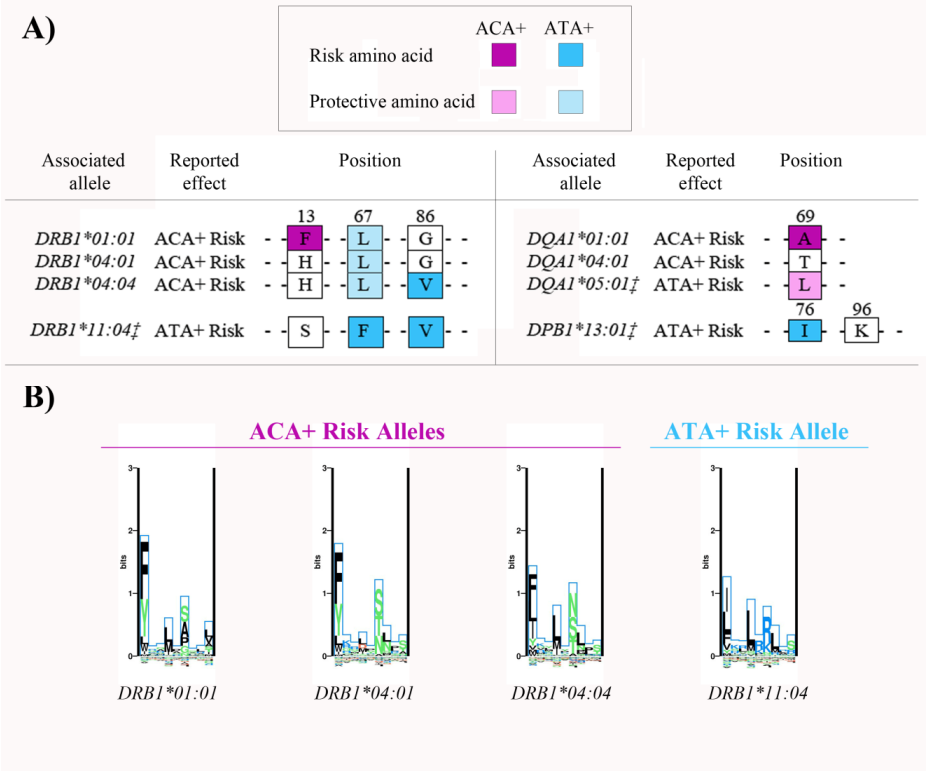


Figure 8, the 6th position in the peptide nonamere corresponds to a neutral amino acid in the ACA+ alleles while it corresponds to a basic amino acid in the ATA+ associated HLA-DRB1*1104 (*Figure 8*). These results add evidence to the already reported ACA+ versus ATA+ subset differences in the HLA region. The serological subset restricted association of the HLA region was evident at the classical HLA allele level; then, it was proposed at the SNP level in Gorlova *et al*. and finally confirmed at the amino acid level in the presented Mayes *et al*. study [23, 66, 72].

In addition to amino acid positions our HLA model included also independent associations of several SNPs [23]. Interestingly, this SNPs with cis eQTLs showing FDR > 0.05 in Blood eQTL [115]: rs12528892 with *TAP2*, *PSMB9* and *HLA-DMB*, rs17500468 with *TAP2*, *PSMB9*, *TAP1*; rs9277052 with *HLA-DPB1*, *HSD17B8*, *HLA-DMA*, *PSMB9*; rs24427196, with *HCG27*; rs4713605 with *HLA-DPB1*, *HSD17B8*, *HLA-DMA*, *HLA-DOA*, *TAP2*, *BRD2*; rs443623 with *B3GALT4*, *TAP2* and *HLA-DMB*. The existence of reported eQTLs for genes such as *HLA-DMOA*, *HLA-DMA* and *HLA-DMB*, but also for *TAP1* and *TAP2* and even immunoproteasome subunits such as *PSMB9* suggests that

SSc susceptibility may not lean only in HLA class II alleles but also imply the whole process of peptide degradation and loading in the HLA complexes [135-139]. Moreover, it suggests a possible role for HLA class I unidentified to date. In addition, the *HSD17B8 locus*, which is involved the regulation concentration of biologically active estrogens and androgens, may provide a clue for the already described SSc sex bias [140].

We are aware that despite the fact that ACA+ and ATA+ are the major autoantibody-established SSc subsets of patients, an important limitation of our approach was not to include those patients that may not have these particular antibodies but other less predominant. In this line, a recent study has suggested an oncogenic origin for the SSc onset in patients showing antibodies against the RNA polymerase III (ARA+). SSc ARA+ patients are known to be more prone to cancer than the other serological subsets and now Joseph *et al*. have reported a plausible biological reason [141]. Their data suggest that mutations in the *POLR3A* gene that encodes the RPC1 subunit of the RNA polymerase III occur in neoplasias [141]. Then, the mutant RPC1 acts as an immunogen initiating an anti-RPC1 immune response diversification if the immune response

may produce antibody cross-reactivity and stimulate T cell responses to the wild type auto-antigen [141]. Nevertheless, additional genetic risk factors should be required to generate and maintain the observed SSc tissue injury [141].

Considering all the above, we hope that our findings may help to identify the actual origin of the ACA and ATA antibodies, their role as pathogenic players and/or disease biomarkers and even give clues of the biological differences among the autoantibody subsets and the rationale behind the variable clinical outcomes of the subgroups.

GWAS data mining and replication

Two GWAS follow-up studies have lead to the identification of *IL12RB2* as a new piece in the SSc puzzle and to the confirmation *TNIP1* as a strong SSc risk factor, both association signals reaching genome-wide significance level [25, 27]. In the case of *IL12RB2*, despite the hypothesis-free approximation of GWASs, applying a functional prioritization of the second tier or "grey zone" suggestive GWAS associations (p-value < $5x10^{-5}$) has been a successful data mining strategy in our case. Although

increasing sample size, analyzing more homogeneous patient groups and avoiding population stratification would improve the power of GWASs to detect previously undetected risk *loci* [57, 59], we propose the integration of biological function and pathway information as a discerning tool for suggestive regions of association. On the other hand, regarding our GWAS replication effort to address the novel findings in Allanore *et al*. [63], our results have also highlighted the need of independent replication of GWAS findings, especially those signals that did not reach *p*-value $< 5x10^{-8}$. Studies based on high throughput genotyping platforms imply the analysis of multiple tests and restrictive significance thresholds and replication are mandatory to prevent false positives. Three novel SSc risk *loci* (*TNIP1*, *RHOB* and *PSORS1C1*) were proposed in the previously mentioned GWAS, but only *TNIP1* reached the standard genome-wide significance threshold after the intra-study replication phase [63]. Finally, in our data, the signal *RHOB* was found to be a false positive association and the *PSORS1C1* signal was proven to be dependent on the HLA region [25]. These findings highlight the utility of independent replication. Moreover, both *IL12RB2* and *PSORS1C1* are located nearby very relevant autoimmune-related *loci*, *IL23R* and HLA genes,

respectively [25, 27]. Therefore, our follow-up studies included a comprehensive analysis of the dependence relations between the target *loci* and previously identified autoimmune or SSc susceptibility *loci*. On the one hand, our analyses resulted in the identification of *IL12RB2* as an independent associated gene from *IL23R*. *IL23R* encodes the one of the IL-23 receptor subunits. IL-23R is highly expressed on activated/memory T cells, T-cell clones, and NK cells, but also at low levels on monocytes, macrophages, and DC populations. IL-23 promotes Th17 cells, which produce the proinflammatory IL-17 [142]. IL-17 secretion by Th17 cells induces the production of IL-1, IL-6, TNFα, NOS-2, and estimulates the inflammatory axis [142]. Moreover, *IL23R* polymorphisms have been associated with different ADs such as RA, IBD, PS and AS [143]. Nevertheless, no significant signal was discovered in the *IL23R locus* in our SSc data [27]. This lack of association with SSc in *IL23R* region was consistent with previous reports and the existence of a recombination hotspot between both *loci* supported the independence towards real independence [27, 144, 145]. On the other hand, the tested association in the *PSORS1C1* rs3130573 variant was found to be explained by different known classical HLA alleles in the *HLA-DPB1*, *HLA-DRB1* and especially in *HLA-DQA1*

[25, 66]. Consequently, we think that a carefully analysis including testing for dependence should be carried out when proposing a new candidate *locus* inside the HLA region.

Fine-mapping of novel and known associated regions

One of the main advantages of the Immunochip platform is that it comprises the fine-mapping of 186 immune-related *loci* without considering MAF thresholds [88]. The identification of the SSc-associated variant in the *DNASE1L3* gene is a clear result of this design since this variant had not been included in previous high through-put genotyping platforms, due to its modest MAF and low LD with surrounding variants [23]. Nevertheless, the inclusion of this variant in the Immunochip has also unveiled its association with RA [90]. The *DNASE1L3 locus* is located in a region (chromosome 3 from 57,900,000bp to 58,250,000bp) which especially enriched in immune related *loci* and comprises *PXK*, *ABHD6* and *DNASE1L3* itself.

DNASE1L3 is the coding gene for the deoxyribonuclease 1-like 3 protein, a member of the human DNase 1 family expressed mainly in the liver and secreted by macrophages, which is cleaves

DNA in a Ca^{2+}-dependent manner during apoptosis and may play a role in the immunoglobulin gene reorganization [146-149]. Interestingly, the most associated variant in our SSc Immunochip cohort, rs35677470, corresponded to an arginine to cystein substitution at the 206th in the mature DNASE1L3 protein. The observed association reached the most significant non-HLA association with SSc described to date, reaching a *p*-value = $4.25x10^{-31}$ and a meaningful OR = 2.03 in the ACA+ subset of patients [23]. Moreover, this change alters a phylogenetically conserved residue and has been proven to produce an inactive enzyme [150, 151]. Furthermore, among all the non-synonymous SNPs described in *DNASE1L3*, the rs35677470 variant was the only one that could be classified as abolishing the enzyme catalytic activity [150]. The cause behind this loss of function may be the lack of a hydrogen bond (predicted by PolyPhen-2) or the formation of an aberrant disulfide bond between this residue and additional cysteine residues in the 194th or 231st positions [110, 150, 151]. These changes might change the tertiary protein structure and lead to an inactive molecule [150, 151]. Considering the results of our dependence analyses, that probed that no independent associations remained after conditioning by rs35677470, we consider that

despite the high LD in the region we have found the most likely causal variant for the association in the region [23]. What is more, different genetic and functional studies have shown that *DNASE1L3* is a shared risk factor between different autoimmune and immune-related diseases such as SLE, asthma, hypocomplementemic urticarial vasculitis syndrome (HUVS) and now SSc and RA [23, 90, 152-154].

The *PXK* gene has been found to be associated with SLE in white patients and with RA [155-157]. Interestingly, the recent SSc/SLE pan meta-analysis identified *PXK* as a suggestive risk *locus* for SSc and confirmed its association with SLE [103]. *PXK* polymorphisms are associated with autoantibody production in SLE patients [23, 103, 158]. *PXK* is a PX domain containing serine/threonine kinase composed of a unique PX domain, a protein kinase-like domain, and a WASP homology 2 domain [159]. Interestingly, PX domains bind phosphatidylinositol 3-phosphate and PX containing molecules have been related with internalization of cell surface proteins into the endosomal compartment [159]. Consequently, PXK has been shown to localize in the endosomal membrane and to be involved in the internalization and degradation

of epidermal growth factor receptors (EGFRs) [159]. Epidermal growth factor (EGF) is essential in dermal wound healing and promotes stimulation, proliferation, and migration of keratinocyte, endothelial cells, and fibroblasts [160]. Nevertheless, EGF is also used as a cancer therapy target [160]. Therefore, *PXK* was considered as a good candidate to be the causal gene for the observed associations *a priori* [103, 155]. However, recent findings suggest that the SLE PXK signal may be in fact due to the close *ABHD6* gene [161]. The *ABHD6* gene codes for the abhydrolase domain-containing protein 6, which in involved in the endocannabinoid signalling regulation [161]. *ABDH6* and *DNASE1L3* have been reported to be upregulated in B cells induced by Epstein–Barr virus' EBNA and increased *ABDH6* expression is promoted also after stimulation with estradiol [162, 163]. Moreover, Oparina *et al*. found that the SLE-associated genetic variants in this region act as eQTLs for the *ABHD6* gene in lymphoblastoid cell lines but not in adipose or skin tissues, which further suggests an immune related function for this gene [161]. Thus, we consider that our results added to the findings in Oparina *et al*. point out the benefits of targeted fine-mapping in AD genetic studies to determine the real causal variants in associated regions in each

disease and even discard wrong initial hypotheses that may lead to inconclusive experiments or unfruitful clinical trials [23, 161].

In the case of other known SSc risk factors that reached the genome-wide level significance threshold in Mayes *et al*., *STAT4* and *IRF5*, we observed that the lead SNPs in these *loci* in our data rs11893432 and rs62478615 were equivalent to already reported signals (rs3821236 and rs10488631, respectively) [61]. Moreover, we detected no independent signals in these *loci,* which narrowed down the associated regions. Using the results of this fine-mapping, we will be further discuss the possible functional implications of the SSc-related *loci*, which may provide new insights into the SSc pathogenesis and the shared genetic background of SSc with other ADs.

Rare variant interrogation

GWASs are neither powered nor designed to detect variation under the rare variant model [164, 165]. Moreover, despite the effort in cataloguing and studying rare variants in reference panels as relevant as the 1000 Genomes Project, rare variants are challenging to impute due to their low LD with surrounding variants

and need to be directly genotyped [166]. Moreover, PCA and genomic control are not enough to control stratification at rare variants [166]. These facts underline the requirement for geographical match between cases and controls and for replication [166]. Contrary to the GWAS approach, our Immunochip-based genotyping strategy included a wide range of low frequency and rare variants [23]. What is more, we started from a relatively well-powered discovery cohort to detect even rare variants (MAF < 0.01) supposing intense variant effects in these markers as suggested by the rare variant model (power reached 85% for a SNP with a MAF = 0.01 considering the additive model a predicted OR = 3 and prevalence = 1/10,000 at the $\alpha = 5x10^{-8}$) [164, 165]. Additionally, we considered that our large and geographically matched discovery cohorts and the availability of a wide replication set would guarantee the identification of false positive findings. Therefore, we did not exclude from the replication phase a rare variant, rs77583790, located in the *IL12A* promoter region (MAF in discovery phase = 0.01). This strategy resulted fruitful since we confirmed this association as a novel risk factor for SSc, reaching an OR = 2.81 (p-value = $1.53x10^{-11}$) in the lcSSc subset. However, we could not identify a variety of highly penetrant rare variants in

our data. These observations are concordant with previous reports that suggest that rare variants alone may not be the answer to the observed missing heritability in ADs and other complex traits [164, 166].

Nevertheless, in Mayes *et al*. we performed only single marker tests [23] and diverse locus-specific statistical approaches have been developed to analyze multi-marker association of rare variants with disease (reviewed in Bansal *et al*. and Asimit *et al*. [165, 167]): collapsing methods based on summary statistics (Cohort Allelic Sum Test; Combined Multivariate and Collapsing Test; Weighted Sum Test; Variable-Threshold Approach); methods based on similarities among individual sequences (Kernel Based Association Test; Sequence Kernel Association Test); and regression models that use collapsed sets of variants and other factors as predictors (collapsing test using proportion of rare variants; AdaptiveSum Test; LASSO and Ridge Regression). Therefore, for the discussion of this thesis we have performed multi-marker tests following as implemented in CCRaVAT. This software enabled us to perform a large-scale analysis of low MAF polymorphisms by pooling rare variants within defined regions into a single "super-*locus*". By these

means we identified regions that contained a significantly higher proportion of rare minor alleles in the disease cases or controls. Collapsing methods as this one, combine multiple rare minor alleles into a single locus across predefined regions and can substantially increase power for detecting association [168, 169]. We defined rare variants as SNPs showing MAF<0.05. Moreover, we collapsed all the variants that mapped within the coding region of a gene and 50kb upstream and downstream. The regions that reached an initial *p*-value < $1x10^{-5}$ underwent a 100,000 permutation correction. Those variants in the extended HLA region (chr 6: 20,000,000-40,000,000) were not considered since we could not control for the known associations in this region (no covariates could be included in this analysis). We performed this analysis separately in the two Immunochip study discovery cohorts and observed that the *VAV3 locus* remained suggestively significant (*p*-value < $1x10^{-5}$) in the US cohort after correction and the *FAM172A* region in the Spanish set. Among the SSc risk factors *NOTCH4*, *TNFSF4*, *TNFAIP3*, *ITGAM* and *JAZF1* were nominally significant in the US population but not in the Spanish cohort (*Table 5*). However, the US suggestively associated *loci* were not significant in Spain (*p*-value > 0.05) and the *FAM172A* was not significant in the US cohort (*Table 5*).

Furthermore, the frequencies of these variants were very different among populations (*Table 5*). The lack of replication between cohorts may be due to a lack of statistical power. However, these contradictory results could be expected if we consider that it has been already reported that ethnicity greatly affects rare variant association. In this line, rare variants may have been undergone differential selective pressures in the different populations. Consequently, the underlying pathways of the disease may be similar across different populations and common variants are more prone to be shared among cohorts but rare variants may affect specific *loci* in the different populations and will be hard to replicate trans-ethnically. In any case, these results should be considered carefully since they do not reach genome-wide significance, the statistical power of our cohorts was limited and the identified *loci* need to be further investigated. We consider these regions as good candidates for sequencing approaches.

It should be noted that GWASs are designed to capture causal variation indirectly, while next-generation sequencing (NGS) can directly identify the causal variants [170]. Moreover, as stated previously common diseases could be influenced by rare mutations

in many different genes[170]. Consequently, both whole-genome sequencing (WGS) and whole-exome sequencing (WES) are powerful and nowadays more affordable approaches for detecting genetic variation within an individual [171]. In addition, deep targeted exome resequencing of susceptibility *loci* has been proven to be a successful strategy in IBD [172, 173].

Patients with rare diseases and cancer are the most frequent beneficiaries of NGS techniques [171]. However, exome sequencing or deep sequencing provide complementary information for complex traits by comprehensively assessing the role of all variation, both common and rare [174]. Furthermore, there are many more rare variants than common ones and direct sequencing uncovers additional rare variants continuously [174]. In fact, as sample size increases, the number of identified variants increases faster than expected [174]. Recently, a number of statistical tests have been designed for NGS derived rare variant analysis [174]. Some of them are Combined Multivariate and Collapsing methods, as the one applied to our SSc data in this section, but they are more powerful when data is extracted from NGS platforms and more variants are integrated [174]. Additional, methods include the

weighted-sum statistic (WSS), the kernel-based adaptive cluster (KBAC) test, the variable threshold (VT) approach, SKAT, the rare variant weighted aggregate statistic (RWAS), the likelihood ratio test (LRT) and analyses that integrate previously computed predictions such as PLINK/SEQ [174]. Nevertheless, the sources of spurious experimental artifacts in sequencing data are yet less well understood, the optimal depth of the sequencing is still not clear and may vary depending on the aims of the study, although prediction and interpretation of coding variants is better known, the knowledge about regulatory and non-coding regions of the genome is still scarce [170]. Therefore, our current ability to interrogate the genome as a whole for variants is limited, especially in rare diseases such as SSc [170]. However, these limitations will be solved in the future using appropriate measures to control for the accuracy of sequence data, genetic and bioinformatic tools to calculate prior distinctions among variants in probability of influencing disease, and functional characterization of the variants in both *in vivo* and *in vitro* models [170].

Table 5. Results of the collapsing method analysis in the United States and Spain discovery cohorts in Mayes *et al*. 2014.

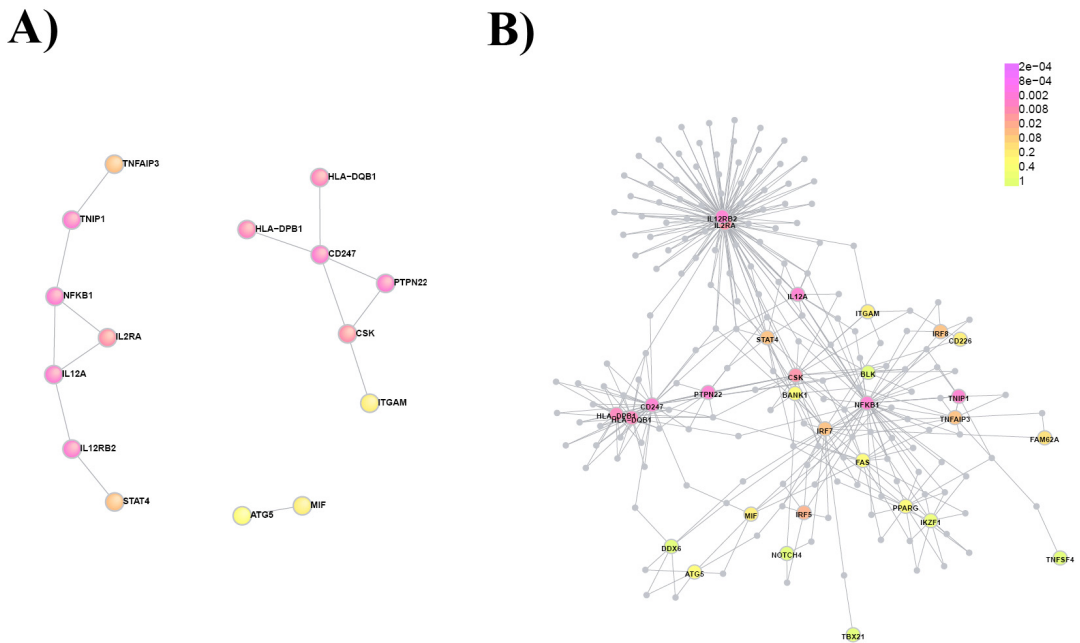
		US			Spain		
Locus	**Position (Chr: start bp- end bp)**	**SSc (a/A)**	**Controls (a/A)**	***P***	**SSc (a/A)**	**Controls (a/A)**	***P***
VAV3	1: 107,865,310-108,359,061	0/963	52/2,672	1.9E-07	138/757	109/691	0.3
FAM172A	5: 92,894,799-93,005,544	0/958	2/2,700	0.4	23/867	1/826	3E-06
NOTCH4	6: 32,220,598-32,349,822	445/502	1,488/1,267	1.90E-04	335/548	306/494	0.896
TNFSF4	1: 171,369,493-171,493,094	576/372	1,758/926	8.80E-03	572/303	520/280	0.873
TNFAIP3	6: 138,180,274-138,296,142	721/225	2,166/541	0.013	626/259	571/210	0.282
ITGAM	16: 31,128,789-31,301,714	598/358	1,798/899	0.021	575/324	509/303	0.585
JAZF1	7: 27,786,717-28,236,962	614/324	1,840/823	0.04	622/258	534/266	0.082

a: presence of the rare variant collapsed allele; A: presence of the rare variant collapsed allele.

Genetic connections and biological relevance of the novel loci

Considering the possible functional implications of the studied *loci*, we think that our results widen the scope of biological process that may undergo pathogenic alterations that lead to SSc onset and prognosis. As it can be observed in *Figure 5*, all the described SSc genetic susceptibility factors show a great overlap in the published literature (Pubmed data mining was carried out using GRAIL as previously mentioned) [79]. It should be noted that the gene list including the firm SSc risk factors in the following paragraphs comprises all the *loci* described in the Introduction section and the novel SSc susceptibility markers described in the presented studies [23-27]. Moreover, we interrogated protein-protein interaction (PPI) databases using the Disease Association Protein-Protein Link Evaluator software (DAPPLE version 2) and confirmed that physical interaction between SSc risk factors were reported in several cases (*Figure 9*) [175]. The rationale behind our DAPPLE analysis is based on the assumption that the causal genetic mutations will affect specific mechanisms that can be detected using protein-protein interactions [175].

Figure 9. Protein-protein interaction analysis results of the firmest systemic sclerosis genetic susceptibility factors as implemented in DAPPLE version 2.



Following the pipeline implemented in the software, at an initial step we define the most likely protein coding gene related to every SNP associated with SSc susceptibility, in our case defined according to our GRAIL results. Then, protein-protein networks were built representing proteins as nodes connected by an edge if there is in vitro evidence of interaction [175]. In vitro interaction data is extracted from the high-confidence InWeb dataset [176], which included 33 out of our 36 query proteins (*DNASE1L3*, *SAMD9L*, *KIAA0319L* could not be included in further analysis). Finally, direct networks (in which any two associated proteins can be connected by exactly one edge) and indirect networks (in which associated proteins can interact through a common interactor protein not encoded in associated *loci* with which the associated proteins each share an edge) were created (*Figure 9*) [175]. We were able to connect 14 out of 33 SSc risk factors via direct connections. Then, different network connectivity parameters and properties were calculated. The clearest metric in this analysis is the direct network connectivity parameter defined as the number of edges in the direct network, this is, the frequency with which the different proteins bind to each. In our case, we observed 14 edges, which was significantly higher than expected (p-value = 9.99×10^{-4}).

Regarding the node properties, the associated protein direct connectivity was significantly more frequent than expected by chance (p-value = $2x10^{-3}$). We conclude that, as previously reported in RA and CD, SSc risk loci directly interact at the protein level more than expected randomly, which supports the idea of similar underlying processes [175]. All but three *loci* were connected in an expanded indirect network. The associated protein indirect connectivity was significantly enriched (p-value = $2.30x10^{-2}$), but the average number of proteins in distinct loci bound by common interactors in indirect networks (common interactor connectivity) did not reach statistical significance (p-value = 0.05). Thus, despite the evidence for an expanding PPI network in SSc, we consider that important connecting nodes and intermediate risk factors will be needed to complete this interaction picture properly.

The results above pointed toward common altered pathways; therefore, we accomplished a molecular pathway enrichment analysis using the Gene Set Enrichment Analysis (GSEA) and DAVID approaches [177-179]. Using GSEA and DAVID we evaluated the overlap of a specific gene set (the set of SSc-related risk factors) with the MSigDB collections (Biocarta,

KEGG and Reactome collections) in a very simplistic manner. When gene sets share genes, their overlap suggests common processes, pathways, and underlying biological themes. *Table 6* shows all the gene sets that showed a significant enrichment *p*-value (*p*-value < 0.05) after False Discovery Rate (FDR) correction (GSEA) or Bonferroni correction (DAVID) with the MSigDB collections. It should be noted that DAVID provided more restringing results than GSEA (*Table 6*), but both algorithms lead to very overlapping results. The interrogation of the three databases had coincident results that highlight the relevance of T-cells in the disease, especially in the Biocarta set in which 8 out of 11 significantly enriched pathways were directly related to T-cells (*Table 6*). In this regard, it should be taken into account that experimental and clinical data backup the enriched pathways based on genetic evidence. Remarkably, T-cells have been shown to present abnormalities in SSc patients and are supposed to play a relevant role in SSc pathology [13]. Remarkably, several SSc mouse models have demonstrated that T-cell depletion reduces fibrosis [180-182]. Furthermore, multiple of T-cell directed therapies (alemtuzumab, anti-human thymocyte globulin, basiliximab…) have have shown promising results that confirm the

essential role of T-cells in the disease [183-186]. It is also known that T-cells are present in the SSc fibrotic *foci* in the skin and several T-cell populations (activated and memory CD4+ T-cells, CD4+CD25+ Tregs, CD4+CD8+ double positive T-cells, CD8+ T-cells) have been reported to have altered numbers and responsiveness in SSc patients [13]. Moreover, T-cells produce IL-6 that is clearly pro-inflammatory but also directly pro-fibrotic and are able to release IL-6 soluble receptor. Thus, T-cells produce the pro-fibrotic cytokine and also release the soluble IL-6 receptor increasing the responsiveness of the local microenvironment. A Th2 response is also a major characteristic of SSc [187]. The Th2 response promotes tissue repair and Th2 cytokines (IL-4, IL-13) are predominantly pro-fibrotic [13, 188]. Moreover, the initial IL-4/IL-13 trigger is of unknown origin although alternatively activated macrophages have been proposed as the link between the innate and adaptative immune responses [13, 189, 190]. Interestingly, the Th2 response inhibits the Th1 response and vice versa. Therefore, T-cell polarization regulators may be suitable as drug targets in SSc. In this direction, *STAT4* a major non-HLA SSc genetic marker for SSc has been described as an interesting target for disease treatment [65, 191]. STAT4 knock-out mice (stat4-/-) have been reported to

Table 6. Gene sets that showed a significant enrichment *p*-value (*p*-value < 0.05) after False Discovery Rate (FDR) correction in the GSEA-based analysis or Bonferroni correction in the DAVID-based analysis.

Database	Gene Set	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
Biocarta	NO2-dependent IL 12 Pathway in NK cells*	17	3	0.18	2.28E-07	4.94E-05
	IL12 and Stat4 Dependent Signaling Pathway in Th1 Development*	23	3	0.13	5.92E-07	6.42E-05
	HIV Induced T Cell Apoptosis*	11	2	0.18	2.74E-05	1.98E-03
	CD40L Signaling Pathway	15	2	0.13	5.22E-05	2.27E-03
	CTL mediated immune response against target cells	15	2	0.13	5.22E-05	2.27E-03
	TNFR2 Signaling Pathway	18	2	0.11	7.60E-05	2.75E-03
	NF-kB Signaling Pathway	23	2	0.09	1.25E-04	3.71E-03
	Activation of Csk by cAMP-dependent Protein Kinase Inhibits Signaling through the T Cell Receptor	24	2	0.08	1.37E-04	3.71E-03
	Keratinocyte Differentiation	46	2	0.04	5.07E-04	1.22E-02
	T Cell Receptor Signaling Pathway	49	2	0.04	5.76E-04	1.25E-02

*Significant after Bonferroni correction in the DAVID-based analysis.

Table 6 (Cont).

Database	Gene Set	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
Biocarta	HIV-I Nef: negative effector of Fas and TNF	58	2	0.03	8.06E-04	1.59E-02
	Leishmania infection	72	4	0.06	2.19E-07	4.07E-05
	Cell adhesion molecules (CAMs)*	134	4	0.03	2.65E-06	1.63E-04
	Allograft rejection*	38	3	0.08	2.80E-06	1.63E-04
	Graft-versus-host disease*	42	3	0.07	3.80E-06	1.63E-04
	Type I diabetes mellitus*	44	3	0.07	4.38E-06	1.63E-04
	Autoimmune thyroid disease*	53	3	0.06	7.71E-06	2.39E-04
	RIG-I-like receptor signaling pathway*	71	3	0.04	1.86E-05	4.95E-04
	Cytokine-cytokine receptor interaction*	267	4	0.02	3.99E-05	9.28E-04
	Toll-like receptor signaling pathway*	102	3	0.03	5.52E-05	1.14E-03
	Asthma	30	2	0.07	2.15E-04	4.00E-03
	Chemokine signaling pathway	190	3	0.02	3.46E-04	5.86E-03
	Intestinal immune network for IgA production	48	2	0.04	5.52E-04	8.56E-03
KEGG	Cytosolic DNA-sensing pathway	56	2	0.04	7.52E-04	1.08E-02

*Significant after Bonferroni correction in the DAVID-based analysis.

Table 6 (Cont).

Database	Gene Set	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
KEGG	NOD-like receptor signaling pathway	62	2	0.03	9.20E-04	1.22E-02
	Epithelial cell signaling in Helicobacter pylori Infection	68	2	0.03	1.11E-03	1.37E-02
	Viral miocarditis	73	2	0.03	1.27E-03	1.48E-02
	Pathways in cáncer	328	3	0.01	1.68E-03	1.83E-02
	Apoptosis	88	2	0.02	1.84E-03	1.84E-02
	Antigen processing and presentation	89	2	0.02	1.88E-03	1.84E-02
	T cell receptor signaling pathway	108	2	0.02	2.75E-03	2.56E-02
	Neurotrophin signaling pathway	126	2	0.02	3.72E-03	3.30E-02
	Natural killer cell mediated cytotoxicity	137	2	0.01	4.38E-03	3.70E-02
	Systemic lupus erythematosus	140	2	0.01	4.57E-03	3.70E-02
	Jak-STAT signaling pathway*	155	2	0.01	5.57E-03	4.32E-02

*Significant after Bonferroni correction in the DAVID-based analysis.

Table 6 (Cont).

Database	Gene Set	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
Reactome	Genes involved in Immune System*	933	11	0.01	2.93E-11	1.98E-08
	Genes involved in Interferon gamma signaling	63	4	0.06	1.27E-07	4.22E-05
	Genes involved in Phosphorylation of CD3 and TCR zeta chains	16	3	0.19	1.88E-07	4.22E-05
	Genes involved in PD-1 signaling	18	3	0.17	2.73E-07	4.60E-05
	Genes involved in Interferon Signaling	159	4	0.03	5.22E-06	7.04E-04
	Genes involved in TCR signaling	54	3	0.06	8.16E-06	9.17E-04
	Genes involved in Costimulation by the CD28 family	63	3	0.05	1.30E-05	1.15E-03
	Genes involved in Interferon alpha/beta signaling	64	3	0.05	1.36E-05	1.15E-03
	Genes involved in RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways	73	3	0.04	2.03E-05	1.52E-03
	Genes involved in Adaptive Immune System	539	5	0.01	3.94E-05	2.46E-03

*Significant after Bonferroni correction in the DAVID-based analysis.

Table 6 (Cont).

Database	Gene Set	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
Reactome	Genes involved in Cytokine Signaling in Immune system	270	4	0.01	4.17E-05	2.46E-03
	Genes involved in Translocation of ZAP-70 to Immunological synapse	14	2	0.14	4.53E-05	2.46E-03
	Genes involved in Innate Immune System	279	4	0.01	4.74E-05	2.46E-03
	Genes involved in Generation of second messenger molecules	27	2	0.07	1.74E-04	8.35E-03
	Genes involved in Negative regulators of RIG-I/MDA5 signaling	31	2	0.06	2.29E-04	1.03E-02
	Genes involved in Downstream TCR signaling	37	2	0.05	3.28E-04	1.38E-02
	Genes involved in Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	70	2	0.03	1.17E-03	4.64E-02

*Significant after Bonferroni correction in the DAVID-based analysis.

show less dermal thickness along with less collagen deposition and lower levels of myofibroblast markers after bleomycin injection than wild-type mice [191]. In addition to its role in fibrosis, STAT4 was found to influence inflammation in the bleomycin model. Stat4-/- mice showed a reduction in infiltrating CD4+ and CD8+ T-cells (not B-cells or monocytes) and of proliferating T-cells, which suggests that STAT4 may control T-cell proliferation in SSc skin [191]. *STAT4* ablation ameliorated the bleomycin induced IFN-γ, IL-6 and IL-2 elevation, but did not alter the IL-4, IL-5 and TGF-β levels [191]. Thus, besides its clear association with increased SSc susceptibility, *STAT4* is involved in matrix deposition, T-cell proliferation and cytokine production, making it a suitable drug therapy candidate [191]. Considering that STAT4 is essential for IL-12 response in T-cells and NK cells, the promising results in Stat4-/- mice, the identification of two genome-wide level associations in the IL-12 pathway in this thesis (*IL12RB2* and *IL12A*) and the implication of IL-12 in the Th1-Th2 response, we propose this pathway as a major candidate for SSc treatment [116, 191-193]. Interestingly, IL-12 inhibitors have been long used in PS and psoriatic arthritis which are typical Th1 diseases [194]. However, in the case of the Th2 polarized SSc immune response, an

increase in IL-12 production may help restore the Th1/Th2 balance. Remarkably, dichloroacetate (a promising anticancer drug that restores mitochondrial function and allows cancer cells to undergo apoptosis) has been recently reported to activate the IFNγ/IL-12 axis and promote a Th1 response [195]. Furthermore, the IFNγ pathway was found to be over-represented in our pathway-enrichment analysis (*Table 6*), which provides further genetic evidence for the relevance of this pathway.

Additional T-cell co-stimulatory pathways such as those including *CD226* and *TNFSF4*, also identified by our group, should also be considered as suitable targets [26, 120]. Regarding *CD226*, this molecule is expressed by NK cells and CD8+ T-cells promotes adhesion and enhances cytolysis. Several CD226 (DNAM-1) monoclonal antibodies have been generated and the absence of CD226 has been shown to increase T_{reg} expansion and the suppressive function of these cells in graft-versus-leukemia after bone marrow transplantation [196, 197]. This evidence suggests a possible therapeutic use of CD226 inhibition to promote tolerance in diseases characterized by T-cell activation [197]. The OX40-OX40L (TNFSF4) interaction is critical for the generation of

productive effector and memory T-cell functions [198]. Several OX40 agonists that mimic TNFSF4 binding have been described to stimulate T-cell proliferation and IFNγ production and Tregs expanded by both OX40 and IL-2 are potent suppressor cells [199, 200]. On the other hand, Th2 induction upon Nod1 and Nod2 activation is dependent on OX40 ligand, on dendritic cells. Thus, TNFSF4 role and possible therapies exploiting its properties might be considered carefully [201].

Besides the relevance of T-cells, the innate immune system compartment seems to be both genetically and functionally important for SSc pathogenesis. Interestingly, several SSc genetic risk factors such as *IRF5*, *IRF7* and *IRF8*, belong to a family of transcription factors that is activated after type I interferon induction [65, 202, 203]. Consequently, we also found the type I IFN as an enriched pathway in our Reactome database search as implemented in GSEA (*Table 6*). Moreover, SSc has been demonstrated to have an IFN-signature, *i.e*. an increased expression of type I IFN-induced genes [202]. Interestingly, IRF5 turns on an interferon-induced gene cascade that starts with TLR signaling and that codes for signaling molecules, proteins involved in adhesion,

cell-cycle regulation, apoptosis and early immune response [204, 205]. *IRF5* knock-out mice models (IRF-5-/-) have demonstrated the implication of this factor in TNF-α and IL-6 production and it has also been involved in caspase pathways and apoptosis [206]. IRF-7, is a key regulator of IFN- α/β production [207]. Interestingly, targeted knockdown of *IRF5* and *IRF7* (among others) using siRNAs in monocytic human cell lines have shown that both of them regulate IL-6 production and different sets of IFN-regulated molecules [208]. On the one hand, reduction of *IRF5* was reported to decrease the levels of the pro-inflammatory IFN_{β}, IP-10, MCP1 and RANTES [208]. On the other hand, *IRF7* regulated IFN_{β} and IP-10 [208]. Thus, inhibition of these proteins may be a promising treatment strategy in SSc and other related ADs [208-210].

We will also like to emphasize that pathway enrichment analyses may be an appropriate basis to prioritize among multiple suggestive signals. However, our naïve method included only well-known SSc risk factors and would hardly generate novel hypotheses that lead to unexpected inferences. For example: *DNASE1L3* and *ATG5 loci*, which are not easily linked to the rest of risk factors or

participate of the same cell mechanisms (*Figure 5*, *Figure 8*, *Table 6*), were not powerful enough to reveal new pathways in our analyses. Nevertheless, we speculate that debris clearance and autophagy should be considered in the future as novel target pathways to take into account in SSc pathology.

To overcome this lack of novelty, several groups have proposed multiple GSEA variations to apply pathway-based approaches to GWAS [211, 212]. In these strategies, all the variants are considered together and overlooked pathways maybe detected [211]. However, GWAS produce genetic marker lists, not gene lists; thus, the reliable conversion of SNPs to genes remains a drawback in the process [212]. Taking advantage of new advances, we have applied a Meta-Analysis Geneset Enrichment of variaNT Associations (MAGENTA) approach to our Immunochip data [213]. MAGENTA is able to combine variant association p-values into gene scores and to correct for different confounding factors (gene size, variant number, and linkage disequilibrium properties) [213]. This software has been successfully used in RA and has been reported to detect associations likely missed by single-marker analysis [213, 214]. Therefore, we conducted a molecular

enrichment analysis using MAGENTA software and adopting GO, Panther, Panther Molecular Function, Ingenuity, KEGG, Biocarta and Reactome databases as pathway information resources using our Immunochip study complete results [23, 213]. We defined significance based on FDR correction after 10^5 permutations [214]. Although several pathways showed initial significance after FDR correction only the PPARα/RXRα activation pathway from Ingenuity database and BCR, NFAT and TCR pathways from Biocarta remained significantly associated (p-value < 0.05). It should be taken into account that the study by Okada *et al*. comprised 29,880 RA cases and 73,758 controls, while our study discovery cohorts reached 1,833 SSc patients and 3,466 controls [214]. Therefore, despite the very modest results of our analysis compared to those in Okada *et al*., we consider that this approach remains an interesting data-mining method since no genotype information is needed and different populations may be meta-analyzed in the future providing more powerful results [213, 214]. In any case, we think that the recent identification of PPARγ as a SSc risk factor and the relevance of the NFAT pathway in T-cell development and function, together with the described results, propose than PPARα/γ-mediated lipid metabolism and the control

of thymocyte development, T-cell differentiation and tolerance mechanisms mediated by NFAT proteins should be taken into account in future SSc research [86, 215, 216].

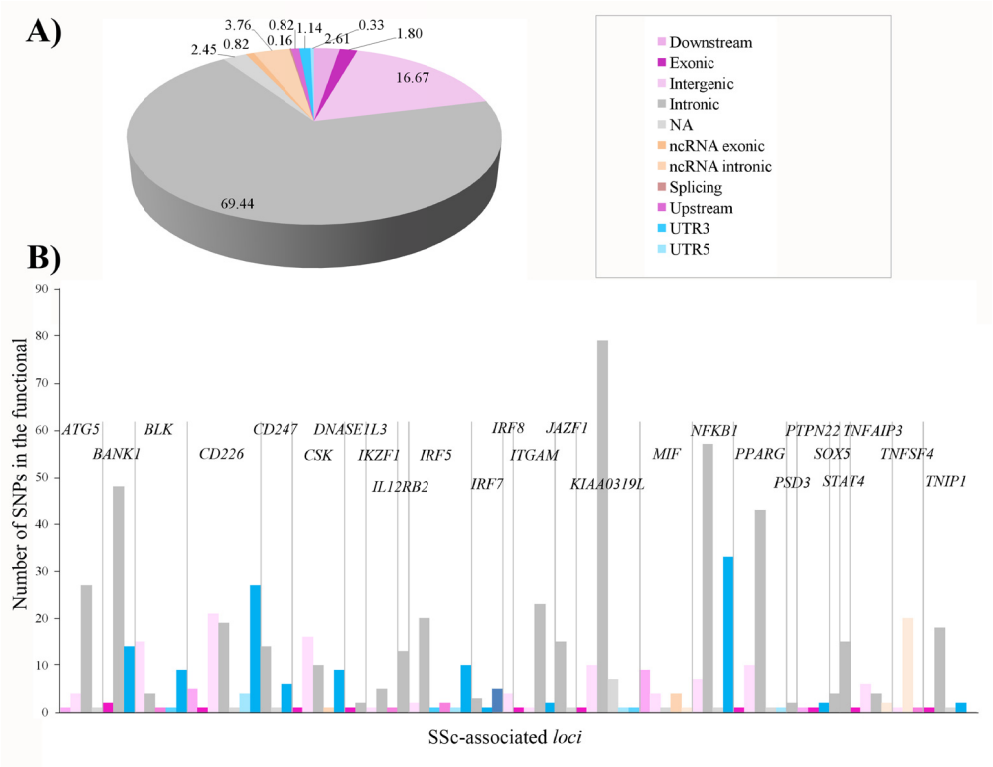
The causal variant quest

It is necessary for geneticist not only to report those regions in the genome that are associated with complex traits but also to identify the causal variant/s for the reported associations. Therefore, we have interrogated a wide number of publically available databases in order to suggest the most likely causal variant/variants for each firm SSc susceptibility *locus*. First, we conducted functional annotation of SSc risk SNPs using ANNOVAR [217]. We included in our analysis not only the associated variants but also all the SNPs that were in high linkage disequilibrium with the associated variants in the 1000 Genomes Project CEU population ($R^2 \geq 0.8$) [214]. The identified variant in *IL12A* was not included in this analysis due to its very low frequency and the inconsistent LD patterns due to this low MAF. ANNOVAR implements the annotation of SNPs and insertions/deletions and the examination of their functional consequence on genes, the inference of cytogenetic bands, the

report of functional importance scores, the finding of variants in conserved regions and the identification of variants reported in the 1000 Genomes Project and dbSNP [217]. Although only 1.80% of the selected variants were located in coding regions, a 40% of SSc associated *loci* were or tagged at least one exonic variant (*Figure 10*). However, this proportion may result misleading since only 3 *loci* were could be linked to missense mutations: *PTPN22*, in which the well-known rs2476601 encodes a damaging Trp to Arg amino acid change which has been related to several ADs but especially RA; *CD226*, with rs763361 that is located in the cytoplasmic tail of the protein; *DNASE1L3*, in which rs35677470 encodes an Arg to Cys change with unclear structural consequences that lead to an inactive protein [23, 26, 218]. As it could be expected, the majority of selected SNPs mapped in intronic regions (69.44%) and almost all the *loci* included this kind of variants (84%) (*Figure 10*). Intronic regions are larger than coding regions and have less *a priori* probabilities of playing a functional role. Despite that other functional categories (UTR3', UTR5', splicing sites, upstream and downstream variants and non-coding RNA located variants) were represented among the selected variants, the second most frequent category comprised that intergenic variants that may play a role in

gene expression control (*Figure 10*). Therefore, using the large data set in Westra *et al*. (peripheral blood mononuclear cells from 5,311 individuals), we analyzed the presence of cis and trans-eQTLs among the identified variants [115]. We observed that a 18.79% of the SSc-associated or proxy SNPs were significantly linked with cis-eQTLs for the established gene, 27.78% with cis-eQTLs for either the *a priori* affected gene or other/s and only 0.82% were found to act as trans-eQTLs ($p<0.05$ after FDR correction). Cis-eQTLs were described for the following SSc genetic susceptibility factors: *BANK1*, *BLK*, *CD226*, *CD247*, *CSK*, *IRF5*, *IRF7*, *ITGAM*, *NFKB1*, *PTPN22* and *TNIP1* (*Figure 10*). Moreover, rs4963128, which is a cis-eQTL for *IRF7*, was the only *locus* that was found to be a significant trans-eQTL for the *PRDM2 locus* (that may act as a specific estrogen activator) and other proteins involved in lipid and glucose metabolism (*ZNF512*), metalloproteinase inhibition (*ITGA7*) and other gene desert regions [219-221]. We consider that this interesting role as trans-eQTL for *IRF7* region variants is logical due to its previously discussed function as IFN response modulator and should be further explored. However, Carmona *et al*. suggested that the association of this variant may be due to a non-synonymous variant in *IRF7* (rs1131665), which is linked but not

Figure 10. Overall functional classification of the systemic sclerosis associated SNPs and their proxies (A). Functional classification of the SNPs in the different SSc associated *loci* (B).



a proxy for rs4963128 [222]. Consequently, we consider that the role of *IRF7* variants needs to be further studied.

Establishing the causal variant/s for an observed association can be a tricky aim as recently described by Smemo *et al*. [223]. The authors of this report redefined the long known and unsuccessfully functionally characterized genetic association of *FTO* with T2D to the distant *IRX3* l*ocus*, after a comprehensive promoter genomic interaction profiling of the region and the generation of a Irx3-deficient mice model [223]. Nevertheless, being aware of the limitations of our approach, we consider that in depth genetic analysis of the SSc-associated regions (summarized in *Figure 10* and *Appendix 1*) may provide useful as *a priori* clues for the future functional research that will be needed to either confirm or discard the implication of the selected *loci* in SSc pathogenesis.

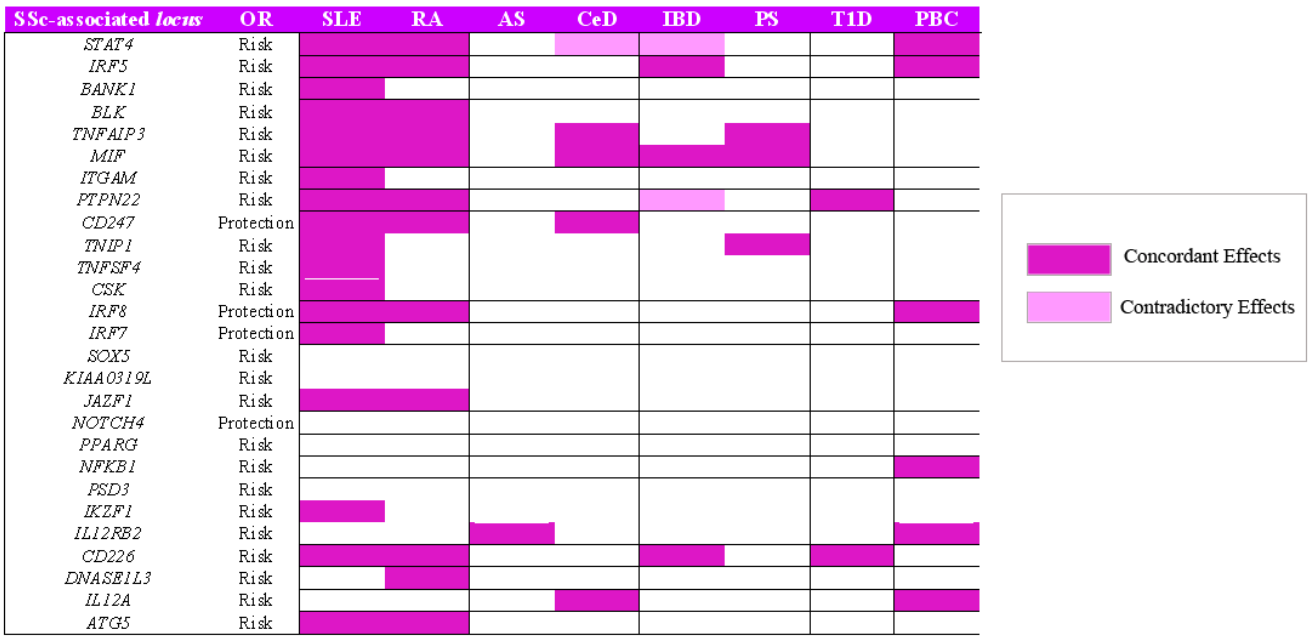
Pleiotropy with other phenotypes and ADs

As commented in the Introduction section, ADs share a common genetic background [102]. This genetic overlap was not unexpected since ADs share some clinical and immunological features [224]. Moreover, co/polyautoimmunity (an individual with

two or more ADs) is common in ADs, for example a 25.7% of polyautoimmunity prevalence has been observed in SSc patients [225]. Interestingly, 71 *loci* have been reported to show associations at the genome-wide level ($p < 5\times10^{-8}$) with two or more ADs [224]. Since this pleiotropic effect of some *loci* has long been acknowledged, Immunochip studies have tried to make use of this particularity [88]. It should not be forgotten that the main aims of the Immunochip studies were the replication of AD-related GWAS results and fine-mapping of confirmed susceptibility *loci* [88, 224]. Therefore, considering the publication of several Immunochip studies in several ADs (including SSc) and the growing number of SSc risk factors identified to date, we would like to update the picture of the genetic overlap of SSc with other ADs. Therefore, we used again the firmest associations outside the HLA region with SSc susceptibility (defined previously) and analyzed which if them have been clearly related to one or more ADs (such as SLE, RA, AS, CeD, IBD, PS, T1D and PBC). *Figure 11* illustrates the shared genetic component between SSc and the previously mentioned ADs. For this relationship matrix we used the recent revision by Parkes *et al*., the novel study in RA that identified 101 RA *loci* by Okada *et al*., the recently published PBC Immunochip studies and

the SLE genetics review by Delgado-Vega *et al*. [94, 95, 214, 224, 226]. As it can be observed in the figure, we found that 81.48% of SSc genetic factors showed pleiotropy with at least another AD and 51.85% were shared between SSc and more than one additional AD. As reported previously, the major overlap was observed for SSc and SLE [65]. Moreover, the recent SSc/SLE pan-meta GWAS identified a novel common *locus KIAA0319L* and a novel *locus* for SSc that had already been reported to be associated with SLE, *JAZF1* (the *PXK* association has been previously addressed) [103]. However, the overwhelming identification of 40 new RA *loci* has been unveiled an extense overlap between both diseases[214]. Therefore, we think that evidence points towards an increasing shared component between SSc and SLE and RA as our power and accuracy to detect novel associations increases. Nevertheless, we consider that the approximately 10% of SSc exclusive associations should be analyzed comprehensively since most of them have been associated with fibrotic processes (*SOX5*, *NOTCH4*, *PPARG*) and may bear the key to the essential fibrotic nature of SSc [227-231]. In addition, we would like to highlight the important overlap between SSc and other autoimmune fibrotic disease such as

Figure 11. Shared genetic component between systemic sclerosis and other autoimmune diseases: systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), ankylosing spondylitis (AS), celiac disease (CeD), inflammatory bowel disease (IBD), psoriaris (PS), type 1 diabetes (T1D) and primary biliary cirrhosis.



PBC, with 22.22% of SSc risk factors being common. It is also interesting that PBC has been reported to account for 4% of SSc coautoimmunity cases, which may suggest common pathogenic mechanisms for both diseases [232]. Consequently, we consider than following a pan-meta GWAS strategy between SSc and other systemic or organ specific fibrotic diseases may shed light into the origin and maintenance of fibrosis in SSc patients.

Systemic sclerosis heritability

Integrating the above mentioned increasing number of SSc genetic susceptibility factors and our high-throughput genotyping data of large SSc cohorts. We think that it is now possible for us to address a topic that has remained controversial since the initial genetic interrogation of SSc patients, the estimation of SSc heritability. Heritability is broadly defined as the proportion of phenotypic variance due to genetic factors [233]. This parameter is assumed to be important in complex diseases with a large genetic component. However, in SSc the available twin studies provided disappointingly reduced heritability estimates, likely due to a clear lack of power [51]. In fact, SSc concordance was estimated to be similar in monozygotic (MZ) and dizygotic (DZ) twins (r_{MZ} = 0.042

and r_{DZ} = 0.056), but the concordance of positive titers of antinuclear autoantibodies (ANA) was found to be very high (r_{MZ} = 0.95 and r_{DZ} = 0.65) [51]. Therefore, according to Falconer's formula of broad sense heritability [$h^2 = 2*(r_{MZ} - r_{DZ})$], SSc heritability may range from 0 to 0.70 [234]. In any case, estimating heritability in a complex trait is a difficult aim [233]; but Visscher and collaborators have described a method for exploiting the variation in genetic concordance among siblings using GWAS genotype data to estimate heritability, GREML implemented in the GCTA package [235, 236]. Moreover, the same group has extended their method to use the same strategy with unrelated individuals I case-control studies [237]. In their approach, for any complex trait more genotype sharing should mean greater phenotypic concordance [235]. Since unrelated individuals don't share the same environment, it is possible to separate genetic effects from environmental factors [235]. Despite its limitations, we consider that this new method may help calculate the proportion of SSc heritability that has been explained to date and even more, it may help us to quantify the contribution of the Immunochip fine-mapping to the missing heritability quest in SSc patients.

As described previously, we will carry out our estimations using case-control data. Therefore, our scale will be set as binary (case/control status) not as a quantitative trait, we will control for ascertainment (higher proportion of cases in our sample set than in the general population) and we will only include SNPs that passed the stringent QC filters described elsewhere [23, 61, 237]. In *Table 7* we report the heritability on the observed scale (h_o^2), which is the genetic variance, but we also transform the proportion of variance in case-control status that is explained by all SNPs on the observed scale to the proportion of variance in disease liability that is captured by the SNPs at population level on the unobserved underlying scale of disease liability (h_l^2). Thus, we will translate the data obtained in our samples to the real disease liability, *i.e.* not only genetic susceptibility but also the whole combination of external circumstances that makes an individual more or less likely to develop the disease. This parameter depends on the prevalence of the disease in the population and the proportional of cases in the sample.

For the following estimations we used our imputed GWAS (the HapMap Project phase 3 data was set as reference

panel) and our Immunochip genotype data, calculations were performed under the previously described GREML method and establishing disease prevalence as 1/10,000 and a case proportion of 32.40% in the GWAS cohorts and 33.97% for the Immunochip cohorts [235, 237]. The using GCTA-GREML Power Calculator the probability of our GWAS cohorts to detect a $h_l^2 \geq 0.1$ reached 99.99%, assuming the above mentioned disease prevalence, $\alpha=5*10^{-8}$ and a variance of the SNP-derived genetic relationships=$2*10^{-5}$ [238]. Under the same conditions, the power of the Immunochip cohorts reached 68.82%. However, adjusting the variance of the SNP-derived genetic relationships as recommend by the authors, this is, using empirical variance of the off-diagonals of the GRM = $5*10^{-4}$, power increased to a virtual 100% [238]. It should be noted that the effective number of SNPs in both sets, *i.e.* the number of independent variants as defined elsewhere [239], was greatly different between the GWAS and the Immunochip approaches 30,839 SNPs and 13,268 SNPs in linkage equilibrium, respectively. This parameters are in concordance with previous reports [240]. The first 3 principal components were included as covariates to control for population stratification [237]. As it can be

Table 7. Systemic sclerosis heritability estimation based on our Immunochip and GWAS genotyping data.

Platform (Cases / Controls)	SNP subset	Number of SNPs	h2o (SE)	LR b	h2l (SE)
GWAS (2,716 / 5,666)					
	All SNPs	311,615	0.39 (0.03)	272.43	0.09 (0.007)
	MAF > 5%	292,850	0.36 (0.03)	249.12	0.09 (0.007)
	No HLA	306,076	0.34 (0.03)	189.52	0.08 (0.007)
	MAF > 5% no HLA	287,761	0.31 (0.03)	171.02	0.07 (0.007)
	Non-HLA known regions	460	0.03 (0.007)	36.47	0.007 (0.002)
	Non-HLA known and novel regions	512	0.03 (0.007)	38.79	0.007 (0.002)
Immunochip (1,833 / 3,466)					
	All SNPs	124,305	0.44 (0.03)	406.8	0.10 (0.007)
	MAF > 5%	74,518	0.34 (0.02)	366.22	0.08 (0.006)
	No HLA	116,306	0.38 (0.03)	305.99	0.09 (0.007)
	MAF > 5% no HLA	68,730	0.28 (0.02)	269.56	0.07 (0.006)
	Non-HLA known regions	461	0.05 (0.01)	50.66	0.01 (0.002)
	Non-HLA known and novel regions	540	0.05 (0.01)	54.2	0.01 (0.002)

SE: Standard error; a) Estimate of genetic variance proportional to the total phenotypic variance on the observed scale; b) Likelihood-ratio test statistic; c) Transformed genetic variance proportional to the total phenotypic variance on the liability scale under the assumption that the population prevalence is 1/10,000

observed in *Table 7*, the proportion of variance in case-control status that was explained by all SNPs ranged 0.39 in the GWAS data and 0.44 in the Immunochip dataset. In the disease liability scale our genotyping platforms accounted for 0.09 and 0.10 of SSc heritability in the general population, respectively (*Table 7*). These estimates are lower than those observed in other ADs such as CD, UC and T1D [237, 240]. However, we recognize that our SSc sample size is smaller than those reached in the mentioned diseases which are more prevalent is the Western population. Therefore, the gap between our estimates and the real disease liability heritability may be large. In any case, our results establish a lower bound for heritability of SSc liability and we can confirm that SSc heritability in the broad sense was underestimated in the initial twin studies. It should be remarked that the HLA region accounted for approximately 10% of the explained liability heritability in both platforms (*Table 7*). Considering the known SSc-associated regions including the *loci* described in the present thesis resulted in a local explained a heritability of SSc liability reaching 0.7% for GWAS data and 1% for the Immunochip (*Table 7*). In order not to lose possible secondary signals or epistatic effects, our local estimations comprised all the variants located 500kb upstream and downstream

the associated variants. This window size was established according to empirical data from a previous report by Gusev *et al*. which quantified missing heritability at known GWAS *loci* and established a 1Mb window as a stable window for heritability estimation [241]. Our results were in the line of those reported in the mentioned report for RA, T1D and UC, but lower than the calculated estimates for T2D and MS [241]. However, our data included several populations and LD-residual correction was not applied.

It can also be inferred from our data that the HLA region association is comparable and even greater than all the remaining *loci* together. Interestingly, if we had only considered the HLA region and the known *loci* we would have only estimated an explained SSc liability heritability of around 2%, but including all the SNPs in each platform lead us to a higher but firm estimation reaching 10%. Therefore, highly significant and well-replicated SNPs identified to date account only for ~20% of the estimated explained heritability for SSc liability. Hence, 80% approximately of the variation due to already tested SNPs identified in our analysis has been undetected in published GWASs. This overlooked SNP effects may be too small to be statistically significant or arise due to

the lack of awareness of the complex genetic interactions between them [235, 240]. However, the followed strategy accumulates the variance explained by all SNPs and is not limited by the need for individual SNPs to pass stringent significance tests [235]. The Immunochip contains rare variants directly genotyped which may give us a clue that the real causal variant/s have lower frequency and some of the association peaks that have been described may be caused by synthetic associations with this rarer causal variants that remain unknown [237]. As expected, the removal of low frequency and rare variants affected more severely to the Immunochip than to the GWAS estimates (*Table 7*) [240]. Therefore, SNPs with low MAF explained proportionally more variance in the Immunochip than in the GWAS data [240]. Additionally, the Immunochip fine-mapping lead to a higher explained heritability of liability in the known regions (*Table 7*), suggesting that the causal variants were better tagged in this platform. However, considering the overall results, little divergence was observed between the estimates from both platforms despite the differences in SNP numbers and sample size. Then, the Immunochip data explained at least as much variation as, but not substantially more than, the corresponding regions in the GWAS data. This observation is consistent with

previous in CD and UC results. However, Chen *et al*. found that compared to the GWAS data, the Immunochip suffered substantial information loss from inadequate tagging of the complete genome due to the uneven density of its marker distribution [240]. In our case, both platforms were equivalent. However, we would like to remark that, as previously pointed out, our sample sizes are still modest for these estimations. Moreover, our estimations may be biased simply by the use of a finite number of SNPs, the differential LD patterns among populations, etc [235, 237]. In addition, SSc was not one of the ADs used in the Immunochip design but controls. [88]. Thus, it is likely that larger GWASs with denser marker sets and more adequate custom fine-mapping platforms may lead to higher explained disease heritability. Moreover, previous reports have documented overlap between causal variants from different ADs, which is supported by our Immunochip data. The difference between the heritability of SSc liability due to known factors and the rest due to the remaining variants in the Immunochip, which are clearly immune-biased, shows that several shared autoimmune risk factors are involved in SSc missing heritability but yet uncovered [102, 241]. Thus, we propose pleiotropic *loci* as good follow-up candidates.

The method that we have presented in this section should not be misinterpreted as a method for estimating the heritability of SSc. It is in fact an estimation of the variance in SSc explained by the SNPs.

Although the heritability explained by the known SSc risk factors may seem modest, we would like to point out that it is indeed a great advance that was not possible a decade ago. Anyway, we would like to notice that explaining all the missing SSc heritability was not the main focus of our studies. Otherwise, our aim was to detect *loci* that are associated with SSc and provide new insights about the disease genetics that explain its pathognomonic characteristics and support further genetic, functional and clinical research.

CONCLUSIONS

1. The novel identified *loci DNASE1L3*, *ATG5*, *SCHIP1*/*IL12A* and *IL12RB2* are associated with susceptibility to systemic sclerosis.

2. The *TNIP1 locus*, but not *RHOB* and *PSORS1C1*, was confirmed as an SSc genetic risk factor.

3. A six amino acid and seven SNP model explains all the associations with SSc susceptibility observed in the HLA region in our data.

4. A three variant haplotype in the *CD226* gene is associated with increased frequency of pulmonary fibrosis in SSc patients.

5. The previous findings of association of the *KCNA5 locus* with higher frequency of pulmonary arterial hypertension in SSc cases were not replicated.

6. Pathway analysis suggests that T-cells, especially via IL12 signaling, and the type I interferon pathway are essential in SSc pathogenesis.

7. Using a combined strategy combining public resources and novel *in silico* tools, we have proposed probable causal variants for several SSc risk *loci* either in the coding regions or in regulatory elements.

8. Our data have highlighted an increasing overlap of SSc with other autoimmune diseases, especially systemic lupus erythematosus, rheumatoid arthritis and other fibrotic disorders such as primary biliary cirrhosis.

9. Our SSc heritability estimates using high-throughput genotype data confirm that SSc heritability in the broad sense was underestimated in the previous SSc twin studies.

CONCLUSIONES

1. Los genes *DNASE1L3*, *ATG5*, *SCHIP1*/*IL12A* e *IL12RB2* están asociados a la susceptibilidad a la esclerosis sistémica.

2. El locus del gen TNIP1, pero no RHOB ni PSORS1C1, fue confirmado como factor de riesgo genético a la esclerodermia.

3. Un modelo compuesto por seis aminoácidos y siete polimorfismos de un solo nucleótido explica todas las asociaciones con susceptibilidad a esclerosis sistémica observadas en la región del HLA en nuestros datos.

4. Un haplotipo de tres variantes en el gen CD226 se asocia con una mayor predisposición a la aparición de fibrosis pulmonar en pacientes de esclerosis sistémica.

5. La evidencia previa de asociación del locus KCNA5 a una mayor frecuencia de aparición de hipertensión arterial pulmonar en pacientes de esclerosis sistémica no fue replicada en nuestras cohortes.

6. Los análisis de rutas bioquímicas realizados en la presente tesis sugieren que los linfocitos T, especialmente mediante la

ruta de señalización de la IL-12, y la ruta del interferón de tipo I son esenciales en la patogénesis de la esclerosis sistémica.

7. Mediante una estrategia combinada usando bases de datos públicas así como novedosas herramientas bioinformáticas, proponemos variantes causales probables, tanto en las regiones codificantes como en zonas reguladoras, para varios *loci* asociados al riesgo a padecer esclerosis sistémica.

8. Nuestros datos ha puesto de manifiesto el creciente solapamiento de la esclerosis sistémica con otras enfermedades autoinmunes, especialmente el lupus eritematoso, la artritis reumatoide y otros desórdenes fibróticos como la cirrosis biliar primaria.

9. Nuestra estimación de la heredabilidad de la esclerodermia usando datos de genotipado de alto rendimiento ha confirmado que la heredabilidad de la esclerodermia en sentido amplio fue subestimada previamente en los estudios llevados a cabo en pares de gemelos con la enfermedad.

BIBLIOGRAPHY

1. Cho, J.H. and P.K. Gregersen, Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med*, 2011. **365**: p. 1612-23.
2. Moroni, L., I. Bianchi, and A. Lleo, Geoepidemiology, gender and autoimmune disease. *Autoimmun Rev*, 2012. **11**: p. A386-92.
3. Miller, F.W., et al., Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop. *J Autoimmun*, 2012. **39**: p. 259-71.
4. Craig, J., Complex diseases: Research and applications, in *Nature Education*. 2008. p. 184.
5. Wandstrat, A. and E. Wakeland, The genetics of complex autoimmune diseases: non-MHC susceptibility genes. *Nat Immunol*, 2001. **2**: p. 802-9.
6. Katsumoto, T.R., M.L. Whitfield, and M.K. Connolly, The pathogenesis of systemic sclerosis. *Annu Rev Pathol*, 2011. **6**: p. 509-37.
7. Gabrielli, A., E.V. Avvedimento, and T. Krieg, Scleroderma. *N Engl J Med*, 2009. **360**: p. 1989-2003.
8. Abraham, D.J., et al., Overview of pathogenesis of systemic sclerosis. *Rheumatology (Oxford)*, 2009. **48 Suppl 3**: p. iii3-7.
9. Gu, Y.S., et al., The immunobiology of systemic sclerosis. *Semin Arthritis Rheum*, 2008. **38**: p. 132-60.
10. Matucci-Cerinic, M., B. Kahaleh, and F.M. Wigley, Review: evidence that systemic sclerosis is a vascular disease. *Arthritis Rheum*, 2013. **65**: p. 1953-62.
11. Rabquer, B.J. and A.E. Koch, Angiogenesis and vasculopathy in systemic sclerosis: evolving concepts. *Curr Rheumatol Rep*, 2012. **14**: p. 56-63.
12. Manetti, M., S. Guiducci, and M. Matucci-Cerinic, The origin of the myofibroblast in fibroproliferative vasculopathy: does the endothelial cell steer the pathophysiology of systemic sclerosis? *Arthritis Rheum*, 2011. **63**: p. 2164-7.
13. O'Reilly, S., T. Hugle, and J.M. van Laar, T cells in systemic sclerosis: a reappraisal. *Rheumatology (Oxford)*, 2012. **51**: p. 1540-9.
14. Bosello, S., et al., B cells in systemic sclerosis: a possible target for therapy. *Autoimmun Rev*, 2011. **10**: p. 624-30.

15. Baraut, J., et al., Relationship between cytokine profiles and clinical outcomes in patients with systemic sclerosis. *Autoimmun Rev*, 2010. **10**: p. 65-73.
16. Gourh, P., et al., Plasma cytokine profiles in systemic sclerosis: associations with autoantibody subsets and clinical manifestations. *Arthritis Res Ther*, 2009. **11**: p. R147.
17. Ranque, B. and L. Mouthon, Geoepidemiology of systemic sclerosis. Autoimmun Rev. **9**(5): p. A311-8.
18. Mayes, M.D., et al., Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum*, 2003. **48**: p. 2246-55.
19. Reveille, J.D., Ethnicity and race and systemic sclerosis: how it affects susceptibility, severity, antibody genetics, and clinical manifestations. *Curr Rheumatol Rep*, 2003. **5**: p. 160-7.
20. Arnett, F.C., et al., Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. *Arthritis Rheum*, 1996. **39**: p. 1362-70.
21. Pennell, L.M., C.L. Galligan, and E.N. Fish, Sex affects immunity. *J Autoimmun*, 2012. **38**: p. J282-91.
22. Selmi, C., et al., The X chromosome and the sex ratio of autoimmunity. *Autoimmun Rev*, 2012. **11**: p. A531-7.
23. Mayes, M.D., et al., Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet*, 2014. **94**: p. 47-61.
24. Bossini-Castillo, L., et al., KCNA5 gene is not confirmed as a systemic sclerosis-related pulmonary arterial hypertension genetic susceptibility factor. *Arthritis Res Ther*, 2012. **14**: p. R273.
25. Bossini-Castillo, L., et al., Confirmation of TNIP1 but not RHOB and PSORS1C1 as systemic sclerosis risk factors in a large independent replication study. *Ann Rheum Dis*, 2013. **72**: p. 602-7.
26. Bossini-Castillo, L., et al., A multicenter study confirms CD226 gene association with systemic sclerosis-related pulmonary fibrosis. *Arthritis Res Ther*, 2012. **14**: p. R85.
27. Bossini-Castillo, L., et al., A GWAS follow-up study reveals the association of the IL12RB2 gene with systemic sclerosis in Caucasian populations. *Hum Mol Genet*, 2012. **21**: p. 926-33.
28. Ostensen, M., P.M. Villiger, and F. Forger, Interaction of pregnancy and autoimmune rheumatic disease. *Autoimmun Rev*, 2012. **11**: p. A437-46.
29. Goldblatt, F. and S.G. O'Neill, Clinical aspects of autoimmune rheumatic diseases. *Lancet*, 2013. **382**: p. 797-808.

30. 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**: p. 581-90.
31. LeRoy, E.C. and T.A. Medsger, Jr., Criteria for the classification of early systemic sclerosis. *J Rheumatol*, 2001. **28**: p. 1573-6.
32. LeRoy, E.C., et al., Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol*, 1988. **15**: p. 202-5.
33. Steen, V.D. and T.A. Medsger, Changes in causes of death in systemic sclerosis, 1972-2002. *Ann Rheum Dis*, 2007. **66**: p. 940-4.
34. Elhai, M., et al., Trends in mortality in patients with systemic sclerosis over 40 years: a systematic review and meta-analysis of cohort studies. *Rheumatology (Oxford)*, 2012. **51**: p. 1017-26.
35. Thomas, E., et al., National study of cause-specific mortality in rheumatoid arthritis, juvenile chronic arthritis, and other rheumatic conditions: a 20 year followup study. *J Rheumatol*, 2003. **30**: p. 958-65.
36. Tani, C., et al., Systemic sclerosis: a critical digest of the recent literature. *Clin Exp Rheumatol*, 2013. **31**: p. 172-9.
37. Badesch, D.B., et al., Diagnosis and assessment of pulmonary arterial hypertension. *J Am Coll Cardiol*, 2009. **54**: p. S55-66.
38. Wipff, J., et al., Association of a KCNA5 gene polymorphism with systemic sclerosis-associated pulmonary arterial hypertension in the European Caucasian population. *Arthritis Rheum*, 2010. **62**: p. 3093-100.
39. Steele, R., et al., Clinical decision rule to predict the presence of interstitial lung disease in systemic sclerosis. *Arthritis Care Res (Hoboken)*, 2012. **64**: p. 519-24.
40. Winklehner, A., et al., Screening for interstitial lung disease in systemic sclerosis: the diagnostic accuracy of HRCT image series with high increment and reduced number of slices. *Ann Rheum Dis*, 2012. **71**: p. 549-52.
41. Koenig, M., M. Dieude, and J.L. Senecal, Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. *Autoimmun Rev*, 2008. **7**: p. 588-93.
42. Nihtyanova, S.I. and C.P. Denton, Autoantibodies as predictive tools in systemic sclerosis. *Nat Rev Rheumatol*, 2010. **6**: p. 112-6.
43. Marie, I., et al., Prospective study to evaluate the association between systemic sclerosis and occupational exposure and review of the literature. *Autoimmun Rev*, 2014. **13**: p. 151-6.
44. Rocha-Parise, M., et al., Lymphocyte activation in silica-exposed workers. *Int J Hyg Environ Health*, 2014. **217**:586-91

45. Mora, G.F., Systemic sclerosis: environmental factors. *J Rheumatol*, 2009. **36**: p. 2383-96.
46. Patterson, R. and D. Germolec, Review article toxic oil syndrome: review of immune aspects of the disease. *J Immunotoxicol*, 2005. **2**: p. 51-8.
47. Moroncini, G., et al., Role of viral infections in the etiopathogenesis of systemic sclerosis. *Clin Exp Rheumatol*, 2013. **31**: p. 3-7.
48. Grossman, C., et al., Do infections facilitate the emergence of systemic sclerosis? *Autoimmun Rev*, 2011. **10**: p. 244-7.
49. Lunardi, C., et al., Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. *Nat Med*, 2000. **6**: p. 1183-6.
50. Adams Waldorf, K.M. and J.L. Nelson, Autoimmune disease during pregnancy and the microchimerism legacy of pregnancy. *Immunol Invest*, 2008. **37**: p. 631-44.
51. Feghali-Bostwick, C., T.A. Medsger, Jr., and T.M. Wright, Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis Rheum*, 2003. **48**: p. 1956-63.
52. Assassi, S., et al., Clinical, immunologic, and genetic features of familial systemic sclerosis. *Arthritis Rheum*, 2007. **56**: p. 2031-7.
53. Arnett, F.C., et al., Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. *Arthritis Rheum*, 2001. **44**: p. 1359-62.
54. Lobo, I. Multifactorial inheritance and genetic disease. *Nature Education*, 2008. **1**: 5.
55. Nussbaum R.L. , M.R.R., Willard H.F., *Thompson & Thompson Genetics in Medicine*, 2007. 7th Edition ed.
56. Cardenas-Roldan, J., A. Rojas-Villarraga, and J.M. Anaya, How do autoimmune diseases cluster in families? A systematic review and meta-analysis. *BMC Med*, 2013. **11**: p. 73.
57. Visscher, P.M., et al., Five years of GWAS discovery. *Am J Hum Genet*, 2012. **90**: p. 7-24.
58. Raychaudhuri, S. and S.S. Rich, Autoimmunity: insights from human genomics. *Curr Opin Immunol*, 2012. **24**: p. 513-5.
59. Jia, X., et al., Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One*, 2013. **8**: p. e64683.
60. Spencer, C.C., et al., Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. *PLoS Genet*, 2009. **5**: p. e1000477.

61. Radstake, T.R., et al., Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet*, 2010. **42**: p. 426-9.
62. Dieude, P., et al., Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. *Ann Rheum Dis*, 2011. **70**: p. 1695-6.
63. Allanore, Y., et al., Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS Genet*, 2011. **7**: p. e1002091.
64. Tobon, G.J., et al., Are autoimmune diseases predictable? *Autoimmun Rev*, 2012. **11**: p. 259-66.
65. Martin, J.E., L. Bossini-Castillo, and J. Martin, Unraveling the genetic component of systemic sclerosis. *Hum Genet*, 2012. **131**: p. 1023-37.
66. Arnett, F.C., et al., Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann Rheum Dis*, 2010. **69**: p. 822-7.
67. Loubiere, L.S., et al., HLA allelic variants encoding DR11 in diffuse and limited systemic sclerosis in Caucasian women. *Rheumatology (Oxford)*, 2005. **44**: p. 318-22.
68. Vlachoyiannopoulos, P.G., et al., Systemic scleroderma in Greece: low mortality and strong linkage with HLA-DRB1*1104 allele. *Ann Rheum Dis*, 2000. **59**: p. 359-67.
69. Vargas-Alarcon, G., et al., Association of HLA-DR5 (DR11) with systemic sclerosis (scleroderma) in Mexican patients. *Clin Exp Rheumatol*, 1995. **13**: p. 11-6.
70. Morel, P.A., et al., Severe systemic sclerosis with anti-topoisomerase I antibodies is associated with an HLA-DRw11 allele. *Hum Immunol*, 1994. **40**: p. 101-10.
71. Rands, A.L., et al., MHC class II associations with autoantibody and T cell immune responses to the scleroderma autoantigen topoisomerase I. *J Autoimmun*, 2000. **15**: p. 451-8.
72. Gorlova, O., et al., Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS Genet*, 2011. **7**: p. e1002178.
73. Karp, D.R., et al., Novel sequence feature variant type analysis of the HLA genetic association in systemic sclerosis. *Hum Mol Genet*, 2010. **19**: p. 707-19.
74. Simeon, C.P., et al., Association of HLA class II genes with systemic sclerosis in Spanish patients. *J Rheumatol*, 2009. **36**: p. 2733-6.

75. Gilchrist, F.C., et al., Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun*, 2001. **2**: p. 76-81.
76. Muller-Hilke, B., HLA class II and autoimmunity: epitope selection vs differential expression. *Acta Histochem*, 2009. **111**: p. 379-81.
77. Assassi, S., et al., Genetics of scleroderma: implications for personalized medicine? *BMC Med*, 2013. **11**: p. 9.
78. Gourh, P., et al., Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. *J Autoimmun*, 2010. **34**: p. 155-62.
79. Raychaudhuri, S., et al., Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet,* 2009. **5**: p. e1000534.
80. Mi, H., A. Muruganujan, and P.D. Thomas, PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res*, 2013. **41**(Database issue): p. D377-86.
81. Mi, H. and P. Thomas, PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol*, 2009. **563**: p. 123-40.
82. Stranger, B.E., E.A. Stahl, and T. Raj, Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, 2011. **187**: p. 367-83.
83. Duggal, P., et al., Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics*, 2008. **9**: p. 516.
84. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007. **447**: p. 661-78.
85. Martin, J.E., et al., Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. *Hum Mol Genet*, 2012. **21**: p. 2825-35.
86. Lopez-Isac, E., et al., A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility. *Arthritis Res Ther*, 2014. **16**: p. R6.
87. Schaub, M.A., et al., Linking disease associations with regulatory information in the human genome. *Genome Res*, 2012. **22**: p. 1748-59.
88. Cortes, A. and M.A. Brown, Promise and pitfalls of the Immunochip. *Arthritis Res Ther*, 2011. **13**: p. 101.
89. Trynka, G., et al., Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet*, 2011. **43**: p. 1193-201.

90. Eyre, S., et al., High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet*, 2012. **44**: p. 1336-40.
91. Kim, K., et al., High-density genotyping of immune loci in Koreans and Europeans identifies eight new rheumatoid arthritis risk loci. *Ann Rheum Dis*, 2014. **In press**.
92. Cooper, J.D., et al., Seven newly identified loci for autoimmune thyroid disease. *Hum Mol Genet*, 2012. **21**: p. 5202-8.
93. Tsoi, L.C., et al., Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet*, 2012. **44**: p. 1341-8.
94. Liu, J.Z., et al., Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet*, 2012. **44**: p. 1137-41.
95. Juran, B.D., et al., Immunochip analyses identify a novel risk locus for primary biliary cirrhosis at 13q14, multiple independent associations at four established risk loci and epistasis between 1p31 and 7q32 risk variants. *Hum Mol Genet*, 2012. **21**: p. 5209-21.
96. Hinks, A., et al., Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet*, 2013. **45**: p. 664-9.
97. Liu, J.Z., et al., Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet*, 2013. **45**: p. 670-5.
98. Faraco, J., et al., ImmunoChip study implicates antigen presentation to T cells in narcolepsy. *PLoS Genet*, 2013. **9**: p. e1003270.
99. Ellinghaus, D., et al., High-density genotyping study identifies four new susceptibility loci for atopic dermatitis. *Nat Genet*, 2013. **45**: p. 808-12.
100. Cortes, A., et al., Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet*, 2013. **45**: p. 730-8.
101. Saruhan-Direskeneli, G., et al., Identification of multiple genetic susceptibility loci in Takayasu arteritis. *Am J Hum Genet*, 2013. **93**: p. 298-305.
102. Zhernakova, A., C.C. van Diemen, and C. Wijmenga, Detecting shared pathogenesis from the shared genetics of immune-related diseases. *Nat Rev Genet*, 2009. **10**: p. 43-55.
103. Martin, J.E., et al., A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. *Hum Mol Genet*, 2013. **22**: p. 4021-9.

104. Carmona, F.D., et al., The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. *PLoS One*, 2013. **8**: p. e54419.
105. Raychaudhuri, S., et al., Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet*, 2012. **44**: p. 291-6.
106. Bossini-Castillo, L., et al., A genome-wide association study of rheumatoid arthritis without antibodies against citrullinated peptides. *Ann Rheum Dis*, 2014. **In press**.
107. Han, B., et al., Fine Mapping Seronegative and Seropositive Rheumatoid Arthritis to Shared and Distinct HLA Alleles by Adjusting for the Effects of Heterogeneity. *Am J Hum Genet*, 2014. **94**: p. 522-32
108. Pillai, N.E., et al., Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. *Hum Mol Genet*, 2014. **In press**.
109. Ng, P.C. and S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res*, 2003. **31**: p. 3812-4.
110. Adzhubei, I.A., et al., A method and server for predicting damaging missense mutations. *Nat Methods*, 2010. **7**: p. 248-9.
111. Nicolae, D.L., et al., Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet*, 2010. **6**: p. e1000888.
112. Zhong, H., et al., Integrating pathway analysis and genetics of gene expression for genome-wide association studies. *Am J Hum Genet*, 2010. **86**: p. 581-91.
113. Lappalainen, T., et al., Transcriptome and genome sequencing uncovers functional variation in humans. *Nature*, 2013. **501**: p. 506-11.
114. Fehrmann, R.S., et al., Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet*, 2011. **7**: p. e1002197.
115. Westra, H.J., et al., Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*, 2013. **45**: p. 1238-43.
116. Paunovic, V., et al., Signalling, inflammation and arthritis: crossed signals: the role of interleukin (IL)-12, -17, -23 and -27 in autoimmunity. *Rheumatology (Oxford)*, 2008. **47**: p. 771-6.
117. Ito, I., et al., Association of the FAM167A-BLK region with systemic sclerosis. *Arthritis Rheum*, 2010. **62**: p. 890-5.
118. Coustet, B., et al., C8orf13-BLK is a genetic risk locus for systemic sclerosis and has additive effects with BANK1: results

from a large french cohort and meta-analysis. *Arthritis Rheum*, 2011. **63**: p. 2091-6.
119. Gourh, P., et al., Association of TNFSF4 (OX40L) polymorphisms with susceptibility to systemic sclerosis. *Ann Rheum Dis*, 2010. **69**: p. 550-5.
120. Bossini-Castillo, L., et al., A replication study confirms the association of TNFSF4 (OX40L) polymorphisms with systemic sclerosis in a large European cohort. *Ann Rheum Dis*, 2011. **70**: p. 638-41.
121. Coustet, B., et al., Association study of ITGAM, ITGAX, and CD58 autoimmune risk loci in systemic sclerosis: results from 2 large European Caucasian cohorts. *J Rheumatol*, 2011. **38**: p. 1033-8.
122. Carmona, F.D., et al., Association of a non-synonymous functional variant of the ITGAM gene with systemic sclerosis. *Ann Rheum Dis*, 2011. **70**: p. 2050-2.
123. Beretta, L., et al., Analysis of Class II human leucocyte antigens in Italian and Spanish systemic sclerosis. *Rheumatology (Oxford)*, 2012. **51**: p. 52-9.
124. Wucherpfennig, K.W. and D. Sethi, T cell receptor recognition of self and foreign antigens in the induction of autoimmunity. *Semin Immunol*, 2011. **23**: p. 84-91.
125. Painter, C.A. and L.J. Stern, Conformational variation in structures of classical and non-classical MHCII proteins and functional implications. *Immunol Rev*, 2012. **250**: p. 144-57.
126. Rammensee, H.G., T. Friede, and S. Stevanoviic, MHC ligands and peptide motifs: first listing. *Immunogenetics*, 1995. **41**: p. 178-228.
127. Castellino, F., G. Zhong, and R.N. Germain, Antigen presentation by MHC class II molecules: invariant chain function, protein trafficking, and the molecular basis of diverse determinant capture. *Hum Immunol*, 1997. **54**: p. 159-69.
128. Sette, A., et al., Capacity of intact proteins to bind to MHC class II molecules. *J Immunol*, 1989. **143**: p. 1265-7.
129. Nielsen, M., et al., MHC class II epitope predictive algorithms. *Immunology*, 2010. **130**: p. 319-28.
130. Brown, J.H., et al., Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature*, 1993. **364**: p. 33-9.
131. Tollefsen, S., et al., Structural and functional studies of trans-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. *J Biol Chem*, 2012. **287**: p. 13611-9.

132. Dai, S., et al., Crystal structure of HLA-DP2 and implications for chronic beryllium disease. *Proc Natl Acad Sci U S A*, 2010. **107**: p. 7425-30.
133. Rapin, N., et al., MHC motif viewer. *Immunogenetics*, 2008. **60**: p. 759-65.
134. Nielsen, M., et al., Quantitative predictions of peptide binding to any HLA-DR molecule of known sequence: NetMHCIIpan. *PLoS Comput Biol*, 2008. **4**: p. e1000107.
135. Van Kaer, L., Accessory proteins that control the assembly of MHC molecules with peptides. *Immunol Res*, 2001. **23**: p. 205-14.
136. Brocke, P., et al., HLA-DM, HLA-DO and tapasin: functional similarities and differences. *Curr Opin Immunol*, 2002. **14**: p. 22-9.
137. Pos, W., D.K. Sethi, and K.W. Wucherpfennig, Mechanisms of peptide repertoire selection by HLA-DM. *Trends Immunol*, 2013. **34**: p. 495-501.
138. Hulpke, S. and R. Tampe, The MHC I loading complex: a multitasking machinery in adaptive immunity. *Trends Biochem Sci*, 2013. **38**: p. 412-20.
139. Fruh, K., et al., Displacement of housekeeping proteasome subunits by MHC-encoded LMPs: a newly discovered mechanism for modulating the multicatalytic proteinase complex. *EMBO J*, 1994. **13**: p. 3236-44.
140. Rotinen, M., et al., Type 10 17beta-hydroxysteroid dehydrogenase expression is regulated by C/EBPbeta in HepG2 cells. *J Steroid Biochem Mol Biol*, 2010. **122**: p. 164-71.
141. Joseph, C.G., et al., Association of the autoimmune disease scleroderma with an immunologic response to cancer. *Science*, 2014. **343**: p. 152-7.
142. Rong, C., et al., Interleukin-23 as a potential therapeutic target for rheumatoid arthritis. *Mol Cell Biochem*, 2012. **361**: p. 243-8.
143. van Wanrooij, R.L., et al., Genetic variations in interleukin-12 related genes in immune-mediated diseases. *J Autoimmun*, 2012. **39**: p. 359-68.
144. Farago, B., et al., Functional variants of interleukin-23 receptor gene confer risk for rheumatoid arthritis but not for systemic sclerosis. *Ann Rheum Dis*, 2008. **67**: p. 248-50.
145. Rueda, B., et al., The interleukin 23 receptor gene does not confer risk to systemic sclerosis and is not associated with systemic sclerosis disease phenotype. *Ann Rheum Dis*, 2009. **68**: p. 253-6.
146. Baron, W.F., et al., Cloning and characterization of an actin-resistant DNase I-like endonuclease secreted by macrophages. *Gene*, 1998. **215**: p. 291-301.

147. Shiokawa, D. and S. Tanuma, Characterization of human DNase I family endonucleases and activation of DNase gamma during apoptosis. *Biochemistry*, 2001. **40**: p. 143-52.
148. Errami, Y., et al., Apoptotic DNA fragmentation may be a cooperative activity between caspase-activated deoxyribonuclease and the poly(ADP-ribose) polymerase-regulated DNAS1L3, an endoplasmic reticulum-localized endonuclease that translocates to the nucleus during apoptosis. *J Biol Chem*, 2013. **288**: p. 3460-8.
149. Okamoto, M., et al., Involvement of DNase gamma in the resected double-strand DNA breaks in immunoglobulin genes. *Biochem Biophys Res Commun*, 2005. **327**: p. 76-83.
150. Ueki, M., et al., Evaluation of all non-synonymous single nucleotide polymorphisms (SNPs) in the genes encoding human deoxyribonuclease I and I-like 3 as a functional SNP potentially implicated in autoimmunity. *FEBS J*, 2014. **281**: p. 376-90.
151. Ueki, M., et al., Caucasian-specific allele in non-synonymous single nucleotide polymorphisms of the gene encoding deoxyribonuclease I-like 3, potentially relevant to autoimmunity, produces an inactive enzyme. *Clin Chim Acta*, 2009. **407**: p. 20-4.
152. Baines, K.J., et al., Sputum gene expression signature of 6 biomarkers discriminates asthma inflammatory phenotypes. J Allergy Clin Immunol. **133**(4): p. 997-1007.
153. Al-Mayouf, S.M., et al., Loss-of-function variant in DNASE1L3 causes a familial form of systemic lupus erythematosus. *Nat Genet*, 2011. **43**: p. 1186-8.
154. Ozcakar, Z.B., et al., DNASE1L3 mutations in hypocomplementemic urticarial vasculitis syndrome. *Arthritis Rheum*, 2013. **65**: p. 2183-9.
155. Harley, J.B., et al., Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. *Nat Genet*, 2008. **40**: p. 204-10.
156. Lee, Y.H., et al., Associations between PXK and TYK2 polymorphisms and systemic lupus erythematosus: a meta-analysis. *Inflamm Res*, 2012. **61**: p. 949-54.
157. Stahl, E.A., et al., Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat Genet*, 2010. **42**: p. 508-14.
158. Yu, B., et al., Polymorphisms of PXK are associated with autoantibody production, but not disease risk, of systemic lupus erythematosus in Chinese mainland population. *Lupus*, 2011. **20**: p. 23-7.

159. Takeuchi, H., et al., Characterization of PXK as a protein involved in epidermal growth factor receptor trafficking. *Mol Cell Biol*, 2010. **30**: p. 1689-702.
160. Bodnar, R.J., Epidermal Growth Factor and Epidermal Growth Factor Receptor: The Yin and Yang in the Treatment of Cutaneous Wounds and Cancer. *Adv Wound Care (New Rochelle)*, 2013. **2**: p. 24-29.
161. Oparina, N.Y., et al., PXK locus in systemic lupus erythematosus: fine mapping and functional analysis reveals novel susceptibility gene ABHD6. *Ann Rheum Dis*, 2014. **In press**.
162. Maier, S., et al., Cellular target genes of Epstein-Barr virus nuclear antigen 2. *J Virol*, 2006. **80**: p. 9761-71.
163. Boverhof, D.R., et al., Inhibition of estrogen-mediated uterine gene expression responses by dioxin. *Mol Pharmacol*, 2008. **73**: p. 82-93.
164. Gibson, G., Rare and common variants: twenty arguments. *Nat Rev Genet*, 2012. **13**: p. 135-45.
165. Asimit, J. and E. Zeggini, Rare variant association analysis methods for complex traits. *Annu Rev Genet*, 2010. **44**: p. 293-308.
166. Panoutsopoulou, K., I. Tachmazidou, and E. Zeggini, In search of low-frequency and rare variants affecting complex traits. *Hum Mol Genet*, 2013. **22**: p. R16-21.
167. Bansal, V., et al., Statistical analysis strategies for association studies involving rare variants. *Nat Rev Genet*, 2010. **11**: p. 773-85.
168. Li, B. and S.M. Leal, Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am J Hum Genet*, 2008. **83**: p. 311-21.
169. Morris, A.P. and E. Zeggini, An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet Epidemiol*, 2010. **34**: p. 188-93.
170. Goldstein, D.B., et al., Sequencing studies in human genetics: design and interpretation. *Nat Rev Genet*, 2013. **14**: p. 460-70.
171. Boycott, K.M., et al., Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*, 2013. **14**: p. 681-91.
172. Rivas, M.A., et al., Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nat Genet*, 2011. **43**: p. 1066-73.
173. Momozawa, Y., et al., Resequencing of positional candidates identifies low frequency IL23R coding variants protecting against inflammatory bowel disease. *Nat Genet*, 2011. **43**: p. 43-7.

174. Kiezun, A., et al., Exome sequencing and the genetic basis of complex traits. *Nat Genet*, 2012. **44**: p. 623-30.
175. Rossin, E.J., et al., Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. *PLoS Genet*, 2011. **7**: p. e1001273.
176. Lage, K., et al., A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*, 2007. **25**: p. 309-16.
177. Subramanian, A., et al., Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 2005. **102**: p. 15545-50.
178. Huang da, W., B.T. Sherman, and R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*, 2009. **4**: p. 44-57.
179. Huang da, W., B.T. Sherman, and R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 2009. **37**: p. 1-13.
180. Phelps, R.G., et al., Induction of skin fibrosis and autoantibodies by infusion of immunocompetent cells from tight skin mice into C57BL/6 Pa/Pa mice. *J Autoimmun*, 1993. **6**: p. 701-18.
181. Huaux, F., et al., Eosinophils and T lymphocytes possess distinct roles in bleomycin-induced lung injury and fibrosis. *J Immunol*, 2003. **171**: p. 5470-81.
182. Westermann, W., et al., Th2 cells as effectors in postirradiation pulmonary damage preceding fibrosis in the rat. *Int J Radiat Biol*, 1999. **75**: p. 629-38.
183. Isaacs, J.D., et al., Monoclonal antibody therapy of diffuse cutaneous scleroderma with CAMPATH-1H. *J Rheumatol*, 1996. **23**: p. 1103-6.
184. Stratton, R.J., H. Wilson, and C.M. Black, Pilot study of anti-thymocyte globulin plus mycophenolate mofetil in recent-onset diffuse scleroderma. R*heumatology (Oxford)*, 2001. **40**: p. 84-8.
185. Kuwana, M., T.A. Medsger, Jr., and T.M. Wright, T and B cell collaboration is essential for the autoantibody response to DNA topoisomerase I in systemic sclerosis. *J Immunol*, 1995. **155**: p. 2703-14.
186. Becker, M.O., et al., The monoclonal anti-CD25 antibody basiliximab for the treatment of progressive systemic sclerosis: an open-label study. *Ann Rheum Dis*, 2011. **70**: p. 1340-1.
187. Mavalia, C., et al., Type 2 helper T-cell predominance and high CD30 expression in systemic sclerosis. *Am J Pathol*, 1997. **151**: p. 1751-8.

188. Allen, J.E. and T.A. Wynn, Evolution of Th2 immunity: a rapid repair response to tissue destructive pathogens. *PLoS Pathog*, 2011. **7**: p. e1002003.
189. Higashi-Kuwata, N., et al., Characterization of monocyte/macrophage subsets in the skin and peripheral blood derived from patients with systemic sclerosis. *Arthritis Res Ther*, 2010. **12**: p. R128.
190. Higashi-Kuwata, N., et al., Alternatively activated macrophages (M2 macrophages) in the skin of patient with localized scleroderma. *Exp Dermatol*, 2009. **18**: p. 727-9.
191. Barnes, J. and S.K. Agarwal, Targeting STAT4 in systemic sclerosis: a promising new direction. *Expert Rev Clin Immunol*, 2011. **7**: p. 445-8.
192. Gately, M.K., et al., The interleukin-12/interleukin-12-receptor system: role in normal and pathologic immune responses. *Annu Rev Immunol*, 1998. **16**: p. 495-521.
193. Wang, K.S., D.A. Frank, and J. Ritz, Interleukin-2 enhances the response of natural killer cells to interleukin-12 through up-regulation of the interleukin-12 receptor and STAT4. *Blood*, 2000. **95**: p. 3183-90.
194. Kivelevitch, D., B. Mansouri, and A. Menter, Long term efficacy and safety of etanercept in the treatment of psoriasis and psoriatic arthritis. *Biologics*, 2014. **8**: p. 169-182.
195. Badr, M.M., et al., Dichloroacetate modulates cytokines toward T helper 1 function via induction of the interleukin-12-interferon-gamma pathway. *Onco Targets Ther*, 2014. **7**: p. 193-201.
196. Yamashita, Y., et al., Establishment and characterization of a novel anti-DNAM-1 monoclonal antibody. *Monoclon Antib Immunodiagn Immunother*, 2013. **32**: p. 60-4.
197. Koyama, M., et al., Promoting regulation via the inhibition of DNAM-1 after transplantation. *Blood*, 2013. **121**: p. 3511-20.
198. Gough, M.J. and A.D. Weinberg, OX40 (CD134) and OX40L. *Adv Exp Med Biol*, 2009. **647**: p. 94-107.
199. Pratico, E.D., B.A. Sullenger, and S.K. Nair, Identification and characterization of an agonistic aptamer against the T cell costimulatory receptor, OX40. *Nucleic Acid Ther*, 2013. **23**: p. 35-43.
200. Xiao, X., et al., New insights on OX40 in the control of T cell immunity and immune tolerance in vivo. *J Immunol*, 2012. **188**: p. 892-901.
201. Magalhaes, J.G., et al., Nucleotide oligomerization domain-containing proteins instruct T cell helper type 2 immunity through stromal activation. *Proc Natl Acad Sci U S A*, 2011. **108**: p. 14896-901.

202. Wu, M. and S. Assassi, The Role of Type 1 Interferon in Systemic Sclerosis. *Front Immunol*, 2013. **4**: p. 266.
203. Honda, K. and T. Taniguchi, IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol*, 2006. **6**: p. 644-58.
204. Barnes, B.J., et al., Multiple regulatory domains of IRF-5 control activation, cellular localization, and induction of chemokines that mediate recruitment of T lymphocytes. *Mol Cell Biol*, 2002. **22**: p. 5721-40.
205. Schoenemeyer, A., et al., The interferon regulatory factor, IRF5, is a central mediator of toll-like receptor 7 signaling. *J Biol Chem*, 2005. **280**: p. 17005-12.
206. Takaoka, A., et al., Integral role of IRF-5 in the gene induction programme activated by Toll-like receptors. *Nature*, 2005. **434**: p. 243-9.
207. Ning, S., J.S. Pagano, and G.N. Barber, IRF7: activation, regulation, modification and function. *Genes Immun*, 2011. **12**: p. 399-414.
208. Sweeney, S.E., Targeting interferon regulatory factors to inhibit activation of the type I IFN response: implications for treatment of autoimmune disorders. *Cell Immunol*, 2011. **271**: p. 342-9.
209. Cham, C.M., K. Ko, and T.B. Niewold, Interferon regulatory factor 5 in the pathogenesis of systemic lupus erythematosus. *Clin Dev Immunol*, 2012. **2012**: p. 780436.
210. Xu, W.D. and D.Q. Ye, Interferon regulatory factor (IRF)-5: a potential therapeutic target for ankylosing spondylitis. *Rheumatol Int*, 2012. **32**: p. 4065-7.
211. Wang, K., M. Li, and M. Bucan, Pathway-based approaches for analysis of genomewide association studies. *Am J Hum Genet*, 2007. **81**: p. 1278-83.
212. Hong, M.G., et al., Strategies and issues in the detection of pathway enrichment in genome-wide association studies. *Hum Genet*, 2009. **126**: p. 289-301.
213. Segre, A.V., et al., Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet*, 2010. **6**: e1001058.
214. Okada, Y., et al., Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature*, 2014. **506**: p. 376-81.
215. Macian, F., NFAT proteins: key regulators of T-cell development and function. *Nat Rev Immunol*, 2005. **5**: p. 472-84.
216. Kersten, S., et al., Peroxisome proliferator-activated receptor alpha mediates the adaptive response to fasting. *J Clin Invest*, 1999. **103**: p. 1489-98.

217. Wang, K., M. Li, and H. Hakonarson, ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 2010. **38**: p. e164.
218. Diaz-Gallo, L.M., et al., Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. *Ann Rheum Dis*, 2011. **70**: p. 454-62.
219. Kraja, A.T., et al., A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. *Diabetes*, 2011. **60**: p. 1329-39.
220. Abbondanza, C., et al., Identification of a functional estrogen-responsive enhancer element in the promoter 2 of PRDM2 gene in breast cancer cell lines. *J Cell Physiol*, 2012. **227**: p. 964-75.
221. Tan, L.Z., et al., Integrin alpha7 binds tissue inhibitor of metalloproteinase 3 to suppress growth of prostate cancer cells. *Am J Pathol*, 2013. **183**: p. 831-40.
222. Carmona, F.D., et al., Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. *Ann Rheum Dis*, 2012. **71**: p. 114-9.
223. Smemo, S., et al., Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, 2014. **507**: p. 371-5.
224. Parkes, M., et al., Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet*, 2013. **14**: p. 661-73.
225. Elhai, M., et al., Systemic sclerosis at the crossroad of polyautoimmunity. *Autoimmun Rev*, 2013. **12**: p. 1052-7.
226. Delgado-Vega, A., et al., Recent findings on genetics of systemic autoimmune diseases. *Curr Opin Immunol*, 2010. **22**: p. 698-705.
227. Lefebvre, V., R.R. Behringer, and B. de Crombrugghe, L-Sox5, Sox6 and Sox9 control essential steps of the chondrocyte differentiation pathway. *Osteoarthritis Cartilage*, 2001. **9 Suppl A**: p. S69-75.
228. Hardie, W.D., et al., Genomic profile of matrix and vasculature remodeling in TGF-alpha induced pulmonary fibrosis. *Am J Respir Cell Mol Biol*, 2007. **37**: p. 309-21.
229. Ghosh, A.K., et al., Disruption of transforming growth factor beta signaling and profibrotic responses in normal skin fibroblasts by peroxisome proliferator-activated receptor gamma. *Arthritis Rheum*, 2004. **50**: p. 1305-18.
230. Burgess, H.A., et al., PPARgamma agonists inhibit TGF-beta induced pulmonary myofibroblast differentiation and collagen production: implications for therapy of lung fibrosis. *Am J Physiol Lung Cell Mol Physiol*, 2005. **288**: p. L1146-53.

231. Kulkarni, A.A., et al., PPAR-gamma ligands repress TGFbeta-induced myofibroblast differentiation by targeting the PI3K/Akt pathway: implications for therapy of fibrosis. *PLoS One*, 2011. **6**: p. e15909.
232. Hudson, M., et al., Polyautoimmunity and familial autoimmunity in systemic sclerosis. *J Autoimmun*, 2008. **31**: p. 156-9.
233. Tenesa, A. and C.S. Haley, The heritability of human disease: estimation, uses and abuses. *Nat Rev Genet*, 2013. **14**: p. 139-49.
234. Falconer DS, M.T., Introduction to Quantitative Genetics. *Harlow, Essex, UK: Longmans Green*, 1996. 4 ed.
235. Yang, J., et al., Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*, 2010. **42**: p. 565-9.
236. Yang, J., et al., GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*, 2011. **88**: p. 76-82.
237. Lee, S.H., et al., Estimating missing heritability for disease from genome-wide association studies. *Am J Hum Genet*, 2011. **88**: p. 294-305.
238. Lee, S.H. and N.R. Wray, Novel genetic analysis for case-control genome-wide association studies: quantification of power and genomic prediction accuracy. *PLoS One*, 2013. **8**: p. e71494.
239. Purcell, S.M., et al., Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*, 2009. **460**: p. 748-52.
240. Chen, G.B., et al., Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum Mol Genet*, 2014. **In press**.
241. Gusev, A., et al., Quantifying missing heritability at known GWAS loci. *PLoS Genet*, 2013. **9**: p. e1003993.