New findings in the genetic landscape of systemic sclerosis

Doctoral Thesis

Elena López Isac, 2017

Programa de Doctorado en Biomedicina

Tesis Doctoral

New findings in the genetic landscape of systemic sclerosis

Memoria presentada por la licenciada en biología Elena López Isac para optar al grado de Doctor Internacional por la Universidad de Granada.

Programa de Doctorado en Biomedicina.

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

Instituto de Parasitología y Biomedicina López-Neyra, CSIC.
Granada, junio de 2017

Editor: Universidad de Granada. Tesis Doctorales
Autora: Elena López Isac
ISBN: 978-84-9163-330-3
URI: http://hdl.handle.net/10481/47538

La doctoranda / *The doctoral candiate* **Elena López Isac** y el director de la tesis / *and the thesis supervisor* **Javier Martín Ibáñez**

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección del director 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.

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected

Asimismo, la doctoranda **Elena López Isac** 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.

Lugar y fecha / *Place and date*:

Granada, 22 de mayo de 2017

Director de la Tesis / *Thesis supervisor*

Firma / *Signed*

Doctorando / *Doctoral candidate*

Firma / *Signed*

A Aarón,

A mi familia,

A mis amigos y compañeros

INDEX

ABBREVIATIONS

ACA: Anticentromere antibody

AD: Autoimmune disease

AKT**:** Protein kinase B

ANA: Antinuclear antibody

ARA: Anti–RNA polymerase III antibody

ATA: Anti–topoisomerase 1 antibody

CDDO: Oleanane 2-cyano-3,12-dioxoolean-1,9-dien-28-oic

CeD: Celiac disease

CMV: Cytomegalovirus

CNV: Copy number variantion

dcSSc: Diffuse cutaneous systemic sclerosis

EBV: Epstein-Barr virus

EC: Endothelial cell

ECM: Extracellular matrix

ENCODE: Encyclopedia of DNA Elements

eQTLs: Expression quantitative trait loci

GTEx: Genotype-Tissue Expression

GWAS: Genome-wide association study

HLA: human leukocyte antigen

IBD: Inflammatory bowel disease

IFN: Interferon

IMD: Immune-mediated disease

IRF: IFN regulatory factor

JAK-STAT

lcSSc: Limited cutaneous systemic sclerosis

LD: Linkage disequilibrium

MAF: minor allele frequency

MMP: Matrix metalloproteinase

ncRNA: non-coding RNA

NGS: Next generation sequencing

OR: Odds ratio

PAH: Pulmonary arterial hypertension

PDGF: Platelet-derived growth factor

PF: Pulmonary fibrosis

PI3K: fosfoinositol 3-quinasa

PPI: Protein-protein interaction

RA: Rheumatoid arthritis

ROS: Reactive oxygen species

SCR: Scleroderma renal crisis

SLE: Systemic-lupus erythematosus

SNP: Single-nucleotide polymorphism

sQTLs: Splicing quantitative trait *loci*

SSC: Systemic sclerosis or scleroderma

TF: Transcription factor (TF)

TGF-β: transforming growth factor β

Th1: T helper 1

TNF- α: tumor necrosis factor α

T1D: Type 1 diabetes

GENES

ATG5: Autophagy related 5

BLK: BLK proto-oncogene, Src family tyrosine kinase

CD247: T cell receptor zeta-chain

CSK: C-Src

DNASE1L3: Deoxyribonuclease I-like 3

GRB10: Growth factor receptor bound protein 10

GSDMA: Gasdermin A

GSDMB: Gasdermin B

IKZF3: IKAROS family zinc finger 3

IL12A: Interleukin 12A

IL12RB1: Interleukin 12 receptor, beta 1

IL12RB2: Interleukin 12 receptor, beta 2

IRF4: IFN regulatory factor 4

IRF5: IFN regulatory factor 5

IRF7: IFN regulatory factor 7

IRF8: IFN regulatory factor 8

ITGAM: Integrin subunit alpha M

JAZF1: JAZF zinc finger 1

MHC: major histocompatibility complex

NFKB1: Nuclear factor kappa B subunit 1

PPARG: Peroxisome proliferator-activated receptor gamma

PRDM1: PR/SET domain 1

PSORS1C1: psoriasis susceptibility 1 candidate 1

PTPN22: Protein tyrosine phosphatase, non-receptor type 22

RHOB: ras homolog family member B

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

STAT4: Signal Transducer and activator of transcription 4

TIMP4: TIMP metallopeptidase inhibitor 4

TNFAIP3: Tumor necrosis factor alpha-induced protein

TNFSF4: Tumor necrosis factor ligand superfamily member 4

TNIP1: TNFAIP3-interacting protein

TYK2: tyrosine kinase 2

RESUMEN

La esclerosis sistémica (SSc, del inglés “systemic sclerosis”) es una enfermedad compleja autoinmune que afecta al tejido conectivo y que presenta una expresión clínica muy heterogénea. La enfermedad está caracterizada por la presencia de fibrosis en la piel y órganos internos. En las primeras fases de la enfermedad aparecen diferentes eventos vasculares, como el síndrome de Raynaud y edema. Los principales mecanismos fisiopatológicos que subyacen a la enfermedad son el daño vascular, el desequilibrio inmunológico (incluyendo la presencia de auto-anticuerpos) y un excesivo depósito de colágeno y otros componentes de la matriz extracelular. La etiología de la enfermedad es en su mayor parte desconocida, pero se piensa que su aparición está relacionada con la combinación de factores de predisposición genética y factores ambientales.

La presente tesis doctoral se centró en el estudio del componente genético subyacente a la SSc. Previo al comienzo de esta tesis, nuestro grupo de investigación publicó el primer estudio de asociación de genoma completo (GWAS, del inglés “genome-wide association study”) en la SSc para población europea. A su vez, de forma simultánea al desarrollo de esta tesis, nuestro grupo también publicó el primer estudio de Immunochip. Ambos estudios se llevan a cabo con plataformas de genotipado de alto rendimiento y, además de identificar nuevas señales de asociación, se caracterizan porque definen la denominada “zona gris”, en la que posibles señales de asociación quedan enmascaradas por falta de poder estadístico. De este modo, tres de los estudios que comprenden la presente memoria consisten en estudios de seguimiento de la zona gris del GWAS e Immunochip. Con ello, hemos podido identificar tres nuevos *loci* de susceptibilidad

asociados con la enfermedad: *PPARG*, *IL12RB1* y *TYK2*. Estos tres genes presentan grandes implicaciones para el conocimiento de las vías patogénicas que subyacen a la SSc. En primer lugar, el gen *PPARG* codifica un receptor nuclear con una potente actividad anti-fibrótica. En segundo lugar, el hallazgo de *IL12RB1* y *TYK2*, junto con estudios previos, refuerzan el importante papel de la ruta de la IL-12 e IL-23 en la SSc. Nuestros resultados evidencian que esta vía de señalización podría ser interesante como nueva diana terapéutica para la enfermedad.

Por otro lado, también se llevó a cabo el análisis del componente genético compartido entre la SSc y la artritis reumatoide (RA, del inglés "rheumatoid arthritis") mediante un meta-GWAS que combinó datos de GWAS para ambas enfermedades. Este estudio nos permitió identificar al gen *IRF4* como nuevo *locus* de susceptibilidad compartido. Además, confirmamos otros *loci* comunes previamente descritos.

A su vez, aprovechando nuestras amplias cohortes de SSc procedentes de distintos países europeos y de Estados Unidos, quisimos replicar la asociación descrita para una variante exónica y rara en el gen *ATP8B4* identificada mediante secuenciación de genoma completo. Nuestros resultados descartan el posible papel de dicha variante en la predisposición a la SSc y resaltan la importancia de los estudios de replicación para la confirmación de forma robusta de las asociaciones genéticas.

Finalmente, para la discusión de la presente memoria se ha indagado en las señales de asociación comunes a SSc y RA mediante la caracterización funcional de las variantes asociadas en los *loci* compartidos por ambas enfermedades. Estos análisis revelan un extenso solapamiento no sólo en los factores de susceptibilidad sino

también en los mecanismos etiopatogénicos. Asimismo, en la discusión se ofrece una visión global del componente genético conocido hasta la fecha para la SSc, y se integra este conocimiento con la información disponible a través de diversas bases de datos públicas para caracterizar funcionalmente las asociaciones.

INTRODUCTION

1. Systemic sclerosis, a complex autoimmune disease

The immune system has two major cellular components to create the adaptive immune response: B cells and T cells. Both types of cells have the ability to recognize antigens, including autoantigens. The ability to distinguish between pathogens and self-antigens is critical to the success of the adaptive immune system and there are several self-tolerance immune mechanisms to protect against self-reactive B and T cells. A breakdown in this immunological tolerance can trigger autoimmunity, leading to erroneous immune responses that damage healthy tissues (1).

More than 80 autoimmune diseases (ADs) have been described to date and, unfortunately, the vast majority of them are chronic, debilitating and have no cure. ADs are also heterogeneous with regard to prevalence, manifestations, and pathogenesis. Overall, the estimated prevalence is 4.5%, although there is a considerable epidemiological variability for different ADs, ranging from common (such as rheumatoid arthritis or Hashimoto's thyroiditis) to rare diseases (such as systemic sclerosis or systemic lupus erythematosus) Moreover, prevalence of ADs is normally higher in women than men (2.7% for males and 6.4% for females) and varies by ethnicity (2, 3).

There are two groups of ADs: organ-specific diseases, if they affect particular targets of the body (for example, type 1 diabetes and Graves' disease); or systemic diseases, if they affect multiple organs and tissues (such as rheumatoid arthritis, systemic sclerosis or systemic

lupus erythematosus). Moreover, ADs can be Mendelian disorders (monogenic or polygenic diseases), such as autoimmune polyendocrine syndrome 1, in which mutations in the *AIRE* gene (that encodes the transcription factor autoimmune regulator) lead to a relaxed negative selection of self-reactive T cells in the thymus; or complex diseases, in which both genetic and environmental factors play role in the development of the disorder (Figure 1), as in the case of systemic sclerosis or scleroderma (SSc) (4).

SSc is a chronic and heterogeneous AD that affects the connective tissue. Its pathogenesis involves extensive fibrosis of skin and internal organs, vascular alterations, and immune imbalance (5-7). The main vascular abnormalities include Raynaud's phenomenon, renal crisis and pulmonary arterial hypertension (PAH). The lungs, heart, kidneys and esophagus are the main internal organs affected by fibrosis, although this complex disease can cause severe dysfunction and failure of almost any internal organ. While esophageal dysfunction is the most common visceral complication, lung involvement (both pulmonary hypertension and pulmonary fibrosis) is the leading cause of death (8). Immune imbalance includes altered lymphocyte activation that leads to autoantibody production, aberrant cytokine release, and deregulation of the innate immune system (5, 6).

Although SSc is a very heterogeneous disease, patients are usually classified in two main subgroups: limited cutaneous systemic sclerosis (lcSSc) and diffuse cutaneous systemic sclerosis (dcSSc), with a prevalence of approximately 65% and 35%, respectively. In lcSSc, fibrosis is mainly restricted to the skin of hands, arms, and face. Raynaud's phenomenon appears several years before fibrosis, and PAH is frequent. dcSSc is characterized by a more aggressive, generalized

and rapidly progressing fibrosis that affect skin of all body and one or more visceral organs (9-12).

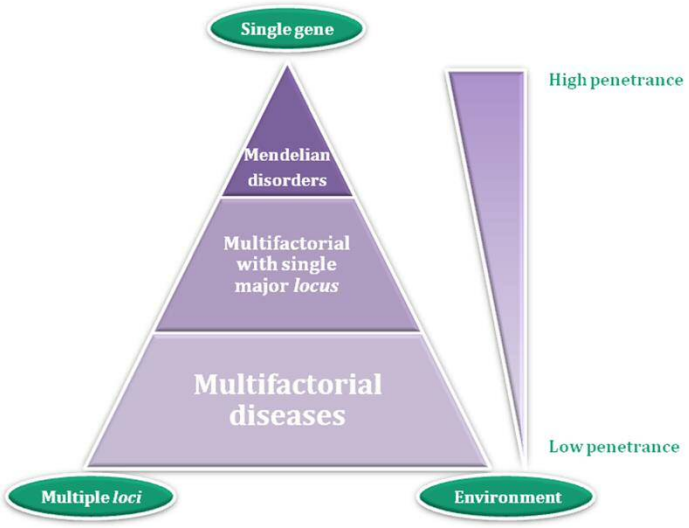


Figure 1. Classification of diseases according to the genetic component. Most of autoimmune diseases belong to the first category 'Multifactorial diseases'.

Different types of autoantibodies can be observed in patients with SSc, including antinuclear antibodies (ANAs), anti-fibroblast and anti-endothelial cell antibodies. ANAs are present in around a 90% of patients with SSc. There are three major subclasses of ANAs: anticentromere antibodies (ACAs), anti–topoisomerase 1 antibodies (ATAs) and anti–RNA polymerase III antibodies (ARAs). These subclasses of SSc autoantibodies tend to be mutually exclusive (5). Although the potential role of ANAs in the pathogenesis of SSc is unclear, they are a well-recognized clinical tool for both diagnosis and classification of SSc patients. Approximately 70% of lcSSc patients present ACAs, which are well-correlated with risk of PAH (5); while ATAs are more frequently observed in dcSSc patients (in around a 20% of dcSSc) and it is correlated with worse prognosis and increased risk of

pulmonary fibrosis (PF) (13) . ARAs are present in around 20% of dcSSc patients and their presence is well-correlated with risk of scleroderma renal crisis (SCR) (14, 15).

As most ADs, SSc affects women more frequently than men, commonly in a ratio ranging from 3:1 to 12:1 (16). SSc typically appears in middle-aged individuals, although it can also affect children, and the elderly. It is a rare disease that has one of the highest mortality rates in comparison with other rheumatic diseases (17). The estimated prevalence ranges from 7 to 700 cases per million among the different studies due to a noteworthy variation across geographic regions and ethnic differences (18, 19). Thus, black population has a higher risk for SSc than white and Asian populations (17, 20). To date, the highest prevalence has been observed in a Native American tribe, the Choctaw population, with 660 cases per million (21). Moreover, an increasing North-to-South gradient in European population has been reported (22). In Spain, specifically, a 19-year epidemiologic study carried out in Northwestern Spain estimated a prevalence of 277 SSc cases per million and an incidence of 23 cases per million and year (23). Mean age at diagnosis and mean age at disease onset were estimated at 51.2±15.1 and 44.9±15.8 years, respectively, according to the Registry of the Spanish Network for Systemic Sclerosis (24).

The etiology and pathogenic mechanisms underlying SSc remain poorly understood, but the disease is thought to be caused by a complex interplay among vascular, immune and fibrotic altered processes in association with a genetic susceptibility. The traditional proposed model suggests that microvascular injury and endothelial cell (EC) activation are the primary events in SSc (25-27) (Figure 2). This hypothesis arises from the observation that vascular damage

(Raynaud's phenomenon and edema) is the earliest feature that takes place in the disease. The factors that promote the vascular injury are unknown, but could be viral infections, autoantibodies, toxins or oxidative stress. In the initial states of the disease, there is increased vascular permeability that facilitates mononuclear cell infiltration, leading to perivascular inflammation. Vascular damage evolves with vascular intimal thickening, vessel narrowing and obliteration, leading to tissue ischemia. Vessels lose their elasticity and become fibrotic, ultimately leading to organ dysfunction (25-27). In addition, there is an increased expression of adhesion molecules that allows the recruitment of inflammatory cells. Activated ECs also secrete vasoconstricting agents, such as endothelin 1, along with a decreased expression of vasodilating agents and platelet activation. Deregulation of angiogenesis, vasculogenesis and altered cytokine production are also observed in SSc (5).

The vascular inflammatory phase is more prominent in the earlier stages of SSc, but it is gradually replaced by fibroblast activation and excessive collagen deposition, ultimately leading to fibrosis (Figure 2). On this base, it is thought that the inflammatory environment plays a relevant role in the development of fibrosis (26). Deregulation of both innate and adaptive immunity is observed in SSc patients. Several immune cells are observed in skin and lung infiltrate of patients in initial states of the disease, including T cells, macrophages, plasmacytoid dendritic cells and mast cells (28, 29). Interestingly, these cells show a prominent type I interferon (IFN) signature (30-32). Moreover, T cells in SSc express an activated phenotype and signatures of antigen-driven cell expansion (33-35). Most of T cells observed in SSc infiltrates are T helper 2 (Th2) cells, characterized by secretion of profibrotic mediators (such as IL-4, IL-13, IL-6 and TGFβ) (36). Several

in vitro and *in vivo* studies, along with studies in animal models, support the prominent role of Th2 cytokines in SSc pathogenesis (36, 37). Moreover, a Th1-Th2 cytokine imbalance toward increased levels of Th2 cytokines has been described in patients with SSc (38). Th1 cytokine levels (such as IFNγ, TNFα, IL-2 and IL-12) are also altered in SSc and promote inflammation (39). The role of Th17 cells remains controversial, since some studies support a role of IL-17 in fibrosis, while other studies indicate an anti-fibrotic effect (40). B cells are also present in skin and lung infiltrate of patients with SSc and show dysregulated homeostasis (41). The role of B cells in SSc is not only restricted to the production of autoantibodies, since activated B cells secrete IL-6, which directly stimulate fibroblasts. This cytokine is also produced by T cells, fibroblast and ECs, and different *in vitro* and mice model studies clearly support its role in the induction of inflammation and fibrosis in SSc (42-44).

The vascular injury, EC activation and the uncontrolled and altered immune reaction ultimately give rise to fibroblast activation and fibrosis (Figure 2). The activated fibroblasts undergo differentiation into myofibroblast, leading to an excessive accumulation of collagen and other components of the extracellular matrix (ECM) (26). In SSc, TGF-β, that stimulates collagen synthesis in fibroblast and myofibroblast, is considered a master regulator of fibrosis (45). Deregulated TGF-β signaling in fibroblast and myofibroblast has been reported in various studies in patients with SSc (46). In addition, fibroblast and myofibroblast in SSc show particular traits, such an altered expression and deregulated responses to cytokines, an increased proliferation and decreased apoptosis (47). Finally, recent studies also implicate reactive oxygen species (ROS) in the pathogenesis of SSc (48). Tissue ischemia and activated fibroblast can

lead to the generation of these chemical species. In fact, high levels of ROS have been observed in SSc, and several *in vitro* and mouse model studies further support the profibrogenic effect of these chemical species in fibroblast (48-51) (Figure 2).

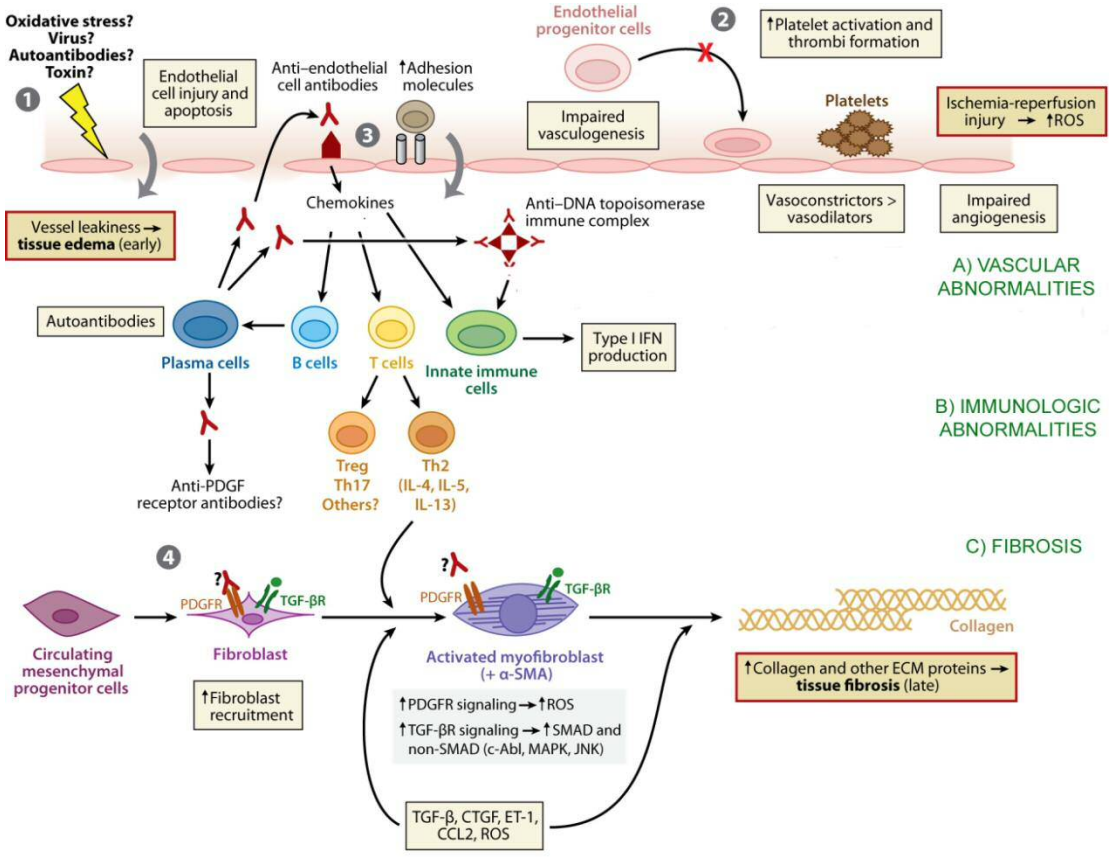


Figure 2. Summary of the proposed pathogenic mechanisms involved in systemic sclerosis (Modified from Katsumoto et al, Annu Rev Pathol. 2011).

2. Etiology of systemic sclerosis

As we stated in the previous section, the etiology of SSc remains unclear, although it is thought to result from a complex interaction between multiple genetic risk factors, each of them with a modest effect in the disease susceptibility, and environmental features, triggering the onset of the disease and affecting its severity and progression (Figure 3). The current knowledge of both etiological aspects will be reviewed in this section, although the genetic component of SSc will be further addressed in section 4.

2.1 Environmental factors

Overall, the environment may affect the development of the disease by: 1) directly activating the immune response; 2) modifying self-antigens by external substances and triggering the loss of tolerance to self; or 3) molecular mimicry, a process in which a foreign antigen shares sequence or structural similarities with self-antigens leading to cross-reaction in the immune response.

The importance of environmental factors in the onset of SSc is not robustly established, mainly due to the absence of good sample size and other methodological limitations in the studies. Here we describe some of the most investigated agents.

A) Chemical agents.

Among this category, the exposure to silica is the most studied factor. The first evidences of the involvement of this chemical agent in SSc date back to 1914, when Bramwell described SSc in Scottish stonemasons (52). In 1957, Erasmus reported that gold miners exposed

to crystalline silica had higher risk for SSc than the general population (53). On the light of these evidences, several case-control studies have been performed in order to evaluate and quantify the risk that the exposure to silica confers for the development of SSc. All of them have found a positive correlation between silica and SSc (54-58). The mechanisms underlying this association have not been elucidated, although there are some experimental studies that have shown an imbalance of the immune response caused by the exposure to this agent (including T and B cells activation, autoimmunity-related apoptosis and increased fibroblast proliferation) (59-61).

Occupational exposition to organic solvents, like trichloroethylene, benzene, xylene, chlorinated solvents, aromatic solvents, white spirit, ketones and other molecules sharing structural characteristics, has been reported to increase SSc risk. The results of different studies attempting to determine the risk increase for suffering SSc caused by organic solvents have not obtained very conclusive findings (62, 63). Nevertheless, occupational exposition to organic solvents has been proven to have a predictive value of SSc severity (63). For example, it has been observed that SSc patients who were exposed to organic solvents exhibited more frequently dSSc and microangiopathy (64). As in the case of silica, the pathogenic mechanisms through which these chemical factors influence disease onset remain unknown, although -on the light of several experimental studies- it has been postulated that organic solvents may link with nucleic acids and proteins, leading to altered immune response (62).

Welding fumes, asbestos, vinyl chloride, epoxy resins and formaldehyde are other industrial agents that have been related to SSc (65). Interestingly, a massive chemical intoxication that took place in central and northwest provinces in Spain in 1981-1982 caused an SSc-like illness that was called toxic oil syndrome (TOS). Several agents

were investigated and a significant association was found with the consumption of rapeseed oil contaminated with 1,2-di-oleyl ester (DEPAP) and oleic anilide (66). Some of the main effects of these chemical components were a non-necrotizing vasculitis in multiple organs, liver disease and pulmonary hypertension. The immunological analyses of the patients pointed out to altered activation of T-cells and cytokine production among the immune mechanisms underlying the disease. Moreover, several studies have reported a significant association of disease severity with certain HLA-DR2 alleles and polymorphisms in metabolism and immune response genes (66).

B) Infections.

Several viruses and other infectious agents have been proposed and investigated as potential environmental triggers. Some of them are also related with other ADs, as in the case of parvovirus B19. Interestingly, the presence of the parvovirus B19 has been detected in bone marrow biopsy of 57 % of SSc patients and they have the ability to persistently infect SSc fibroblasts (67, 68). Members of the herpesvirus family, such us cytomegalovirus (CMV) and Epstein-Barr virus (EBV) are other infectious agents related to SSc etiology. Some SSc patients have serum autoantibodies that bind the UL94 epitope of CMV and induce ECs apoptosis (69). Moreover, Farina *et al.* found RNA from EBV in fibroblasts, myofibroblasts and ECs in the skin of SSc patients (70). In the same study, the authors reported that EBV was able to persistently infect human SSc fibroblasts *in vitro* and induce a dysregulated innate immune response. Another study linked certain retroviruses with SSc after the observation of high homology between retrovirus antigens and the terminal end of DNA topoisomerase I (71).

C) Pregnancy. Other factors.

Pregnancy related events, such as fetal microchimersm (this is, the presence of DNA of the offspring in maternal blood), have also been considered an environmental factor involved in SSc (72).

Life-style, diet and dietary contaminants or physical agents (such as ionizing radiation) could also be related to the development of SSc, but this hypothesis requires further studies since there are no robustly confirmed data.

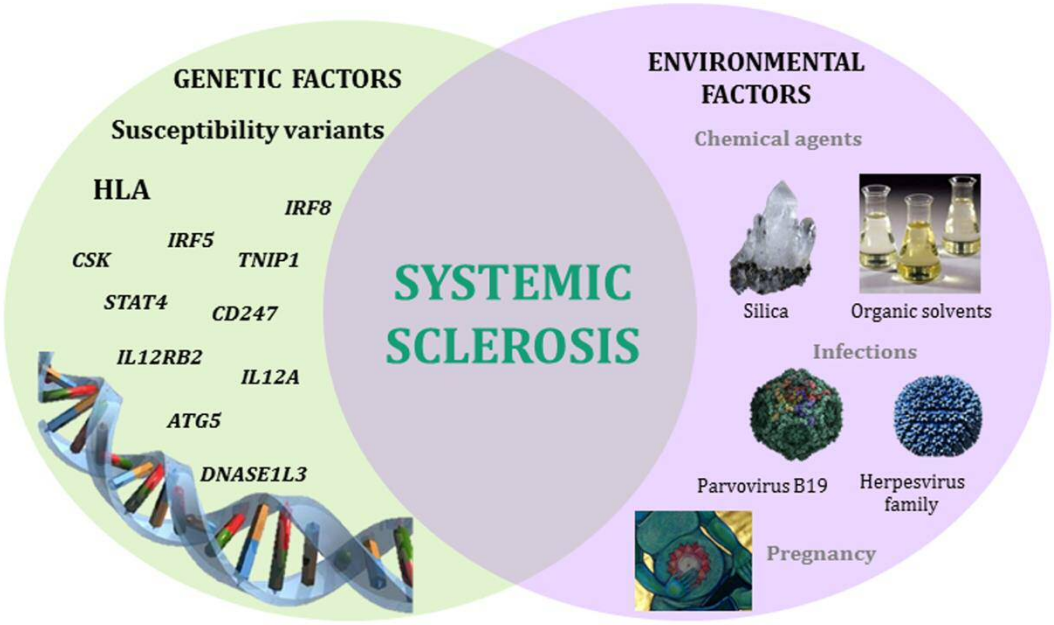


Figure 3. Etiology of systemic sclerosis, a complex autoimmune disease determined by both genetic and environmental factors.

2.2 Genetic component

One of the evidences that support the genetic component of SSc is the familial occurrence observed in the disease. In fact, a positive familial history of SSc is the major risk factor reported to date (73). Familial and twin studies in SSc have described a high concordance of autoantibody production (74, 75). Moreover, in a study that comprised 703 families, first-degree relatives of patients with SSc showed relative risk ranging from 10-16-fold higher than in the general population, and siblings showed relative risk ranging from 10–27-fold higher than normal (73). In addition, the prevalence variation that it is observed among populations also supports the role of the genetic component in SSc.

The estimation of heritability of complex diseases is a challenging task. In the case of SSc, efforts in the estimation of disease heritability have not provided conclusive reports, mainly due to limited sample sizes (76). The large numbers of SNPs provided by GWAS and Immunochip has offered new opportunities to develop new methodologies for predicting genetic risk of complex diseases in a more accurate way, as in the case of GREML (77-79). This method has been applied in SSc and is based on the assumption that more genotype sharing between unrelated individuals in case-control studies should mean greater phenotypic concordance for any complex trait (77-79) . GRELM estimated SSc heritability on the observed scale (h_o^2) of 0.39 and 0.44 with the GWAS and Immunochip data performed by our group, respectively (76). h_o^2 is the proportion of variance in case-control status that is explained by all SNPs, and can be transformed into the unobserved underlying scale of disease liability (h_l^2) if the prevalence of the disease in the population and the proportion of cases are

considered. The estimation reported an $h_l^2 = 0.09$ and $h_l^2 = 0.10$ in the GWAS and Immunochip data, respectively (76). It is important to note that this estimation cannot be understood as a direct measure of SSc heritability, but it represents the variance in SSc that was explained by SNPs at the population level.

The significant SNPs described to date for SSc only account for a ~20% of the estimated SSc heritability (76). Thus, it is expected that additional SSc risk *loci* remain to be discovered. The results of the genetic association studies that have been performed to date with the aim to identify SSc susceptibility *loci* are addressed in section 4.

3. The study of the genetic component of complex diseases

Nowadays, the starting point for understanding the genetic bases of complex diseases is the identification of genetic markers associated with the interrogated phenotype. There are several genetic markers that can be used for this purpose, such as microsatellites, copy number variations (CNVs) or single-nucleotide polymorphisms (SNPs). Among them, SNPs are the most widely used in genetic association studies.

A SNP is a variation at a single position in a DNA sequence among individuals, leading to different alleles (commonly bi-allelic SNPs). By definition, the frequency of the minor allele (MAF) has to be higher than 1% in the overall population. To date, around 38 million of SNPs have been reported and validated, and it is estimated that they account for around a 90% of the human genome genetic variation (80). The association of a SNP with a phenotype is determined by genetic

association studies, which consist on case-control studies that compare the MAF of one or more SNPs between cases (individuals diagnosed with the disease) and controls (unaffected individuals). If the difference between the MAF in cases and controls is statistically significant, that is the p-value for association is below the significance threshold, then the SNP is considered as associated with the disease (81).

The Human Genome Project was a crucial event for genomic research. For the first time, a full reference genome sequence was announced, providing a wealth of data that would help to determine the genetic contribution to human traits (82, 83). With this, the identification of genes associated with disease started. Initially, genetic association studies for complex traits appeared shaped like candidate gene studies, in which a relatively small number of variants of a specific gene or genes are tested for association. The selection of genes is always orchestrated under the previous hypothesis of being plausible genes implicated in the pathogenesis of the disease (for the functional role of the gene or for a previous study in which an association with a related disease has been reported).

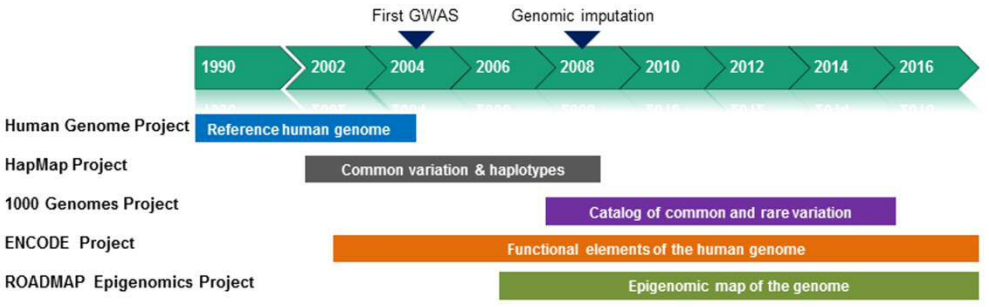


Figure 4. Relevant events in genomic development.

The International HapMap project was the next big event in genomic development (Figure 4) (84-86). This project is a systematic effort to characterize genetic variation, recombination hotspots, haplotypes and linkage disequilibrium (LD) patterns in different populations. The knowledge of genome-wide SNP data and haplotype structure allowed the design of genotyping arrays that capture a large proportion of common genetic variation, while genotyping only a few hundred thousand of SNPs in a cost-effective manner. Since the first genome-wide association study (GWAS) in 2005 (87), thousands of SNPs have been associated with hundred traits. Considering the multiple tests performed in a GWAS, a stringent threshold for statistical significance is needed in order to correct for false-positive associations. The standard significance threshold usually applied in GWAS is $p < 5 \times 10^{-08}$, which corresponds to the so-called Bonferroni correction for testing 1,000,000 independent common SNPs (88). In addition, GWASs have two other distinctive traits: they are hypothesis-free studies, because the selected SNPs are included in the array without the previous idea of being plausible SNPs; and hypothesis generating, since the novel discovered *loci* may pinpoint new molecular pathways involved in the pathogenesis of the disease (Figure 5).

Later on, the 1000 Genomes Project was launched with the goal of providing further characterization of the human genome variation, including not only common but also rare variation (Figure 4). This project included 14 populations and applied whole-genome sequencing and exome sequencing, along with SNP genotyping (80). With this deeper knowledge, genotyping arrays have evolved to much more efficient platforms and they can currently contain more than 2 million SNPs.

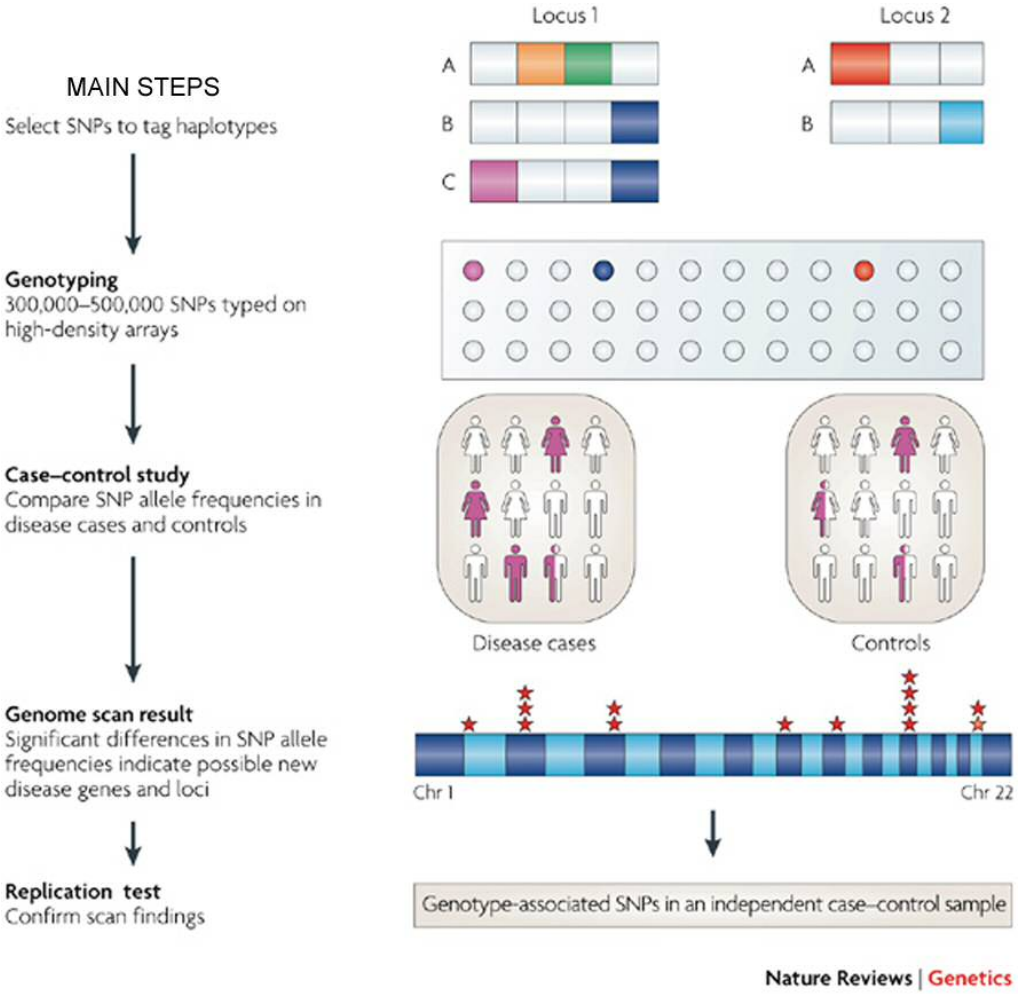


Figure 5. Main steps for genome-wide association studies (Modified from Mathew CG, Nat Rev Genet, 2008)

The large population-specific reference panels provided by The International HapMap project and 1000 Genomes Project also gave rise to genotype imputation algorithms. Genotype imputation is a process that enables to infer missing genotypes (non-genotyped variants) based on nearby observed genotypes and using a reference panel (89). There are various imputation methods available, although all of them are

based on the comparison of haplotypes between the individuals of the study and the reference panel (Figure 6). This process boosts the number of SNPs tested for association, increasing the power of the study and the ability to fine-map associations and to identify the causal SNP. Moreover, imputation facilitates combining association results from different GWASs in a meta-analysis (meta-GWAS). Meta-GWAS strategy has been widely applied and has produced a large amount of new *loci* since they increase the statistical power and consequently the chance to identify significant association signals (90).

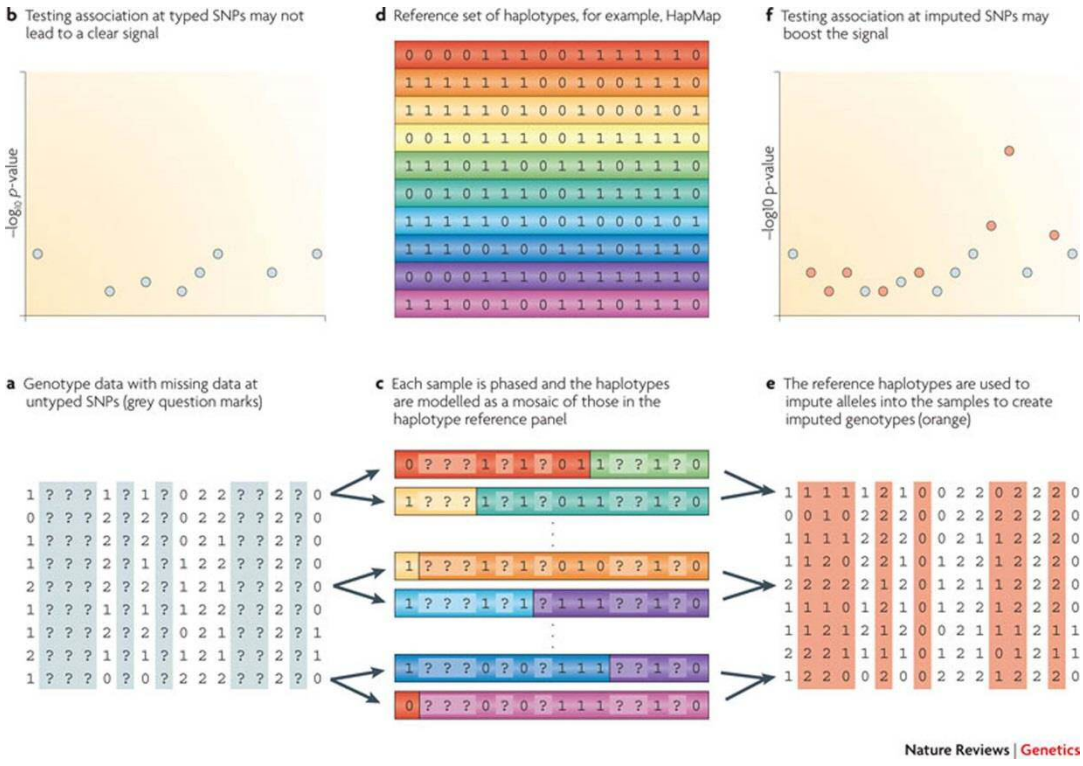


Figure 6. Main features of imputation process (Marchini and Howie, Nat Rev Genet, 2010). Imputation methods attempt to impute the missing data (?) in study individuals (represented by each row in 'a') by the identification of sharing between the underlying haplotypes and the haplotypes in the reference set.

Despite the success provided by GWASs, the majority of associated SNPs have modest effect sizes, with odds ratios (ORs) ranging from 1.1 to 1.3, and they tend to explain a relatively small proportion of the heritability of many complex diseases (91). Therefore, it has been proposed that this 'missing heritability' may lie on non-well covered region from arrays or on rare variants with large effect sizes. This issue has led to a next phase of genotyping arrays with the goal of fine-mapping GWAS associated *loci.* The Immunochip is an example of that, which was created and designed for fine-mapping of *loci* associated with ADs or immune-mediated diseases (IMDs) (92). This custom genotyping platform contains 196,524 variants (including SNPs and small insertions-deletions) for fine-mapping 186 autoimmunity *loci,* and a dense coverage of the HLA region. The array includes all markers described for white population by 1000 Genomes Project, dbSNP and additional sequencing projects (92).

Once a *locus* is identified as a risk factor for a disease, the next step is to (attempt to) identify the causal variant/s and to provide biological sense of the association. As it was stated above, fine-mapping strategy facilitates the first task. In addition, there are a number of bioinformatics approaches and publicly available databases that provide further help for SNP prioritization and functional annotation. If the associated SNP is a coding variant, the functional characterization of the association is fairly straightforward with tools such as SIFT and PolyPhen-2 (93, 94). However, the vast majority of GWAS-associated SNPs lie on non-coding regions and they probably affect regulatory mechanisms. In this case, functional annotation is mandatory. Two main events have made possible the hard task of integrate functional information: 1) The Encyclopedia of DNA Elements (ENCODE), launched with the goal of providing a deep characterization of

functional elements in the human genome, including elements that act at the protein and RNA levels, and regulatory elements; and 2) The NIH Roadmap Epigenomics Project, which aims to produce resource of human epigenomic data, including DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts (95, 96). These very valuable projects are very useful for follow-up of disease-associated variants and for unraveling the biological mechanisms underlying associations.

4. Genetic component of systemic sclerosis: previous studies

As in the case of other complex ADs, the interrogation of the genetic bases underlying SSc started with candidate gene studies, which used to comprise relatively small cohorts. Despite the limited sample size, some of them were able to identify susceptibility genes that are currently considered as firmly associated genes with the disease, such as *STAT4, IRF5* and the HLA region (97). In 2010, the first GWAS in SSc in European population was published (98). Our group was involved in the study, which identified *CD247* as a novel gene associated with SSc risk, and confirmed the previously reported associations in the HLA region, *STAT4* and *IRF5*. Interestingly, the findings on *CD247* were independently replicated by Dieudé *et al*. (99). One year later, a second GWAS in SSc was published by Allanore *et al.*, which comprised only one French cohort and had lower statistical power than the first SSc GWAS (100). Our group was involved in the independent replication of the findings from this second GWAS, which reported *TNIP1*, *RHOB* and *PSORS1C1* as novel susceptibility *loci*. We could confirm *TNIP1* signal,

but *RHOB* and *PSORS1C1* were discarded as genetic risk factors for SSc (101). This fact highlights the relevance of a high statistical power in GWAS, since GWAS results tend to present inflated effect sizes -also called the winner's curse-. In addition, another SSc GWAS has been performed in Korean population (102). This study also comprised a relatively small cohort and did not identify significant new *loci,* although the results showed the strong association in the HLA region with SSc.

Another distinctive trait of GWASs is the so-called grey zone, where SNPs with tier 2 associations (*P*-values between 5×10^{-8} and 5×10^{-3}) are located. Follow-up studies focused on this grey zone constitute one of the most useful GWAS data mining methods, since possible real association signals could be masked in that area due to a lack of statistical power. In SSc, these types of studies have been successful in the identification of new risk *loci*. Bossini-Castillo *et al*. performed a follow-up focused on *IL12RB2*, a *locus* that showed suggestive signal of association in the first SSc GWAS (103). They analyzed the signal in a large independent European cohort and reported the association of *IL12RB2* at the genome-wide significance level. Later on, Martin *et al*. carried out a large follow-up of the GWAS that included 768 polymorphisms selected from the grey zone and they could identify *CSK* as a genetic risk factor for SSc and confirmed previously reported associations (104). Taking advantage of our GWAS data, we also performed a follow-up of the grey zone from the French GWAS. The results of this work are part of the present thesis (105).

The Immunochip has gathered important achievement on the genetic component of IMDs (106). Applying this fine-mapping approach, our group identified several new SSc susceptibility *loci* (*DNASE1IL3*, *IL12A* and *ATG5*) that implicated new biological pathways

into the pathogenesis of the disease, such as apoptosis and autophagy (107). In addition, an extensive analysis of the HLA region was performed. As in most ADs, the HLA region is the major genetic association reported to date for SSc. The dense coverage of this genomic region in the Immunochip, along with novel imputation methods that enable the inference of classical HLA alleles and even polymorphic amino acid positions from genetic data, allowed our group to describe a comprehensive model that explained all the observed associations in the region. The model includes six polymorphic amino acid positions in HLA-DRB1, HLA-DQA1 and HLA-DPB1, and seven SNPs independently associated. The analysis also confirmed the divergent HLA allele associations between ACA-positive and ATA-positive serological subgroups (107). Later on, a second SSc Immunochip performed in an Australian cohort with relatively small sample size confirmed some of the reported associations (108).

The Immunochip platform has significantly contribute to the idea of a shared genetic component among IMDs (106). With the aim of delving into the common genetic bases of ADs, the scientific community developed another approach that lies in combining genome-wide genotype data from two autoimmune diseases (cross-phenotype GWAS). This systematic approach has been widely applied during the past five years and has showed encouraging results (109-114). Our group performed a combined-phenotype GWAS with SSc and systemic-lupus erythematosus (SLE), another AD that shares several genetic susceptibility *loci* and clinical features with SSc. This study identified three new shared susceptibility *loci*, increasing the knowledge of the genetic overlap in ADs (113). During the period of this thesis, we have applied this approach for SSc and rheumatoid arthritis (RA) in a study that comprised more than 8,000 SSc patients, 16,000 RA patients and

43,300 controls. This cross-disease meta-GWAS has been consequently included in the present PhD dissertation (115).

Table 1 summarizes the *loci* that have been reported for SSc at the genome-wide significant level, excluding the findings presented in this thesis. The interrogation of the biological pathways that are involved in SSc according to these firm genetic risk factors showed the prominent role of the immune imbalance in the susceptibility to SSc (Figure 7). Furthermore, some of the SSc risk *loci* play a role in other processes such as angionenesis (through platelet-derived growth factor (PDGF) signaling pathway) and TGF-β activation (through integrin signalling pathway). Thus, the genetic component of SSc also reflects the complex molecular network underlying the disease.

Despite GWASs and Immunochip have reached great advances in understanding the genetic bases of SSc, the number of well-established susceptibility *loci* is relatively low in comparison with other ADs, such as RA and SLE, for which 101 and 43 *loci* have been validated, respectively (116, 117). The low prevalence of SSc makes difficult the recruitment of large cohorts required to reach a high statistical power and to effectively detect association signals. Over the past seven years, our group has coordinated collaborative efforts that have allowed us to gather close to 10,000 patients of SSc and more than 16,000 healthy controls. The present PhD dissertation clearly reflects the successful of this collaborative network that has allowed us to continue increasing our knowledge of the genetic predisposition to SSc.

Table 1. Non-HLA *loci* associated with systemic sclerosis susceptibility that have reached the genome-wide significance level

Locus	Chr	Gene name	Approach	SNP	References
Innate immunity, interferon signature and inflammation					
IRF5	7	IFN regulatory factor 5	Candidate gene (Confirmed in GWAS and Immunochip)	rs2004640, rs10954213,rs2280714, s10488631,rs12537284,rs4728142, rs3757385	Dieude 2009a, 2010a;Ito 2009; Radstake 2010; Allanore 2011; Mayes 2014
IRF8	16	IFN regulatory factor 8	GWAS	rs11642873	Gorlova 2011
TNIP1	5	TNFAIP3-interacting protein	GWAS	rs2233287, rs4958881, rs3792783	Allanore 2011;Bossini-Castillo 2013
Adaptative immune response: B and T cell proliferation, survival and cytokine production					
CD247	1	T cell receptor zeta-chain	GWAS	rs2056626	Radstake 2010; Dieude 2010a; Allanore 2011
CSK	15	C-Src	GWAS follow-up study	rs1378942	Martin 2012
STAT4	2	Signal Transducer and activator of transcription 4	Candidate gene (Confirmed in GWAS and Immunochip)	rs7574865,rs11889341,rs8179673, rs10181656, rs6752770, rs3821236	Rueda 2009; Tsuchiya 2009; Gourh 2009; Dieude 2009b; Radstake 2010;Allanore 2011; Mayes 2014
IL12A	3	Interleukin 12A	Immunochip	rs77583790	Mayes 2014
IL12RB2	1	Interleukin 12 receptor, beta 2	GWAS follow-up study	rs3790567	Bossini-Castillo 2011
Other processes					
DNASE1L3	3	Deoxyribonuclease I-like 3	Immunochip	rs35677470	Mayes 2014, Zochling 2014
ATG5	6	Autophagy related 5	Immunochip	rs9373839	Mayes 2014

Note: This table does not include the findings from the publications that comprise the present thesis, and the two novel *loci* (*GSDMA*, *PRDM1*) identified in a trans-ethnic meta-GWAS contemporary to the preparation of the present PhD dissertation (173).

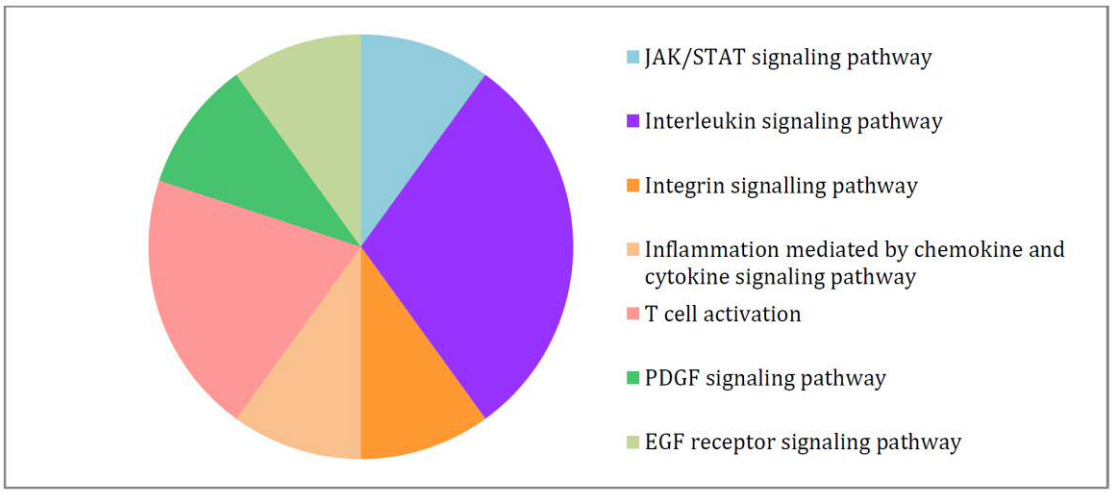


Figure 7. Pathway**s** involved in systemic sclerosis according to the well-established susceptibility *loci* for the disease (Table 1), excluding the novel *loci* reported in this thesis. Pathway analysis was performed by means of Panther Classification System.

OBJECTIVES

The global objective of the present thesis was to further explore the genetic component of systemic sclerosis by means of genetic association studies.

The specific objectives were:

1. To identify novel *loci* associated with the susceptibility to systemic sclerosis to keep increasing our understanding of the pathological mechanisms underlying the onset and progression of the disease.

2. To perform follow-up studies of the grey zone from the first SSc genome-wide association study in European population and the first SSc Immunochip.

3. To further explore the common genetic component between systemic sclerosis and rheumatoid arthritis.

4. To independently replicate in large SSc cohorts the recently reported association of a functional rare variant at *ATP8B4* with the predisposition to SSc.

PUBLICATIONS

V

Publication 1: A genome-wide association study follow-up suggests a possible role for *PPARG* in systemic sclerosis susceptibility

López-Isac et al. Arthritis Research & Therapy 2014, 16:R6
http://arthritis-research.com/content/16/1/R6

arthritis research&therapy

RESEARCH ARTICLE Open Access

A genome-wide association study follow-up suggests a possible role for *PPARG* in systemic sclerosis susceptibility

Elena López-Isac[1], Lara Bossini-Castillo[1], Carmen P Simeon[2], María Victoria Egurbide[3], Juan José Alegre-Sancho[4], Jose Luis Callejas[5], José Andrés Roman-Ivorra[6], Mayka Freire[7], Lorenzo Beretta[8], Alessandro Santaniello[8], Paolo Airó[9], Claudio Lunardi[10], Nicolas Hunzelmann[11], Gabriela Riemekasten[12], Torsten Witte[13], Alexander Kreuter[14], Jörg H W Distler[15], Annemie J Schuerwegh[16], Madelon C Vonk[17], Alexandre E Voskuyl[18], Paul G Shiels[19], Jacob M van Laar[20], Carmen Fonseca[21], Christopher Denton[21], Ariane Herrick[22], Jane Worthington[22], Shervin Assassi[23], Bobby P Koeleman[24], Maureen D Mayes[23], Timothy RDJ Radstake[25], Javier Martin[1*] and Spanish Scleroderma Group

Abstract

Introduction: A recent genome-wide association study (GWAS) comprising a French cohort of systemic sclerosis (SSc) reported several non-HLA single-nucleotide polymorphisms (SNPs) showing a nominal association in the discovery phase. We aimed to identify previously overlooked susceptibility variants by using a follow-up strategy.

Methods: Sixty-six non-HLA SNPs showing a *P* value $<10^{-4}$ in the discovery phase of the French SSc GWAS were analyzed in the first step of this study, performing a meta-analysis that combined data from the two published SSc GWASs. A total of 2,921 SSc patients and 6,963 healthy controls were included in this first phase. Two SNPs, *PPARG* rs310746 and *CHRNA9* rs6832151, were selected for genotyping in the replication cohort (1,068 SSc patients and 6,762 healthy controls) based on the results of the first step. Genotyping was performed by using TaqMan SNP genotyping assays.

Results: We observed nominal associations for both *PPARG* rs310746 ($P_{MH} = 1.90 \times 10^{-6}$, OR, 1.28) and *CHRNA9* rs6832151 ($P_{MH} = 4.30 \times 10^{-6}$, OR, 1.17) genetic variants with SSc in the first step of our study. In the replication phase, we observed a trend of association for *PPARG* rs310746 (*P* value = 0.066; OR, 1.17). The combined overall Mantel-Haenszel meta-analysis of all the cohorts included in the present study revealed that *PPARG* rs310746 remained associated with SSc with a nominal non-genome-wide significant *P* value ($P_{MH} = 5.00 \times 10^{-7}$; OR, 1.25). No evidence of association was observed for *CHRNA9* rs6832151 either in the replication phase or in the overall pooled analysis.

Conclusion: Our results suggest a role of *PPARG* gene in the development of SSc.

* Correspondence: martin@ipb.csic.es
[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18016-Armilla, Granada, Spain
Full list of author information is available at the end of the article

Introduction

Systemic sclerosis (SSc) is a complex autoimmune disease with heterogeneous clinical manifestations characterized by extensive fibrosis in the skin and multiple internal organs, vascular damage, and immune imbalance with autoantibody production [1]. SSc patients are commonly classified in two major subtypes: limited cutaneous SSc (lcSSc) and diffuse cutaneous SSc (dcSSc), the latter with more progressive fibrosis of the skin, lungs, and other internal organs and, ultimately, with worse prognosis [2].

The etiology of this disorder is still unclear. However, epidemiologic and genetic studies clearly reflect the existence of a complex genetic component together with the influence of environmental factors [1]. During recent years, great advances have been made in our knowledge of the genetic basis of SSc [3,4], in part, thanks to the two independent genome-wide associations studies (GWASs) conducted in Caucasian populations that have been recently published [5,6], and several consequent follow-up studies [7-10].

However, despite these advances, the number of currently known *loci* explaining the genetic component of SSc is limited. To date, 13 *loci* have been identified as genetic risk factors for SSc at the genome-wide significance level. In other autoimmune diseases with multifactorial inheritance, such as Crohn disease, ulcerative colitis, or systemic lupus erythematosus, individual GWAS scans and follow-up meta-analyses have identified more than 71, 47, and 35 susceptibility *loci*, respectively [11-13]. Therefore, it is expected that additional risk factors for SSc remain to be discovered, and further meta-analyses and large replication studies are needed to identify part of the missing heritability of this disease.

Follow-up studies focused on the so-called grey zone of the GWASs, where SNPs with tier 2 associations (*P* values between 5×10^{-8} and 5×10^{-3}) are located, constitute one of the most useful GWAS data-mining methods, because possible real association signals could be masked in that area because of a lack of statistical power. On this basis, we aimed to perform a follow-up study of the SNPs located in the grey zone of the GWAS by Allanore *et al.* [6], taking advantage of our GWAS data sets. We hypothesize that using a larger cohort would increase the statistical power and might lead to the identification of new suitable SSc genetic risk factors.

Methods

Study design

In the first step of this study, we focused on the 90 GWAS-genotyped SNPs that reached a *P* value $< 10^{-4}$ in the discovery phase of the GWAS carried out by Allanore *et al.* [6]. Then, we analyzed the SNPs overlapping with those included in Radstake *et al.* [5]. After excluding those SNPs located within *MHC* genes or in previously associated *loci*, data for 66 SNPs were selected. A meta-analysis including these 66 SNPs was performed on the combined data set from the two SSc GWASs, showing only two SNPs (rs310746 *PPARG* and rs6832151 *CHRNA9* genetic variants) with a *P* value $< 10^{-5}$ (see later). These two genetic variants were genotyped in independent replication cohorts. Finally, we performed a meta-analysis for these two selected SNPs combining genotype data from both first and replication steps.

Study population

The first step of the study comprised a total of 2,921 SSc patients and 6,963 healthy controls of Caucasian ancestry from two previously published GWASs (European, USA, and French) [5,6]. The replication cohort was composed of 1,068 SSc patients and 1,490 healthy controls from two case–control sets of European ancestry (Italy and United Kingdom). We also included 5,272 extra English controls from The Wellcome Trust Case Control Consortium for the replication step comprising a total of 6,762 controls for this stage.

All SSc patients fulfilled the classification criteria by LeRoy *et al.* [2]. Approval from the local ethical committees (Comité de Bioética del Consejo Superior de Investigaciones Científicas, U.O. Comitato di Etica e Sperimentazione Farmaci Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico di Milano, Comitato Etico Azienda Ospedaliera Universitaria Integrata di Verona, The Ethics Committee of the Spedali Civili, Brescia, Royal Free Hospital and Medical School Research Ethics Committee, Manchester University Research Ethics Committee, Local Research Ethics Committee at Glasgow Royal Infirmary, Newcastle University Ethics Committee, Ethical Committee of the University Erlangen-Nuremberg, Local Ethics Committee of the Radboud University Nijmegen Medical Centre, Medical Ethics Review Committee of the VU University, Medische Ethische Commissie Leids Universitair Medisch Centrum, Ethics Review Board of the Ruhr University Bochum, Ethics Committee of the University of Cologne, Ethical Committee from the Charité University Hospital, Ethik-Kommission der MHH, Internal Review Board of Texas University), and written informed consents from all participants were obtained in accordance with the tenets of the Declaration of Helsinki.

Genotyping

In the first stage, genotype data for the 66 selected SNPs were obtained from both published SSc GWASs [5,6]. QC filters and principal component analysis were applied to the GWASs data, as described in Radstake *et al.* [5] and Allanore *et al.* [6].

In the replication phase, DNA from patients and controls was obtained by using standard methods. Genotyping was performed by using TaqMan 5′ allele discrimination

López-Isac *et al. Arthritis Research & Therapy* 2014, **16**:R6
http://arthritis-research.com/content/16/1/R6

Table 1 Meta-analysis of 66 GWAS-genotyped SNPs in scleroderma (SSc) patients and healthy controls of Caucasian origin

Chr	*Locus*	SNP	Minor/major	MAF cases	MAF controls	P_{MH}	OR (CI 95%)[a]	P_{BD}
3	*PPARG*	rs310746	C/T	0.108	0.086	1.90E-06	1.28 [1.12-1.47]	0.33
4	*CHRNA9 | RHOH*	rs6832151	G/T	0.315	0.281	4.30E-06	1.17 [1.075-1.27]	0.05
2	*DYSF*	rs11692280	A/G	0.195	0.220	2.31E-04	0.86 [0.80-0.93]	0.04
4	*PGDS*	rs17021463	T/G	0.393	0.421	2.45E-04	0.89 [0.83-0.94]	0.13
22	*DGCR6*	rs2543958	G/T	0.127	0.109	4.98E-04	1.18 [1.09-1.28]	0.09
17	*ORMDL3/GSDML*	rs8079416	C/T	0.435	0.461	1.13E-03	0.90 [0.85-0.94]	0.01
1	-	rs6679637	A/G	0.100	0.116	2.35E-03	0.85 [0.77-0.94]	0.03
1	*CSFR3*	rs4653210	G/T	0.111	0.122	3.71E-03	0.86 [0.78-0.95]	0.02
3	*PPARG/ TSEN2*	rs9855622	T/C	0.124	0.110	4.75E-03	1.14 [1.01-1.30]	2.94E-03
7	*CACNA2D1*	rs1544461	A/G	0.429	0.409	5.14E-03	1.09 [1.02-1.16]	7.81E-04
18	*CNDP2*	rs2241508	G/A	0.421	0.401	5.26E-03	1.09 [1.03-1.15]	2.94E-04
11	*PHF21A/CREB3L1*	rs7128538	A/G	0.491	0.470	5.51E-03	1.09 [1.02-1.16]	4.62E-03
14	*NPAS3*	rs1299512	G/A	0.228	0.211	7.76E-03	1.10 [1.00-1.21]	0.05
13	*RFC3*	rs7335534	G/A	0.398	0.415	8.61E-03	0.91 [0.84-0.99]	7.34E-04
8	*DDEF1*	rs7817803	A/C	0.437	0.421	0.012	1.08 [1.01-1.15]	2.01E-03
8	*DDE/*	rs3057	C/T	0.439	0.423	0.012	1.08 [1.01-1.15]	1.13E-03
17	*TMEM132E/CCDC16*	rs887081	T/G	0.115	0.129	0.013	0.88 [0.810.95]	5.33E-03
5	*CDH18*	rs1911856	T/C	0.059	0.048	0.013	1.18 [1.03-1.35]	3.24E-03
7	*CAV1*	rs2402091	A/G	0.110	0.122	0.014	0.88 [0.80-0.97]	0.02
7	-	rs1228966	A/G	0.222	0.208	0.015	1.09 [1.01-1.18]	2.29E-03
7	*SEMA3A*	rs1228870	T/G	0.222	0.208	0.017	1.09 [1.01-1.18]	3.15E-03
5	*LOC389293*	rs7708428	G/A	0.401	0.418	0.019	0.92 [0.87-0.98]	0.01
9	*XPA*	rs2808699	A/C	0.403	0.423	0.021	0.92 [0.87-0.98]	2.38E-04
8	*DDEF1*	rs7839523	G/T	0.440	0.425	0.021	1.07 [1.011-1.14]	1.10E-03
9	*XPA*	rs2805790	A/G	0.403	0.422	0.022	0.92 [0.87-0.98]	3.37E-04
3	*IRAK2*	rs11706450	T/C	0.465	0.482	0.024	0.93 [0.86-1.01]	9.00E-03
9	*XPA*	rs2805815	A/G	0.403	0.422	0.024	0.93 [0.87-0.99]	3.02E-04
2	*NOL10*	rs4668690	A/G	0.067	0.059	0.026	1.15 [1.01-1.30	3.10E-03
7	-	rs1029541	T/C	0.230	0.219	0.028	1.08 [1.01-1.17]	8.07E-04
9	*XPA*	rs2668797	A/G	0.071	0.112	0.029	0.933 [0.87-0.99]	2.48E-04
14	-	rs1036570	A/G	0.322	0.335	0.032	0.92 [0.85-1.014]	2.13E-03
3	-	rs4128236	T/C	0.322	0.306	0.034	1.07 [0.98-1.17]	3.87E-04
1	-	rs10925871	A/G	0.193	0.181	0.038	1.08 [1.00-1.17]	6.22E-03
7	*CADPS2*	rs2501439	G/A	0.418	0.432	0.042	0.93 [0.87-0.99]	7.49E-03
7	-	rs757747	T/C	0.229	0.218	0.047	1.07 [1.00-1.16]	1.29E-03
7	*WBSCR17*	rs4585627	T/C	0.323	0.308	0.051	1.07 [1.00-1.14]	4.70E-03
10	-	rs1254860	C/T	0.110	0.100	0.064	1.09 [0.99-1.21]	0.03
8	*DDEF1*	rs6470805	G/A	0.333	0.344	0.069	0.94 [0.88-1.00]	4.88E-03
9	*LCN9*	rs541131	G/A	0.400	0.385	0.071	1.06 [0.99-1.12]	1.78E-03
5	*CDH18*	rs2202798	T/C	0.080	0.069	0.078	1.11 [0.98-1.24]	9.17E-04
3	*TDGF1*	rs6799581	G/T	0.260	0.268	0.080	0.94 [0.86-1.03]	6.85E-04
5	*CDH18*	rs12655266	A/G	0.074	0.065	0.111	1.10 [0.98-1.24]	1.09E-03
4	*NPY2R*	rs2880417	G/A	0.292	0.281	0.117	1.05 [0.99-1.13]	2.07E-05

Table 1 Meta-analysis of 66 GWAS-genotyped SNPs in scleroderma (SSc) patients and healthy controls of Caucasian origin *(Continued)*

7	-	rs10272701	T/C	0.192	0.183	0.130	1.06 [0.98-1.15]	7.73E-03
8	*FBX032*	rs3739284	C/T	0.219	0.224	0.148	0.94 [0.87-1.01]	3.71E-03
6	*ASCC3*	rs7771570	C/T	0.492	0.482	0.153	1.04 [0.98-1.11]	2.97E-04
18	*PHLPP*	rs2877745	T/C	0.091	0.084	0.158	1.08 [0.98-1.18]	3.49E-03
21	-	rs2831511	T/C	0.393	0.403	0.175	0.95 [0.90-1.01]	1.54E-04
11	*OPCML*	rs10894623	T/G	0.291	0.281	0.175	1.04 [0.95-1.14]	8.98E-04
6	*ASCC3*	rs6919745	T/C	0.476	0.467	0.190	1.042 [0.98-1.11]	1.88E-04
4	*NPY2R*	rs13138293	G/T	0.308	0.299	0.195	1.05 [0.97-1.11]	8.20E-05
15	*SMAD3*	rs4147358	A/C	0.237	0.243	0.203	0.95 [0.86-1.04]	4.76E-04
17	*TMEM132E*	rs4795032	T/C	0.351	0.341	0.232	1.04 [0.95-1.13]	6.65E-04
9	-	rs10756265	A/G	0.342	0.352	0.251	0.96 [0.90-1.02]	2.18E-04
9	-	rs443042	G/A	0.366	0.376	0.261	0.96 [0.90-1.02]	2.01E-04
9	*SUSD3*	rs9696357	T/C	0.154	0.160	0.268	0.95 [0.87-1.03]	2.11E-03
12	*SFRS8*	rs10794423	C/T	0.439	0.442	0.282	0.96 [0.89-1.04]	1.22E-03
2	*ATP6V1C2*	rs7422405	A/G	0.428	0.433	0.385	0.97 [0.91-1.03]	1.93E-04
3	*RBMS3*	rs35883	A/G	0.457	0.4564	0.556	1.019 [0.94-1.10]	1.04E-04
21	*CHODL/ PRSS7*	rs2248200	C/T	0.484	0.4815	0.598	1.01 [0.96-1.07]	1.35E-03
21	*CHODL/PRSS7*	rs1688165	A/G	0.485	0.4820	0.628	1.01 [0.96-1.07]	8.62E-04
16	*ZNF423*	rs1477020	T/C	0.121	0.1235	0.649	0.97 [0.86-1.10]	8.30E-06
1	*C1QB*	rs631090	C/T	0.073	0.0702	0.729	1.02 [0.90-1.15]	3.73E-04
16	*ZNF423*	rs1990629	G/A	0.128	0.1308	0.736	0.98 [0.87-1.10]	2.13E-05
11	*OPCML*	rs11223273	T/C	0.275	0.2721	0.759	1.01 [0.92-1.10]	1.40E-04
19	*TSPAN16*	rs322151	T/C	0.252	0.2527	0.990	0.99 [0.94-1.06]	3.37E-04

N, 2,921 SSc/6,963 controls. [a]Odds ratio for the minor allele. *Chr*, chromosome; *CI*, confidence interval; *MAF*, minor allele frequency; *OR*, odds ratio; P_{BD}, Breslow–Day test *P* value; P_{MH}, allelic Mantel-Haenszel fixed effects model *P* value; *SNP*, single nucleotide polymorphism.

predesigned assays from Applied Biosystems (rs310746 ID: C___8756618_10; rs6832151 ID: C__29224385_10, Foster City, CA, USA) in both 7900HT Fast Real-Time PCR System (Applied Biosystems, *Foster City, CA, USA)*, and LightCycler 480 Real-Time PCR System (Roche Applied Science, Mannheim, Germany). Genotyping call rate was > 98% for both genotyped SNPs.

Statistical analysis

Association analyses of the genotype data was carried out with StatsDirect V.2.6.6 (StatsDirect, Altrincham, UK) and PLINK V.1.07 [14] software. Statistical significance was calculated by 2×2 contingency tables and χ^2 or Fisher Exact test, when necessary, to obtain *P* values, odds ratios (ORs), and 95% confidence intervals (CIs) in the population-specific analyses. Mantel-Haenszel tests under fixed effects or random effects, when appropriate, were performed to meta-analyze the combined data. Breslow–Day method (BD) was used to assess the homogeneity of the associations among the different populations (Breslow–Day *P* values <0.05 were considered statistically significant). Hardy-Weinberg equilibrium (HWE) was tested for all cohorts (HWE *P* values lower than 0.01 were considered to show significant deviation from the equilibrium). None of the included cohorts showed significant deviation from HWE for the two genotyped SNPs. Since the analyses were performed by using GWAS data, the statistical threshold for considering a *P* value as a significant *P* value in the allelic association analyses was set at 5×10^{-8}.

The statistical power of the combined analysis was 70% for the *PPARG* rs310746 and 100% for the *CHRNA9* rs6832151 to detect associations with OR = 1.3 and a statistical significance of 5×10^{-8}, according to Power Calculator for Genetic Studies 2006 software [15].

Results

Table 1 shows the results of the 66 GWAS-genotyped SNPs selected for the combined meta-analysis of the two GWAS data sets performed in the first step of this study (see Additional file 1: Table S1 provides the results from both GWASs and the combined meta-analysis for the 66 selected SNPs). Two SNPs showed a *P* value lower than 10^{-5} (*PPARG* rs310746: $P_{MH} = 1.90 \times 10^{-6}$; OR, 1.28; CI,

López-Isac *et al. Arthritis Research & Therapy* 2014, **16**:R6
http://arthritis-research.com/content/16/1/R6

Table 2 Analysis of rs310746 and rs6832151 minor allele frequencies in the GWASs, replication, and combined cohorts

Cohort, *N* (cases/controls)	Chr	*Locus*	SNP	Minor/major	MAF cases	MAF controls	P_{MH}	OR (CI 95%)[a]	P_{BD}
GWASs	3	*SYN2|PPARG*	rs310746	C/T	0.108	0.087	1.90E-06	1.28 [1.12-1.47]	0.334
2921/6963	4	*CHRNA9|RHOH*	rs6832151	G/T	0.315	0.282	4.30E-06	1.17 [1.075-1.27]	0.054
Replication	3	*SYN2|PPARG*	rs310746	C/T	0.099	0.103	0.066	1.17 [0.99-1.38]	0.231
1068/6762	4	*CHRNA9|RHOH*	rs6832151	G/T	0.296	0.280	0.962	0.99 [0.89-1.11]	0.934
Combined	3	*SYN2|PPARG*	rs310746	C/T	0.106	0.094	5.00E-07	1.25 [1.15-1.37]	0.324
3989/13725	4	*CHRNA9|RHOH*	rs6832151	G/T	0.310	0.281	1.07E-04[b]	1.12 [1.06-1.19]	0.017

[a]Odds ratio for the minor allele. [b]*P* value from meta-analysis under random effects = 0.051; OR = 1.10 (0.99-1.22).
Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; P_{BD}, Breslow–Day test *P* value; P_{MH}, allelic Mantel-Haenszel fixed-effects model P value; SNP, single-nucleotide polymorphism.

95%, 1.12 to 1.47; and *CHRNA9* rs6832151: $P_{MH} = 4.30 \times 10^{-6}$, OR, 1.17; CI 95%, 1.08 to 1.27), and presented no significant Breslow-Day *P* values (P_{BD}) showing homogeneity in the ORs among populations. Therefore, these two SNPs were selected to genotype in independent cohorts. Patients and healthy controls were found to be in HWE at 1% significance level for both selected SNPs.

In the replication phase, we observed a trend of association for the *PPARG* rs310746 genetic variant (*P* value = 0.066; OR = 1.17; CI 95%, 0.99 to 1.38) in the combined analysis of the two replication cohorts (Table 2, upper rows). However, no evidence of association was observed for *CHRNA9* rs6832151 either in the pooled analysis (Table 2, upper rows) or in the analysis of each individual population (see Additional file 2: Table S2).

Finally, we combined the results from both steps of the study and performed a Mantel-Haenszel meta-analysis observing that the *PPARG* genetic variant showed suggestive evidence of association with SSc ($P_{MH} = 5.00 \times 10^{-7}$; OR = 1.25; CI, 95%, 1.15 to 1.37) (Table 2, lower rows; Figure 1). However, *CHRNA9* rs6832151 showed no evidence of association with the disease when the meta-analysis was performed either under a random-effects model (heterogeneity of the ORs was observed for this SNP; *P* value = 5.10×10^{-2}, OR = 1.10; CI 95%, 0.99 to 1.22), or a fixed-effects model (*P* value = 1.07×10^{-4}; OR = 1.12; CI 95%, 1.06 to 1.19) (Table 2, lower rows; Figure 1).

Discussion

In this study we conducted a meta-analysis combining previously published SSc GWASs data for 66 SNPs and analyzed the possible role of two selected SNPs, *PPARG* rs310746 and *CHRNA9* rs6832151, in SSc risk by using independent replication cohorts.

Meta-analyses are a useful tool to increase the statistical power of genetic studies, thus improving the accuracy of the estimations of statistical significance. Of

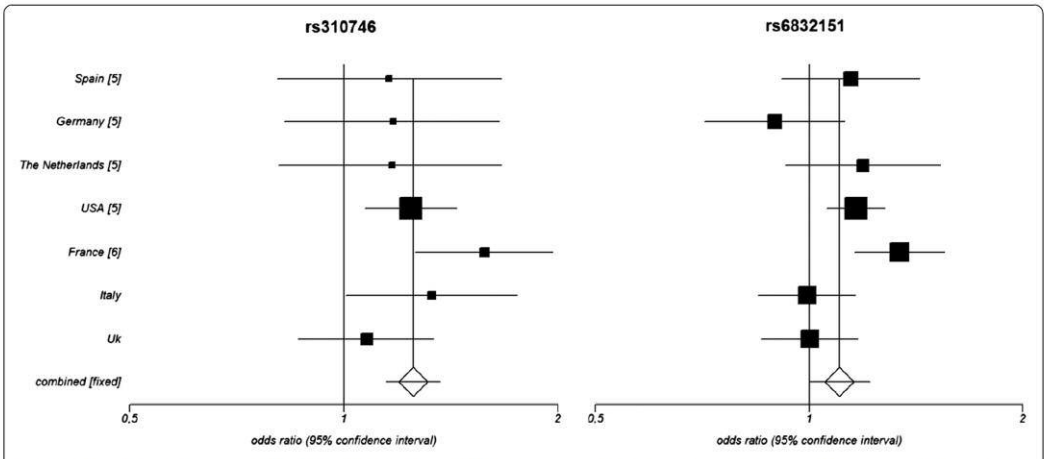


Figure 1 Forest plots of *PPARG* rs310746 and *CHRNA9* rs6832151. Forest plots showing the odds ratios and confidence intervals of both *PPARG* rs310746 (under a fixed-effects model) and *CHRNA9* rs6832151 (under a random-effects model) in all the populations included in the combined analysis.

note, associations identified from a single GWAS often tend to have inflated effect sizes [16]. On this basis, our data suggest that most signals from the grey zone observed in the discovery phase of the GWAS by Allanore *et al.* [6] presented inflated effect sizes, also called the winner's curse. In fact, this effect was already observed in the replication study conducted by our group for the novel SSc genetic risk factors identified by Allanore *et al.* [6], in which we could not replicate the association described for *RHOB* [17].

Our overall combined meta-analysis showed that the association of the *PPARG* rs310746 genetic variant with SSc remained with a nominal but non-genome-wide significant *P* value. This SNP is located upstream of *PPARG*, which encodes the peroxisome proliferator-activated receptor gamma (PPARG). *PPARG* was initially identified in adipose tissue, where this nuclear receptor plays important roles in adipogenesis, insulin sensitivity, and homeostasis [18]. Interestingly, during recent years, several studies have identified a novel role of PPARG as an antifibrotic effector. Thus, it has been reported that fibroblasts exposure to pharmacologic PPARG ligands give rise to suppression of collagen synthesis, myofibroblast differentiation, and other TGF-β-induced fibrotic responses *in vitro* [19-21]. Moreover, functional studies showed that PPARG agonist attenuated dermal fibrosis in mice with bleomycin-induced scleroderma [22,23].

These findings are remarkable in SSc, in which fibrosis is one of the main hallmarks of the disease. In this regard, Wei *et al.* [24] demonstrated that PPARG expression and function are impaired in SSc patients. Therefore, defects in *PPARG* expression may influence the uncontrolled progression of fibrosis in SSc. In addition, *PPARG* has been associated with other autoimmune diseases, such as inflammatory bowel disease [25,26] and psoriatic arthritis [27], and it is also a confirmed susceptibility *locus* in type 2 diabetes mellitus [28].

Although *PPARG* was the most likely biologic candidate gene for the reported suggestive association signal, we could not rule out *TIMP4* as another possible gene for this signal. Further analyses are required to elucidate the functional implication of the reported signal.

Regarding the *CHRNA9* genetic variant, despite the suggestive association found in the first step of the present study, the overall combined meta-analysis did not show evidence of association with SSc. Moreover, the effect size of the analyzed genetic variant was heterogeneous between the different populations. Although our data showed heterogeneity and lack of association in this *locus*, a slight or modest effect of *CHRNA9* cannot be ruled out, and further studies will be required to determine whether this region is associated with SSc.

It is worth mentioning that the analyzed *CHRNA9* SNP has been previously associated with Graves disease (first, through a GWAS performed in the Chinese Han population [29], and subsequently, in a replication study performed in a Polish Caucasian population [30]), but this is the only reported association between this gene and an autoimmune disease.

Conclusion

In conclusion, we report a suggestive association between *PPARG* rs310746 and SSc. However, further studies are needed to establish this *locus* firmly as a new susceptibility SSc genetic risk factor.

Additional files

Additional file 1: GWASs results from Allanore *et al.* [6] and Radstake *et al.* [5], and combined meta-analysis. Description: this file contains Additional file 1: Table S1 showing the results for the 66 selected SNPs in Allanore *et al.* and Radstake *et al.* GWASs, followed by the results of the combined meta-analysis performed in the present study.

Additional file 2: Genotype and minor allele frequencies of rs310746 and rs6832151 SNPs in two European cohorts (Replication-step). Description: this file contains: Additional file 2: Table S2 showing the genotype and allele distributions of rs310746 and rs6832151 genetic variants in two European cohorts (1032 SSc cases and 6700 controls).

Abbreviations

BD test: Breslow-day test; CHRNA9: cholinergic receptor nicotinic, Alpha 9; CI: Confidence interval; dcSSc: diffuse cutaneous systemic sclerosis; DNA: deoxyribonucleic acid; GWAS: genome-wide association study; HLA: Human leukocyte antigen; HWE: Hardy-Weinberg equilibrium; lcSSc: limited cutaneous systemic sclerosis; MAF: minor allele frequency; MHC: major histocompatibility complex; OR: Odds ratio; PCR: polymerase chain reaction; PPARG: peroxisome proliferator-activated receptor gamma; SNP: single nucleotide polymorphism; SSc: systemic sclerosis; TGF: transforming growth factor.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

ELI and LBC contributed to the analysis and interpretation of data and the drafting the manuscript. CPS and AH participated in the acquisition of data and the drafting of the manuscript. JM contributed to the conception and design of the study and critically revised the manuscript. MVE, JJAS, JLC, JARI, MF, LB, AS, PA, CL, NH, GR, TW, AK, JHWD, AJS, MCV, AEV, PGS, JMvL, CF, CD, JW, SA, BPK, MDM, TRDJR, 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.

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), who supplied part of the control DNA samples.
This work was supported by the following grants: JM was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009-11110 and SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness, CTS-4977, and CTS-180 from Junta de Andalucía, and is sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). This study was also funded by PI-0590-2010, from Consejería de Salud y Bienestar Social, Junta de Andalucía, Spain. JLCR and JM are funded by Consejería de Salud, Junta de Andalucía, through PI-0590-2010. ELI was supported by Ministerio de Educación, Cultura y Deporte through the program FPU. TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis

Foundation (National Reumafonds). TW was granted by DFG WI 1031/6.1. Study on USA samples were supported by US National Institutes of Health and National Institute of Arthritis and Musculoskeletal Diseases (NIH-NIAMS) R01-AR-055258, Two-Stage Genome Wide Association Study in Systemic Sclerosis (MDM) and by the NIH-NIAMS Center of Research Translation (CORT) in SSc (P50AR054144) (MDM, FCA, FKT), the NIH-NIAMS SSc Family Registry and DNA Repository (N01-AR-0-2251) (MDM), NIH-KL2RR024149 (SA), K23AR061436 (SA), and the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111) (MDM).

Spanish Scleroderma Group

Spanish Scleroderma Group: Norberto Ortego-Centeno and Raquel Ríos, Unidad de Enfermedades Sistémicas Autoinmunes, Department of Internal Medicine, Hospital Clínico Universitario San Cecilio, Granada; Nuria Navarrete, Department of Internal Medicine, Hospital Virgen de las Nieves, Granada; Rosa García Portales, Department of Rheumatology, Hospital Virgen de la Victoria, Málaga; María Teresa Camps, Department of Internal Medicine, Hospital Carlos Haya, Málaga; Antonio Fernández-Nebro, Department of Rheumatology, Hospital Carlos Haya, Málaga; María F. González-Escribano, Department of Immunology, Hospital Virgen del Rocío, Sevilla; Julio Sánchez-Román, Francisco José García-Hernández and Mª Jesús Castillo, Department of Internal Medicine, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Department of Rheumatology, Hospital Reina Sofía/IMIBIC, Córdoba; Patricia Carreira, Department of Rheumatology. Hospital Universitario 12 de Octubre, Madrid; Benjamín Fernández-Gutiérrez and Luis Rodríguez-Rodríguez, Department of Rheumatology, Hospital Clínico San Carlos, Madrid; Esther Vicente, Department of Rheumatology, Hospital La Princesa, Madrid; José Luis Andreu and Mónica Fernández de Castro, Department of Rheumatology, Hospital Puerta de Hierro Majadahonda, Madrid; Paloma García de la Peña, Department of Rheumatology, Hospital Madrid Norte Sanchinarro, Madrid; Francisco Javier López-Longo and Lina Martínez, Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Madrid; Vicente Fonollosa and Carmen Pilar Simeón, Department of Internal Medicine, Hospital Valle de Hebrón, Barcelona; Gerard Espinosa, Department of Internal Medicine, Hospital Clinic, Barcelona; Iván Castellví, Department of Rheumatology, Hospital de la Santa Creu i Sant Pau, Barcelona; Carlos Tolosa, Department of Internal Medicine, Hospital Parc Tauli, Sabadell; Anna Pros, Department of Rheumatology, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Department of Internal Medicine, Hospital Universitari Mútua Terrasa, Barcelona; Francisco Javier Narváez, Department of Rheumatology, Hospital Universitari de Bellvitge, Barcelona; Manel Rubio Rivas, Department of Internal Medicine, Hospital Universitari de Bellvitge, Barcelona; Vera Ortiz Santamaría, Department of Rheumatology, Hospital General de Granollers, Granollers; Miguel Ángel González-Gay, Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, IFIMAV, Santander; Bernardino Díaz and Luis Trapiella, Department of Internal Medicine, Hospital Central de Asturias, Oviedo; Adrián Sousa, Unidad de Trombosis y Vasculitis, Department of Internal Medicine, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; Patricia Fanlo Mateo, Department of Internal Medicine Hospital Virgen del Camino, Pamplona; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Department of Internal Medicine, Hospital Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Department of Rheumatology, Hospital Universitario de Canarias, Tenerife; Emma Beltrán, Department of Rheumatology, Hospital General Universitario de Valencia, Valencia; and Francisco J. Blanco García and Natividad Oreiro, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña.

Author details

[1]Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18016-Armilla, Granada, Spain. [2]Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona, Spain. [3]Department of Internal Medicine, Hospital Universitario Cruces, Barakaldo, Spain. [4]Department of Rheumatology, Hospital del Doctor Peset, Valencia, Spain. [5]Department of Internal Medicine, Hospital Clínico San Cecilio, Granada, Spain. [6]Rheumatology Department, Hospital Universitario La Fe, Valencia, Spain. [7]Department of Internal Medicine, Thrombosis and Vasculitis Unit, Complexo Hospitalario Universitario de Vigo, Vigo, Spain. [8]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggioire Policlinico di Milano, Milan, Italy. [9]UO Reumatologia ed Immunologia Clinica, Spedali Civili, Brescia, Italy. [10]Department of Medicine, Università degli Studi di Verona, Verona, Italy. [11]Department of Dermatology, University of Cologne, Cologne, Germany. [12]Department of Rheumatology and Clinical Immunology, Charité University Hospital, and German Rheumatism Research Centre, a Leibniz Institute, Berlin, Germany. [13]Hannover Medical School, Hannover, Germany. [14]Ruhr University of Bochum, Bochum, Germany. [15]Department of Internal Medicine, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen, Germany. [16]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands. [17]Department of Rheumatology, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands. [18]Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands. [19]Section of Epigenetics, Inst. Cancer Sciences, MVLS, University of Glasgow, Glasgow, UK. [20]Institute of Cellular Medicine, Newcastle University, Newcastle, UK. [21]Centre for Rheumatology, Royal Free and University College Medical School, London, UK. [22]Arthritis Research UK Epidemiology Unit and NIHR Manchester Musculoskeletal Biomedical Research Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK. [23]The University of Texas Health Science Center–Houston, Houston, TX, USA. [24]Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands. [25]Department of Rheumatology and Clinical Immunology, University Medical Center, Utrecht, The Netherlands.

Received: 12 July 2013 Accepted: 18 December 2013
Published: 9 January 2014

References

1. Gabrielli A, Avvedimento EV, Krieg T: **Scleroderma.** *N Engl J Med* 2009, **360:**1989–2003.
2. 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.
3. Martin JE, Bossini-Castillo L, Martin J: **Unraveling the genetic component of systemic sclerosis.** *Hum Genet* 2012, **131:**1023–1037.
4. Assassi S, Radstake TR, Mayes MD, Martin J: **Genetics of scleroderma: implications for personalized medicine?** *BMC Med* 2013, **11:**9.
5. 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.
6. Allanore Y, Saad M, Dieude P, Avouac J, Distler JH, Amouyel P, Matucci-Cerinic M, Riemekasten G, Airo P, Melchers I, Hachulla E, Cusi D, Wichmann HE, Wipff J, Lambert JC, Hunzelmann N, Tiev K, Caramaschi P, Diot E, Kowal-Bielecka O, Valentini G, Mouthon L, Czirjak L, Damjanov N, Salvi E, Conti C, Muller M, Muller-Ladner U, Riccieri V, Ruiz B, *et al*: **Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis.** *PLoS Genet* 2011, **7:**e1002091.
7. 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.
8. Bossini-Castillo L, Martin JE, Broen J, Gorlova O, Simeon CP, Beretta L, Vonk MC, Callejas JL, Castellvi I, Carreira P, Garcia-Hernandez FJ, Fernandez Castro M, Coenen MJ, Riemekasten G, Witte T, Hunzelmann N, Kreuter A, Distler JH, Koeleman BP, Voskuyl AE, Schuerwegh AJ, Palm O, Hesselstrand R, Nordin A, Airo P, Lunardi C, Scorza R, Shiels P, van Laar JM, Herrick A, *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:**926–933.
9. Martin JE, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, Van't Slot R, Rodriguez-Rodriguez L, Vicente E, Fonollosa V, Ortego-Centeno N, Gonzalez-Gay MA, Garcia-Hernandez FJ, De la Pena PG, Carreira P, Voskuyl AE, Schuerwegh AJ, van Riel PL, Kreuter A, Witte T, Riemekasten G, Airo P, Scorza R, Lunardi C, Hunzelmann N, Distler JH, Beretta L, van Laar J, Chee MM,

Worthington J, *et al*: **Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up.** *Hum Mol Genet* 2012, **21**:2825–2835.
10. Sharif R, Mayes MD, Tan FK, Gorlova OY, Hummers LK, Shah AA, Furst DE, Khanna D, Martin J, Bossini-Castillo L, Gonzalez EB, Ying J, Draeger HT, Agarwal SK, Reveille JD, Arnett FC, Wigley FM, Assassi S: **IRF5 polymorphism predicts prognosis in patients with systemic sclerosis.** *Ann Rheum Dis* 2012, **71**:1197–1202.
11. Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, Ahmad T, Lees CW, Balschun T, Lee J, Roberts R, Anderson CA, Bis JC, Bumpstead S, Ellinghaus D, Festen EM, Georges M, Green T, Haritunians T, Jostins L, Latiano A, Mathew CG, Montgomery GW, Prescott NJ, Raychaudhuri S, Rotter JI, Schumm P, Sharma Y, Simms LA, Taylor KD, Whiteman D, *et al*: **Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.** *Nat Genet* 2010, **42**:1118–1125.
12. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, Lee JC, Goyette P, Imielinski M, Latiano A, Lagace C, Scott R, Amininejad L, Bumpstead S, Baidoo L, Baldassano RN, Barclay M, Bayless TM, Brand S, Buning C, Colombel JF, Denson LA, De Vos M, Dubinsky M, Edwards C, Ellinghaus D, Fehrmann RS, Floyd JA, Florin T, Franchimont D, *et al*: **Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47.** *Nat Genet* 2011, **43**:246–252.
13. Sestak AL, Furnrohr BG, Harley JB, Merrill JT, Namjou B: **The genetics of systemic lupus erythematosus and implications for targeted therapy.** *Ann Rheum Dis* 2011, **70**:i37–i43.
14. 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.
15. 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.
16. Ioannidis JP, Thomas G, Daly MJ: **Validating, augmenting and refining genome-wide association signals.** *Nat Rev Genet* 2009, **10**:318–329.
17. Bossini-Castillo L, Martin JE, Broen J, Simeon CP, Beretta L, Gorlova OY, Vonk MC, Ortego-Centeno N, Espinosa G, Carreira P, Garcia De La Pena P, Oreiro N, Roman-Ivorra JA, Castillo MJ, Gonzalez-Gay MA, Saez-Comet L, Castellvi I, Schuerwegh AJ, Voskuyl AE, Hoffmann-Vold AM, Hesselstrand R, Nordin A, Lunardi C, Scorza R, Van Laar JM, Shiels PG, Herrick A, Worthington J, Fonseca C, Denton C, Denton C, *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**:602–607.
18. Tontonoz P, Spiegelman BM: **Fat and beyond: the diverse biology of PPARgamma.** *Annu Rev Biochem* 2008, **77**:289–312.
19. Ghosh AK, Bhattacharyya S, Lakos G, Chen SJ, Mori Y, Varga J: **Disruption of transforming growth factor beta signaling and profibrotic responses in normal skin fibroblasts by peroxisome proliferator-activated receptor gamma.** *Arthritis Rheum* 2004, **50**:1305–1318.
20. Burgess HA, Daugherty LE, Thatcher TH, Lakatos HF, Ray DM, Redonnet M, Phipps RP, Sime PJ: **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**:L1146–L1153.
21. Kulkarni AA, Thatcher TH, Olsen KC, Maggirwar SB, Phipps RP, Sime PJ: **PPAR-gamma ligands repress TGFbeta-induced myofibroblast differentiation by targeting the PI3K/Akt pathway: implications for therapy of fibrosis.** *PLoS One* 2011, **6**:e15909.
22. Wei J, Zhu H, Komura K, Lord G, Tomcik M, Wang W, Doniparthi S, Tamaki Z, Hinchcliff M, Distler JH, Varga J: **A synthetic PPAR-gamma agonist triterpenoid ameliorates experimental fibrosis: PPAR-gamma-independent suppression of fibrotic responses.** *Ann Rheum Dis* 2013. Epub ahead of print.
23. Wu M, Melichian DS, Chang E, Warner-Blankenship M, Ghosh AK, Varga J: **Rosiglitazone abrogates bleomycin-induced scleroderma and blocks profibrotic responses through peroxisome proliferator-activated receptor-gamma.** *Am J Pathol* 2009, **174**:519–533.
24. Wei J, Ghosh AK, Sargent JL, Komura K, Wu M, Huang QQ, Jain M, Whitfield ML, Feghali-Bostwick C, Varga J: **PPARgamma downregulation by TGFss in fibroblast and impaired expression and function in systemic sclerosis: a novel mechanism for progressive fibrogenesis.** *PLoS One* 2010, **5**:e13778.
25. Glas J, Seiderer J, Markus C, Pfennig S, Wetzke M, Paschos E, Goke B, Ochsenkuhn T, Muller-Myhsok B, Diegelmann J, Roeske D, Brand S: **Role of PPARG gene variants in inflammatory bowel disease.** *Inflamm Bowel Dis* 2011, **17**:1057–1058.
26. Zhang ZF, Yang N, Zhao G, Zhu L, Wang LX: **Association between the Pro12Ala polymorphism of peroxisome proliferator-activated receptor gamma 2 and inflammatory bowel disease: a meta-analysis.** *PLoS One* 2012, **7**:e30551.
27. Butt C, Gladman D, Rahman P: **PPAR-gamma gene polymorphisms and psoriatic arthritis.** *J Rheumatol* 2006, **33**:1631–1633.
28. Lyssenko V, Jonsson A, Almgren P, Pulizzi N, Isomaa B, Tuomi T, Berglund G, Altshuler D, Nilsson P, Groop L: **Clinical risk factors, DNA variants, and the development of type 2 diabetes.** *N Engl J Med* 2008, **359**:2220–2232.
29. Chu X, Pan CM, Zhao SX, Liang J, Gao GQ, Zhang XM, Yuan GY, Li CG, Xue LQ, Shen M, Liu W, Xie F, Yang SY, Wang HF, Shi JY, Sun WW, Du WH, Zuo CL, Shi JX, Liu BL, Guo CC, Zhan M, Gu ZH, Zhang XN, Sun F, Wang ZQ, Song ZY, Zou CY, Sun WH, Guo T, *et al*: **A genome-wide association study identifies two new risk loci for Graves' disease.** *Nat Genet* 2011, **43**:897–901.
30. Szymanski K, Bednarczuk T, Krajewski P, Ploski R: **The replication of the association of the rs6832151 within chromosomal band 4p14 with Graves' disease in a Polish Caucasian population.** *Tissue Antigens* 2012, **79**:380–383.

doi:10.1186/ar4432
Cite this article as: López-Isac *et al.*: A genome-wide association study follow-up suggests a possible role for *PPARG* in systemic sclerosis susceptibility. *Arthritis Research & Therapy* 2014 **16**:R6.

V

Publication 2: Identification of *IL12RB1* as a novel systemic sclerosis susceptibility locus

ARTHRITIS & RHEUMATOLOGY
Vol. 66, No. 12, December 2014, pp 3521–3523

CONCISE COMMUNICATION

DOI 10.1002/art.38870

Identification of *IL12RB1* as a novel systemic sclerosis susceptibility locus

Genome-wide association studies (GWAS) have identified several immune-related loci associated with systemic sclerosis (SSc), which clearly supports the idea that the immune system plays an important role in the disease etiology (1,2). Using gene set enrichment analysis and DAVID algorithms in the Biocarta pathway collection, we found that the most enriched pathways in SSc corresponded to the nitric oxide synthase 2–dependent interleukin-12 (IL-12) pathway in natural killer cells and the IL-12/STAT-4–dependent signaling pathway in Th1 development (see Supplementary Table 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract). Moreover, several studies have implicated IL-12 in autoimmune inflammatory processes (3). Interestingly, in our recent large-scale fine mapping Immunochip study in SSc (4), we observed suggestive association signals in the *IL12RB1* locus, which encodes the β1 subunit of the IL-12 receptor. Consequently, we aimed to evaluate for the first time the genetic contribution of the *IL12RB1* region in SSc through a followup study.

Forty-six single-nucleotide polymorphisms (SNPs) within *IL12RB1* were screened in 1,871 SSc cases and 3,636 controls from the SSc Immunochip discovery cohorts (4). In this first phase, 11 of the 46 SNPs showed nominal association signals (see Supplementary Table 2 and Supplementary Figure 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract). After conditional logistic regression, we selected 4 SNPs (rs8109496, rs2305743, rs436857, and rs11668601) as the genetic variants that better explained the observed signals in the *IL12RB1* region (see Supplementary Table 3). These SNPs were selected for genotyping in 6 independent replication cohorts of subjects of European ancestry (Spain, Germany, The Netherlands, Italy, Sweden, and the UK) totaling 3,181 SSc patients and 5,076 controls. All SSc patients fulfilled previously described classification criteria for SSc (4).

A genome-wide significance level was achieved for 1 SNP, rs2305743, in the independent replication cohorts ($P = 3.936 \times 10^{-8}$ by Mantel-Haenszel fixed-effects model, odds ratio [OR] 0.79) (see Supplementary Table 4, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract). Interestingly, the combined analysis (5,052 SSc patients and 8,712 controls) showed that the 4 selected SNPs were associated with SSc at the genome-wide significance level (Table 1 and Supplementary Table 5), providing robust evidence for the implication of *IL12RB1* in SSc development.

Despite the fact that dependence analysis did not yield discernable differences between variants (see Supplementary Table 6, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract), our in silico functional analysis showed that minor alleles of both the rs436857 promoter variant and rs2305743 were in an *IL12RB1 cis*–expression quantitative trait locus that decreased *IL12RB1* expression ($P = 2.4 \times 10^{-81}$, $Z = -19.10$ and $P = 9.6 \times 10^{-80}$, $Z = -18.91$, respectively) (available online at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract).

Additionally, using data from the Encyclopedia of DNA Elements database, we found evidence that rs436857 affected the binding of several transcription factors (available online at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract), such as POLR2A (the largest subunit of RNA polymerase II) and YY-1 (a ubiquitously distributed transcription factor that interacts with POLR2A). Therefore, we hypothesize that this promoter SNP should be the best candidate for driving the reported association,

Table 1. *IL12RB1* SNP minor allele frequencies in the overall combined analysis*

SNP, minor/major allele	Genotype, no (%)			MAF, %	Allele test		
	1/1	1/2	2/2		*P* vs. controls†	*P* vs. controls‡	OR (95% CI)§
rs8109496, C/G							
Controls (n = 8,697)	329 (3.78)	2,736 (31.46)	5,632 (64.76)	19.51	7.347×10^{-9}	0.32	–
SSc patients (n = 5,036)	148 (2.94)	1,343 (26.67)	3,545 (70.39)	16.27			0.82 (0.77–0.88)
rs2305743, A/G							
Controls (n = 8,697)	353 (4.06)	2,796 (32.15)	5,548 (63.79)	20.13	4.294×10^{-10}	0.10	–
SSc patients (n = 5,032)	166 (3.30)	1,358 (26.99)	3,508 (69.71)	16.79			0.81 (0.76–0.87)
rs436857, A/G							
Controls (n = 8,652)	297 (3.43)	2,652 (30.65)	5,703 (65.92)	18.76	3.938×10^{-9}	0.23	–
SSc patients (n = 4,924)	142 (2.88)	1,245 (25.28)	3,537 (71.83)	15.53			0.81 (0.76–0.87)
rs11668601, C/T							
Controls (n = 8,682)	539 (6.21)	3,302 (38.03)	4,841 (55.76)	25.22	5.612×10^{-9}	0.06	–
SSc patients (n = 4,962)	245 (4.94)	1,690 (34.06)	3,027 (61.00)	21.97			0.84 (0.79–0.89)

* SNP = single-nucleotide polymorphism; MAF = minor allele frequency; 95% CI = 95% confidence interval; SSc = systemic sclerosis.
† By Mantel-Haenszel fixed-effects model.
‡ By Breslow-Day test.
§ Odds ratio (OR) for the minor allele.

narrowing down the signal to the promoter region. The protective OR and the decrease in *IL12RB1* expression related to these variants are consistent with a reduced IL-12 response and lower SSc susceptibility. Moreover, the coexpression of IL12R*B*1 and IL12R*B*2 is necessary to form the high-affinity IL-12 receptor, and IL-12 binding leads to the activation of STAT-4. Of note, coding genes for these proteins (*IL12RB2*, *IL12A*, *STAT4*) are well-established risk factors for SSc (1,4,5).

We thus report for the first time the association of *IL12RB1* with SSc and add a novel IL-12 pathway–related gene to the list of SSc susceptibility loci. These results highlight the special relevance of this pathway in SSc pathophysiology and its integration in the SSc genetic susceptibility context (see Supplementary Figure 2, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.38870/abstract), and suggest that blocking this pathway could be a possible new therapeutic target in an orphan disease such as scleroderma.

Dr. López-Isac's work was supported by Ministerio de Educación, Cultura y Deporte through the program Formación de Profesorado Universitario. Dr. Bossini-Castillo's work was supported by the Formación de Personal Investigador program at the Spanish Ministry of Economy and Competitiveness. Dr. Assassi's work was supported by NIH grant KL2-RR-024149-04, NIH National Center for Research Resources grant 3UL1-RR-024148, and NIH National Institute of Allergy and Infectious Diseases grants U01-1U01-AI09090 and K23-AR-061436 for the study of US samples. Dr. Mayes' work was supported by NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Centers of Research Translation (CORT) grant P50-AR-054144, NIH NIAMS SSc Family Registry and DNA Repository grant N01-AR-0-2251, Department of Defense grant PR1-206877, and NIH NIAMS grant R01-AR-055258 for the study of US samples. Dr. Ortego-Centeno's work was supported by grant PI-0590-2010 from Consejería de Salud y Bienestar Social, Junta de Andalucía, Spain. Dr. Radstake's work was supported by a Vidi grant from the Netherlands Organisation for Scientific Research and National Reumafonds. Dr. Martin's work was supported by a GEN-FER grant from the Spanish Society of Rheumatology, grants SAF2009-11110 and SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness, and grant CTS-4977 from Junta de Andalucía. Dr. Ortego-Centeno has received consulting fees, speaking fees, and/or honoraria from Actelion (less than $10,000). Dr. Distler has received consulting fees from Actelion, Bayer, GlaxoSmithKline, Boehringer Ingelheim, Sanofi-Aventis, and UCB (less than $10,000 each) and owns stock or stock options in 4D Science. Ms López-Isac and Dr. Bossini-Castillo contributed equally to this work.

Elena López-Isac
Lara Bossini-Castillo, PhD
Institute of Parasitology and Biomedicine López-Neyra
Consejo Superior de Investigaciones Cientificas and Parque Tecnologico de la Salud
Granada, Spain
Sandra G. Guerra, PhD
Christopher Denton, PhD, FRCP
Carmen Fonseca, MD, PhD
Royal Free Hospital and University College Medical School
London, UK
Shervin Assassi, MD
Xiaodong Zhou, MD
Maureen D. Mayes, MD, MPH
University of Texas Health Science Center at Houston
Carmen Pilar Simeón, MD, PhD
Valle de Hebrón Hospital
Barcelona, Spain
Norberto Ortego-Centeno, PhD
Clinic University Hospital
Granada, Spain
Iván Castellví, MD
Santa Creu i Sant Pau University Hospital
Barcelona, Spain
Patricia Carreira, MD
Hospital University 12 de Octubre
Madrid, Spain
The Spanish Scleroderma Group
(Multiple institutions throughout Spain)
Olga Gorlova, PhD
Dartmouth College
Lebanon, NH
Lorenzo Beretta, MD
Alessandro Santaniello, MD
Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano
Milan, Italy
Claudio Lunardi, MD
Università degli Studi di Verona
Verona, Italy
Roger Hesselstrand, MD, PhD
Lund University
Lund, Sweden
Annika Nordin, MD, PhD
Karolinska University Hospital and Karolinska Institute
Stockholm, Sweden
Gabriela Riemekasten, MD
Charité University Hospital and German Rheumatism Research Center
Berlin, Germany
Torsten Witte, MD
Hannover Medical School
Hannover, Germany
Nicolas Hunzelmann, MD
University of Cologne
Cologne, Germany
Alexander Kreuter, MD
HELIOS St. Elisabeth Hospital
Oberhausen, Germany
Jörg H. W. Distler, MD
University of Erlangen-Nuremberg
Erlangen, Germany
Alexandre E. Voskuyl, MD, PhD
VU University Medical Center
Amsterdam, The Netherlands
Jeska de Vries-Bouwstra, MD, PhD
Leiden University Medical Center
Leiden, The Netherlands
Bobby P. Koeleman, PhD
University Medical Center Utrecht
Utrecht, The Netherlands
Ariane Herrick, MD
Jane Worthington, PhD
Arthritis Research UK Epidemiology Unit
NIHR Manchester Musculoskeletal Biomedical Research Unit
University of Manchester and Manchester Academic Health Science Centre
Manchester, UK
Timothy R. D. J. Radstake, MD, PhD
Laboratory of Translational Immunology and University Medical Center Utrecht
Utrecht, The Netherlands
Javier Martin, MD, PhD
Institute of Parasitology and Biomedicine López-Neyra
Granada, Spain

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Dr. Martin had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Fonseca, Mayes, Radstake, Martin.

Acquisition of data. Guerra, Denton, Fonseca, Assassi, Zhou, Mayes, Simeón, Ortego-Centeno, Castellví, Carreira, Gorlova, Beretta, Santaniello, Lunardi, Hesselstrand, Nordin, Riemekasten, Witte, Hunzelmann, Kreuter, Distler, Voskuyl, de Vries-Bouwstra, Koeleman, Herrick, Worthington, Radstake.

Analysis and interpretation of data. López-Isac, Bossini-Castillo.

1. Martin JE, Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. Hum Genet 2012;131:1023–37.
2. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010;42:426–9.
3. Vignali DA, Kuchroo VK. IL-12 family cytokines: immunological playmakers. Nat Immunol 2012;13:722–8.
4. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet 2014;94:47–61.
5. Bossini-Castillo L, Martin JE, Broen J, Gorlova O, Simeon CP, Beretta 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;21:926–33.

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

1. Pathway enrichment analysis

We accomplished a molecular pathway enrichment analysis using the Gene Set Enrichment Analysis (GSEA) and DAVID approaches testing the set of SSc-related risk factors (*STAT4, IRF5, BANK1, BLK, TNFAIP3, MIF, ITGAM, PTPN22, HLA-DPB1, HLA-DQB1, CD247, TNIP1, TNFSF4, CSK, IRF8, IRF7, SOX5, KIAA0319L, JAZF1, NOTCH4, NFKB1, PSD3, IKZF1, PPARG, CCR6, TBX21, DDX6, DNASE1L3, ATG5, IL12A, IL12RB2, CD226, IL2RA, SAMD9L, FAS, FAM62A*) in the MSigDB Biocarta collection (1-3). The polymorphism/*locus* correspondence was established as implemented in Gene Relationships Across Implicated Loci, GRAIL software (4), using the release 18 of the Human Genome and the Pubmed text of 2012. Supplementary Table 1 illustrates all the pathways 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, but both algorithms lead to very overlapping results.

2. SNPs prioritization and genotyping

In the first phase of our study we performed a screening of the *IL12RB1* region in the discovery cohort of the Immunochip (5). We included 28 kpb spanning the complete *IL12RB1* gene and 19.6 kpb upstream and 1.7 downstream from this *locus*, from base pair 18,168,674 to 18,217,277 in chromosome 19. After QC filters and principal component analysis as described in (5), genotyping data for 46 single-nucleotide polymorphisms (SNPs) were available (Supp. Fig. 1). These 46 SNPs tagged an 88.9% of the haplotype blocks of the *IL12RB1* region, considering those SNPs with a Minor

Allele Frequency (MAF) > 0.01 according to the 1000 Genomes Project CEU population (6).

In the replication phase 4 SNPs were selected for genotyping based on both nominal association signals (*P*-values between $5x10^{-3}$ and $5x10^{-5}$) and the results of the dependence analyses in the discovery cohort of the Immunochip (dependence analyses considering the lowest *P*-values for each SNP were performed in order to use the powerful subgroup for each of them). Genotyping was performed using both TaqMan 5' allele discrimination predesigned and custom assays from Applied Biosystems (rs8109496 ID: AHRSGB3; rs2305743 ID: C__16191629_20; rs436857 ID: C____795468_1_; rs11668601 ID: AHZAEC0, Foster City, CA, USA) in a LightCycler® 480 Real-Time PCR System (Roche Applied Science, Mannheim, Germany). Genotyping call rate was > 98% for all the genotyped SNPs.

Approval from the local ethical committees and written informed consent from all participants were obtained in accordance with the tenets of the Declaration of Helsinki.

3. Statistical analysis

Genotype association analyses were carried out using PLINK V.1.07 software (7). Statistical significance was calculated by 2x2 contingency tables and χ^2 or Fisher's exact test, when necessary, to obtain *P*-values, odds ratios (ORs) and 95% confidence intervals (CIs) in the population specific analyses. Cochran-Mantel-Haenszel tests were performed to meta-analyze the combined data. *P*-values < 0.05 and *P*-values $<5x10^{-8}$ were considered statistically significant for the replication phase and the combined analysis, respectively. Breslow–Day method (BD) was used to assess the homogeneity of the associations among the different populations (Breslow–Day *P*-values < 0.05 were considered statistically significant). Hardy-Weinberg equilibrium (HWE) was tested for all cohorts (HWE *P*- values < 0.05 were considered to show significant deviation from

the equilibrium). None of the included control cohorts showed significant deviation from HWE for all the genotyped SNPs. Logistic regression and conditional logistic regression analyses considering the different cohorts as covariates were performed as implemented in PLINK. In the dependence analyses, *P*-values < 0.01 were considered independent associations. Regional association plot for *IL12RB1* region was performed using LocusZoom V1.1 software (http://csg.sph.umich.edu/locuszoom/) (8). The HapMap Project Phase I, II and III (CEU+TSI populations) was used to define the linkage disequilibrium (LD) pattern across *IL12RB1* region and Haploview V4.2 software (http://www.broadinstitute.org/haploview/haploview) was used to perform the LD plot.

The statistical power of the combined analysis was >99% for all the genotyped SNPs to detect associations with OR=1.2 and assuming an additive model, according to Power Calculator for Genetic Studies 2006 software (9).

4. *In silico* analysis for functional prioritization

We searched for possible functionality associations using annotations of gene regulatory regions from ENCODE (Encyclopedia of DNA elements) through RegulomeDB search tool. Blood eQTL browser was used to explore the evidence of potential expression Quantitative Trait *Loci* (eQTLs) (10).

We also investigated the functional connectivity of the new reported association with previously SSc associated *loci* using a computational approach. For that purpose, we performed a GRAIL analysis to identify functionally related genes in a systematic and objective manner according to textual relationships within the published literature (4). This analysis included all described SSc susceptibility *loci* with an association *P*-value of $< 5 \times 10^{-4}$. The HapMap release 18 of the human genome and the PubMed text of 2012 were used.

SUPPLEMENTARY REFERENCES

1. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* 2005;102: 15545-50.

2. Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4: 44-57.

3. Huang da W, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37: 1-13.

4. Raychaudhuri S, Plenge RM, Rossin EJ, et al. Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet* 2009;5:e1000534.

5. Mayes MD, Bossini-Castillo L, Gorlova O, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. *Am J Hum Genet* 2014;94:47-61.

6. Abecasis GR, Altshuler D, Auton A, et al. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061-73.

7. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.

8. Pruim RJ, Welch RP, Sanna S, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336-7.

9. 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.

10. Westra HJ, Peters MJ, Esko T, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet* 2013;45:1238-1243.

SUPPLEMENTARY TABLES

Supp. Table 1. GSEA and DAVID significantly enriched Biocarta pathways.

Pathway	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	p-value	FDR q-value
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
HIV-I Nef: negative effector of Fas and TNF	58	2	0.03	8.06E-04	1.59E-02

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

Supp. Table 2. Logistic regression analysis of eleven *IL12RB1* nominally associated SNPs in the Immunochip discovery cohort.

				SSc		ACA		ATA		lcSSc		dcSSc	
Chr	SNP	BP	Minor allele	P_{log}	OR	P_{log}	OR	P_{log}	OR	P_{log}	OR	P_{log}	OR
19	**rs436857**	18197635	A	3.00E-03	0.84	**3.72E-05**	0.69	0.131	0.84	1.14E-04	0.76	0.980	1.00
19	**rs2305743**	18193191	A	2.23E-03	0.85	**6.71E-05**	0.71	0.024	0.78	**8.72E-05**	0.77	0.837	0.98
19	**rs8109496**	18190344	G	2.50E-03	0.84	1.06E-04	0.71	0.044	0.80	**8.47E-05**	0.76	0.817	0.98
19	rs425648	18202112	T	0.006	0.85	1.35E-04	0.71	0.119	0.84	2.65E-04	0.78	0.958	1.01
19	rs2305742	18191441	C	3.75E-03	0.85	1.56E-04	0.71	0.048	0.80	1.35E-04	0.77	0.818	0.98
19	rs404733	18169997	T	0.042	0.92	7.05E-04	0.81	0.242	0.91	2.21E-03	0.86	0.720	0.98
19	rs365179	18181278	T	0.061	0.92	7.62E-04	0.79	0.531	0.95	6.19E-04	0.83	0.136	1.11
19	rs376008	18189568	T	0.102	0.93	1.51E-03	0.80	0.634	0.96	1.36E-03	0.84	0.111	1.12
19	rs383483	18171886	G	0.129	0.94	1.53E-03	0.82	0.692	1.03	3.09E-03	0.87	0.584	0.97
19	rs372889	18173603	C	0.165	0.94	2.05E-03	0.83	0.800	1.02	3.55E-03	0.87	0.496	0.96
19	**rs11668601**	18214521	C	4.77E-04	0.84	3.63E-03	0.81	**3.15E-04**	0.70	1.14E-03	0.82	0.029	0.85

ACA, anti-centromere antibodies; ATA, anti-topoisomerase antibodies; BP, base pair; Chr, chromosome; dcSSc, diffuse cutaneous SSc; lcSSc, limited cutaneous SSc; P_{log}, logistic regression *P*-value; OR, logistic regression odds ratio; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

Supp. Table 3. Conditional logistic regression analysis of eleven *IL12RB1* nominally associated SNPs in the Immunochip discovery cohort.

				ACA		ATA		lcSSc				
Chr	SNP	BP	Minor allele	*P-value: add to rs436857	**P-value rs436857 add to SNP	*P-value: add to rs11668601	**P-value rs11668601 add to SNP	*P-value: add to rs8109496	**P-value rs8109496 add to SNP	r^2 with rs436857	r^2 with rs11668601	r^2 with rs8109496
19	**rs436857**	18197635	A	NA	NA	0.387	9.30E-04	0.959	0.145	NA	0.39	0.88
19	**rs2305743**	18193191	A	0.406	0.420	0.770	4.45E-03	0.656	0.394	0.89	0.34	0.92
19	**rs8109496**	18190344	G	0.420	0.387	0.996	2.71E-03	NA	NA	0.88	0.34	NA
19	rs425648	18202112	T	NA	NA	0.389	9.03E-04	0.844	0.214	0.99	0.38	0.88
19	rs2305742	18191441	C	0.521	0.319	0.909	2.48E-03	NA	NA	0.89	0.34	0.99
19	rs404733	18169997	T	0.060	1.70E-03	0.401	3.33E-04	0.158	2.49E-03	0.18	0.02	0.20
19	rs365179	18181278	T	0.358	6.61E-03	0.608	2.87E-04	0.400	0.020	0.46	0.13	0.51
19	rs376008	18189568	T	0.410	5.88E-03	0.528	3.58E-04	0.534	0.014	0.46	0.13	0.51
19	rs383483	18171886	G	0.084	1.60E-03	0.916	2.27E-04	0.165	3.46E-03	0.18	0.02	0.20
19	rs372889	18173603	C	0.111	1.19E-03	0.953	2.57E-04	0.195	2.68E-03	0.19	0.02	0.21
19	**rs11668601**	18214521	C	0.608	2.51E-03	NA	NA	0.213	0.013	0.39	NA	0.33

* Single locus test *P*-value when SNP conditioned on rs436857/rs11668601/ rs8109496.
** Lead SNP single locus test *P*-value when conditioned on each analyzed SNP in logistic regression analyses.
ACA, anti-centromere antibodies; ATA, anti-topoisomerase antibodies; BP, base pair; Chr, chromosome; lcSSc, limited cutaneous SSc; NA, not applicable; SNP, single nucleotide polymorphism.

Supp. Table 4. Analysis of *IL12RB1* SNPs minor allele frequencies in Caucasian SSc patients and healthy controls from the replication cohort.

			Genotype, N (%)				Allele test		
SNP	**Minor/major**	**Subgroup (N)**	**1/1**	**1/2**	**2/2**	**MAF (%)**	**P_{MH}**	**P_{BD}**	**OR* [CI 95%]**
rs8109496	C/G	Controls (n=5070)	197 (3.89)	1648 (32.50)	3225 (63.61)	20.14			
		SSc (n=3168)	104 (3.28)	867 (27.37)	2197 (69.35)	16.97	7.251E-07	0.29	0.81 [0.74-0.88]
		lcSSc (n=1954)	68 (3.48)	549 (28.10)	1337 (68.42)	17.53	0.0010	0.47	0.85 [0.77-0.94]
		dcSSc (n=886)	29 (3.27)	220 (24.83)	637 (71.90)	15.69	8.831E-06	0.32	0.73 [0.63-0.84]
		ACA+ (n=1122)	33 (2.94)	300 (26.74)	789 (70.32)	16.31	0.0002	0.65	0.79 [0.70-0.90]
		ATA+ (n=738)	22 (2.98)	190 (25.75)	526 (71.27)	15.85	0.0001	0.13	0.74 [0.64-0.86]
rs2305743	A/G	Controls (n=5074)	209 (4.12)	1681 (33.13)	3184 (62.75)	20.68			
		SSc (n=3162)	108 (3.42)	873 (27.61)	2181 (68.98)	17.22	3.936E-08	0.08	0.79 [0.73-0.86]
		lcSSc (n=1948)	70 (3.59)	544 (27.93)	1334 (68.48)	17.56	6.322E-05	0.13	0.82 [0.75-0.90]
		dcSSc (n=885)	30 (3.39)	228 (25.76)	627 (70.85)	16.27	9.54E-06	0.30	0.73 [0.64-0.84]
		ACA+ (n=1117)	35 (3.13)	299 (26.77)	783 (70.10)	16.52	4.359E-05	0.52	0.77 [0.68-0.88]
		ATA+ (n=739)	24 (3.25)	196 (26.52)	519 (70.23)	16.51	0.0002	0.08	0.75 [0.65-0.87]
rs436857	A/G	Controls (n=5066)	178 (3.51)	1582 (31.23)	3306 (65.26)	19.13			
		SSc (n=3074)	97 (3.16)	783 (25.47)	2194 (71.37)	15.89	2.772E-07	0.23	0.80 [0.73-0.87]
		lcSSc (n=1905)	65 (3.41)	482 (25.30)	1358 (71.29)	16.06	7.029E-05	0.26	0.82 [0.74-0.90]

		dcSSc (n=853)	27 (3.17)	209 (24.50)	617 (72.33)	15.42	0.0002	0.42	0.76 [0.66-0.88]
		ACA+ (n=1096)	33 (3.01)	268 (24.45)	795 (72.54)	15.24	0.0001	0.38	0.78 [0.68-0.88]
		ATA+ (n=714)	18 (2.52)	174 (24.37)	522 (73.11)	14.71	5.823E-05	0.08	0.72 [0.62-0.85]
rs11668601	C/T	Controls (n=5067)	322 (6.35)	1921 (37.91)	2824 (55.73)	25.31			
		SSc (n=3094)	165 (5.33)	1026 (33.16)	1903 (61.51)	21.91	2.737E-06	0.36	0.83 [0.77-0.90]
		lcSSc (n=1919)	95 (4.95)	645 (33.61)	1179 (61.44)	21.76	4.964E-05	0.07	0.83 [0.76-0.91]
		dcSSc (n=863)	52 (6.03)	283 (32.79)	528 (61.18)	22.42	0.0132	0.19	0.85 [0.75-0.97]
		ACA+ (n=1099)	56 (5.10)	362 (32.94)	681 (61.97)	21.57	0.0003	0.59	0.81 [0.73-0.91]
		ATA+ (n=724)	47 (6.49)	231 (31.91)	446 (61.60)	22.44	0.0148	0.23	0.85 [0.74-0.97]

*Odds ratio for the minor allele

ACA, anti-centromere antibodies; ATA, anti-topoisomerase antibodies; CI, confidence interval; dcSSc, diffuse cutaneous SSc; lcSSc, limited cutaneous SSc; MAF, minor allele frequency; OR, odds ratio; P_{BD}, Breslow–Day test P-value; P_{MH}, allelic Mantel-Haenszel fixed effects model *P*-value; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

Supp. Table 5. Analysis of *IL12RB1* SNPs minor allele frequencies in the overall combined analysis.

SNP	Minor/major	Subgroup (N)	Genotype, N (%)			MAF (%)	Allele test		
			1/1	1/2	2/2		P_{MH}	P_{BD}	OR* [CI 95%]
rs8109496	C/G	Controls (n=8697)	329 (3.78)	2736 (31.46)	5632 (64.76)	19.51			
		SSc (n=5036)	148 (2.94)	1343 (26.67)	3545 (70.39)	16.27	7.35E-09	0.32	0.82 [0.77-0.88]
		lcSSc (n=3096)	92 (2.97)	821 (26.52)	2183 (70.51)	16.23	5.65E-07	0.49	0.82 [0.76-0.89]
		dcSSc (n=1468)	47 (3.20)	387 (26.36)	1034 (70.44)	16.38	0.0003	0.04	0.82 [0.74-0.91]
		ACA+ (n=1801)	43 (2.39)	453 (25.15)	1305 (72.46)	14.96	1.47E-07	0.64	0.76 [0.69-0.84]
		ATA+ (n=1096)	33 (3.01)	271 (24.73)	792 (72.26)	15.37	1.36E-05	0.23	0.76 [0.67-0.86]
rs2305743	A/G	Controls (n=8697)	353 (4.06)	2796 (32.15)	5548 (63.79)	20.13			
		SSc (n=5032)	166 (3.30)	1358 (26.99)	3508 (69.71)	16.79	4.29E-10	0.10	0.81 [0.76-0.87]
		lcSSc (n=3091)	101 (3.27)	825 (26.69)	2165 (70.04)	16.61	2.38E-08	0.23	0.80 [0.74-0.87]
		dcSSc (n=1467)	52 (3.54)	399 (27.20)	1016 (69.26)	17.14	0.0003	0.03	0.83 [0.74-0.92]
		ACA+ (n=1797)	50 (2.78)	455 (25.32)	1292 (71.90)	15.44	1.28E-08	0.60	0.75 [0.68-0.83]
		ATA+ (n=1097)	36 (3.28)	280 (25.52)	781 (71.19)	16.04	1.08E-05	0.20	0.76 [0.67-0.86]
rs436857	A/G	Controls (n=8652)	297 (3.43)	2652 (30.65)	5703 (65.92)	18.76			
		SSc (n=4924)	142 (2.88)	1245 (25.28)	3537 (71.83)	15.53	3.94E-09	0.23	0.81 [0.76-0.87]
		lcSSc (n=3035)	90 (2.97)	745 (24.55)	2200 (72.49)	15.24	3.62E-08	0.39	0.80 [0.73-0.86]

		dcSSc (n=1431)	44 (3.07)	377 (26.35)	1010 (70.58)	16.25	0.0035	0.11	0.85 [0.76-0.95]
		ACA+ (n=1767)	44 (2.49)	411 (23.26)	1312 (74.25)	14.12	2.67E-08	0.40	0.74 [0.67-0.83]
		ATA+ (n=1067)	27 (2.53)	262 (24.55)	778 (72.91)	14.81	3.11E-05	0.11	0.76 [0.67-0.87]
rs11668601	C/T	Controls (n=8682)	539 (6.21)	3302 (38.03)	4841 (55.76)	25.22			
		SSc (n=4962)	245 (4.94)	1690 (34.06)	3027 (61.00)	21.97	5.61E-09	0.06	0.84 [0.79-0.89]
		lcSSc (n=3062)	138 (4.51)	1063 (34.72)	1861 (60.78)	21.86	2.18E-07	0.11	0.83 [0.77-0.89]
		dcSSc (n=1443)	83 (5.75)	479 (33.19)	881 (61.05)	22.35	0.0010	0.01	0.85 [0.77-0.94]
		ACA+ (n=1779)	81 (4.55)	609 (34.23)	1089 (61.21)	21.67	3.68E-06	0.64	0.81 [0.74-0.89]
		ATA+ (n=1080)	59 (5.46)	346 (32.04)	675 (62.50)	21.48	4.47E-05	0.20	0.79 [0.71-0.89]

Odds ratio for the minor allele

ACA, anti-centromere antibodies; ATA, anti-topoisomerase antibodies; CI, confidence interval; dcSSc, diffuse cutaneous SSc; lcSSc, limited cutaneous SSc; MAF, minor allele frequency; OR, odds ratio; P_{BD}, Breslow–Day test P-value; P_{MH}, allelic Mantel-Haenszel fixed effects model *P*-value; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

Supp. Table 6. Conditional logistic regression analysis of *IL12RB1* selected SNPs in the overall combined cohort.

CHR	SNP	Minor allele	P_{log}	OR	*P-value: add to rs2305743	OR[1] add to rs2305743	*P-value: add to rs436857	OR[1] add to rs436857	*P-value: add to rs8109496	OR[1] add to rs8109496	*P-value: add to rs11668601	OR[1] add to rs11668601
19	rs8109496	C	3.466E-11	0.80	0.214	0.87	0.043	0.84	NA	NA	1.72E-04	0.86
19	rs2305743	A	1.484E-11	0.80	NA	NA	0.100	0.86	0.452	0.92	1.32E-04	0.86
19	rs436857	A	2.18E-11	0.80	0.325	0.91	NA	NA	0.521	0.94	1.91E-04	0.85
19	rs11668601	C	1.298E-09	0.83	0.005	0.90	0.017	0.91	0.005	0.90	NA	NA

* Single locus test *P*-value when SNP conditioned on rs2305743/ rs436857/rs11668601/ rs8109496.
1. Single locus test OR when SNP conditioned on rs2305743/rs436857/rs11668601/ rs8109496.
Chr, chromosome; NA, not applicable; P_{log}, logistic regression *P*-value; SNP: single nucleotide polymorphism; OR, odds ratio.

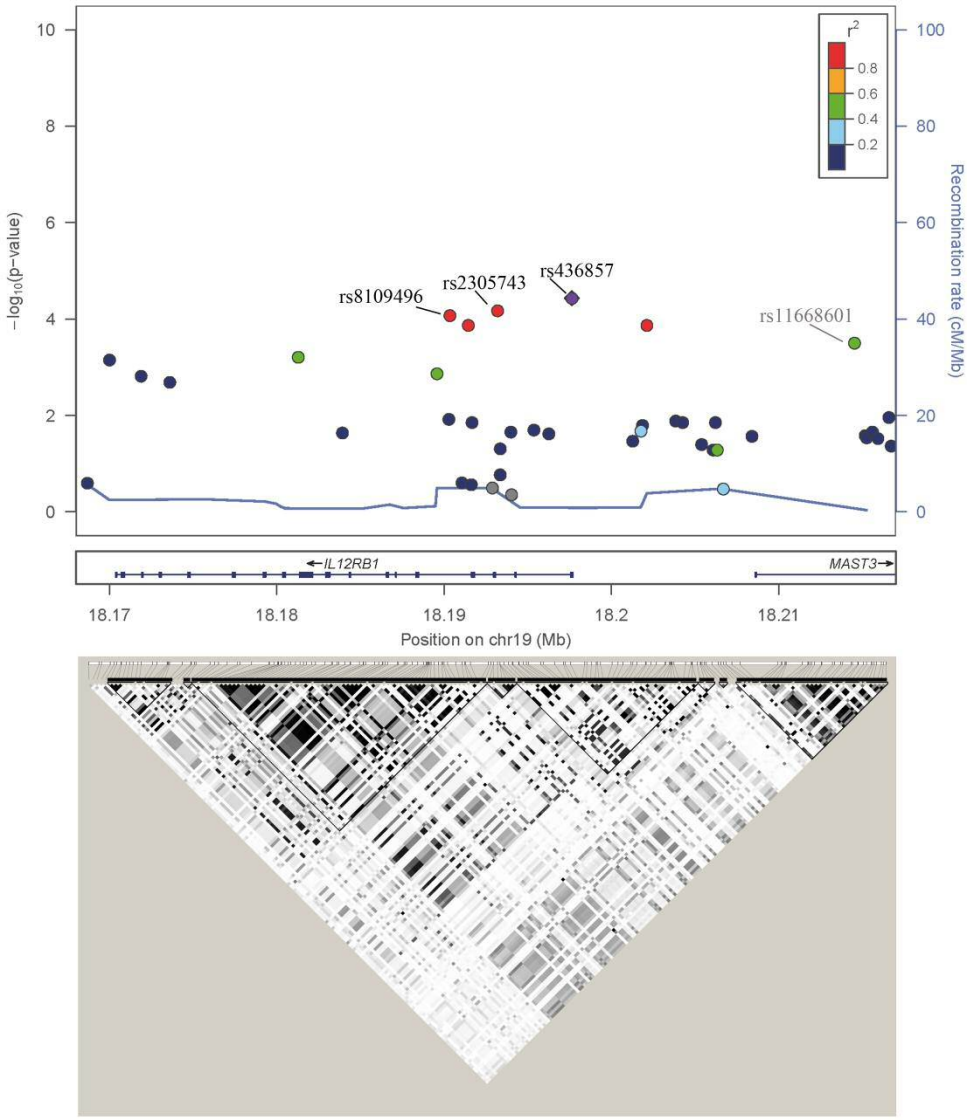
SUPPLEMENTARY FIGURES

FIGURE LEGENDS

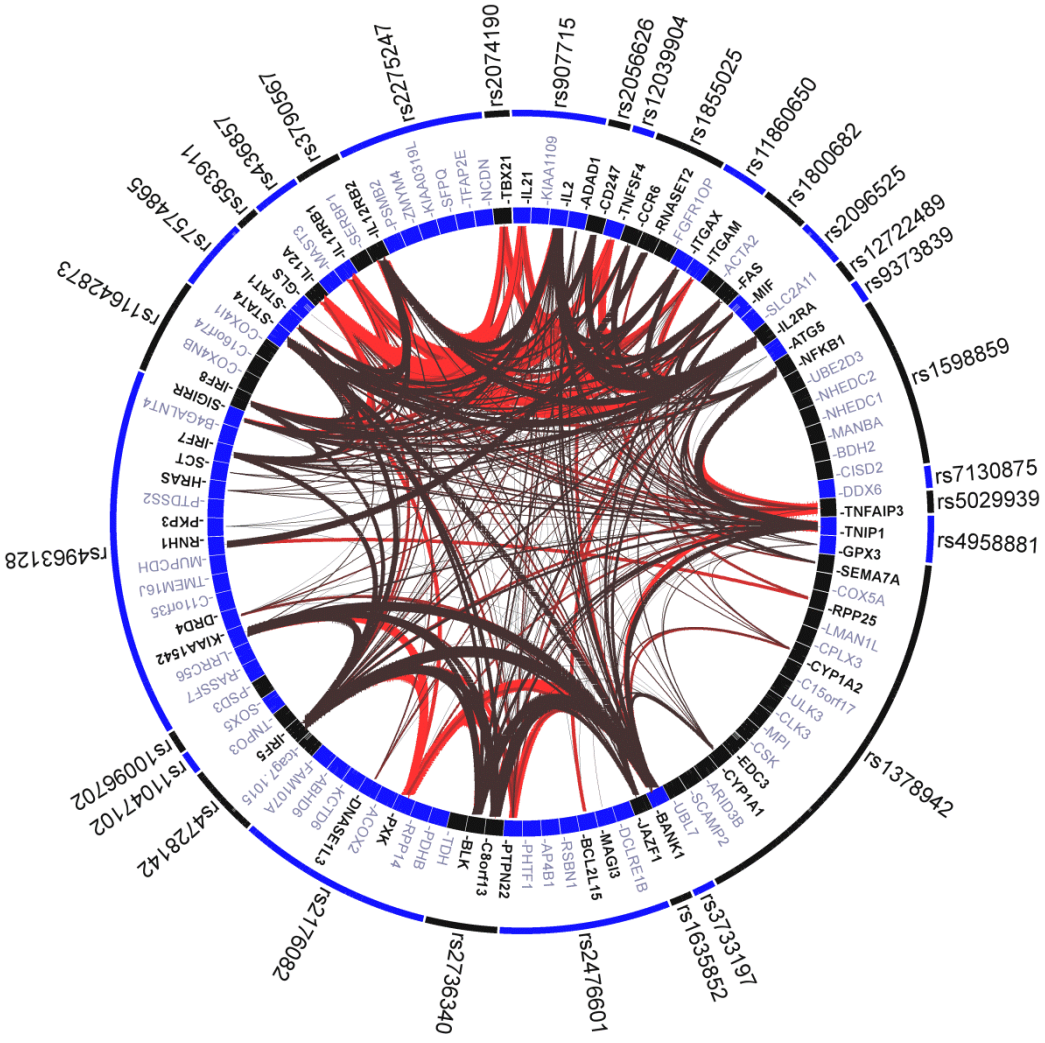
Supplementary figure 1. Association result plot for *IL12RB1* region in the Immunochip screening phase and linkage disequilibrium pattern across the *locus*. (The *P*-values presented in the plot correspond to the subgroup in which each SNP showed the most significant association).

Supplementary figure 2. GRAIL analysis circle plot showing functional connectivity between *loci*. The higher density of established connections is found in the upper part of the circle plot, in which the majority of the IL12-pathway related genes associated with SSc are located (e.g. *STAT4*, *IL12RB2*, *IL12RB1*, *TBX21* and *IL12A*). In addition, the plot shows strong literature-based connectivity between IL12-pathway related genes and several robustly SSc associated *loci,* such as *CD247*, *TNFSF4*, *ITGAM* and *BLK*.

Supplementary figure 1.



Supplementary figure 2.



V

Publication 3: Influence of *TYK2* in systemic sclerosis susceptibility: a new locus in the IL-12 pathway

EXTENDED REPORT

Influence of *TYK2* in systemic sclerosis susceptibility: a new *locus* in the IL-12 pathway

Elena López-Isac,[1] Diana Campillo-Davo,[1] Lara Bossini-Castillo,[1] Sandra G Guerra,[2] Shervin Assassi,[3] Carmen Pilar Simeón,[4] Patricia Carreira,[5] Norberto Ortego-Centeno,[6] Paloma García de la Peña,[7] the Spanish Scleroderma Group, Lorenzo Beretta,[8] Alessandro Santaniello,[8] Chiara Bellocchi,[8] Claudio Lunardi,[9] Gianluca Moroncini,[10] Armando Gabrielli,[10] Gabriela Riemekasten,[11,12] Torsten Witte,[13] Nicolas Hunzelmann,[14] Alexander Kreuter,[15] Jörg HW Distler,[16] Alexandre E Voskuyl,[17] Jeska de Vries-Bouwstra,[18] Ariane Herrick,[19] Jane Worthington,[19] Christopher P Denton,[2] Carmen Fonseca,[2] Timothy RDJ Radstake,[20] Maureen D Mayes,[3] Javier Martín[1]

Handling editor Tore K Kvien

▸ Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2015-208154).

For numbered affiliations see end of article.

Correspondence to
Dr Javier Martin, Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC. Parque Tecnológico Ciencias de la Salud. Avenida del Conocimiento s/n Armilla (Granada) 18016, Spain; martin@ipb.csic.es

Received 29 June 2015
Revised 27 July 2015
Accepted 13 August 2015
Published Online First 2 September 2015

ABSTRACT

Objectives *TYK2* is a common genetic risk factor for several autoimmune diseases. This gene encodes a protein kinase involved in interleukin 12 (IL-12) pathway, which is a well-known player in the pathogenesis of systemic sclerosis (SSc). Therefore, we aimed to assess the possible role of this *locus* in SSc.

Methods This study comprised a total of 7103 patients with SSc and 12 220 healthy controls of European ancestry from Spain, USA, Germany, the Netherlands, Italy and the UK. Four *TYK2* single-nucleotide polymorphisms (V362F (rs2304256), P1104A (rs34536443), I684S (rs12720356) and A928V (rs35018800)) were selected for follow-up based on the results of an Immunochip screening phase of the locus. Association and dependence analyses were performed by the means of logistic regression and conditional logistic regression. Meta-analyses were performed using the inverse variance method.

Results Genome-wide significance level was reached for *TYK2* V362F common variant in our pooled analysis ($p=3.08\times10^{-13}$, OR=0.83), while the association of P1104A, A928V and I684S rare and low-frequency missense variants remained significant with nominal signals ($p=2.28\times10^{-3}$, OR=0.80; $p=1.27\times10^{-3}$, OR=0.59; $p=2.63\times10^{-5}$, OR=0.83, respectively). Interestingly, dependence and allelic combination analyses showed that the strong association observed for V362F with SSc, corresponded to a synthetic association dependent on the effect of the three previously mentioned *TYK2* missense variants.

Conclusions We report for the first time the association of *TYK2* with SSc and reinforce the relevance of the IL-12 pathway in SSc pathophysiology.

CrossMark

To cite: López-Isac E, Campillo-Davo D, Bossini-Castillo L, *et al*. *Ann Rheum Dis* 2016;**75**:1521–1526.

INTRODUCTION

Systemic sclerosis (SSc) is an autoimmune disease that involves extensive fibrosis in the skin and different internal organs, abnormalities of the vascular system and immune imbalance with autoantibody production, particularly anticentromere autoantibodies (ACA) and antitopoisomerase autoantibodies (ATA). The aetiology of the disease is largely unknown, although both environmental and genetic factors are thought to be involved in the disease development.[1]

Large genetic studies, including genome-wide association studies and Immunochip analysis, have identified several immune-related loci underlying the susceptibility to SSc onset.[2 3] Although great advances have been made over the last 7 years, our knowledge of SSc genetic background is still limited, and the numbers of convincingly SSc genetic markers only account for a small proportion of the total genetic variance for the disease.[4 5] Thus, further genetic studies will help to better understand the pathogenic processes implicated in SSc development.

A recent fine-mapping genetic study of a common autoimmunity locus, *TYK2-ICAM*, in rheumatoid arthritis (RA) identified three *TYK2* protein-coding variants as the most likely causal variants responsible for the signal of association in the region. The authors also extended the results into other autoimmune phenotypes, such as systemic lupus erythematosus (SLE), and observed that the three variants are missense mutations predicted to be damaging using functional prediction tools.[6]

TYK2 encodes a tyrosine kinase member of the JAK-STAT family, and mediates signalling of different interleukin 12 (IL-12) family cytokines, such us IL-12 and IL-23. Several polymorphisms in this locus have been associated with other autoimmune diseases, such as psoriasis, multiple sclerosis, Crohn's disease and ulcerative colitis.[7]

Interestingly, SSc Immunochip study[3] found suggestive, but not significant, evidence of association in *TYK2* region (p values ranging from 5×10^{-4} to 5×10^{-2}). Moreover, different functional and genetic studies highlighted the special relevance of IL-12/STAT4 pathway in the disease pathophysiology.[3 4 8 9] Thus, we performed a follow-up study to further investigate whether variations within this genomic region, including

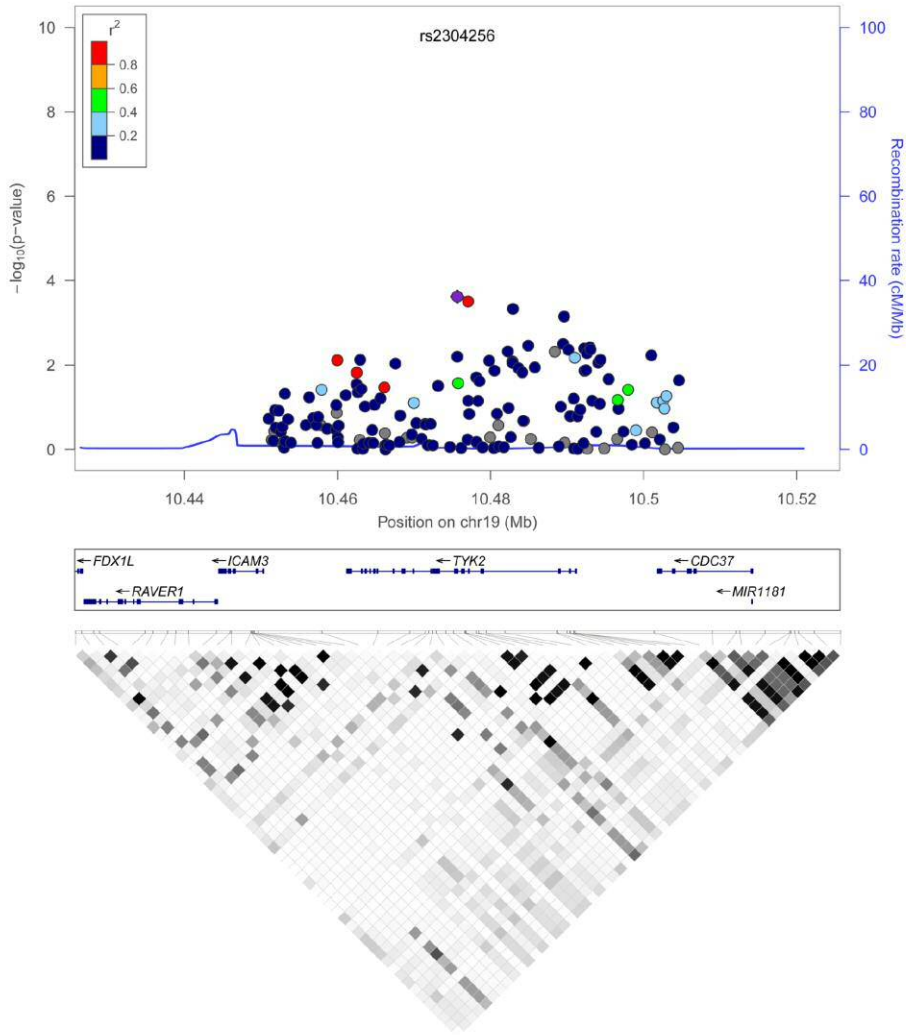


Figure 1 Association result plot for *TYK2* region in the Immunochip screening phase. The p values for association ($-\log_{10}$ values) of each single-nucleotide polymorphism are plotted against their physical position on chromosome 19. The lower panel shows the linkage disequilibrium pattern at the TYK2 *locus* (r^2 values are indicated by colour gradient).

the three variants responsible for the association in RA and other autoimmune phenotypes, are also involved in SSc susceptibility.

METHODS

Study population

This study comprised a total of 7103 patients with SSc and 12 220 healthy controls of European ancestry. The 2118 patients with SSc and 4742 healthy controls from Spain and USA enrolled in the SSc Immunochip screening phase were obtained from the previously published SSc Immunochip study[3] and additional Immunochip data for Spanish patients with SSc and healthy controls. The validation cohort included 4985 SSc cases and 7478 controls from independent case–control sets of European ancestry (Germany, the Netherlands, Italy, UK and USA).

Patients with SSc fulfilled the 1980 American College of Rheumatology classification criteria for this disease or the criteria proposed by LeRoy and Medsger for early SSc.[10 11] In addition, patients were classified as having limited cutaneous SSc or diffuse cutaneous SSc as described in LeRoy *et al*.[12] Patients were also subdivided by autoantibody status according to the presence of ACA or ATA.

Approval from the local ethics committees and written informed consent from all participants were obtained in accordance with the tenets of the Declaration of Helsinki.

Study design

SSc Immunochip screening phase

An initial evaluation of *TYK2* region was performed in the SSc Immunochip screening phase. We included 30 kpb spanning the complete *TYK2* gene and 10 kpb upstream and downstream

Table 1 Inverse variance meta-analysis of four *TYK2* SNPs in seven different cohorts of patients with SSc and healthy controls (7103 patients with SSc and 12 220 controls)

						Inverse variance test		
Chr	SNP	Minor/major	Comment	MAF cases	MAF controls	p Value	OR (95% CI)*	Q
19	rs34536443 (P1104A)	C/G	missense Pro >Ala	0.023	0.026	2.28E-03	0.80 (0.69 to 0.92)	0.13
19	rs35018800 (A928V)	A/G	missense Ala >Val	0.004	0.008	1.27E-03	0.59 (0.42 to 0.81)	0.34
19	rs12720356 (I684S)	C/A	missense Ile >Ser	0.067	0.078	2.63E-05	0.83 (0.78 to 0.91)	0.27
19	rs2304256 (V362F)	A/C	missense Val >Phe	0.246	0.279	3.08E-13	0.83 (0.79 to 0.87)	0.69

*OR for the minor allele.
Chr, chromosome; MAF, minor allele frequency; Q, heterogeneity value; SNP, single-nucleotide polymorphism; SSc, systemic sclerosis.

from this locus, from base pair 10 450 993 to 10 504 616 in chromosome 19. The analysed genetic region comprised the linkage disequilibrium (LD) block that completely covers *TYK2* (figure 1). Quality control filters and principal component analysis were applied as described in ref. 3. We performed single-nucleotide polymorphism (SNP) genotype imputation of the *TYK2* region as implemented in IMPUTE2 with the use of the 1000 Genomes Phase 1 reference panel.[13 14] After imputation, genotyping data for 154 SNPs were available.

Follow-up phase

Four *TYK2* missense mutations were selected for validation in independent replication cohorts: one common coding variant (V362F (rs2304256)), two low-frequency coding variants (P1104A (rs34536443), I684S (rs12720356)) and one rare coding variant (A928V (rs35018800)). Finally, we performed meta-analysis for the selected SNPs combining the cohorts from both stages.

Genotyping methods

The genotyping of the SSc cases included in the validation cohorts was performed with both TaqMan SNP genotyping technology and Immunochip platform. For TaqMan genotyping system, we used TaqMan 5′ allele discrimination predesigned assays from Applied Biosystems in a LightCycler 480 Real-Time PCR System (Roche Applied Science, Mannheim, Germany). Genotyping call rate was >95% for all the SNPs. The Immunochip genotyping was performed on the Illumina iScan system, as per Illumina protocols, in the Centre for Genomics and Oncological Research (GENYO, Granada, Spain). Control genotyping data partially overlapped with those from previous Immunochip reports.[15–23] If any of the four selected SNPs was missing in a dataset, imputation was applied. Genotype imputation was performed with IMPUTE2 using the 1000 Genomes Phase 1 reference panel.[13 14] The correspondence between Immunochip (including imputed data) and TaqMan genotyping data was >98% for all the SNPs.

Data analysis

Associations of the SNPs with SSc were evaluated by logistic regression analysis in all the cohorts separately. Meta-analysis was performed with inverse-variance weighting under a fixed-effects model as implemented in PLINK V.1.07 software.[24] The combined analysis, including the two phases of the study, was also performed using the inverse variance method based on population-specific logistic regression analyses. p Values <0.05 were considered statistically significant in the association analyses. Heterogeneity between the datasets was assessed using Cochran's Q test. Q values <0.05 were considered statistically significant. Hardy–Weinberg equilibrium (HWE) was tested for all the validation cohorts (HWE p values <0.01 were considered to show significant deviation from the equilibrium). None of the included control cohorts showed significant deviation from HWE for all the genotyped SNPs.

To test the independence of association between each SNP, we performed conditional logistic regression analyses as implemented in PLINK. To analyse the possible effect of A928V (rs35018800) in conditioning analysis, a generalised null linear model including population origin and two variants (P1104A (rs34536443) and I684S (rs12720356)) as covariates was compared against an alternative model including the same variables and A928V (rs35018800) variant by the means of a likelihood ratio test in R. We also assessed the different allelic combinations using PLINK. Allelic combinations with a frequency <0.5% were excluded from the analysis.

Regional association plot for *TYK2* region was performed using LocusZoom V.1.1 software (http://csg.sph.umich.edu/locuszoom/).[25] The HapMap Project phase I, II and III (CEU populations) was used to define the LD pattern across *TYK2* region, and Haploview V.4.2 software (http://www.

Table 2 Dependence analysis by pairwise conditioning of four *TYK2* SNPs in the overall combined cohort (7103 patients with SSc and 12 220 controls)

SNP	MAF cases/controls	Unconditioned p value	OR	*p Value: add to rs2304256	OR† add to rs2304256	*p Value: add to rs34536443	OR† add to rs34536443	*p Value: add to rs35018800	OR† add to rs35018800	*p Value: add to rs12720356	OR† add to rs12720356
rs34536443	0.023/0.026	2.28E-03	0.80	0.02	0.84	NA	NA	1.51E-03	0.79	1.04E-03	0.78
rs35018800	0.004/0.008	1.27E-03	0.59	9.56E-03	0.65	1.11E-03	0.57	NA	NA	1.20E-03	0.59
rs12720356	0.067/0.078	2.63E-05	0.83	0.176	0.94	5.09E-06	0.82	2.22E-05	0.83	NA	NA
rs2304256	0.246/0.279	3.08E-13	0.83	NA	NA	1.37E-05	0.89	4.75E-12	0.84	1.63E-07	0.86

*Single locus test p value when SNP conditioned on rs2304256/rs34536443/rs35018800/rs12720356.
†Single locus test OR when SNP conditioned on rs2304256/rs34536443/rs35018800/rs12720356.
MAF, minor allele frequency; NA, not applicable; SNP, single-nucleotide polymorphism; SSc, systemic sclerosis.

Table 3 Conditional logistic regression analysis of four *TYK2* SNPs in the overall combined cohort (7103 patients with SSc and 12 220 controls)

SNP	Unconditioned p value	OR	Conditioned to rs2304256, rs35018800, rs12720356 p Value	OR	Conditioned to rs2304256, rs12720356, rs34536443 p Value	OR	Conditioned to rs2304256, rs35018800, rs34536443 p Value	OR	Conditioned to rs34536443, rs12720356 p Value	OR	Conditioned to rs35018800, rs12720356, rs34536443 p Value	OR
rs34536443	2.28E-03	0.80	6.94E-04	0.76	NA	NA	NA	NA	NA	NA	NA	NA
rs35018800	1.27E-03	0.59	NA	NA	8.60E-04	0.56	NA	NA	9.39E-04	0.57	NA	NA
rs12720356	2.63E-05	0.83	NA	NA	NA	NA	4.70E-04	0.84	NA	NA	NA	NA
rs2304256	3.08E-13	0.83	NA	NA	NA	NA	NA	NA	0.091	0.94	0.270	0.97

NA, not applicable; SNP, single-nucleotide polymorphism; SSc, systemic sclerosis.

broadinstitute.org/haploview/haploview) was used to perform the LD plot. The statistical power of the combined analysis is shown in online supplementary table S1, and was calculated according to Power Calculator for Genetic Studies 2006 software under an additive model.[26]

RESULTS

SSc Immunochip initial screening

The initial screening of *TYK2* region performed in the SSc Immunochip study showed several tier-two association signals at this locus (figure 1). A common protein-coding missense variant previously associated with SLE showed the strongest association with the disease (V362F (rs2304256) p value=2.39×10^{-4}, OR=0.85).[27–29] This variant and the three *TYK2* protein-coding variants responsible for the association with RA and SLE according to Diogo *et al*[6] were selected for follow-up in independent validation cohorts to confirm the suggestive evidence of association found in this locus with SSc.

Follow-up phase and meta-analysis

Pooled analysis, including the five validation cohorts, revealed significant associations for the four *TYK2* SNPs with SSc at $p<0.05$ (see online supplementary table S2). The meta-analysis combining both steps showed that *TYK2* V362F (rs2304256) variant achieved the genome-wide significance level ($p=3.08\times10^{-13}$, OR=0.83), while P1104A (rs34536443), A928V (rs35018800) and I684S (rs12720356) remained with significant nominal p values ($p=2.28\times10^{-3}$, OR=0.80; $p=1.27\times10^{-3}$, OR=0.59; $p=2.63\times10^{-5}$, OR=0.83, respectively) (table 1). No significant heterogeneity in the ORs among the seven cohorts was observed. The analyses carried out for the main SSc clinical features revealed that the observed association signal rely on the whole disease (data not shown).

Dependence analyses

We then assessed the independence of associations by conditional logistic regression analyses. Although pairwise conditioning results were not conclusive (table 2), the V362F genome-wide significance association was lost when adding the allelic dosage for rs3453644, rs35018800 and rs12720356 as covariates ($p_{cond}=0.270$) (table 3), supporting that the *TYK2* V362F association was dependent on the three rare and low-frequency missense variants. Although A928V (rs35018800) seemed not to exert an effect on V362F (rs2304256) association, model fitting test showed that the regression model, including this rare variant as covariate had a significantly better likelihood than the model excluding it ($p=1.15\times10^{-4}$). Allelic combination tests also confirmed that the V362F association was driven by the presence of the minor alleles of P1104A, A928V and I684S *TYK2* variants, since no genome-wide significant p value was observed for the allelic model carrying only the minor allele of V362F (see online supplementary table S3).

DISCUSSION

The overall analysis of our study reported genome-wide significance level of association for *TYK2* with SSc, providing robust evidence for the implication of this new locus in SSc development.

The meta-analysis showed strong association for V362F common variant, whereas the rare and low-frequency variants—P1104A, A928V and I684S—remained with significant nominal association signals. Although our study was underpowered to detect associations at the genome-wide level of significance for these three missense variants, dependence analyses clearly supported that V362F association was a spurious signal, driven by P1104A, A928V and I684S. This effect is probably due to the high D values between V362F and the three rare and low-frequency variants.

Our findings are in accordance with the results reported by Diogo *et al*,[6] which narrowed down *TYK2* association to the three missense variants—P1104A, A928V and I684S—in RA and other autoimmune diseases through a fine-mapping strategy. The results are also consistent with the predictions of Polyphen-2 and SIFT tools, since common *TYK2* missense variant, V362F, was predicted to be benign while P1104A, A928V and I684S were damaging mutations.[30 31] In addition, the functional effect of P1104A and I684S variants (located in the kinase domains of the protein) has also been addressed by in vitro studies in primary T cells, B cells and fibroblasts. These studies showed that P1104A and I684S are catalytically impaired, leading to a reduced TYK2 activity and decreasing pro-inflammatory cytokines signalling, such as IL-6 or IL-12.[32 33] Nevertheless, since the three *TYK2* rare and low-frequency variants included in the present study were selected according to the detailed fine-mapping study performed by Diogo *et al* in a large RA study cohort, the genetic effect of additional independent rare and low-frequency *TYK2* variants cannot be ruled out in SSc susceptibility.

Interestingly, several IL-12 pathway-related genes have been reported to be associated with SSc: *IL12RB1* and *IL12RB2* (both IL-12-receptor chains), *IL12A* (p35 subunit of IL-12) and *STAT4* (the transcription factor of the IL-12 signalling axis).[3 4 8 9] Thus, the association of *TYK2* with SSc reported in the present study adds another piece of evidence showing the crucial role of this IL pathway in SSc pathogenesis.

IL-12 is a pro-inflammatory cytokine that induces type 1 helper T cells (Th1) and, in combination with interferon (IFN)-γ, antagonises type 2 helper T cells (Th2) differentiation.[34] Serum levels of IL-12 are significantly increased in patients with SSc, and this overproduction has been associated with renal vascular damage.[35] In addition, functional studies have suggested that Th1 responses may be crucial in mediating early inflammatory processes in SSc. As stated above, P1104A, A928V and I684S missense variants are damaging *TYK2* mutations that ultimately lead to an impaired IL-12 signalling. This effect would be consistent with the protective effect observed for these variants and a lower SSc susceptibility. Thus, target therapies blocking this pathway could be an effective treatment for the disease, such as ustekinumab, an anti-IL-12/23 p40 monoclonal antibody currently approved for the treatment of psoriatic arthritis.[36–38]

Remarkably, pharmaceutical companies are setting their sight on JAK family as therapeutic targets for the treatment of autoimmune diseases, such as RA and type 1 diabetes, given its central role in the signalling pathways of a wide range of cytokines. Drug discovery research is focused on the development of specific JAK protein inhibitors, such as the recently approved JAK3 inhibitor, tofacitinib, for the treatment of RA.[39] TYK2 inhibitors have also been described, although none of these drugs have yet made it to the clinical trials.[40]

In summary, the present study identified *TYK2* as a novel susceptibility factor for SSc. Our results, together with previous findings, reinforce the crucial involvement of IL-12 signalling axis in the disease development; thus, this pathway might represent an attractive therapeutic target for the treatment of SSc.

Author affiliations
[1]Institute of Parasitology and Biomedicine López-Neyra, IPBLN-CSIC, PTS Granada, Granada, Spain
[2]Centre for Rheumatology, Royal Free and University College Medical School, London, UK
[3]Division of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center–Houston, Houston, USA
[4]Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain
[5]Department of Rheumatology, 12 de Octubre University Hospital, Madrid, Spain
[6]Department of Internal Medicine, Clinic University Hospital, Granada, Spain
[7]Department of Rheumatology, Madrid Norte Sanchinarro Hospital, Madrid, Spain
[8]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy
[9]Department of Medicine, Università degli Studi di Verona, Verona, Italy
[10]Clinica Medica, Dipartimento di Scienze Cliniche e Molecolari, Università Politecnica delle Marche and Ospedali Riuniti, Ancona, Italy
[11]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany
[12]German Rheumatism Research Center (DRFZ), a Leibniz Institute, Berlin, Germany
[13]Department of Clinical Immunology, Hannover Medical School, Hannover, Germany
[14]Department of Dermatology, University of Cologne, Cologne, Germany
[15]Department of Dermatology, Venereology, and Allergology, HELIOS St Elisabeth Hospital Oberhausen, Germany
[16]Department of Internal Medicine, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen, Germany
[17]Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands
[18]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[19]Centre for Musculoskeletal Research and NIHR Manchester Musculoskeletal Biomedical Research Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[20]Department of Rheumatology & Clinical Immunology, Laboratory of Translational Immunology, department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

Acknowledgements We thank Sofia Vargas, Sonia García and Gema Robledo for her excellent technical assistance and all the patients and control donors for their essential collaboration. We thank National DNA Bank Carlos III (University of Salamanca, Spain) who supplied part of the control DNA samples. We also would like to thank the following organisations: The EULAR Scleroderma Trials and Research group (EUSTAR), the German Network of Systemic Sclerosis, The Scleroderma Foundation (USA) and RSA (Raynaud's & Scleroderma Association).

Collaborators *Spanish Scleroderma Group*: Raquel Ríos and Jose Luis Callejas, Unidad de Enfermedades Sistémicas Autoinmunes, Department of Internal Medicine, Hospital Clínico Universitario San Cecilio, Granada; José Antonio Vargas Hitos, Department of Internal Medicine, Hospital Virgen de las Nieves, Granada; Rosa García Portales, Department of Rheumatology, Hospital Virgen de la Victoria, Málaga; María Teresa Camps, Department of Internal Medicine, Hospital Carlos Haya, Málaga; Antonio Fernández-Nebro, Department of Rheumatology, Hospital Carlos Haya, Málaga; María F González-Escribano, Department of Immunology, Hospital Virgen del Rocío, Sevilla; Francisco José García-Hernández and Mª Jesús Castillo, Department of Internal Medicine, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Department of Rheumatology, Hospital Reina Sofía/IMIBIC, Córdoba; Benjamín Fernández-Gutiérrez and Luis Rodríguez-Rodríguez, Department of Rheumatology, Hospital Clínico San Carlos, Madrid; Esther Vicente, Department of Rheumatology, Hospital La Princesa, Madrid; José Luis Andreu and Mónica Fernández de Castro, Department of Rheumatology, Hospital Puerta de Hierro Majadahonda, Madrid; Francisco Javier López-Longo and Lina Martínez, Department of Rheumatology, Hospital General Universitario Gregorio Marañón, Madrid; Vicente Fonollosa and Alfredo Guillén, Department of Internal Medicine, Hospital Valle de Hebrón, Barcelona; Iván Castellvi, Department of Rheumatology, Santa Creu i Sant Pau University Hospital, Barcelona; Gerard Espinosa, Department of Internal Medicine, Hospital Clínic, Barcelona; Carlos Tolosa, Department of Internal Medicine, Hospital Parc Taulí, Sabadell; Anna Pros, Department of Rheumatology, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Department of Internal Medicine, Hospital Universitari Mútua Terrasa, Barcelona; Francisco Javier Narváez, Department of Rheumatology, Hospital Universitari de Bellvitge, Barcelona; Manel Rubio Rivas, Department of Internal Medicine, Hospital Universitari de Bellvitge, Barcelona; Vera Ortiz Santamaría, Department of Rheumatology, Hospital General de Granollers, Granollers; Ana Belén Madroñero, Department of Internal Medicine, Hospital General San Jorge, Huesca; Miguel Ángel González-Gay, Department of Rheumatology, Hospital Universitario Marqués de Valdecilla, IFIMAV, Santander; Bernardino Díaz and Luis Trapiella, Department of Internal Medicine, Hospital Central de Asturias, Oviedo; Mayka Freire and Adrián Sousa, Unidad de Trombosis y Vasculitis, Department of Internal Medicine, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; María Victoria Egurbide, Department of Internal Medicine, Hospital Universitario Cruces, Barakaldo; Patricia Fanlo Mateo, Department of Internal Medicine Hospital Virgen del Camino, Pamplona; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Department of Internal Medicine, Hospital Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Department of Rheumatology, Hospital Universitario de Canarias, Tenerife; Emma Beltrán, Department of Rheumatology, Hospital General Universitario de Valencia, Valencia; José Andrés Román-Ivorra and Elena Grau, Department of Rheumatology, Hospital Universitari i Politecnic La Fe, Valencia. Juan José Alegre Sancho, Department of Rheumatology, Hospital del Doctor Peset, Valencia. Francisco J. Blanco García and Natividad Oreiro, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña.

Contributors All of the coauthors have made substantial contributions to the conception or design of the work, or the acquisition, analysis or interpretation of data. They have revised the work critically for important intellectual content. They have provided a final approval of the version published. They are in agreement with being accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding This work was supported by the following grants: JM was funded by GEN-FER from the Spanish Society of Rheumatology, SAF2009–11110 and SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness, and BIO-1395 from Junta de Andalucía. NO-C was funded by PI-0590-2010, from Consejería de Salud y Bienestar Social, Junta de Andalucía, Spain. EL-I was supported by Ministerio de Educación, Cultura y Deporte through the programme FPU. DC was supported by Spanish Ministry of Economy and Competitiveness through the programme FPI. LB-C was funded by the EU/EFPIA Innovative Medicines Initiative Joint Undertaking PRECISESADS (ref: 115565). TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). Study on USA samples were supported by the Institutes of Health (NIH) National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), Centers of Research Translation (CORT) grant P50AR054144 (MDM), the NIH-NIAMS SSc Family Registry and DNA Repository (N01-AR-0-2251) (MDM), NIH-KL2RR024149-04 (SA), NIH-NCRR 3UL1RR024148, US NIH NIAID UO1 1U01AI09090, K23AR061436 (SA), Department of Defense PR1206877 (MDM) and NIH/NIAMS-RO1- AR055258 (MDM).

Competing interests None declared.

Ethics approval The local ethics committees.

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 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.
3 Mayes MD, Bossini-Castillo L, Gorlova O, *et al*. Immunochip analysis identifies multiple susceptibility Loci for systemic sclerosis. *Am J Hum Genet* 2014;94:47–61.
4 Martin JE, Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. *Hum Genet* 2012;131:1023–37.
5 Assassi S, Radstake TR, Mayes MD, *et al*. Genetics of scleroderma: implications for personalized medicine? *BMC Med* 2013;11:9.
6 Diogo D, Bastarache L, Liao KP, *et al*. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. *PLoS ONE* 2015;10:e0122271.
7 Parkes M, Cortes A, van Heel DA, *et al*. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat Rev Genet* 2013;14:661–73.
8 Lopez-Isac E, Bossini-Castillo L, Guerra SG, *et al*. Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus. *Arthritis Rheumatol* 2014;66:3521–3.
9 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.
10 Anon MC. 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;23:581–90.
11 LeRoy EC, Medsger TA, Jr. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001;28:1573–6.
12 LeRoy EC, Black C, Fleischmajer R, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;15:202–5.
13 Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet* 2009;5:e1000529.
14 Abecasis GR, Altshuler D, Auton A, *et al*. A map of human genome variation from population-scale sequencing. *Nature* 2010;467:1061–73.
15 Trynka G, Hunt KA, Bockett NA, *et al*. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* 2011;43:1193–201.
16 Eyre S, Bowes J, Diogo D, *et al*. High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* 2012;44:1336–40.
17 Cooper JD, Simmonds MJ, Walker NM, *et al*. Seven newly identified loci for autoimmune thyroid disease. *Hum Mol Genet* 2012;21:5202–8.
18 Liu JZ, Almarri MA, Gaffney DJ, *et al*. Dense fine-mapping study identifies new susceptibility loci for primary biliary cirrhosis. *Nat Genet* 2012;44:1137–41.
19 Tsoi LC, Spain SL, Knight J, *et al*. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. *Nat Genet* 2012;44:1341–8.
20 Hinks A, Cobb J, Marion MC, *et al*. Dense genotyping of immune-related disease regions identifies 14 new susceptibility loci for juvenile idiopathic arthritis. *Nat Genet* 2013;45:664–9.
21 Liu JZ, Hov JR, Folseraas T, *et al*. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013;45:670–5.
22 Cortes A, Hadler J, Pointon JP, *et al*. Identification of multiple risk variants for ankylosing spondylitis through high-density genotyping of immune-related loci. *Nat Genet* 2013;45:730–8.
23 Saruhan-Direskeneli G, Hughes T, Aksu K, *et al*. Identification of multiple genetic susceptibility loci in Takayasu arteritis. *Am J Hum Genet* 2013;93:298–305.
24 Purcell S, Neale B, Todd-Brown K, *et al*. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559–75.
25 Pruim RJ, Welch RP, Sanna S, *et al*. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010;26:2336–7.
26 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.
27 Cunninghame Graham DS, Morris DL, Bhangale TR, *et al*. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. *PLoS Genet* 2011;7:e1002341.
28 Hellquist A, Jarvinen TM, Koskenmies S, *et al*. Evidence for genetic association and interaction between the TYK2 and IRF5 genes in systemic lupus erythematosus. *J Rheumatol* 2009;36:1631–8.
29 Sigurdsson S, Nordmark G, Goring HH, *et al*. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. *Am J Hum Genet* 2005;76:528–37.
30 Adzhubei IA, Schmidt S, Peshkin L, *et al*. A method and server for predicting damaging missense mutations. *Nat Methods* 2010;7:248–9.
31 Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 2009;4:1073–81.
32 Couturier N, Bucciarelli F, Nurtdinov RN, *et al*. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. *Brain* 2011;134:693–703.
33 Li Z, Gakovic M, Ragimbeau J, *et al*. Two rare disease-associated Tyk2 variants are catalytically impaired but signaling competent. *J Immunol* 2013;190:2335–44.
34 Vignali DA, Kuchroo VK. IL-12 family cytokines: immunological playmakers. *Nat Immunol* 2012;13:722–8.
35 Sato S, Hanakawa H, Hasegawa M, *et al*. Levels of interleukin 12, a cytokine of type 1 helper T cells, are elevated in sera from patients with systemic sclerosis. *J Rheumatol* 2000;27:2838–42.
36 Leonardi CL, Kimball AB, Papp KA, *et al*. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1). *Lancet* 2008;371:1665–74.
37 Papp KA, Langley RG, Lebwohl M, *et al*. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 2). *Lancet* 2008;371:1675–84.
38 Griffiths CE, Strober BE, van de Kerkhof P, *et al*. Comparison of ustekinumab and etanercept for moderate-to-severe psoriasis. *N Engl J Med* 2010;362:118–28.
39 Garber K. Pfizer's first-in-class JAK inhibitor pricey for rheumatoid arthritis market. *Nat Biotechnol* 2013;31:3–4.
40 Liang Y, Zhu Y, Xia Y, *et al*. Therapeutic potential of tyrosine kinase 2 in autoimmunity. *Expert Opin Ther Targets* 2014;18:571–80.

Supplementary tables

Supp. Table 1. Overall statistical power of the study for each analyzed *TYK2* genetic variant accordingly with global disease.

	OR=1.2		OR=1.3		OR=1.4		OR=1.5	
Significance level	**0.05**	**5x10^{-8}**	**0.05**	**5x10^{-8}**	**0.05**	**5x10^{-8}**	**0.05**	**5x10^{-8}**
rs2304256	**1.00**	**0.98**	1.00	1.00	1.00	1.00	1.00	1.00
rs34536443	**0.76**	**0.01**	**0.97**	**0.08**	1.00	0.42	1.00	0.84
rs35018800	0.34	0.01	0.62	0.01	0.83	0.01	**0.94**	**0.03**
rs12720356	**0.99**	**0.24**	1.00	0.93	1.00	1.00	1.00	1.00

OR, odds ratio.

Supp. Table 2. Inverse variance meta-analysis of four *TYK2* SNPs in the validation cohort (4,985 SSc patients and 7,478 controls).

						Inverse variance test		
Chr	**SNP**	**Minor/ Major**	**Comment**	**MAF Cases**	**MAF Controls**	***P*-value**	**OR [CI 95%]***	**Q**
19	rs34536443 (P1104A)	C/G	missense Pro > Ala	0.026	0.029	0.019	0.82 [0.70-0.97]	0.06
19	rs35018800 (A928V)	A/G	missense Ala > Val	0.004	0.008	6.00E-03	0.57 [0.38-0.85]	0.19
19	rs12720356 (I684S)	C/A	missense Ile > Ser	0.068	0.078	9.54E-05	0.81 [0.73-0.90]	0.59
19	rs2304256 (V362F)	A/C	missense Val > Phe	0.250	0.283	2.26E-10	0.82 [0.77-0.87]	0.58

*Odds ratio for the minor allele.

Chr, chromosome; CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; Q heterogeneity value; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

Supp. Table 3. Allelic combination analysis of 4 *TYK2* SNPs in the overall combined cohort.

Chr	From bp	To bp	Allelic Combination	Frequency	*P*-value	OR
19	10463118	10475652	GGCA	0.074	1.24E-06	0.81
19	10463118	10475652	GAAA	0.006	1.45E-04	0.54
19	10463118	10475652	CGAA	0.024	6.35E-06	0.70
19	10463118	10475652	GGAA	0.162	0.014	0.93
19	10463118	10475652	GGAC	0.730	5.95E-12	1.19

*SNP order of the allelic combination: rs34536443, rs35018800, rs12720356 and rs2304256. Minor allele is highlighted in red.

Bp: base-pair ; Chr, chromosome; OR, odds ratio.

V

Publication 4: *IRF4* Newly Identified as a Common Susceptibility *Locus* for Systemic Sclerosis and Rheumatoid Arthritis in a Cross-Disease Meta-Analysis of Genome-Wide Association Studies

ARTHRITIS & RHEUMATOLOGY
Vol. 68, No. 9, September 2016, pp 2338–2344
DOI 10.1002/art.39730

BRIEF REPORT

IRF4 Newly Identified as a Common Susceptibility Locus for Systemic Sclerosis and Rheumatoid Arthritis in a Cross-Disease Meta-Analysis of Genome-Wide Association Studies

Elena López-Isac,[1] Jose-Ezequiel Martín,[1] Shervin Assassi,[2] Carmen P. Simeón,[3] Patricia Carreira,[4] Norberto Ortego-Centeno,[5] Mayka Freire,[6] Emma Beltrán,[7] Javier Narváez,[8] Juan J. Alegre-Sancho,[9] the Spanish Scleroderma Group, Benjamín Fernández-Gutiérrez,[10] Alejandro Balsa,[11] Ana M. Ortiz,[12] Miguel A. González-Gay,[13] Lorenzo Beretta,[14] Alessandro Santaniello,[14] Chiara Bellocchi,[14] Claudio Lunardi,[15] Gianluca Moroncini,[16] Armando Gabrielli,[16] Torsten Witte,[17] Nicolas Hunzelmann,[18] Jörg H. W. Distler,[19] Gabriella Riekemasten,[20] Annette H. van der Helm-van Mil,[21] Jeska de Vries-Bouwstra,[21] Cesar Magro-Checa,[21] Alexandre E. Voskuyl,[22] Madelon C. Vonk,[23] Øyvind Molberg,[24] Tony Merriman,[25] Roger Hesselstrand,[26] Annika Nordin,[27] Leonid Padyukov,[27] Ariane Herrick,[28] Steve Eyre,[28] Bobby P. C. Koeleman,[29] Christopher P. Denton,[30] Carmen Fonseca,[30] Timothy R. D. J. Radstake,[31] Jane Worthington,[28] Maureen D. Mayes,[2] and Javier Martín[1]

Objective. **Systemic sclerosis (SSc) and rheumatoid arthritis (RA) are autoimmune diseases that have similar clinical and immunologic characteristics. To date, several shared SSc–RA genetic loci have been identified independently. The aim of the current study was to systematically search for new common SSc–RA loci through an interdisease meta–genome-wide association (meta-GWAS) strategy.**

Methods. **The study was designed as a meta-analysis combining GWAS data sets of patients with SSc and patients with RA, using a strategy that allowed identification of loci with both same-direction and opposite-direction allelic effects. The top single-nucleotide**

Supported by a grant from the Ministerio de Educación, Cultura y Deporte through the program FPU (to Dr. López-Isac), grant 115565 from the EU/EFPIA Innovative Medicines Initiative Joint Undertaking PRECISESADS (ref. no. 115565) and BIO-1395 from the Junta de Andalucía, grant PI-0590-2010 from the Consejería de Salud y Bienestar Social, Junta de Andalucía, Spain (to Dr. Ortego-Centeno), a VIDI laureate from the Dutch Association of Research and Dutch Arthritis Foundation (to Dr. Radstake), and grant SAF2012-34435 from the Spanish Ministry of Economy and Competitiveness (to Dr. J. Martín). Dr. Assassi's work was supported by grants KL2-RR-024149-04 and K23-AR-061436 from the NIH, grant 3-UL1-RR-024148 from the NIH National Center for Research Resources, and grant U01-1U01AI09090 from the NIH National Institute of Allergy and Infectious Diseases. Dr. Mayes' work was supported by grant P50-AR-054144 from the NIH National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) Centers of Research Translation, grant N01-AR-0-2251 from the NIAMS SSc Family Registry and DNA Repository, grant PR-1206877 from the Department of Defense, and grant R01-AR-055258 from the NIAMS.

[1]Elena López-Isac, MSc, Jose-Ezequiel Martín, PhD, Javier Martín, MD, PhD: Institute of Parasitology and Biomedicine López-Neyra CSIC, Granada, Spain; [2]Shervin Assassi, MD, Maureen D. Mayes, MD, MPH: University of Texas Health Science Center, Houston; [3]Carmen P. Simeón, MD, PhD: Valle de Hebrón Hospital, Barcelona, Spain; [4]Patricia Carreira, MD: 12 de Octubre University Hospital, Madrid, Spain; [5]Norberto Ortego-Centeno, MD: Clinic University Hospital, Granada, Spain; [6]Mayka Freire, MD, PhD: Complexo Hospitalario Universitario de Vigo, Vigo, Spain; [7]Emma Beltrán, MD: Hospital General Universitario de Valencia, Valencia, Spain; [8]Javier Narváez, MD, PhD: Hospital Universitari de Bellvitge, Barcelona, Spain; [9]Juan J. Alegre-Sancho, MD, PhD: Hospital Universitari Doctor Peset, Valencia, Spain; [10]Benjamín Fernández-Gutiérrez, MD, PhD: Hospital Clínico San Carlos, Madrid, Spain; [11]Alejandro Balsa, MD, PhD: Hospital Universitario La Paz, Instituto de Investigación Sanitaria La Paz, Madrid, Spain; [12]Ana M. Ortiz, MD, PhD: Hospital Universitario La Princesa, Instituto de Investigación Sanitaria La Princesa, Madrid, Spain; [13]Miguel A. González-Gay, MD, PhD: University of Cantabria, Santander, Spain; [14]Lorenzo Beretta, MD, Alessandro Santaniello, MD, Chiara Bellocchi, MD: Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy; [15]Claudio Lunardi, MD: Università degli Studi di Verona, Verona, Italy; [16]Gianluca Moroncini, MD, PhD, Armando Gabrielli, MD, PhD: Università Politecnica delle Marche and Ospedali Riuniti, Ancona, Italy; [17]Torsten Witte, MD, PhD: Hannover Medical School, Hannover, Germany; [18]Nicolas Hunzelmann, MD: University of Cologne, Cologne, Germany; [19]Jörg H. W. Distler, MD: University of Erlangen–Nuremberg, Erlangen, Germany; [20]Gabriella Riekemasten, MD, PhD: University of Lübeck, Lübeck, Germany; [21]Annette H. van der Helm-van Mil, MD, PhD, Jeska de Vries-Bouwstra, MD, PhD, Cesar Magro-Checa, MD: Leiden University Medical Center, Leiden, The Netherlands; [22]Alexandre E. Voskuyl, MD: VU University Medical Center, Amsterdam, The Netherlands; [23]Madelon C. Vonk, MD, PhD: Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands; [24]Øyvind Molberg, MD, PhD: Oslo University Hospital Rikshospitalet and Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [25]Tony Merriman, PhD: University of Otago, Otago, New Zealand; [26]Roger Hesselstrand, MD, PhD: Lund University,

polymorphisms were followed up in independent SSc and RA case–control cohorts. This allowed an increase in the sample size to a total of 8,830 patients with SSc, 16,870 patients with RA, and 43,393 healthy controls.

Results. **This cross-disease meta-analysis of the GWAS data sets identified several loci with nominal association signals ($P < 5 \times 10^{-6}$) that also showed evidence of association in the disease-specific GWAS scans. These loci included several genomic regions not previously reported as shared loci, as well as several risk factors that were previously found to be associated with both diseases. Follow-up analyses of the putatively new SSc–RA loci identified *IRF4* as a shared risk factor for these 2 diseases ($P_{combined} = 3.29 \times 10^{-12}$). Analysis of the biologic relevance of the known SSc–RA shared loci identified the type I interferon and interleukin-12 signaling pathways as the main common etiologic factors.**

Conclusion. **This study identified a novel shared locus, *IRF4*, for the risk of SSc and RA, and highlighted the usefulness of a cross-disease GWAS meta-analysis strategy in the identification of common risk loci.**

Genome-wide association studies (GWAS) and immune-focused fine-mapping studies have revolutionized our understanding of the genetic component of complex autoimmune diseases by facilitating the identification of thousands of susceptibility loci associated with autoimmunity (1). The vast majority of these loci are shared risk factors for at least 2 autoimmune diseases, pointing to a common genetic background underlying these autoimmune processes. This genetic overlap was suspected some time ago, given the high rate of co-occurrence of autoimmune diseases and the well-established familial aggregation reported for these immune disorders (1).

Systemic sclerosis (SSc) and rheumatoid arthritis (RA) are complex autoimmune diseases that have similar clinical and immunologic features. Both diseases are rheumatic connective tissue disorders, characterized by an exacerbated inflammatory response, deregulation of innate and adaptive immunity, including autoantibody production, and systemic complications. Because of the establishment of large consortiums and international collaborations, the number of confirmed RA susceptibility factors has increased up to a total of 101 loci associated with the disease at the genome-wide significance level (2). With regard to SSc, GWAS, Immunochip, and candidate gene studies have clearly identified various genetic regions involved in susceptibility to SSc (3). However, the knowledge of the genetic predisposition to this disease is relatively limited, in part due to its low prevalence, which impairs the recruitment of large cohorts required to reach a high statistical power and to effectively detect association signals. Interestingly, a considerable proportion of the SSc susceptibility factors also represent RA risk loci (2,3). In addition, although not very common, co-familiarity and co-occurrence between these 2 rheumatic conditions have been observed (4). These observations provide evidence of a genetic overlap of both diseases. Thus, it is expected that additional shared risk factors remain to be discovered.

One approach that has been developed for the identification of common loci in a cost-effective manner is to perform a combined-phenotype GWAS, that is, to combine genome-wide genotype data for 2 autoimmune diseases. This strategy has been successfully applied to the study of not only closely related phenotypes but also nonrelated phenotypes, and thus far the results have been encouraging (5).

Taking into account all of these considerations, the purpose of the present study was to systematically identify new common risk loci for SSc and RA by applying the combined-phenotype GWAS strategy, followed by replication testing in independent case–control data sets.

PATIENTS AND METHODS

Study population. The first stage of the present study, the discovery phase, included 6,537 patients with either SSc or RA and 8,741 healthy controls. The SSc GWAS panel comprised 4 case–control sets from Spain, Germany, The Netherlands, and the US (2,716 cases and 5,666 controls), whose data had been obtained in previous studies (5–7). The RA case–control GWAS panel included 2 previously published RA GWAS cohorts (the Wellcome Trust Case Control Consortium [WTCCC] and the Epidemiological Investigation of Rheumatoid Arthritis study cohort) from the UK and Sweden (3,821 cases and 3,075 controls) (8).

Subjects included in the second stage of the study, the replication phase, were drawn from independent SSc and RA case–control sets of individuals European ancestry. The SSc

Lund, Sweden; [27]Annika Nordin, MD, PhD, Leonid Padyukov, PhD: Karolinska University Hospital, Karolinska Institutet, Stockholm, Sweden; [28]Ariane Herrick, MD, PhD, Steve Eyre, MD, PhD, Jane Worthington, MD, PhD: University of Manchester, Manchester Academic Health Science Centre, Manchester, UK; [29]Bobby P. C. Koeleman, PhD: University Medical Center Utrecht, Utrecht, The Netherlands; [30]Christopher P. Denton, PhD, FRCP, Carmen Fonseca, MD, PhD: Centre for Rheumatology, Royal Free and University College Medical School, London, UK; [31]Timothy R. D. J. Radstake, MD, PhD: University Medical Center Utrecht, Utrecht, The Netherlands. Members of the Spanish Scleroderma Group are shown in Appendix A.

Address correspondence to Elena López-Isac, MSc, or to Javier Martín, MD, PhD, IPBLN-CSIC, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18016-Armilla, Granada, Spain. E-mail: elenalopezisac@ipb.csic.es or martin@ipb.csic.es.

Submitted for publication March 30, 2016; accepted in revised form April 19, 2016.

replication cohort included 6,114 cases and 8,744 healthy controls from 8 different countries (Spain, Germany, Italy, the UK, The Netherlands, Sweden, Norway, and the US). The healthy controls from the UK and US partially overlapped with the control sets of previously published cohorts (the WTCCC and the second North American Rheumatoid Arthritis Consortium [NARAC2]) (8). The RA replication cohort included 9 case–control collections from North America (US and Canada), Spain, The Netherlands, the UK, Sweden, France, and New Zealand, and comprised a total of 13,049 RA cases and 25,908 healthy controls. Of these, 9,711 cases and 24,253 healthy controls were obtained from several previously published studies, including the Brigham Rheumatoid Arthritis Sequential Study, NARAC1, CANADA, studies from the Rheumatoid Arthritis Consortium International (RACI-US, RACI-i2b2, RACI-UK, RACI-SE-U, and RACI-NL), Consortium of Rheumatology Researchers of North America, Vanderbilt, Dutch studies (Amsterdam Medical Center, Treatment Strategies for RA [BeSt Study], Leiden University Medical Center, and Dutch Rheumatoid Arthritis Monitoring Registry), Research in Active Rheumatoid Arthritis, and the Anti-TNF Response to Therapy collection (ACR-REF: BRAGGSS, BRAGGSS2, ERA, KI, and TEAR) (2). All of the patients with SSc and patients with RA fulfilled previously described classification criteria for each disease (2,5). All individuals enrolled in the present study provided written informed consent, and approval from the local ethics committees was obtained from all of the centers in accordance with the tenets of the Declaration of Helsinki.

Study design. We performed a 2-stage study to systematically identify SSc–RA shared risk factors, with the first stage being the discovery phase for GWAS meta-analysis of each disease separately and a combined-phenotype GWAS meta-analysis, and the second stage being the replication phase (Figure 1).

Discovery phase. We performed GWAS analysis for each disease separately and in a combined-phenotype GWAS analysis. Two different tests were considered for the combined analysis (5). In the first test, in order to detect common signals for SSc and RA with same-direction allelic effects, the meta-analysis considering both diseases was performed as usual. Those SNPs that showed an association at $P < 5 \times 10^{-6}$ in the combined-phenotype analysis (referred to as $P_{combined}$) and also showed nominal significance in the association study for each disease ($P < 0.05$) were selected for follow-up in the replication phase.

In the second test, in order to identify common signals with opposite-direction allelic effects, we flipped the direction of association (1/odds ratio [OR]) in the RA data set for the combined-phenotype meta-analysis. To select SNPs for replication, the same selection criteria as stated above were followed.

For both sorts of meta-analyses, we only considered for follow-up those SNPs that had not been previously reported as genetic risk factors for SSc and RA, or those that had been reported for one disease but not reported for the other.

Replication phase. The SNPs selected were followed-up in independent replication cohorts. Subsequently, we performed a meta-analysis of the initial GWAS screening and replication stages. The SNP signals that 1) reached the genome-wide significance level for association ($P_{combined} < 5 \times 10^{-8}$) in the combined-phenotype meta-analysis (GWAS + Replication phases), and that 2) showed, for each disease separately, nominally significant associations ($P < 0.05$) in the replication phase as well as significant associations ($P < 5 \times 10^{-3}$) in the GWAS + Replication meta-analysis were considered shared risk factors for the 2 analyzed diseases.

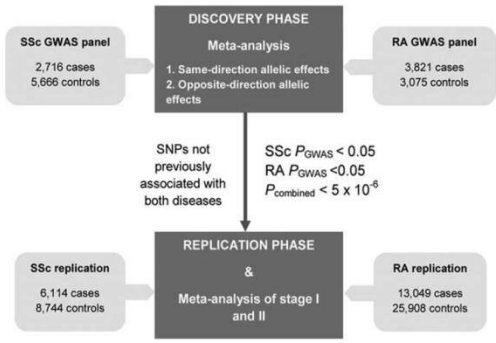


Figure 1. Overall work flow of the present study. SSc = systemic sclerosis; GWAS = genome-wide association study; RA = rheumatoid arthritis; SNPs = single-nucleotide polymorphisms.

Quality control and genotype imputation of GWAS data. We applied stringent quality control criteria in all of the GWAS data sets. Cutoff values for the sample call rate and the SNP call rate were set as 95%. Markers with allele distributions deviating from Hardy-Weinberg equilibrium (HWE) ($P < 0.001$) in controls from any of the populations analyzed separately were excluded. Markers with minor allele frequencies lower than 1% were filtered out. After quality control, we performed whole-genome genotype imputation with IMPUTE2 software (9) using as reference panels the CEU (Utah residents with northern and western European ancestry from the CEPH collection) and TSI (Toscani in Italy) populations of the HapMap Phase 3 project (available at http://www.hapmap.org). Imputed SNP quality was assessed by establishing a probability threshold for merging genotypes at 0.9. Subsequently, stringent quality control was applied to the imputed data using the same criteria as stated above. Thereafter, genome-wide genotyping data were available for a total of 219,756 SNPs.

The first 5 principal components were estimated, and individuals deviating more than 6 SDs from the cluster centroids were considered outliers. In addition, duplicate pairs or highly related individuals among data sets were also removed on the basis of pairwise comparisons, using the Genome function in Plink version 1.7 (see http://pngu.mgh.harvard.edu/purcell/plink/) (Pi-HAT threshold of 0.5).

Follow-up genotyping. The genotyping of the replication cohorts was performed with either TaqMan SNP genotyping technology in a LightCycler 480 Real-Time polymerase chain reaction system (Roche Applied Science) or the GWAS and Immunochip platforms.

For the SSc study, all cases were genotyped using the TaqMan genotyping system, with TaqMan 5′ allele discrimination predesigned assays from Applied Biosystems. The genotyping call rate was >95% for the 3 SNPs. The control samples were also genotyped using this technology, with the exception of the UK and US cohorts. For these 2 control

cohorts, genotyping data were obtained from previously published genome-wide genotyping data sets (from the WTCCC and NARAC2) (8).

RA cases from Spain and New Zealand and the Spanish controls were genotyped by TaqMan technology. Genotype data for the New Zealand healthy controls partially overlapped with those from a previous GWAS report (10). For the remaining RA case–control sets, genotype frequencies and association data were obtained from a previously published study (2). The genotype methods used in these studies were described in detail in the study by Okada et al (2). For those cohorts in which genotyping was performed using the Illumina Immunochip platform, only data for the rs9328192 SNP of the interferon regulatory factor 4 (IRF-4) gene (*IRF4*) were available.

Statistical analysis. All data were analyzed using Plink software. To test for association, we performed logistic regression analysis in each of the SSc and RA GWAS cohorts separately. The first 5 principal components were included as covariates to control for any potential population stratification effects. The replication cohorts were also analyzed by logistic regression analysis. The meta-analyses were performed with the inverse-variance method based on population-specific logistic regression results. Heterogeneity of the ORs across studies was assessed using Cochran's Q test. HWE was tested in all of the validation cohorts genotyped by TaqMan technology (in HWE analyses, $P < 0.01$ was considered to show significant deviation from equilibrium). None of the included control cohorts showed significant deviation from HWE, with the exception of *HNF1A* rs10774577. The cohorts in which HWE was not observed were excluded from the analysis of this specific SNP. The statistical power of the combined-phenotype analysis and the analysis for each disease separately is shown in Supplementary Table 1 (available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39730/abstract).

RESULTS

Discovery analysis. In the first phase of this study, we conducted a cross-disease meta-analysis in order to systematically identify new putatively shared loci between SSc and RA. The overall workflow of the study is illustrated in Figure 1.

The meta-analysis combining both data sets identified various SNPs from 7 distinct genomic regions that showed a significant association at the level of $P < 5 \times 10^{-6}$, as well as a nominal signal of association ($P < 0.05$) in the disease-specific analyses. The strongest association was found in the well-accepted SSc- and RA-associated locus *IRF5* ($P_{combined} = 8.44 \times 10^{-17}$; for SSc, $P_{GWAS} = 1.14 \times 10^{-16}$; for RA, $P_{GWAS} = 7.86 \times 10^{-4}$). Three additional known SSc–RA loci, namely *PTPN22*, *ATG5*, and *BLK*, were also identified at the initial discovery stage (Figure 2) (see also Supplementary Table 2 and Supplementary Figure 1, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39730/abstract). The remaining SNPs were located in 3 different loci, including *FBN2* and *HNF1A*, neither of which has been previously reported as a genetic risk factor for SSc and RA, and *IRF4*, which has been found to be associated with RA in previous studies (Table 1 and Figure 2) (see also Supplementary Figure 2 on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39730/abstract). Interestingly, the regional association plots of the *FBN2*, *IRF4*, and *HNF1A* loci showed that the top SNPs in the combined-phenotype analysis were also the top SNPs in the analyses for SSc and RA separately, or at least were in high linkage disequilibrium with the top signal observed for each disease (see Supplementary Figure 2).

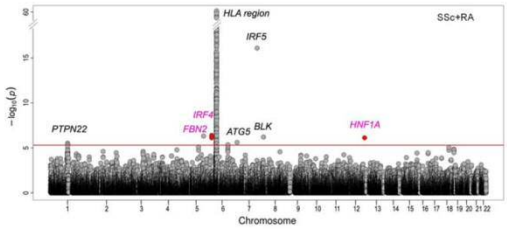


Figure 2. Manhattan plot showing the results of the cross-disease meta–genome-wide association study. The $-\log_{10}$ of the combined-phenotype meta-analysis P values are plotted against their physical chromosomal position. The plot displays the $-\log_{10}$ P values from the same-direction meta-analysis of associations with systemic sclerosis (SSc) and rheumatoid arthritis (RA). The signals from the opposite-direction meta-analysis that reached the selection criteria are also plotted (red points). The red line represents the threshold of significance at $P < 5 \times 10^{-6}$. Those loci with single-nucleotide polymorphisms that reached the selection criteria for the replication phase are plotted (loci selected for follow-up are highlighted in pink).

These new putatively shared SNPs were selected for follow-up in additional SSc and RA replication cohorts. For *IRF4*, 3 SNPs met our criteria for being selected for validation in the replication phase. In this case, we selected the SNP with the lowest P value for association (see Supplementary Table 2).

Replication phase and meta-analysis. According to the established thresholds discussed above in Patients and Methods, we identified 1 new association signal shared between SSc and RA at SNP rs9328192 of *IRF4* ($P_{combined} = 3.29 \times 10^{-12}$). Furthermore, this *IRF4* SNP almost reached genome-wide significance in the meta-analysis for each disease separately (for SSc, $P_{GWAS + Replication} = 2.78 \times 10^{-7}$, OR 0.90; for RA, $P_{GWAS + Replication} = 1.44 \times 10^{-6}$, OR 1.08) (Table 1).

Regarding the *HNF1A* and *FBN2* genetic variants, despite the initial suggestive association signals found in the first stage, these loci did not show genome-wide significance in our combined-phenotype meta-analysis.

Table 1. Results of the cross-disease meta-GWAS of disease associations with 3 selected SNPs*

Locus (Chr.)	SNP	Ref. allele	*P* for SSc–RA GWAS	SSc						RA						SSc–RA GWAS + Replication	
				GWAS		Replication		GWAS + Replication (meta-GWAS)		GWAS		Replication		GWAS + Replication (meta-GWAS)			
				P	OR†	*P*	OR†	*P*	OR†	*P*	OR†	*P*	OR†	*P*	OR†	*P*	Current status
FBN2 (5)	rs6897611	T	4.79×10^{-7}	2.85×10^{-3}	1.16	0.641	0.98	0.165	1.04	3.15×10^{-5}	1.24	0.684	0.99	0.650‡	1.02‡	0.018	–
IRF4 (6)	rs9328192	G	4.06×10^{-7}	8.86×10^{-6}	0.86	1.89×10^{-3}	0.93	2.78×10^{-7}	0.90	7.26×10^{-3}	1.10	5.22×10^{-5}	1.07	1.44×10^{-6}	1.08	3.29×10^{-12}	SSc–RA; RA
HNF1A (12)	rs10774577§	T	7.53×10^{-7}	8.62×10^{-4}	0.89	0.036	0.94	1.64×10^{-4}	0.91	2.50×10^{-4}	1.14	0.290	1.03	0.208‡	1.05‡	1.59×10^{-6}	–

* GWAS = genome-wide association study; SNPs = single-nucleotide polymorphisms; Chr. = chromosome.
† Represents the odds ratio (OR) for the reference (Ref.) allele.
‡ *P* values and ORs were derived from meta-analysis under random effects due to heterogeneity of the ORs among cohorts.
§ The rheumatoid arthritis (RA) and systemic sclerosis (SSc) replication cohorts from Spain, and the SSc replication cohorts from Italy and The Netherlands were excluded from the analysis of rs10774577 due to issues with Hardy-Weinberg equilibrium.

Nevertheless, *HNF1A* rs10774577 showed suggestive evidence of association in the meta-analysis performed in the SSc data set ($P_{Replication} = 0.036$, OR 0.94; $P_{GWAS + Replication} = 1.64 \times 10^{-4}$, OR 0.91), and showed an association at the level of $P = 1.59 \times 10^{-6}$ in the combined-phenotype meta-analysis. Considering that this SNP was not included in those cohorts that were genotyped with Immunochip, the present study had a lower statistical power for the analysis of this genomic region. Therefore, the possibility of a slight or modest genetic effect of *HNF1A* rs10774577 on both diseases cannot be ruled out, and further studies will be required to establish whether this locus is a shared SSc–RA risk factor.

DISCUSSION

In the present study, we identified a novel non-HLA susceptibility locus that is shared between SSc and RA, namely *IRF4*, using a combined-phenotype GWAS strategy in large case–control cohorts of patients with SSc and those with RA. This locus, *IRF4*, was already reported to be involved in RA susceptibility, but had not been previously associated with SSc (2).

The cross-disease meta-analysis performed with the SSc and RA GWAS data sets identified various SNPs from 7 different loci that met our stringent selection criteria for the replication phase ($P_{combined} < 5 \times 10^{-6}$; for SSc and for RA, each $P_{GWAS} < 0.05$). Four of the 7 SNPs were already known risk factors for SSc and RA (*PTPN22*, *ATG5*, *IRF5*, and *BLK*), thus providing support for the effectiveness of this strategy in the identification of shared risk loci (2,3). It is worth mentioning that these loci were detected by the 2 different tests used in the first phase, which were performed in order to detect both same-direction and opposite-direction allelic effects. In fact, the shared *IRF4* SNP newly identified in this study showed opposite effects for SSc and for RA (protective effect and risk effect, respectively). This discrepancy might be attributable to the fact that the actual causal variants for the associations in each disease could be different, and that *IRF4* rs9328192 is tagging them. This discordant phenomenon is particularly common between autoimmune diseases (1). However, to completely understand these discordant effects, the interaction with other genetic variants contributing to disease susceptibility should be considered, in addition to analyzing the precise biologic impact of the associations.

The associated *IRF4* SNP (rs9328192) showed modest effect sizes for SSc and RA. However, we were able to capture this association in our meta-analysis because of the large cohort used in this study, together with the combined-phenotype approach, which allowed us to increase the statistical power. This highlights the capability of the combined-phenotype approach in the identification of shared variants with low penetrance, whose associations might have been missed in disease-specific GWAS due to a lack of power (11).

IRF-4 belongs to the IRF family of transcription factors and plays a pivotal role in the development and function of several autoimmune-associated cells (12). Various genetic and functional studies have pointed to IRF-4 as a master regulator for autoimmunity (12,13). It has been demonstrated that IRF-4 is a crucial factor for the editing and L-chain rearrangements of the B cell receptor, and for pre–B cell expansion, both of which are processes directly related to the development of autoimmunity (14). In addition, IRF-4 is a critical controller of Th17 cell differentiation and the production of interleukin-17 (IL-17) and IL-21 (12), which are components of the immune system that play a key role in the pathogenesis of SSc and RA.

The results of the present study add another *IRF* to the list of *IRF*s associated with SSc (*IRF4*, *IRF5*, *IRF7*, and *IRF8*) and RA (*IRF4*, *IRF5*, and *IRF8*) (2,3), thus providing genetic support for the type I interferon (IFN) signature described in patients with SSc and those with RA (15). Moreover, our pathway enrichment analysis also identified the type I IFN signaling pathway as one of the most relevant common pathways between SSc and RA on the basis of their common genetic background (see Supplementary Methods, Supplementary Table 3, and Supplementary Figure 3, available on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.39730/abstract). Therefore, deregulation of this signaling pathway might be a biologic process that underlies the onset of these 2 autoimmune rheumatic conditions.

In summary, through a cross-disease meta-analysis of GWAS for SSc and RA, we were able to identify *IRF4* as a new shared susceptibility locus for these 2 autoimmune diseases. The results of the present study, taken together with the findings from previous studies, reinforce the idea of a common genetic background between SSc and RA. The identification of these pleiotropic autoimmunity loci may point to common pathogenic pathways, which ultimately may represent a clinical advantage in that it may provide support for drug repositioning on the basis of the true understanding of the pathogenic mechanisms of SSc and RA.

ACKNOWLEDGMENTS

We thank Sofia Vargas, Sonia García, and Gema Robledo for their excellent technical assistance, and all of the

patients and control donors for their essential collaboration. We thank National DNA Bank Carlos III (University of Salamanca, Spain) for supplying part of the control DNA samples. We also would like to thank the following organizations: the European League Against Rheumatism Scleroderma Trials and Research Group, the German Network of Systemic Sclerosis, the US Scleroderma Foundation, and the US Raynauds & Scleroderma Association.

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Drs. López-Isac and J. Martín had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. Fonseca, Radstake, Worthington, Mayes, J. Martín.

Acquisition of data. Assassi, Simeón, Carreira, Ortego-Centeno, Freire, Beltrán, Narváez, Alegre-Sancho, Fernández-Gutiérrez, Balsa, Ortiz, González-Gay, Beretta, Santaniello, Bellocchi, Lunardi, Moroncini, Gabrielli, Witte, Hunzelmann, Distler, Riekemasten, van der Helm-van Mil, de Vries-Bouwstra, Magro-Checa, Voskuyl, Vonk, Molberg, Merriman, Hesselstrand, Nordin, Padyukov, Herrick, Eyre, Koeleman, Denton, Fonseca, Radstake, Worthington, Mayes, J. Martín.

Analysis and interpretation of data. López-Isac, J.-E. Martín.

REFERENCES

1. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet 2013;14:661–73.
2. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376–81.
3. Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. J Autoimmun 2015;64:53–65.
4. Elhai M, Avouac J, Kahan A, Allanore Y. Systemic sclerosis at the crossroad of polyautoimmunity. Autoimmun Rev 2013;12: 1052–7.
5. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum Mol Genet 2013;22:4021–9.
6. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet 2010;42:426–9.
7. Martin JE, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, et al. Identification of CSK as a systemic sclerosis genetic risk factor through genome wide association study follow-up. Hum Mol Genet 2012;21:2825–35.
8. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, Thomson BP, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 2010;42:508–14.
9. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 2009;5:e1000529.
10. Consortium TAaNZMSG. Genome-wide association study identifies new multiple sclerosis susceptibility loci on chromosomes 12 and 20. Nat Genet 2009;41:824–8.
11. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet 2011;7:e1001283.
12. Xu WD, Pan HF, Ye DQ, Xu Y. Targeting IRF4 in autoimmune diseases. Autoimmun Rev 2012;11:918–24.
13. Biswas PS, Gupta S, Stirzaker RA, Kumar V, Jessberger R, Lu TT, et al. Dual regulation of IRF4 function in T and B cells is required for the coordination of T-B cell interactions and the prevention of autoimmunity. J Exp Med 2012;209:581–96.
14. Zouali M. Receptor editing and receptor revision in rheumatic autoimmune diseases. Trends Immunol 2008;29:103–9.
15. Ronnblom L, Eloranta ML. The interferon signature in autoimmune diseases. Curr Opin Rheumatol 2013;25:248–53.

APPENDIX A: MEMBERS OF THE SPANISH SCLERODERMA GROUP

Members of the Spanish Scleroderma Group include the following: Raquel Ríos and Jose Luis Callejas (Hospital Clínico Universitario San Cecilio, Granada), José Antonio Vargas Hitos (Hospital Virgen de las Nieves, Granada), Rosa García Portales (Hospital Virgen de la Victoria, Málaga), María Teresa Camps (Hospital Carlos Haya, Málaga), Antonio Fernández-Nebro (Hospital Carlos Haya, Málaga), María F. González-Escribano (Hospital Virgen del Rocío, Seville), Francisco José García-Hernández and Mª Jesús Castillo (Hospital Virgen del Rocío, Seville), Mª Ángeles Aguirre and Inmaculada Gómez-Gracia (Hospital Reina Sofía/IMIBIC, Córdoba), Luis Rodríguez-Rodríguez (Hospital Clínico San Carlos, Madrid), Paloma García de la Peña (Madrid Norte Sanchinarro Hospital, Madrid), Esther Vicente (Hospital La Princesa, Madrid), José Luis Andreu and Mónica Fernández de Castro (Hospital Puerta de Hierro Majadahonda, Madrid), Francisco Javier López-Longo and Lina Martínez (Hospital General Universitario Gregorio Marañón, Madrid), Vicente Fonollosa and Alfredo Guillén (Hospital Valle de Hebrón, Barcelona), Iván Castellví (Santa Creu i Sant Pau University Hospital, Barcelona), Gerard Espinosa (Hospital Clinic, Barcelona), Carlos Tolosa (Hospital Parc Tauli, Sabadell), Anna Pros (Hospital Del Mar, Barcelona), Mónica Rodríguez Carballeira (Hospital Universitari Mútua Terrasa, Barcelona), Francisco Javier Narváez (Hospital Universitari de Bellvitge, Barcelona), Manel Rubio Rivas (Hospital Universitari de Bellvitge, Barcelona), Vera Ortiz-Santamaría (Hospital General de Granollers, Granollers), Ana Belén Madroñero (Hospital General San Jorge, Huesca), Bernardino Díaz and Luis Trapiella (Hospital Central de Asturias, Oviedo), Adrián Sousa (Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo), María Victoria Egurbide (Hospital Universitario Cruces, Barakaldo), Patricia Fanlo Mateo (Hospital Virgen del Camino, Pamplona), Luis Sáez-Comet (Hospital Universitario Miguel Servet, Zaragoza), Federico Díaz and Vanesa Hernández (Hospital Universitario de Canarias, Tenerife), Emma Beltrán (Hospital General Universitario de Valencia, Valencia), José Andrés Román-Ivorra and Elena Grau (Hospital Universitari i Politecnic La Fe, Valencia), Juan José Alegre-Sancho (Hospital Universitari Doctor Peset, Valencia), Francisco J. Blanco García and Natividad Oreiro (INIBIC-Hospital Universitario A Coruña, La Coruña).

SUPPLEMENTARY MATERIAL

SUPPLEMENTARY METHODS

Biological connection across SSc-RA shared loci

In order to gain insight into the common etiopathogenic factors that underlie SSc and RA, we performed functional protein association analyses considering the shared risk factors described to date between SSc and RA. For this purpose, we only included firmly associated loci for both diseases, which were selected on the basis of the following criteria: (1) loci associated at the genome-wide significance level (*P*-value 5 x 10^{-8}), (2) those that reached second tier level associations (*p*-value < $5x10^{-5}$), or (3) that have been replicated in independent studies. In total, 14 well-established SSc-RA loci were included (Supp. Figure 3).

Protein-protein interaction (PPI) analysis. PPIs among the 14 SSc-RA loci were interrogated using STRING V.10 that provides a critical integration of PPIs, including direct (physical) as well as indirect (functional) associations.(1) A confidence score of 0,400 was applied.

Molecular pathway enrichment analysis. We conducted molecular pathway enrichment analyses using the Gene Set Enrichment Analysis (GSEA) and DAVID approaches (2-4). These tools evaluate the overlap of a specific gen set with gen sets from the MSigDB collections. The statistical significance of the overrepresentation of functional annotation terms is calculated on the basis of a hypergeometric testing. Two MSigDB collections (Biocarta and Reactome) were used. A False Discovery Rate (FDR) correction was applied for the results from GSEA-based results, and Bonferroni correction was applied for DAVID-based results.

LEGENDS OF SUPPLEMENTARY FIGURES

Supp. Figure 1. Manhattan plots summarizing the results of the genome-wide association studies of systemic sclerosis (SSc) and rheumatoid arthritis (RA).

Supp. Figure 2. Regional association plots of the three loci selected for replication. The selected SNPs are represented as a purple diamond.

A) *FBN2* associations in the combined-phenotype meta-analysis (same-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).

B) *IRF4* associations in the combined-phenotype meta-analysis (opposite-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).

C) *HNF1A* associations in the combined-phenotype meta-analysis (opposite-direction meta-analysis) (left), SSc genome-wide association study (middle), and RA genome-wide association study (right).

Supp. Figure 3. Protein protein interaction (PPI) network across the 14 well-established SSc-RA *loci* in STRING. The network was significantly enriched in interactions (P-value $< 5 \times 10^{-10}$). Thus the proteins encoded by the SSc-RA risk loci interact with each other more than expected by chance, suggesting common altered pathways in SSc and RA. The plot shows the 'confidence' view. Thicker lines represent stronger associations.

SUPPLEMENTARY TABLES

Supplementary Table 1. Overall statistical power of the study in both the combined analysis and the analysis for each disease separately.

		OR 1.2				OR 1.3			
Study	**Cases/Controls**	**MAF 0.40**	**MAF 0.30**	**MAF 0.20**	**MAF 0.10**	**MAF 0.40**	**MAF 0.30**	**MAF 0.20**	**MAF 0.10**
SSc+RA	23,240/45,853	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
SSc	8,830/14,410	1.00	0.99	0.93	0.42	1.00	1.00	1.00	1.00
RA	16,870/28,983	1.00	1.00	0.99	0.50	1.00	1.00	1.00	0.99

MAF, minor allele frequency; OR, odds ratio. Statistical power calculated according to Power Calculator for Genetic Studies 2006 software under an additive model (5).

Supplementary Table 2. List of SNPs with a combined meta-analysis *P* value lower than $5x10^{-6}$, SSc $P < 0.05$ and RA $P < 0.05$ in the discovery phase.

Chr	Bp	*Locus*	SNP	Ref. Allele	$P_{combined}$	OR[+]	*P* SSc	OR SSc[+]	*P* RA	OR RA[+]	Previous Association
1	114230863	PTPN22 / C1orf178	rs1746860	C	3.06E-06	0.88	1.79E-03	0.89	4.94E-04	0.88	SSc, RA
1	114240474	PTPN22 /AP4B1	rs1217401	G	3.90E-06	0.89	2.13E-03	0.89	5.31E-04	0.88	SSc, RA
5	**128011704**	**FBN2**	**rs6897611**	**T**	**4.79E-07**	**1.20**	2.85E-03	1.16	3.15E-05	1.24	-
6	106833908	*ATG5*	rs3804333	T	4.36E-06	1.15	3.48E-04	1.16	3.66E-03	1.13	SSc, RA
7	128404702	TNPO3/IRF5	rs12531711	G	8.44E-17	1.36	1.14E-16	1.52	7.86E-04	1.20	SSc, RA
8	11381382	BLK	rs2736340	T	6.29E-07	1.15	1.79E-06	1.20	2.51E-02	1.09	SSc, RA
6	378962	IRF4	rs2048698	T	5.83E-07	0.89	1.48E-05	0.86	6.77E-03	1.10	RA
6	**379364**	**IRF4**	**rs9328192**	**G**	**4.06E-07**	**0.88**	8.86E-06	0.86	7.26E-03	1.10	**RA**
6	379915	IRF4	rs4256472	A	7.68E-07	0.89	2.54E-05	0.87	5.91E-03	1.10	RA
12	**119848707**	**HNF1A**	**rs10774577**	**T**	**7.53E-07**	**0.88**	8.62E-04	0.89	2.50E-04	1.14	-

[+] Odds ratio for the reference allele.

Grey shading indicates the signals from the same-direction allelic effects meta-analysis.

Bp, base pair; Chr, chromosome; OR, odds ratio; RA, rheumatoid arthritis; SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

Supplementary Table 3. Gene sets that showed a significant enrichment *P*-value (*P*-value < 0.05) after False Discovery Rate correction in GSEA-based analysis or Bonferroni correction in DAVID-based analysis.

Database	Gene Set	K	k	k/K	*P*-value	FDR *P*-value
Reactome	Genes involved in Immune System	933	9	0.010	1.03E-12	9.19E-10
Reactome	Genes involved in Interferon alpha/beta signaling*	64	4	0.063	3.39E-09	1.51E-06
Biocarta	NO2-dependent IL 12 Pathway in NK cells*	17	3	0.177	1.53E-08	4.53E-06
Biocarta	IL12 and Stat4 Dependent Signaling Pathway in Th1 Development*	23	3	0.130	3.97E-08	8.85E-06
Reactome	Genes involved in Interferon Signaling*	159	4	0.025	1.34E-07	2.40E-05
Reactome	Genes involved in Interferon gamma signaling*	63	3	0.048	8.84E-07	1.31E-04
Reactome	Genes involved in Cytokine Signaling in Immune system	270	4	0.015	1.11E-06	1.42E-04
Biocarta	IL22 Soluble Receptor Signaling Pathway	16	2	0.125	1.03E-05	1.15E-03
Reactome	Genes involved in Negative regulators of RIG-I/MDA5 signaling	31	2	0.065	3.99E-05	3.95E-03
Reactome	Genes involved in Immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell	70	2	0.029	2.06E-04	1.81E-02
Reactome	Genes involved in RIG-I/MDA5 mediated induction of IFN-alpha/beta pathways	73	2	0.027	2.24E-04	1.81E-02
Reactome	Genes involved in Adaptive Immune System	539	3	0.006	5.30E-04	3.94E-02

*Significant in DAVID-based analysis
FDR *P*-value, false discovery rate corrected *P*-value; K, genes in Gene Set; k, genes in overlap

SUPPLEMENTARY REFERENCES

1. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:D447-52.

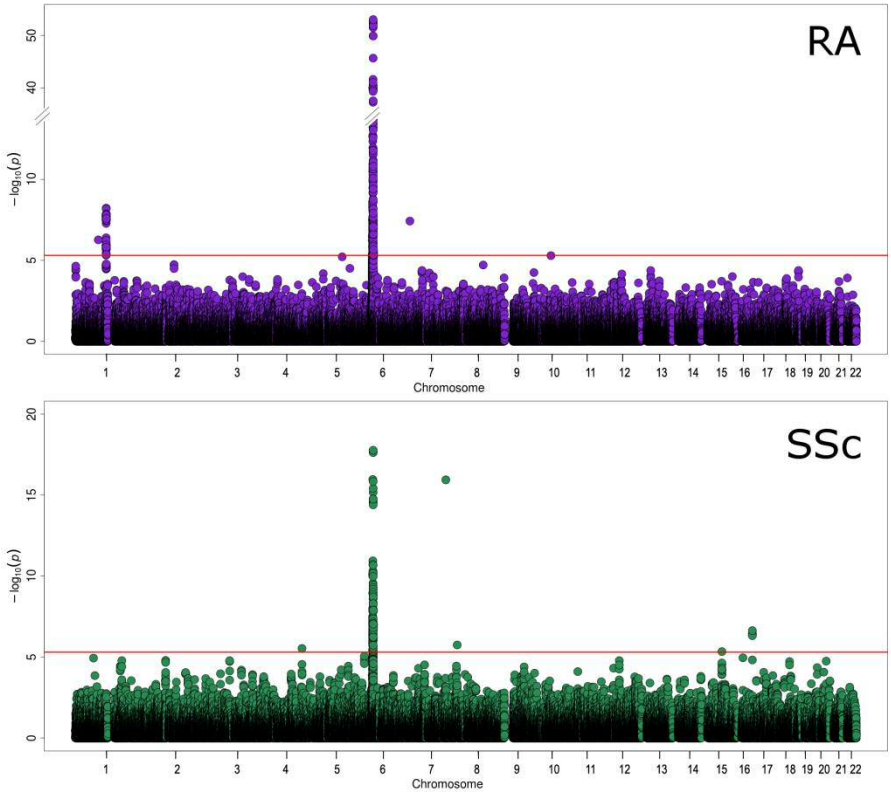
2. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545-50.

3. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44-57.

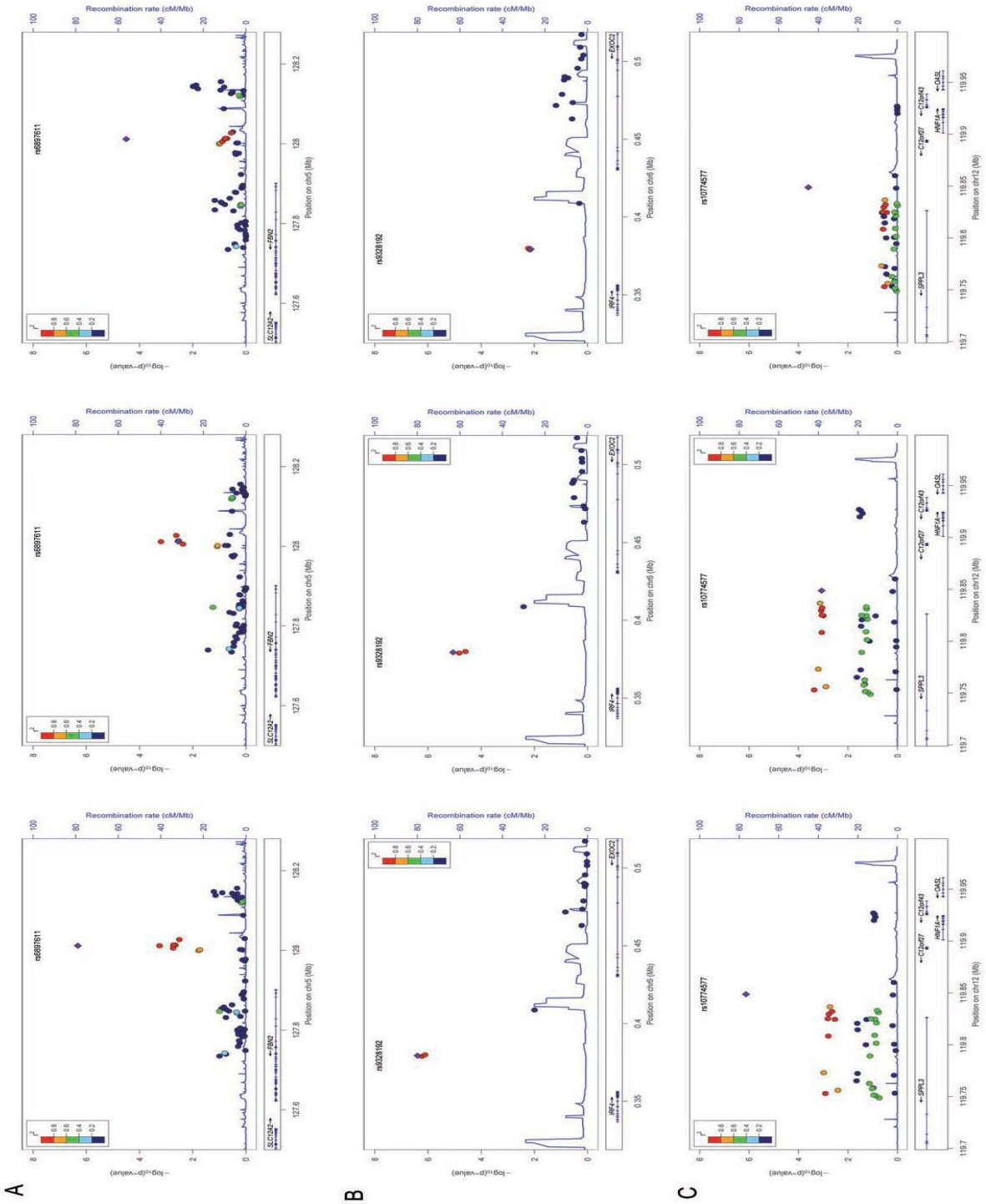
4. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009;37:1-13.

5. 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;38:209-13.

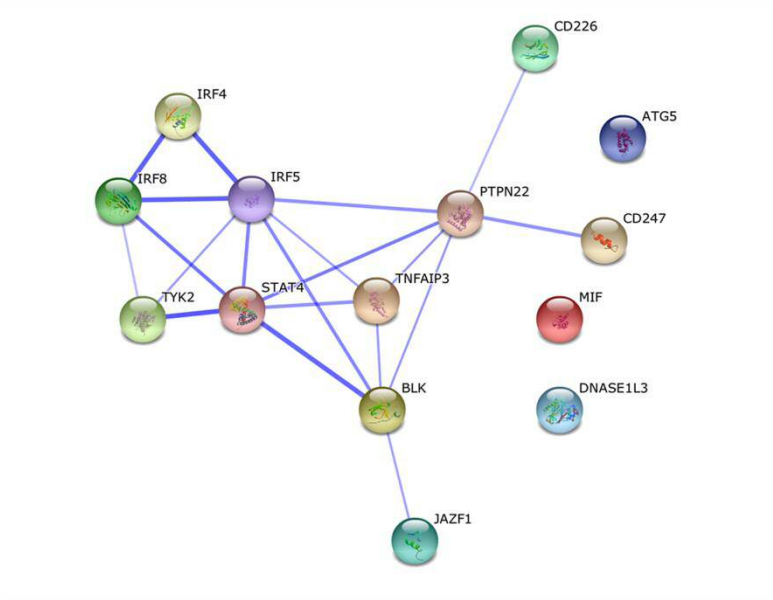
Supp. Figure 1



Supp. Figure 2



Supp. Figure 3



V

Publication 5: Analysis of *ATP8B4* F436L missense variant in a large Systemic sclerosis cohort

ARTHRITIS & RHEUMATOLOGY
Vol. 00, No. 00, Month 2017, pp 00–00

CONCISE COMMUNICATION

DOI 10.1002/art.40058

Analysis of *ATP8B4* F436L missense variant in a large systemic sclerosis cohort

Over the last 7 years, knowledge of the systemic sclerosis (SSc) genetic component has increased considerably, due mainly to large genetic studies including genome-wide association studies (GWAS) and Immunochip analysis. However, there is still a large portion of SSc heritability that remains unexplained, as is the case with most complex traits (1). One hypothesis that has been proposed to explain the missing heritability for complex diseases involves rare and low-frequency variants. These types of genetic variations are not well covered by GWAS, which are mainly focused on common variants. However, the use of next-generation sequencing technologies, such as whole-exome sequencing, has rapidly overcome this problem. In this regard, Gao et al performed, for the first time, whole-exome sequencing in SSc and reported a novel gene, *ATP8B4*, as a risk factor for the disease (2). They suggested a missense rare variant (F436L [rs55687265]) as a potential causal variant for the association signal in *ATP8B4*. We therefore aimed to further evaluate the reported signal of association, taking advantage of our access to large cohorts of patients with SSc.

The *ATP8B4* rare variant rs55687265 was genotyped in 6 independent case–control cohorts of European ancestry (total 7,426 SSc patients and 13,087 healthy controls) (see Supplementary Table 1, on the *Arthritis & Rheumatology* web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40058/abstract). All SSc patients fulfilled the American College of Rheumatology 1980 preliminary classification criteria for the disease (3) or exhibited at least 3 of 5 features of CREST syndrome (calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly, telangiectasias) (4). We first performed association analyses to test whether rs55687265 was associated with SSc susceptibility in each of the cohorts included in the present study (see Supplementary Methods, http://onlinelibrary.wiley.com/doi/10.1002/art.40058/abstract). A trend toward association ($P = 0.071$) was observed in the Spanish case–control set (odds ratio [OR] 1.58) (Table 1). However, we did not observe any suggestive or significant association signal in the remaining cohorts. We also observed opposite effects for the same allele in different populations. The meta-analysis combining all of the sample sets, which was performed using an inverse variance fixed-effects model, showed no significant association with the disease (OR 1.07, $P = 0.484$) (Table 1 and Supplementary Methods). In addition, stratified analysis based on different clinical sub-phenotypes of SSc (limited and diffuse cutaneous subtypes, and the presence of the SSc-specific autoantibodies anticentromere and anti–topoisomerase I) did not show significant associations (data not shown). Thus, we did not find statistically significant differences in the frequency of the *ATP8B4* rs55687265*C allele between the SSc patients and controls enrolled in the study. A meta-analysis combining the results of the present study with the results from the discovery phase of the study reported by Gao et al (2) was also performed, and again no significant *P* value was found (OR 1.36, $P_{random} = 0.212$, heterogeneity q value < 0.01, $I^2 = 82.68$).

The impact of rare variants on the development of autoimmune diseases remains an unanswered and controversial question (5). Moreover, it has long been recognized that the identification of rare variant associations with high-throughput DNA sequencing technologies, such as whole-exome sequencing, is substantially affected by technical artifacts, which may lead to Type I error. This issue becomes especially important when the sample size of the whole-exome sequencing study is not large enough, and when there is a large difference between the case cohort size and the control cohort size (6,7). The present study highlights the importance of validation of whole-exome sequencing results with other sequencing methods, as well as replication of the newly observed associations in independent studies, in order to detect actual disease-causing mutations.

In conclusion, in the present study we could not replicate the association of *ATP8B4* rs55687265 with susceptibility to SSc. However, because we did not attempt to evaluate associations of other rare or common variants with SSc

Table 1. Association analysis of the *ATP8B4* F436L variant in 6 independent systemic sclerosis cohorts, and meta-analysis*

	Minor/major allele	No. of cases/controls	MAF, cases	MAF, controls	OR (95% CI)
Cohort					
Spain	C/G	2,056/2,718	0.008	0.005	1.58 (0.96–2.61)
Germany	C/G	909/486	0.019	0.022	0.87 (0.50–1.50)
The Netherlands	C/G	435/783	0.013	0.007	1.78 (0.75–4.20)
Italy	C/G	1,114/980	0.006	0.011	0.56 (0.28–1.12)
UK	C/G	1,456/5,272	0.012	0.011	1.13 (0.77–1.67)
US	C/G	1,456/2,848	0.015	0.015	1.00 (0.69–1.45)
Meta-analysis†	C/G	7,426/13,087			1.07 (0.88–1.31)

* Self-reported ancestry and genome-wide association study or Immunochip data were used to remove outliers. None of the odds ratios (ORs) for minor allele frequency (MAF) in cases versus controls were statistically significant. In the Spanish cohort there was a trend toward significance ($P = 0.071$). 95% CI = 95% confidence interval.
† Heterogeneity q value = 0.17, $I^2 = 35.73$.

susceptibility, our findings do not eliminate the possibility that this gene plays a role.

Supported by the Spanish Ministry of Economy and Competitiveness (grant SAF2015-66761-P). The authors are grateful to the Spanish Scleroderma Group (whose members are listed in the Supplementary Appendix, on the Arthritis & Rheumatology web site at http://onlinelibrary.wiley.com/doi/10.1002/art.40058/abstract) for valuable contributions to the study.

Elena López-Isac, MSc
Lara Bossini-Castillo, PhD
Ana B. Palma, MSc
Institute of Parasitology and Biomedicine López-Neyra CSIC
Granada, Spain
Shervin Assassi, MD
Maureen D. Mayes, MD, MPH
The University of Texas Health Science Center
Houston, TX
Carmen P. Simeón, MD, PhD
Valle de Hebrón Hospital
Barcelona, Spain
Norberto Ortego-Centeno, MD
Clinic University Hospital
Granada, Spain
Esther Vicente, MD
Hospital La Princesa
Madrid, Spain
Carlos Tolosa, MD
Hospital Parc Tauli
Sabadell, Spain
Manel Rubio-Rivas, MD
Bellvitge University Hospital
Barcelona, Spain
José A. Román-Ivorra, MD
Hospital Universitari i Politecnic La Fe
Valencia, Spain
Lorenzo Beretta, MD
Fondazione IRCCS Ca' Granda Ospedale
Maggiore Policlinico di Milano
Milan, Italy
Gianluca Moroncini, MD, PhD
Università Politecnica delle Marche
and Ospedali Riuniti
Ancona, Italy
Nicolas Hunzelmann, MD
University of Cologne
Cologne, Germany
Jörg H. W. Distler, MD
University of Erlangen-Nuremberg
Erlangen, Germany
Gabriella Riemekasten, MD, PhD
University of Lübeck
Lübeck, Germany
and German Lung Center
Borstel, Germany
Jeska de Vries-Bouwstra, MD, PhD
Leiden University Medical Center
Leiden, The Netherlands
Alexandre E. Voskuyl, MD
VU University Medical Center
Amsterdam, The Netherlands
Timothy R. D. J. Radstake, MD, PhD
University Medical Center Utrecht
Utrecht, The Netherlands
Ariane Herrick, MD, PhD
The University of Manchester
Manchester Academic Health Science Centre
Manchester, UK
Christopher P. Denton, MD, PhD
Carmen Fonseca, MD, PhD
Royal Free and University College Medical School
London, UK
Javier Martín, MD, PhD
Institute of Parasitology and Biomedicine López-Neyra
Granada, Spain

AUTHOR CONTRIBUTIONS

All authors were involved in drafting the article or revising it critically for important intellectual content, and all authors approved the final version to be published. Ms López-Isac had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study conception and design. López-Isac, Martín.

Acquisition of data. Bossini-Castillo, Assassi, Mayes, Simeón, Ortego-Centeno, Vicente, Tolosa, Rubio-Rivas, Román-Ivorra, Beretta, Moroncini, Hunzelmann, Distler, Riemekasten, de Vries-Bouwstra, Voskuyl, Radstake, Herrick, Denton, Fonseca, Martín.

Analysis and interpretation of data. López-Isac, Palma.

1. Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. J Autoimmun 2015;64:53–65.
2. Gao L, Emond MJ, Louie T, Cheadle C, Berger AE, Rafaels N, et al. Identification of rare variants in ATP8B4 as a risk factor for systemic sclerosis by whole-exome sequencing. Arthritis Rheumatol 2016;68:191–200.
3. Subcommittee for Scleroderma Criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.
4. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet 2014;94:47–61.
5. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature 2013;498:232–5.
6. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet 2016;17:333–51.
7. Hu YJ, Liao P, Johnston HR, Allen AS, Satten GA. Testing rare-variant association without calling genotypes allows for systematic differences in sequencing between cases and controls. PLoS Genet 2016;12:e1006040.

SUPPLEMENTARY MATERIAL

Supplementary methods

1. Genotyping methods

The genotyping of *ATP8B4* rs55687265 was performed with TaqMan SNP genotyping technology (assay ID: AHI1051) in a LightCycler® 480 Real-Time PCR System (Roche Applied Science, Mannheim, Germany). Genotype clustering and allele calling accuracies were ensured including control samples representing the three possible genotypes in all the plates. Genotyping call rate was > 95%. All the samples were genotyped by TaqMan assay, with the exception of the UK and USA control cohorts. For these two control sets, genotyping data were obtained from previously published studies: WTCCC (for the UK cohort) and Lung Health Study (for the US cohort), genotyped with the Affymetrix 500K platform and the Illumina Human660W-Quad v.1_A BeadChip, respectively. (1-2)

2. Data analysis

All the statistical analyses were carried out with PLINK v1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). Association tests for the six cohorts were performed in each population by 2x2 contingency tables and $\chi 2$ test. *P*-values lower than 0.05 were considered as statistically significant. The meta-analysis was performed with inverse-variance method under a fixed-effects model. Heterogeneity of the ORs across cohorts was assessed using Cochran's Q test. A meta-analysis combining the results of the present study with the results from the discovery phase of Gao *et al.* was also performed . We could not meta-analyze our data including the replication phase of Gao *et al*. due to the complete overlap between its replication control cohort and the controls included in our US cohort. Self-reported ancestry and GWAS or Immunochip data were used to remove

outliers and control for population stratification. None of the included control sets showed significant deviation from Hardy Weinberg Equilibrium (HWE) (HWE *P*-values < 0.01 were considered to show significant deviation from equilibrium). The statistical power of the combined analysis is shown in Supp. Table 2 and was calculated according to Power Calculator for Genetic Studies 2006 software under an additive model.(3)

Supplementary tables

Supp. Table 1. Study cohorts

COHORT	Patients with SSc	Controls
Spain	2,056	2,718
Germany	909	486
The Netherlands	435	783
Italy	1,114	980
UK	1,456	5,272
USA	1,456	2,848
TOTAL	**7,426**	**13,087**

Supp. Table 2. Overall statistical power of the study for a statistical significance *P*-value = 0.05.

	OR = 1.2	OR = 1.5	OR = 1.8	OR = 2.2	OR = 2.5	OR = 2.8
rs55687265	43%	99%	100%	100%	100%	100%

SUPPLEMENTARY REFERENCES

1. Stahl EA, Raychaudhuri S, Remmers EF, et al. Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet. 2010;42:508-14.

2. Hansel NN, Ruczinski I, Rafaels N, et al. Genome-wide study identifies two loci associated with lung function decline in mild to moderate COPD. Hum Genet. 2013;132:79-90.

3. 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.

DISCUSSION

1. Novel genetic findings in systemic sclerosis

The study of the genetic component of ADs is a very active field of research that has achieved hundreds of signals of association between genomic variants and autoimmunity. Finding the genetic variants that are associated with the susceptibility to a complex AD is essential for understanding the pathological molecular pathways underlying the disease onset and progression (118). The present PhD dissertation was focused on the study of the genetic component of systemic sclerosis, a complex autoimmune disease with low prevalence but high impact on patients' life.

Before the beginning of this thesis, two GWASs in European population were performed in SSc (98, 100). Later on, and simultaneously with the development of the present work, our group published the first SSc Immunochip study (107). These reports achieved a great advance on understanding the genetic bases of the disease, identifying several *loci* at the genome-wide significant level. Moreover, they also provided a considerable number of nominal signals located in the so-called grey zone. Follow-up studies focused on this grey zone (where tier two signals of associations are located) represent a valuable data mining method to identify additional actual risk variants that may have been overlooked due to a lack of statistical power (119). Using this strategy, we have been able to identify three novel SSc susceptibility *loci*: peroxisome proliferator activated receptor gamma (*PPARG*), interleukin 12 receptor subunit beta 1 (*IL12RB1*) and

tyrosine kinase 2 (*TYK2*). The results of these follow-up studies are part of this PhD dissertation (publications 1, 2 and 3) (105, 120, 121).

Besides these pieces of work, a cross-disease meta-GWAS in SSc and RA identified interferon regulatory factor 4 (*IRF4*) as a novel shared susceptibility factor (publication 4). Moreover, in this study we confirmed other SSc-RA common *loci* previously reported: *IRF5*, *PTPN22*, *ATG5* and *BLK* (115).

Finally, this thesis has also contributed to the rejection of a rare variant in *ATP8B4* as a genetic risk factor for SSc (publication 5) (122). In the next sections, we will further discuss the results from the article compendium that integrates this PhD dissertation.

2. *PPARG*: a *locus* pointing out to fibrosis (publication 1)

The vast majority of the genetic factors described for complex diseases have low-to-moderate effects. This means that large cohorts are needed to robustly detect the associations (123). The GWAS performed by the French group in SSc, which included 564 cases and 1,776 healthy controls from France, reported 90 SNPs with *p*-values ranging from 1×10^{-04} to $< 6 \times 10^{-08}$ (100). Under the hypothesis that using additional cohorts would increase the statistical power to detect new suitable SSc genetic risk factors, we performed a follow-up of these SNPs of the grey zone from the French GWAS. The publication 1 of the present thesis integrates the results of this follow-up study.

After excluding the SNPs located within *MHC* genes or in known SSc risk *loci*, 66 SNPs were available. In a first phase, we performed a meta-analysis for the 66 GWAS-genotyped SNPs combining the results

from the French GWAS with the results from the SSc GWAS published by our group, which comprised 2,357 cases and 5,187 controls from Spain, Germany, The Netherlands and USA.

The meta-analysis revealed that 92,4% of the SNPs showed heterogeneity of the ORs. This elevated percentage of SNPs with heterogeneity in their effect size across populations was not observed in the meta-analysis of the four cohort included in the GWAS published by our group, which indicated that most of the observed heterogeneity came from the French cohort. These findings suggest that most of the signals located in the grey zone of the French GWAS present inflated effect sizes (the so-called winner's curse) and highlight the relevance of large cohorts to accurately estimate ORs and statistical significance (119).

The first phase also allowed us to select SNPs for a replication step in independent cohorts. The global analysis, combining the results from the first and the replication phases, led to the identification of a signal very close to the genome-wide significance level. The SNP harboring the signal, rs310746, is an intergenic SNP located in a highly polymorphic region at chromosome 3 that comprises three genes: *PPARG, SYN2* and *TIMP4*.

PPARG encodes the peroxisome proliferator-activated receptor gamma (PPAR-γ), a nuclear receptor involved in various metabolic processes, including adipogenesis, insulin sensitivity, and homeostasis (124). Interestingly, several *in vitro* and functional studies have provided evidences that suggest PPARG as an anti-fibrotic effector, an important aspect in SSc, in which fibrosis is one of the main hallmarks of the disease. In 2004, Ghosh *et al*. showed that PPAR-γ ligands abrogated collagen gene expression induced by TGB-β, myofibroblast transdifferentiation and Smad-dependent promoter activity through *in*

vitro experiments in normal fibroblasts (125). Similar results were obtained in human lung fibroblasts (126, 127). Moreover it has been reported that PPAR-γ ligands can exert their antifibrotic effect by inhibiting the phosphatidylinositol 3 kinase/protein kinase B (PI3K/Akt) pathway (127). Remarkably, experiments with SSc mouse models have shown that synthetic agonists of PPAR-γ, such as rosiglitazone, the triterpenoid oleanane 2-cyano-3,12-dioxoolean-1,9-dien-28-oic (CDDO) and IVA337 attenuate dermal fibrosis. These observations have pointed out PPAR agonists as potential therapeutic targets for the treatment of fibrotic diseases (128-130). It has also been described that the expression and function of *PPARG* are impaired in SSc patients (131). Moreover, a recent candidate association study has reported the association of an intronic variant of *PPARG* with SSc (132). This gene also represents a susceptibility factor for other ADs, such inflammatory bowel disease (IBD) (133, 134) and psoriatic arthritis (135). In the light of these evidences, the rs310746 signal was assigned to *PPARG*. Nonetheless, the putative role of the nearby gene *TIMP4* could not be rule out.

The assignment of disease-associated SNPs to genes due to genomic proximity and biological plausibility has been a common practice in genetic association studies. However, during the past 5 years, the scientific community has been witness to a tremendous revolution on the field of genomics with the development of different ambitious projects that aim to provide a deep characterization of the functional elements of the genome and the epigenome (Figure 1, Introduction). The ENCODE or the NIH Roadmap Epigenomics Projects offer a wide range of public data that help to the prioritization of disease-association signals (95, 96). Moreover, it is well known that gene expression regulation is a complex process that varies across

tissues and cell states. With the idea to resolve this issue, the scientific community launched the Genotype-Tissue Expression (GTEx) project, whose complete resource of data was published in 2015 (136), after the publication of the GWAS follow-up study included in this thesis. The GTEx provides information about human gene expression and regulation and its relationship to genetic variation across multiple tissues. Thus, this project offers the possibility to identify genetic expression quantitative trait loci (eQTLs), which are polymorphisms that are highly correlated with variations in gene expression.

Considering the valuable resource of data, we searched for a putative functionality of the rs310746 in the GTEx database and found that this SNP is an eQTL of *TIMP4* in tibial and aorta arteries (*p*-value = 1.1×10^{-16}, *p*-value = 1.9×10^{-06}, respectively; FDR < 5%) (Figure 8). *TIMP4* encode the metallopeptidase inhibitor 4 (TIMP-4), and belong to the TIMP gene family. The proteins encoded by these genes are inhibitors of the matrix metalloproteinases (MMPs), a group of proteins that regulate the turn-over of the ECM. Thus, MMP/TIMP imbalance is suggested to be important in tissue remodeling during fibrogenic process in response to inflammatory stimuli (137). Interestingly, a clinical study reported higher TIMP-4 serum levels in SSc patients than in healthy controls (138). Altered expression of MMPs and TIMPs has also been reported in IBD (139), in which chronic inflammation and aberrant tissue remodeling are characteristic features. In fact, the potential therapeutic effect of antibodies targeting MMPs has been investigated in IBD mouse models (140). According to GTEx, the minor allele of rs310746 correlates with higher *TIMP4* expression level. This may give rise to an increased inhibition of MMPs, which ultimately might lead to lower collagen degradation. Therefore, rs310746 signal would also be linked to fibrosis through *TIMP4.*

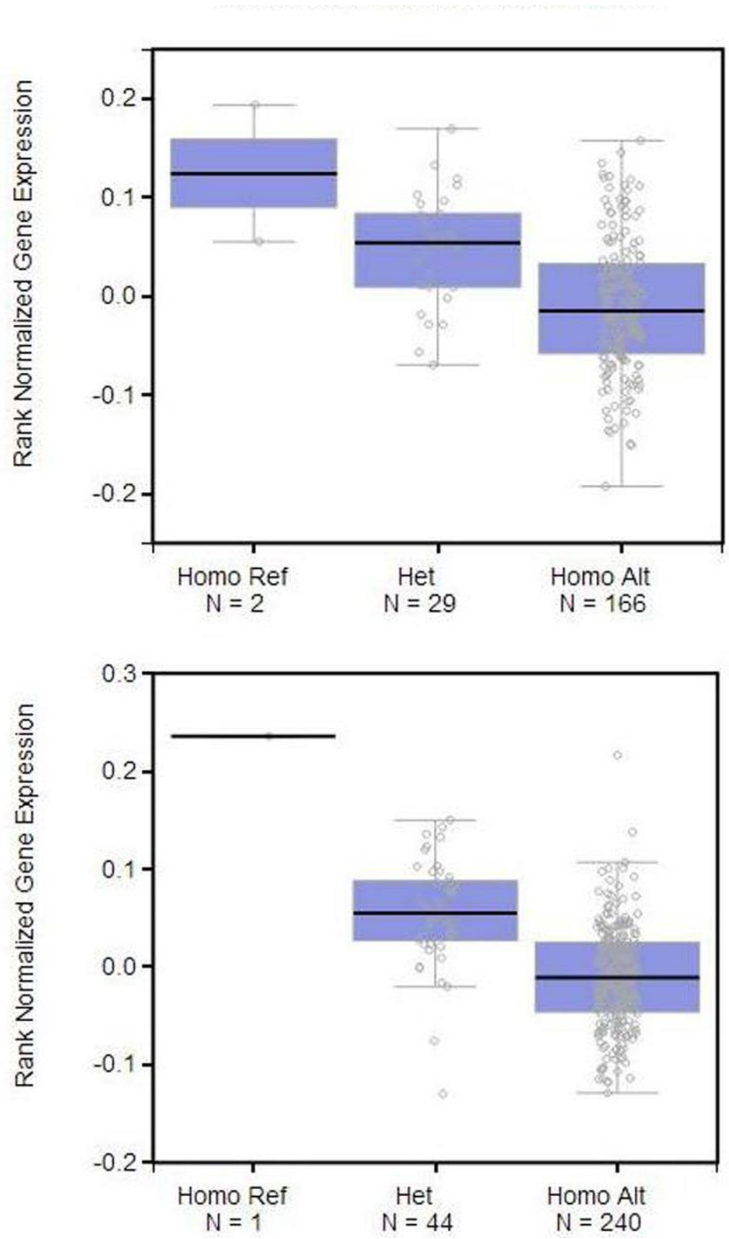


Figure 8. Correlation between rs310746 with gene expression of *TIMP4* in artery aorta (upper panel) and artery tibial (lower panel) (Data obtained from GTEx)

Although the results from the GTEx database offer functional support for *TIMP4* being the causal gene, considering all the evidences that support a role of *PPARG* in fibrosis and SSc, we cannot discard a possible effect of rs310746, or a tagged signal, not described to date in *PPARG* expression. Therefore, we will assign the suggestive rs310746 signal reported in our GWAS follow-up study to '*TIMP4-PPARG*' from now on.

3. Immunochip follow-up and the IL-12/IL-23 pathway in systemic sclerosis (publications 2 and 3)

It is well known that genes do not work independently but in complex interaction networks that form molecular pathways. In genetics of complex ADs, related genes involved in the same pathway are often jointly associated with a particular trait (141). Therefore, functional prioritization by integrating the information of previously associated *loci* and prior knowledge of biological pathways may help to increase the chance to identify new genes and to better define mechanisms underlying disease risk and progression. Good examples of the successful of this strategy are the Immunochip follow-up studies included in this thesis (publications 2 and 3).

Previous studies from our group had yielded evidences for an important role of the IL-12 signalling pathway in SSc pathogenesis (98, 103, 107). Considering this prior information, we prioritized the nominal signals located in the grey zone of the SSc Immunochip study published by our group and performed two follow-up studies on IL-12-related regions. This allowed us to identify two new risk *loci,* namely *IL12RB1* and *TYK2*, reaching the genome-wide significance level. The identification of these new strong signals involved in the same pathway

was also possible thanks to the nature of the Immunochip that contains related sets of SNPs (≈ 200,000 SNPs) from genes implicated or putatively implicated in IMDs (92). Most importantly, this platform was designed for fine-mapping the selected *loci* and includes all the variants described for European population by 1000 Genomes Project February 2010 release, dbSNP and additional sequencing projects (92). This particular characteristic allowed us to perform a comprehensive initial screening of *IL12RB1* and *TYK2* regions in our Immunochip data.

In the case of *IL12RB1* (publication 2), the analyzed region comprised 46 SNPs and 11 of them showed nominal signals. However, through the conditional analyses and the functional prioritization and annotation of the SNPs, we were able to narrow down the signal of association to the promoter region of the gene (rs436857). Moreover, our results are consistent with the results obtained by Takahashi *et al*., which found an *IL12RB1* promoter polymorphism in high LD with rs436857 (rs393548) that had a transcriptional effect in the expression level of *IL12RB1*. According to their results, the minor allele of the mentioned variant was associated with decreased *IL12RB1* mRNA levels. In addition, these authors suggested the presence of enhancer and silencer elements in the 5' region of the gene (142).

In regard with *TYK2* (publication 3), our initial screening included 154 SNPs and allowed us to select for follow-up the strongest signal observed in the region, which was a common protein-coding missense variant previously associated with SLE (V362F (rs2304256))(143-145). In a contemporary study to this thesis, Diogo *et al*. performed a comprehensive fine-mapping genetic study of *TYK2-ICAM* in RA and identified three *TYK2* protein-coding variants as the most likely causal variants responsible for the signal of association in the region. These findings were also extended into other autoimmune

phenotypes (146). Thus, we additionally followed-up these three *TYK2* protein-coding variants: two low-frequency coding variants (P1104A (rs34536443), I684S (rs12720356)) and one rare coding variant (A928V (rs35018800)). Our study showed that V362F (rs2304256) common variant reached the genome-wide significance level. Despite the large cohort enrolled in this study, it was underpowered to detect genome-wide significant associations for the low-frequency and rare variants. However, the dependence analyses clearly supported that the strong signal of V362F was a synthetic association dependent on the presence of the other three selected variants (P1104A, I684S and A928V). These findings highlight the importance of fine-mapping and the complexity to resolve not only candidate genes but also causal variants (118). In fact, considering that P1104A, I684S and A928V were selected according to the observations in RA and SLE, we cannot rule out the genetic effect of additional independent rare and low-frequency *TYK2* variant in SSc susceptibility.

To date, five genes of the IL-12 pathway have been described as firm genetic factors for SSc and all of them are associated at the genome-wide significance level: *IL12RB1* and *IL12RB2* (the genes that encode the IL-12 receptor chains); *IL12A* (which encodes the p35 subunit of IL-12); *TYK2,* the gene that encodes the Jak-STAT tyrosine kinase that activates the signal transducer and activator of transcription; and *STAT4* (transcription factor of the IL-12 signalling axis) (Figure 9) (98, 103, 107, 120, 121). These findings provide genetic evidences for the crucial role of this cytokine pathway in SSc pathogenesis. In addition, two of these five genes are also involved in the IL-23 signalling pathway: *IL12RB1* and *TYK2* (Figure 9).

IL-12 is a cytokine that exerts important pro-inflammatory functions and is a powerful inducer of both Th1 cell differentiation and

responses, thus promoting cell-mediated immunity. On the contrary, IL-12 antagonizes Th2 differentiation in combination with IFNγ (147). Several studies have implicated IL-12 and IFNγ in autoimmune inflammatory processes (147). Interestingly, serum levels of IL-12 are significantly increased in SSc patients, and this overproduction has been associated with renal vascular damage (148). Functional studies have suggested that Th1 responses may be crucial in mediating early inflammatory processes in SSc, while Th2 responses actively promote fibrotic processes by inducing secretion of profibrotic cytokines such as TGF-*β (149)*. In this regard, IL-12 is known to have an anti-fibrotic effect in fibroblast and its serum-level elevation has been correlated with improvement in skin fibrosis in SSc (150, 151).Therefore, the implication of IL-12 in SSc seems to be complex taking into account its dual effect as pro-inflammatory cytokine and anti-fibrotic effector.

On the other hand, IL-23 mediates chronic inflammation by promoting the survival and maintenance of Th17 cells as well as by stimulating the production of IL-17. Concentrations of IL-23 and IL-17 have been reported to be increased in SSc patients (152, 153). Several studies have been performed to understand the functional significance of the Th17 cell imbalance in SSc and some of them proposed a potential role of Th17 cells in promoting inflammatory responses (154). However, the implication of Th17 in SSc remains controversial, since some studies support a role of IL-17 in fibrosis, while other studies indicate an anti-fibrotic effect for this cytokine (40, 154, 155).

Our studies revealed for the first time the association of *IL12RB1* and *TYK2* with SSc (publications 2 and 3). The associations signals for both *loci* showed protective effects (OR < 1), meaning that the minor alleles of the analyzed variants are significantly less frequent in SSc patients than in healthy controls. Functional annotation of the

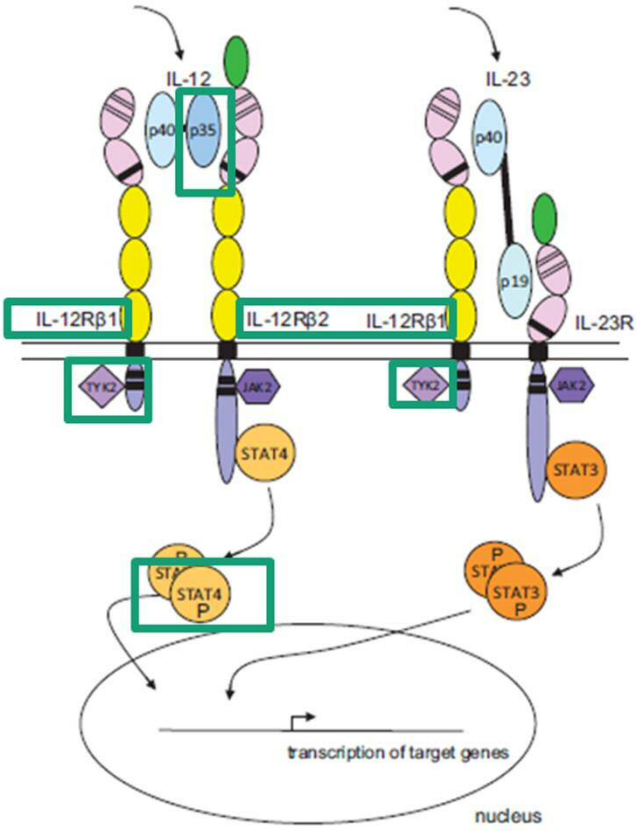


Figure 9. Representation of the Il-12/IL-23 pathway. SSc-associated genes are highlighted in green. (Modified from van de Vosse *et al.,* Hum Mutat, 2013).

associations revealed that *IL12RB1* rs436857 promoter variant is a *cis*-eQTL whose minor allele correlates with decreased *IL12RB1* expression (p-value = 2.4 x 10^{-81}, Z-score = -19.10), which ultimately may lead to a decrease IL-12/IL-23 signalling. In the same way, the three missense variants responsible for the V362F (rs2304256) *TYK2* association are predicted to be damaging mutations (93, 94). Moreover, the effect of two of them, P1104A and I684S (located in the kinase and pseudokinase domains of the protein, respectively), has been characterized by *in vitro* studies that have demonstrated that P1104A

is catalytically impaired, it leads to a reduced TYK2 activity and decrease type I IFN, IL-23 and IL-12 signaling (156-158).These functional findings of the *IL12RB1* and *TYK2* variants would be consistent with the protective effect and a lower SSc susceptibility.

In the light of these evidences, target therapies to this pathway could be an effective treatment for the disease. In fact, an anti-IL-12/23 p40 monoclonal antibody (ustekinumab) is currently approved for the treatment of psoriatic arthritis (159-161). Remarkably, Jak–STAT pathway has become an eminent drug target for pharmaceutical companies given the big amount of data linking these molecules to autoimmune diseases and its central role in the immune system (162-164). However, the effect of a Jak inhibitor in the treatment of ADs could be counterbalanced by the increasing risk of certain infections. In fact, human TYK2 or IL-12Rβ1 deficiencies lead to autosomal recessive immunodeficiency syndromes characterized by predisposition to recurrent mycobacterial and/or viral infections, caused by impaired cytokine responses (165, 166).

A very recent study has provided a comprehensive analysis of the role of TYK2 in autoimmunity that highlights that understanding the specific meaning of a genetic association is a necessary step in identifying drug targets (158). Their genetic study showed that, as for SSc, P1104A (rs34536443) protects against ten ADs (including psoriatic arthritis, RA, SLE, ankylosing spondylitis, Crohn's disease and ulcerative colitis). The authors performed an extensive characterization of the implication of this variant in TYK2 activity that showed that the protective effect arises from the minor allele homozygosity at rs34536443, which leads to a near-complete loss of TYK2 catalitic function, and consequently an impairment on cytokine signaling.

Moreover the protective effect is genotype-dependent, since a relatively modest protective effect of rs34536443 heterozygosity was observed.

Interestingly, according to the results of this study, rs34536443 minor allele homozygosity allows for the minimal amount of TYK2-mediated cytokine signaling needed to prevent immunodeficiency. Therefore, the perfect drug would be a molecule capable to mimic the impact of rs34536443 minor allele on TYK2 function that would guarantee an optimum balance between autoimmunity and immunodeficiency.

In summary, the results of publications 2 and 3 have provided further evidence for the important role of IL-12/IL-23 pathway in SSc susceptibility, by identifying *IL12RB1* and *TYK2* as new *loci* firmly associated with the disease. It has been demonstrated that selecting genetically supported drug targets can have a considerable impact on drug success (167). Thus, considering the functional consequences of the analyzed genetic variants and their protective effect in the predisposition to the disease, our results may have therapeutic interest since they provide support for the promising effect of TYK2 inhibitors in SSc treatment.

4. Shared *loci* between systemic sclerosis and rheumatoid arthritis (publication 4)

The existence of a common genetic background underlying ADs has long been suspected given the overlap in clinical and immunological characteristics, along with the familial clustering and co-ocurrence observed among these diseases. GWAS and, specially, Immunochip findings have widely contributed to the idea of this shared genetic component among IMDs (4, 106, 168, 169). To date, there are more than 70 loci associated at the genome-wide significance level with two or more ADs (106). The increasing evidences of this genetic overlap led to the development of a new strategy, namely cross-disease meta-GWAS, which lies in combining genome-wide genotype data from two complex diseases to systematically identify new shared *loci*. Thus, the methodology represents a systematic approach to further explore common etiopathological pathways.

We have applied this strategy for the interrogation of the common genetic risk factors between SSc and RA (publication 4). The cross-disease meta-GWAS identified *IRF4* as a new SSc-RA shared *locus*. In addition, we also confirmed common risk factors previously described for both diseases, such as *IRF5, PTPN22, ATG5, BLK* and the *HLA* region, which provides support for the ability of this strategy to resolve overlapping associations. We performed a protein-protein interaction (PPI) analysis taking into account all the shared susceptibility factors in SSc and RA. The SSc-RA network showed a significant enrichment in interactions, which implies that the proteins encoded by the SSc-RA risk *loci* interact with each other more than expected by chance (Figure 10), suggesting common altered pathways in SSc and RA. Consistently, the molecular pathway enrichment analyses identified significant overrepresentation of several gene sets

mainly involved in the immune response, especially the type I interferon (IFN) signaling pathway.

More than 50% of the SSc susceptibility *loci* are also genetic risk markers for RA, and it is clear that both autoimmune rheumatic conditions are closely related phenotypes that share a substantial portion of their pathological pathways (170). However, there are many nuances that should be considered in regard with their genetic component. A genetic overlap can comprise a shared *locus* for which

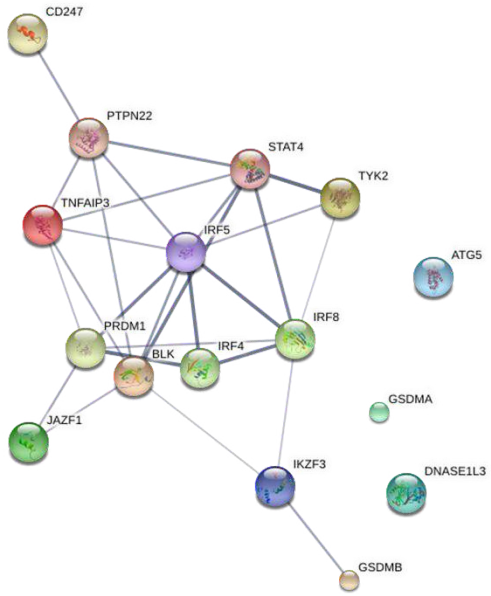


Figure 10. Protein-protein interaction network of risk *loci* shared between systemic sclerosis and rheumatoid arthritis.

the same allele can have the same or opposite effect among different diseases (concordant *vs*. discordant effect, respectively). A good example of this observation is the *PTPN22* non-synonymous variant R620W that confers risk to RA, SLE and type 1 diabetes (T1D), but is protective against Crohn's disease (106). Moreover, the true causal SNP

(or haplotype) for each disease at a given shared *locus* can be independent ('non-corraleted' signal), which may involve different downstream effects underlying the associations. Thus initial evidence of extensive genetic overlap may not always being translated into the same functional consequences (106). Therefore, we would like to update the picture of SSc-RA shared *loci* and to further characterize this genetic overlap taking into account the functional downstream effects of the associated variants. For this aim, we first updated the list of common SSc-RA *loci* and selected the index SNPs in each *locus* for both diseases. To maximize the scope of the analysis, we not only considered those *loci* that reached the genome-wide significance level but also those *loci* confirmed in at least two independent studies and/or showing a *p*-value of association < 5×10^{-5} in European population. The most powerful association studies for each disease were considered to select the top SNPs. In the case of RA, the reference work was the extensive meta-GWAS performed by Okada *et al*. (116), except for *CD247* and *PRDM1* (the index SNPs for these two *loci* were obtained from Teruel *et al*. and Raychaudhuri *et al*., respectively (171, 172). In the case of SSc, the index SNPs were selected from the initial phase of a large SSc meta-GWAS that our group is currently performing, which initially includes 7,828 SSc patients and 14,523 healthy controls of European ancestry. In addition, the Immunochip results from our group and a very recently published trans-ethnic meta-GWAS in SSc were also considered (107, 173).

In total, there are 14 well-established non-HLA shared *loci* for SSc and RA. As it can be observed in Table 2, the index SNP was different for 13 out of the 14 *loci*. Next we assessed whether the SNPs, for each of the 13 *loci*, were correlated by calculating the LD between them and found that in four *loci* (*ATG5*, *BLK, TNFAIP3* and *IKZF3*-

GSDMA) the top SSc and RA variants were in strong LD with each other ($r^2 > 0.8$). Intermediate LD ($0.2 < r^2 < 0.8$) was observed for another five *loci*, while for the remaining four *loci* the SSc and RA SNPs were in low LD with each other ($r^2 < 0.2$) (*PTPN22*, *IRF5*, *JAZF1*, *IRF8*). These markers in very low LD with each other ($r^2 < 0.2$) can be considered independent or non-correlated SNPs. Therefore, although the top signals in each *locus* can differ between SSc and RA, most of the SSc-RA common associations (71%) are driven by correlated variants. In regard with the protective (OR < 1) or risk effect (OR > 1) of the variants, the associations were concordant (same effect) for ten out of the 14 *loci*. The discordant effects were observed for *PTPN22*, *PRDM1*, *JAZF1*, and *IRF8* (Table 2).

We next explored the impact on gene function for the 14 *loci* to evaluate whether the associated SNPs in each disease trigger similar or different downstream effects. As it was expected, most of the associated SNPs were non-coding variants, thus it is more likely that they have a regulatory effect -for instances, affecting gene expression- rather than an effect on protein function itself. Therefore, we performed *cis*-eQTL analysis using blood eQTL data from Westra *et al.*(174), the Geuvadis dataset, which contains expression data from lymphoblastoid cell lines (175), and the GTEx project (136). Since we also found protein-coding missense variants among the top associated variants in each disease, our functional annotation also integrated previously reported impact of these variants on protein function (see Table2).

Interestingly, we found different downstream effects among some of the shared *loci* with low LD between SSc and RA SNPs (i.e. those harboring independent signals in each disease). One of the most curious examples was observed for *PTPN22*. The functional variant R620W (rs2476601) of *PTPN22* is a common risk factor for ADs, including RA. The minor allele of this variant leads to an amino acid

substitution that disrupts binding between *PTPN22*-encoded protein (LYP) and an intracellular kinase called Csk, which gives rise to impaired dephosphorylation and inactivation of its substrate, thus increasing T cell receptor signalling and activation (176). In RA, rs2476601 and its tag-SNP (rs6679677) are the most associated variants within *PTPN22*, and the effect size of the association is 1.8 (116). In SSc, the index SNP at this *locus*, which is not in LD with R620W (rs2476601), significantly correlates with an eQTL that increases *PTPN22* expression (Table 2). Candidate gene studies have previously

Table 2. Well-established non-*HLA* shared *loci* for systemic sclerosis and rheumatoid arthritis.

			Systemic sclerosis					Rheumatoid arthritis				
CHR	SSc-RA Locus	r^2	Reported index SNPs	SNP Function	OR for the minor allele	eQTL	Other downstream effect	Reported index SNPs	SNP Function	OR for the minor allele	eQTL	Other downstream effect
1	*CD247*	0.61	rs2056626	intron	0.82	CD247 (Increased expression)		rs864537	intron	0.9	CD247 (Increased expression)	
1	*PTPN22*	**0.05**	rs1970559	intron	0.88	PTPN22 (Increased expression); HIPK1|AP4B1, RSBN1 (decreased expression)		rs2476601	missense	1.8	PTPN22 (Increased expression)	Disrupts binding between PTPN22-Csk
2	*STAT4*	0.69	rs3821236	intron	1.35	GLS (Increased expression); STAT1, STAT4		rs11889341	intron	1.12	STAT1, STAT4	
3	*DNASE1L3*	0.54	rs35677470	missense	1.47; 2.03 in ACA+	PXK (Increased expression); RP11-802O23.3 (decreased expression)	The minor allele results in an inactive enzyme	rs73081554	intron	1.18	PXK (Increased expression); RP11-359I18.5, RP11-802O23.3, RPP14 (decreased expression)	
6	*TNFAIP3*	**0.12** **1**	rs2230926	missense	1.44	TNFAIP3		rs17264332 rs7752903	intron intergenic	1.17 1.41	No eQTL TNFAIP3	McGovern A *et al*. (2016)**
6	*ATG5*	**0.95**	rs9373839	intron	1.19	ATG5, PRDM1		rs9372120	intron	1.1	ATG5	
6	*IRF4*	1 0.39	rs9328192-G	intergenic	0.9	DUSP22, EXOC2, IRF4, OR2W1	rs9328192-G	intergenic rs9378815-G	1.08 intergenic	 0.92	DUSP22, EXOC2, IRF4, OR2W1 No eQTL	
6	*PRDM1*	0.66	rs4134466	intron	0.85	ATG5, PRDM1		rs548234	Downstream *PRDM1*	1.11	ATG5, PRDM1	
7	*IRF5*	**0.16**	rs36073657	intron	1.4091	IRF5 (decreased expression)		rs3778753-G	intron	1.12	IRF5 (Increased expression)	
7	*JAZF1*	**0.06**	rs849139-C	intron	1.14	JAZF1 (decreased expression)		rs67250450	intron	0.9	JAZF1 (decreased expression)	

Table 2. Well-established non-*HLA* shared *loci* for systemic sclerosis and rheumatoid arthritis.

			Systemic sclerosis					Rheumatoid arthritis				
CHR	**SSc-RA Locus**	**r^2**	**Reported index SNPs**	**SNP Function**	**OR for the minor allele**	**eQTL**	**Other downstream effect**	**Reported index SNPs**	**SNP Function**	**OR for the minor allele**	**eQTL**	**Other downstream effect**
8	*BLK*	**0.97**	rs2736340	Regulatory region	1.26	BLK (decreased expression); FAM167A, C8orf13|C8orf12 (Increased expression)		rs2736337	intergenic	1.09	BLK (decreased expression); FAM167A, C8orf13|C8orf12 (Increased expression)	
16	*IRF8*	**0.12**	rs11117420	Downstream *IRF8*	0.81	COX4NB (decreased expression); IRF8		rs13330176	Downstream *IRF8*	1.12	IRF8	
17	*GSDMA*	**0.86**	rs883770	intron *GSDMB*	1.14	GSDMA, RP11-387H17.4 (Increased expression); GSDMB, ORMDL3, PGAP3, ZPBP2 (decreased expression)		rs59716545-G	intron ZPBP2	1.09	GSDMA, RP11-387H17.4 (Increased expression); GSDMB, ORMDL3, PGAP3, ZPBP2, PSMD3 (decreased expression)	
		0.44	rs3894194 (From Terao et al.)	missense *GSDMA*	0.85	GSDMB, ORMDL3, PGAP3, ZPBP2 (Increased expression); GSDMA, IKZF3, MED24, PSMD3 (decreased expression)						
19	*TYK2*	1	rs34536443	missense	0.8		Dendrou *et al.* (2016)***	rs34536443	missense	0.68		Dendrou *et al.* (2016)***

**This SNP is in tight LD with the most likely causal variant in this region, rs6927172. Capture Hi-C experiments have shown that the DNA fragment containing these SNPs interacts through chromatin looping not only with TNFAIP3, but also with IL20RA. The risk allele of rs6927172 correlates with and increased expression of IL20RA.

***(158)

reported contradictory results for the role of R620W variant in SSc and the described effect size is lower than for RA (OR ≈ 1.13) (177-181). The results from the initial phase of our large meta-GWAS did not show significant results for this functional variant, thus we could conclude that R620W is not associated with SSc. However, considering that rs2476601 minor allele has a low frequency (MAF between 5-13%), we should be cautious: despite the large cohort, our study is underpowered to detect genome-wide significant associations for low-frequency variant with low effect size. Therefore, much larger cohort would be needed to convincingly identify or discard R620W as a susceptibility factor for SSc. Another interesting result was observed for *IRF5*: the rs36073657 minor allele (which confers risk for SSc) is significantly correlated with decreased expression of *IRF5*, while rs3778753 (which confers risk to RA and is not in LD with rs3778753) increases the expression of *IRF5*. Even when similar downstream effects are observed, the effect on each disease (risk *vs*. protective effect) can differ. This is the case of *JAZF1*: the RA SNP (rs67250450) is not associated with SSc and is not correlated with the SSc index SNP (rs849139) (independent signals); the minor alleles of both top SNPs correlate with decreased expression of *JAZF1*. However, the rs849139 minor allele confers risk for SSc, while rs67250450 minor allele confers protection for RA.

The identification of shared *loci* can help to discover common pathways across different ADs, which may represent a clinical advantage, thus providing support for drug repurposing of current therapies. However, our results and similar studies underline the importance of careful inspection of the functional effects driven by the associated variants in each disease, since shared *loci* commonly associated with two related diseases could actually contribute to disease susceptibility through different regulatory mechanisms.

As stated before, 71% of the SSc-RA shared *loci* harbor correlated signals. In addition, despite the aforementioned differences, most of the regulatory effects observed in our *cis*-eQTL analysis are also similar in both diseases, as well as the effect that the variants confer to disease susceptibility. Therefore, our observations suggest an extensive overlap both in their genetic component and in the etiopathogenic pathways underlying SSc and RA. Similar studies have been performed in other ADs and more discordant results were obtained, as in the case of RA and celiac disease (CeD)(182).

We are aware that our analyses have considered the top SNPs for each disease, which are not necessarily the causal SNPs of the associations in each disease. Therefore, performing similar analyses considering the most likely causal SNPs would get more conclusive results. In the same way, it should be taken into account that most of the eQTL data are derived from blood tissue and/or lymphoblastoid cell lines. eQTL analyses considering disease-relevant cell types and cell states would be a critical next step to identify disease-specific effects (118, 183). Overall, our results highlight the complexity on understanding the mechanisms leading to autoimmune processes and the complex interplay among different susceptibility variants described in autoimmunity.

5. Role of next generation sequencing in systemic sclerosis (publication 5)

The first study that applied next generation sequencing (NGS) to the genetic component of SSc was published in 2015 by Gao *et al.*(184). The whole-exome sequencing (WES) study reported a novel gene, namely *ATP8B4*, as a risk factor for SSc and pointed out a missense rare variant (F436L [rs556872659]) as the most plausible causal variant

underlying the observed genetic association (184). Consequently, we performed a replication study to further address the role of the reported rare variant in SSc susceptibility. The results of this study comprised the publication 5 included in this thesis.

Our replication study included 7,426 SSc patients and 13,087 healthy controls, a much larger cohort than the included in Gao *et al*. (78 patients and 3,179 controls in their WES study; 415 patients and 2,848 controls in the validation cohort). However, we did not replicate the reported association. In addition, we performed a meta-analysis of our results with the discovery phase of Gao *et al*. that also showed no significant differences for the minor allele of rs55687265 between SSc cases and healthy controls. Therefore, our results robustly discarded a role for this missense variant in SSc.

The lack of replication raised concerns about several limitations of the WES study by Gao *et al*. First, their discovery cohort was not large enough and they did not validate their WES results with a different sequencing method. It is well known that the identification of rare variant associations with high throughput DNA sequencing technologies is substantially affected by technical artifacts, which may lead false positive findings (185)). This issue becomes especially important when cases and controls have not been exactly sequenced in the same way (another important limitation of the study), which also leads to high type I error rates in detecting rare variant associations (186, 187). In this regard, a recent method for the analysis of rare variant associations from NGS has been developed with the aim to control for differential sequencing qualities between cases and controls (187).

Rare protein-coding variants are more likely to affect the protein function, thus, *a priori*, they are supposed to have larger effect

size on disease susceptibility(188). This idea gave rise to the hypothesis that rare variants may explain at least a considerable portion of the so-called 'missing heritability' for complex traits. However, the impact of rare variants in the development of ADs remains an unanswered and controversial question (189). In this regard, some studies have illustrated that the identification of rare variants is more useful to dissect the target gene within a susceptibility *locus* than for the discovery of rare variants with large effect sizes and high impact on disease heritability (146, 190). Therefore, taking into account the current high cost of NGS approaches, we consider that exon sequencing of SSc candidate genes could be a fruitful approach to differentiate true causal genes within associated *loci* comprising several genes, thus shedding light into the pathogenic mechanisms underlying the disease.

6. Overview of the genetic component of systemic sclerosis: functional implication of associated *loci*

In the previous sections we have summarized and discussed the results from the five publications that comprise the present thesis, which have added four additional *loci* (*TIMP4-PPARG*, *IL12RB1*, *TYK2* and *IRF4*) to the genetic component of SSc. In the last section of this Discussion, we would like to integrate our results together with the remaining SSc susceptibility *loci* and to offer an overview about our current knowledge of the genetic background underlying SSc predisposition. For this purpose, we have considered all the SSc *loci* associated at the genome-wide significant level. In addition, to maximize the scope of the overview, we have also considered previously reported associations that did not reach the genome-wide

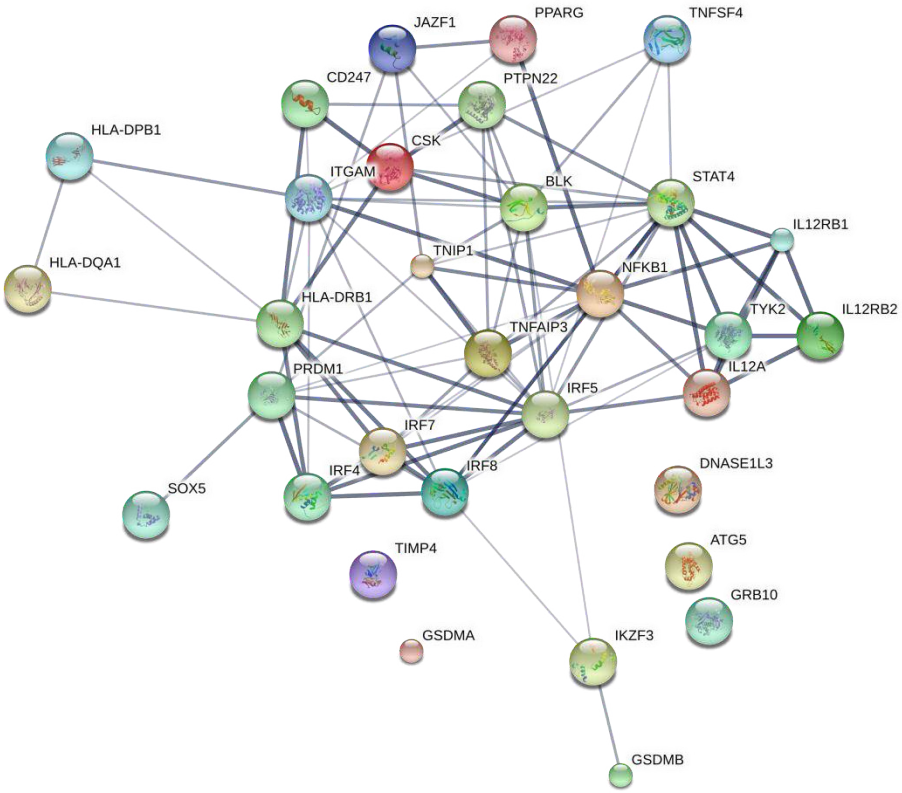
significant level but have been confirmed in the initial phase of our large meta-GWAS (mentioned at point 4 of the Discussion) at least with a *p*-value < 5 x 10^{-5}. Moreover, we have also included the results from a very recently published trans-ethnic meta-GWAS in SSc that has identified *PRDM1* and *GSDMA* as novel susceptibility factors for the disease (173). In total, there are 27 *loci* outside the HLA region that fulfilled the inclusion criteria (Table 3). Most of them point out to genes that are closely related according to their function and that form complex interaction networks, as can be observed in the PPI analysis performed by means of STRING V10.0 (Figure 11) (191). These molecular networks point toward biological pathways involved in SSc onset and progression.

One of the pathways represented by SSc risk factors is the type I IFN signalling pathway. It is noteworthy that four IFN regulatory factor (*IRF*) genes are part of the genetic component of SSc: *IRF4*, *IRF5*, *IRF7* and *IRF8* (Table 3). These IRFs belong to a family of transcription factors that are activated after type I IFN induction (192, 193). In this regard, over the past years, there has been increasing evidence for the implication of type I IFN deregulation in the pathogenesis of SSc. Increased expression and activation of IFN-inducible genes have been observed in blood and skin of SSc patients (192). Thus SSc, as other ADs, is also known to have the so-called 'IFN signature'.

As it was stated in the Introduction, the excessive inflammatory environment is crucial in the development of SSc. Therefore, it is not surprising that several *loci* directly involved in the regulation of the inflammatory response are also SSc susceptibility factors (Table 3). This is the case of different genes involved in the TNF-induced NF-κB proinflamatory signalling pathway, such as *TNFAIP3*, *TNIP1* and *NFKB1*. TNFAIP3 and TNIP1 negatively regulate the TNF-induced NF-κB

signalling pathway. In addition, they also participate in B cell survival and in the TNF-mediated apoptosiss (194).

Figure 11. Protein-protein interaction network of systemic sclerosis risk *loci* performed by means of SRING V10.0.



The aforementioned pathways highlight the role of deregulated innate immunity compartments in SSc. In addition, as we also stated in the Introduction, several components of the adaptive immune response are also involved in the disease. B and T cell proliferation, differentiation, survival and activation are biological processes represented for several SSc risk factors (Table 3). For example, TNFSF4 is involved in T and B cell proliferation and survival(178); *CD247*

encodes the T cell receptor T3 zeta chain that forms the T cell receptor-CD3 complex (TCR/CD3 complex), which is negatively regulated by CSK and LYP (encoded by *PTPN22*) (195, 196). IL12/23 and Jak/STAT signaling pathways are also overrepresented by the genetic background of SSc as we have previously addressed.

Fibrosis is one of the main hallmarks of SSc. Therefore, it can be expected that genes related to this process are also associated with the disease. This is the case of the *locus TIMP4-PPARG* that we already discussed previously; or *ITGAM*, which encodes the integrin subunit alpha M. This subunit form the leukocyte-specific integrin macrophage receptor 1 ('Mac-1') in combination with the beta 2 chain (ITGB2). Integrins take part of the innate immune response, but are also central activators of the latent complex of TGF-β, which is considered a master regulator of fibrosis in SSc (45, 197).

Most of the SSc risk *loci* encode proteins that are implicated in different processes of the immune system. In addition, there are a number of *loci* that are not easily connected to this 'immune system network' in our PPI analysis, which point toward new biological processes, for example, *DNASE1L3* and *ATG5*. The protein encoded by *DNASE1L3* is a member of the deoxyribonuclease I family and is involved in DNA fragmentation, DNA breakdown during apoptosis and the generation of the resected double-strand DNA breaks in immunoglobulin genes (198-200). Thus, this *locus* represents a link between apoptosis, impaired clearance of degraded DNA and autoimmunity. Regarding ATG5, this protein is involved in autophagosome elongation. Autophagy is a central player in the immune system, since it is involved in different processes such as development, survival and homeostasis of B and T cells, cytokine

Table 3. Susceptibility *loci* for systemic sclerosis outside the *HLA* region.

Locus	Chr	Gene name	Index SNP	SNP function
Innate immunity, interferon signature and inflammation				
IRF5	7	IFN regulatory factor 5	rs36073657	intronic
IRF8	16	IFN regulatory factor 8	rs11117420	intergenic
IRF7	11	IFN regulatory factor 7	rs1131665	exonic
IRF4	6	IFN regulatory factor 4	rs9328192	intergenic
ITGAM	16	Integrin subunit alpha M	rs11859349	intronic
PRDM1	6	PR/SET domain 1	rs4134466	intergenic
TNFAIP3	6	Tumor necrosis factor alpha-induced protein	rs2230926	exonic
TNIP1	5	TNFAIP3-interacting protein	rs3792783	intronic
NFKB1	4	Nuclear factor kappa B subunit 1	rs230534	intronic
Adaptive immune response: B and T cell proliferation, survival and cytokine production				
TNFSF4	1	Tumor necrosis factor ligand superfamily member 4	rs11576547	ncRNA_intronic
CD247	1	T cell receptor zeta-chain	rs2056626	intronic
CSK	15	C-Src	rs1378942	intronic
PTPN22	1	Protein tyrosine phosphatase, non-receptor type 22	rs1970559	ncRNA_intronic
STAT4	2	Signal Transducer and activator of transcription 4	rs3821236	intronic
TYK2	19		rs34536443	exonic
IL12A	3	Interleukin 12A	rs589446, rs77583790	ncRNA_intronic
IL12RB2	1	Interleukin 12 receptor, beta 2	rs3790566	intronic
IL12RB1	19	Interleukin 12 receptor, beta 1	rs436857	UTR5
BLK	8	BLK proto-oncogene, Src family tyrosine kinase	rs2736340	intergenic

Table 3. Susceptibility *loci* for systemic sclerosis outside the *HLA* region.

Locus	Chr	Gene name	Index SNP	SNP function
Apoptosis, autophagy, fibrosis and others				
DNASE1L3	3	Deoxyribonuclease I-like 3	rs35677470	exonic
ATG5	6	Autophagy related 5	rs9373839	intronic
TIMP4-PPARG	3	Peroxisome proliferator-activated receptor gamma	rs310746	intergenic
JAZF1	7	JAZF zinc finger 1	rs849139	intronic
SOX5	6	SRY (sex determining region Y)-box 5	rs10734732	intronic
IKZF3-GSDMB	17	IKAROS family zinc finger 3 Gasdermin B	rs883770	intronic
GSDMA	17	Gasdermin A	rs3894194	exonic
GRB10	7	Growth factor receptor bound protein 10	rs12540874	intronic

Note: This table includes all the SSc *loci* associated at the genome-wide significant level. It also includes previously reported associations that did not reach the genome-wide significance level but have been confirmed in the initial phase of our large meta-GWAS (mentioned at point 4 of the Discussion) at least with a *p*-value < 5×10^{-6}.

production and pathogen elimination (201). Thus, it is not surprising that defects in this destructive mechanism may modulate the onset and outcome of SSc.

Although Table 3 does not include the associations described in the HLA region for SSc, it is important to note that this region represents the strongest association for SSc (107), as well as for other ADs (106). This implies that disease-associated *MHC* alleles must be responsible for an important portion of the susceptibility to autoimmune processes. The proposed mechanisms that may connect the development of autoimmunity and the HLA molecules are diverse: Disease-associated polymorphisms that may enable the presentation of key self-antigens; disease-associated polymorphisms that result in poor presentation of critical epitopes that may give rise to the escape from thymic tolerance mechanisms; key polymorphisms exerting effects on the T cell repertoire; or disease-associated *MHC* alleles that may present not only relevant self-peptides, but also microbial peptides that enable expansion and activation of relevant self-reactive T cells (202).

6.1 Functional characterization of disease-associated variants

The conversion of statistical associations of disease-associated variants into the functional consequences is essential to provide insight into the pathological mechanisms that lead to a certain complex disease. However, connecting risk alleles to molecular traits is not a trivial task. We have previously stated that the lead signal for a *locus* is not necessarily the functional genetic variant, known as the causal variant. Thus the first challenge is to identify the true causal variants that are responsible for the genetic associations. Once the variants are

identified, the next step is to try to connect the genetic markers to likely target genes.

The development of ambitious projects focused on improving the functional characterization of human genome and epigenome, including the ENCODE, the NIH Roadmap Epigenomics and GTEx, has offered a valuable source of data for SNP functional annotation and prioritization (95, 96, 136). Therefore, taking advantage of this useful information, we would like to explore the functional roles of SSc-associated variants. For this purpose, we first conducted functional annotation using wANNOVAR (203). Our input list of SNPs included not only the index SNP for each *locus*, but also all the proxy SNPs showing a $r^2 > 0.8$ in the 1000 Genomes Project CEU population with the top SNP in each genetic region (80). The index SNP was selected according to the results of our ongoing SSc large meta-GWAS. In addition, we also considered the SSc Immunochip results from our group and the trans-ethnic meta-GWAS performed by Terao *et al*. (107, 173). If the lead SNP for a certain *locus* in our SSc meta-GWAS was not in high LD with the lead SNP in the SSc Immunochip or in Terao *et al*., both genetic variants underwent our analysis.

As it can be observed in Figure 12, the vast majority of interrogated SNPs are non-coding variants: 55.3% intronics, 20.89% intergenic variants. Only 3.07% (19 SNPs) of annotated SNPs mapped in exons. Moreover, not all of them affect the amino acid sequence: only 1.84% (12 SNPs) of all the analyzed SNPs represented non-synonymous variants. Interestingly, these 12 non-synonymous variants comprise 8 different *loci* (Figure 13): *IRF7*, *ITGAM*, *IKZF3-GSDMB*, *GSDMA*, *TYK2*, *IL12RB1*, *DNASE1L3* and *TNFAIP3*. This means that approximately 30% of the SSc susceptibility *loci* are linked to missense mutations, which are likely to represent the causal variants. It is important to note that

the functional effect of some of these variants has been addressed. For example, *in vitro* studies have shown that rs35677470 minor allele at *DNASE1L3* leads to an inactive form of the protein that lacks its DNase activity (204). Similarly, *TYK2* rs34536443 minor allele leads to a near-complete loss of TYK2 catalitic function, and consequently it impairs cytokine signaling (158), as we have widely discussed in a previous section. Another example is the non-synonymous SNP at *TNFAIP3* that results in a phenylalanine-to-cysteine change that reduces the inhibitory activity of TNFAIP3 at the NF-κB signalling pathway (205).

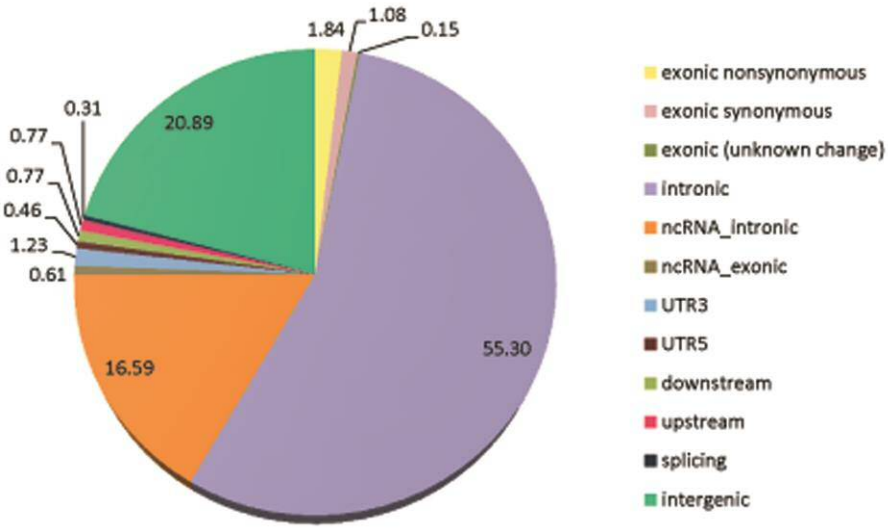


Figure 12. Functional classification of SSc-associated SNPs and their proxies according to the functional annotation performed with wANNOVAR. Numbers indicate the percentage of SNPs in each category.

In addition to variants that mapped in introns and intergenic regions, we also found a number of SNPs located in non-coding RNAs (ncRNAs), in UTR3' or UTR5' regions, upstream or downstream the genes and in splicing sites. In fact, a considerable proportion of the

SNPs (17.20%) mapped in ncRNAs. This suggests that variants affecting mRNA processing or stability may provide additional insights into the regulatory mechanisms affecting expression of disease-implicated *loci*. Our results are in accordance with emerging evidence that suggest a role of ncRNAs in autoimmunity (206). Similarly, there are also increasing evidence for the role of genetic variants underlying transcript splicing (splicing quantitative trait *loci* or sQTLs) in common diseases (207). According to our data, SNPs that control transcript isoforms -for instances, disrupting consensus splice-site sequences – may also being involved into the functional mechanisms underlying some of the genetic associations with the disease, such as *IL12RB1*.

Taking into account that most of the SSc-associated SNPs are linked to regulatory functions rather than affecting the encoded protein function themselves, we wanted to further explore their regulatory effects. For this purpose, we underwent eQTL analysis using the same workflow as we explained in section 4 of the Discussion. In addition, we also interrogated overlap of SNPs with chromatin marks of active enhancers (H3K4me1, H3K27ac) and active promoters (H3K4me3, H3K9ac) (208), DNase hypersensitivity sites and TF-binding sites using data from the NIH Roadmap Epigenomics Project (96). The overlap with chromatin marks and DNase hypersensitivity sites was performed by using relevant cell types for the disease, such as primary T helper cells, primary B cells, primary CD8+ cells, monocytes and fibroblasts (see figure footnote of Figure 14). These complementary analyses are especially relevant for those *loci* that comprise only intronic variants (*IL12RB2*, *SOX5*, *STAT4*, *TNIP1*, *JAZF1*, and *GRB10*) and/or intergenic signals (*NFKB1*, *IRF8*, *IRF4*, and *PRDM1*) (Figure 13).

As it can be observed from Figure 14, the vast majority of the SSc-index SNPs (or their proxies) overlap with promoter and enhancer

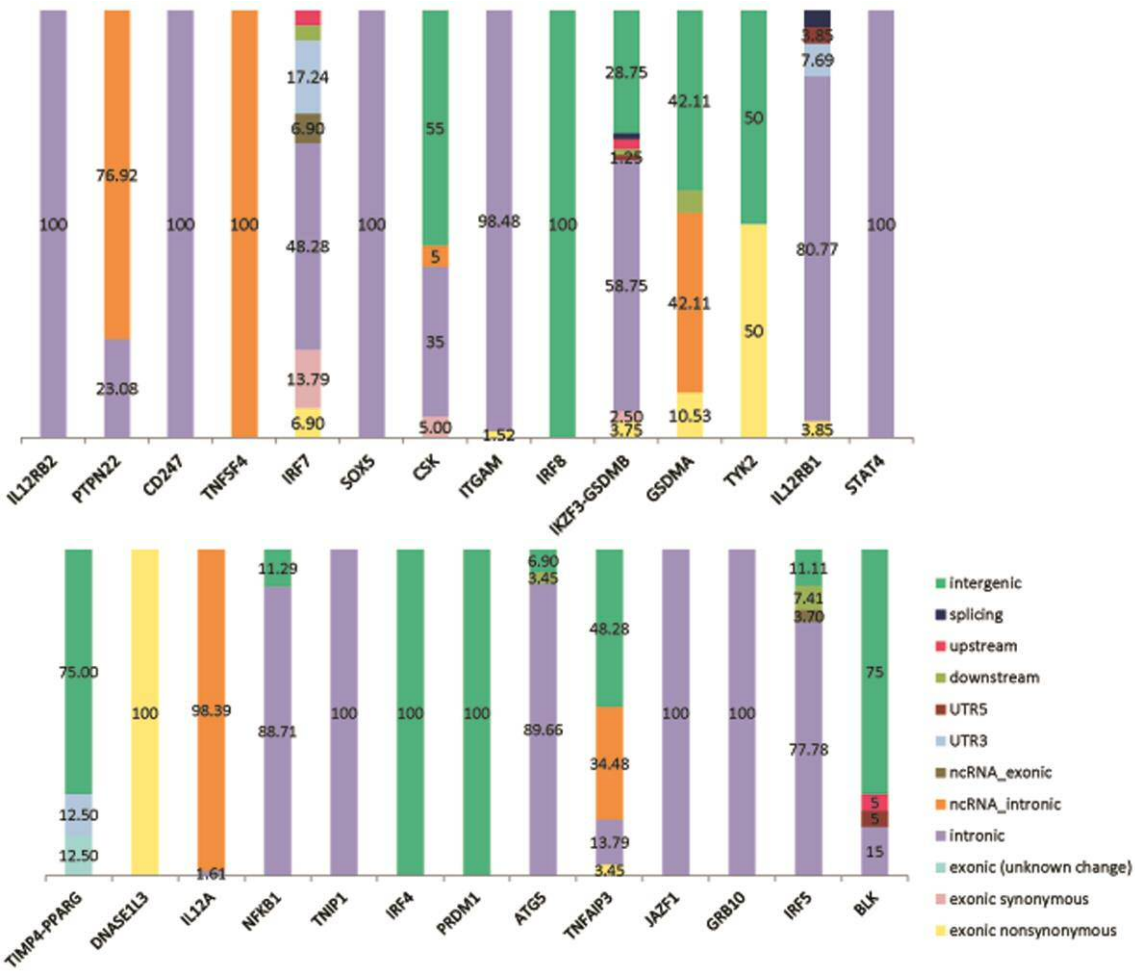


Figure 13. *Locus*-specific functional classification of SSc-associated SNPs and their proxies according to the functional annotation performed with wANNOVAR. Numbers indicate the percentage of SNPs in each category.

histone marks in the interrogated cell types. These results confirm that most of the genetic variations involved in the susceptibility to SSc modulate transcriptional regulatory mechanisms. In this regard, we have found that many of the interrogated variants correlate with eQTLs thus altering gene expression (71.43% of the *loci* has 'eQTl marks'). It is

important to note that many of the SNPs correlate with eQTLs for the *a priori* candidate gene. Nonetheless, we also found genetic variants affecting both *a priori* candidate gene and additional genes, as well as SNPs that were only eQTLs for a different gene from the selected candidate gene. As an example, *DNASE1L3* rs35677470 missense variant - that leads to an inactive enzyme - correlates with eQTL for the neighbour genes *PXK* and RP11-802O23.3. These observations highlight that assigning association signals to the 'closest gene' is not a suitable strategy for some SNPs and that the functional role of certain signals may spread out to different target genes.

In addition, Figure 14 shows that exonic variants can also overlap with epigenetic marks. These observations would mean that certain DNA sequences can have a dual function. In fact, it has been described that there are coding exons that act as coding exons in one tissue and function as enhancers of nearby genes in a different tissue (209).

The index SNPs of some SSc-associated *loci* did not show *a priori* any interesting functional annotation by themselves, such as the top SNPs for *TNFSF4*, *ITGAM* and *SOX5*. Therefore, these SNPs could be initially discarded as the 'causal variants' for the associations. Curiously, in the case of *ITGAM*, the index SNP is in high LD with a missense variant and an intronic deletion. Although the functional consequences of these two other variants are unknown, they would probably add clues to the functional mechanism underlying *ITGAM* association. Interestingly, our functional annotation of the index SNPs and their proxies showed that there are several SSc *loci* that are linked to a number of different 'SNP categories'. For example, the index SNP at *IL12RB1* is a promoter variant that is in high LD with several proxies that involved additional 'SNP categories': exonic non-synonymous,

intronics, UTR3' and splicing variants. The regulatory effect of the promoter index SNP was addressed by Takahashi *et al*. and they showed that the minor allele of the promoter variant decreased *IL12RB1* mRNA levels (142). However, the remaining 'SNP categories' linked to this *locus* are also interesting, especially the splicing variant rs393548 that leads to an alternative 3' acceptor splice site resulting in a different transcript. This example illustrates that the regulatory mechanisms underlying a genetic association is not always straightforward to address (210).

It has been demonstrated that variants associated with the same disease tend to overlap with cell type-specific chromatin marks (183). Considering that our chromatin mark analysis was performed in relevant cell types for SSc, it is not surprising the large colocalization that we have found for our SNP panel and the chromatin marks interrogated. More importantly, the overlap of disease-SNPs and cell-type specific epigenetic marks add valuable information to identify critical cell types for a disease (183). This information will help to direct future functional studies in accurately chosen cell types, thus increasing the experiment success and helping geneticists to move from genetic associations to disease genes and mechanisms.

SSc LOCUS	Chr	Index SNP	SNP function	Exonic nonsynonymous	eQTL	H3K4me1_Enh	H3K27ac_Enh	H3K4me3_Pro	H3K9ac_Pro	DNase	MOTIF
TNFSF4	1	rs11576547	ncRNA_intronic								
CD247	1	rs2056626	intronic								
PTPN22	1	rs1970559	ncRNA_intronic								
IL12RB2	1	rs3790566	intronic								
STAT4	2	rs3821236	intronic								
IL12A	3	rs589446	ncRNA_intronic								
	3	rs77583790	ncRNA_intronic								
DNASE1L3	3	rs35677470	exonic								
TIMP4-PPARG	3	rs310746	intergenic								
NFKB1	4	rs230534	intronic								
TNIP1	5	rs3792783	intronic								
IRF4	6	rs9328192	intergenic								
PRDM1	6	rs4134466	intergenic								
TNFAIP3	6	rs2230926	exonic								
ATG5	6	rs9373839	intronic								
SOX5	6	rs10734732	intronic								
IRF5	7	rs36073657	intronic								
JAZF1	7	rs849139	intronic								
GRB10	7	rs12540874	intronic								
BLK	8	rs2736340	intergenic								
IRF7	11	rs1131665	exonic								
CSK	15	rs1378942	intronic								
IRF8	16	rs11117420	intergenic								
ITGAM	16	rs11859349	intronic								
IKZF3-GSDMB	17	rs883770	intronic								
GSDMA	17	rs3894194	exonic								
TYK2	19	rs34536443	exonic								
IL12RB1	19	rs436857	UTR5								

Figure 14. Functional characterization of SSc-associated *loci*. In each category, dark colours represent overlap with lead SNPs, and light colours indicate overlap with proxy SNPs. The following cells were used to identify overlap of SNPs with chromatin marks of active enhancers (H3K4me1, H3K27ac) and active promoters (H3K4me3, H3K9ac), DNase hypersensitivity sites and TF-binding sites using data from the NIH Roadmap Epigenomics Project: Primary T cells , primary T helper cells , primary T helper 17 cells PMA-I stimulated, primary T helper memory cells , Primary T CD8+ memory cells, primary T helper naive cells, primary T CD8+ naive cells, primary monocytes, primary B cells (all of them from peripheral blood), foreskin fibroblast primary cells skin, dermal fibroblast primary cells, epidermal keratinocyte primary cells.

VI

FUTURE DIRECTIONS

Looking back to the past ten years, our knowledge of the genetic basis underlying SSc has considerably increased. Despite that, we are aware that this field of research needs to keep developing in different directions.

We have widely discussed the importance of enough sample size to reach a high statistical power to robustly detect association signals. Currently, the largest GWAS that has been performed in SSc comprised 2,296 cases and 5,171 healthy controls. Thus, an obvious step is to increase the number of genome-wide genotyped individuals to keep identifying new susceptibility *loci*. In this regard, our group is currently performing a large meta-GWAS for SSc that will include the cohorts of our previous GWAS (98) and more than 6,800 new cases and 11,300 new controls of European ancestry. With this, the large SSc meta-GWAS will reach a sample size of around 9,100 cases and 16,500 controls. Thus, this study will allow us to increase the number of genome-wide genotyped cases and controls for 4 times and 3 times, respectively, considerably increasing the power to detect new association signals.

Figure 15 shows the Manhattan plots from our first GWAS (98) and the initial phase of the large SSc meta-GWAS (which in total comprised 7,828 SSc patients and 14,523 healthy controls). As it can be observed, the number of *loci* reaching the genome-wide significant level has dramatically increased from 3 to 16 signals. Therefore, the final large meta-GWAS will definitely allow us to identify new robust susceptibility *loci* and to confirm or discard risk *loci* previously reported for the disease. Altogether, this large study will help to accurately draw the picture of the genetic basis for SSc.

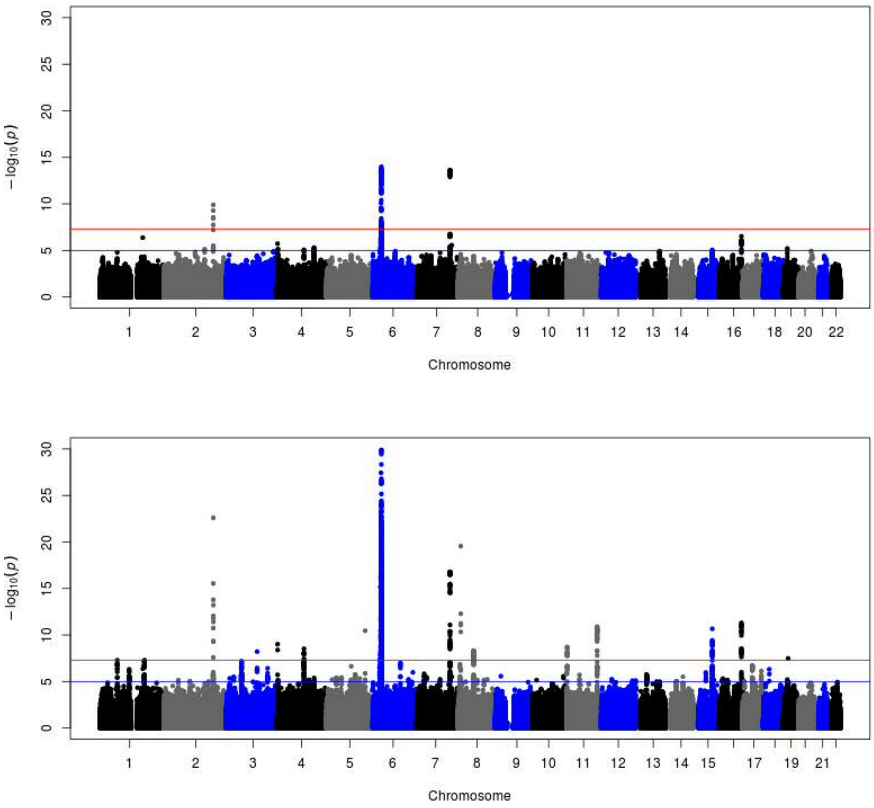


Figure 15. Comparison of Manhattan plots from the first GWAS in systemic sclerosis in European population (2,296 cases and 5,171 healthy controls) and initial phase of our large meta-GWAS (7,828 SSc patients and 14,523 healthy controls). The plots have been truncated at *p*-value <1 x 10^{-30}. Red and blue lines indicate *p*-value thresholds at 5 x 10^{-08} and 5 x 10^{-05}, respectively.

As we mentioned in the Introduction section, SSc is a clinically heterogeneous disease, with two main subgroups (lSSc and dSSc) that differ in the behavior of the disease. It could be expected that the differences among the clinical sub-phenotypes may also being related to specific genetic players in each subtype. In fact, this differential genetic background has already been addressed (211). Therefore, other

interesting directions would be the stratified analysis according to clinical subgroups in order to identify specific susceptibility *loci* for each subtype, which may also help to identify prognostic markers for the disease. In this way, it has been recently demonstrated the existence of a different genetic background between prognosis and susceptibility in Crohn's disease (212). Thus, a similar approach in SSc would be interesting as well.

A critical next step would be the refinement of the association signals to identify the causal SNPs and the target genes. NGS focused on SSc risk *loci* would be a good strategy to capture all genetic variation present on the risk *loci*, along with fine-mapping and accurate imputation with extensive reference panels. In addition, the functional characterization of the most likely causal variants responsible for the association signals is a necessary step in order to translate our statistical genetic findings into pathogenic pathways. Bioinformatic annotation with of all the available functional data will help to prioritize variants, but in turn it may also help to elucidate altered pathways or to pinpoint regions of higher interest.

The assignment of association signals to the 'closest genes' is not always a suitable strategy. This issue is, in part, due to the three-dimensional conformation of the genome, which contains long-range interactions that bring into close proximity distant genomic regions. The investigation of this three-dimensional conformation by approaches such as Capture Hi-C allows clarifying the interactions between risk variants and their functional targets in a specific cell-type and cell-state context. Capture Hi-C has already been applied to autoimmune risk *loci,* and has allowed expanding some of the reported associations to target genes located far away from the association peaks (213, 214).

Overall, by applying these methodologies, we may expect a tremendous amount of information on disease biology that would help to draw predictive models of disease risk, to the identification of diagnostic and prognostic markers, and new putative therapeutic targets. We hope that the efforts behind the present thesis help to achieve these aims, contributing to health care and life quality of the patients in the future.

VII

CONCLUSIONS

1. *PPARG*, *IL12RB1*, *TYK2* and *IRF4* have been identified as novel susceptibility *loci* for systemic sclerosis for the first time through the studies presented in this thesis. Three out of the four *loci* reached the genome-wide significance level ($p < 5 \times 10^{-08}$)

2. The association reported at *PPARG locus* almost reached the genome-wide significance level. eQTL analysis provided functional evidence for the role of *TIMP4* in the pathogenesis of the disease. Both *PPARG* and *TIMP4* represent susceptibility genes related to fibrotic processes.

3. The identification of *IL12RB1* and *TYK2* as susceptibility genes for systemic sclerosis highlights the important role of the IL-12/IL-23 pathway in the predisposition to the disease. Moreover, these findings may give rise to new therapeutic lines for the treatment of the disease.

4. Our results discarded the role of *ATP8B4* F436L missense variant in the susceptibility to systemic sclerosis.

5. The analysis of the shared genetic component between systemic sclerosis and rheumatoid arthritis identified *IRF4* as novel shared risk factor. We also confirmed other SSc-RA common *loci* previously reported. Moreover, the pathway enrichment analysis identified type I interferon as one of the most relevant common pathways between systemic sclerosis and rheumatoid arthritis on the basis of their common genetic background.

6. The analysis of the functional downstream effects of the SSc-RA genetic variants suggests an extensive overlap both in their genetic component and in the etiopathogenic pathways underlying SSc and RA.

7. The functional characterization of the SSc associated variants showed that 30% of the SSc risk *loci* can be linked to exonic missense variants. However, our results demonstrate that most of the genetic variations involved in the susceptibility to the disease modulate several transcriptional regulatory mechanisms.

CONCLUSIONES

1. Los estudios que conforman la presenta tesis doctoral han permitido identificar por primera vez cuatro nuevos *loci* de susceptibilidad para la esclerosis sistémica, a saber, *PPARG*, *IL12RB1*, *TYK2* e *IRF4*. La asociación de tres de ellos, *IL12RB1*, *TYK2* e *IRF4*, alcanzó el nivel de significación establecido para los estudios de asociación de genoma completo, es decir un valor $p < 5 \times 10^{-08}$.

2. La asociación del *locus PPARG* rozó el nivel de significación establecido para los estudios de asociación de genoma completo. El análisis de eQTLs en este *locus* proporcionó evidencias funcionales para la implicación del gen *TIMP4* en la patogénesis de la esclerosis sistémica. Tanto *PPARG* como *TIMP4* representan *loci* the susceptibilidad relacionados con mecanismos fibróticos.

3. Los hallazgos en los *loci IL12RB1* y *TYK2* refuerzan el importante papel de la vía de señalización de la IL-12/IL23 en la predisposición a la esclerosis sistémica. Además, estos resultados ofrecen soporte genético para el interés de esta vía de señalización como nueva diana terapéutica en el tratamiento de la enfermedad.

4. Nuestros resultados descartaron el papel de la variante rara F436L de *ATP8B4* en la susceptibilidad a la esclerosis sistémica.

5. El análisis del componente genético común de la esclerosis sistémica y la artritis reumatoide identificó a *IRF4* como nuevo factor de riesgo compartido por ambas enfermedades autoinmunes, y confirmó varios *loci* comunes previamente descritos. Además, el análisis de rutas bioquímicas identificó la ruta del interferón de tipo I como una de las

vías patogénicas comunes más relevantes de acuerdo al componente genético compartico por ambas enfermedades.

6. El análisis de los efectos reguladores y funcionales de las variantes asociadas a la esclerosis sistémica y la artritis reumatoide localizadas en los factores de riesgo compartidos sugiere un extenso solapamiento no sólo en el componente genético sino también en los mecanismos etiopatogénicos subyacentes a ambas enfermedades.

7. La caracterización funcional de las variantes asociadas a la enfermedad reveló que el 30% de los *loci* de susceptibilidad descritos hasta el momento están ligados a variantes exónicas de pérdida de sentido. No obstante, nuestros análisis demuestran que la mayoría de las variaciones genéticas implicadas en la predisposición a la esclerosis sistémica ejercen su efecto a través de la modulación de la expresión génica.

VIII

BIBLIOGRAPHY

1. Rosenblum MD, Remedios KA, Abbas AK. Mechanisms of human autoimmunity. J Clin Invest. 2015;125:2228-33.

2. Cooper GS, Bynum ML, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun. 2009;33:197-207.

3. Hayter SM, Cook MC. Updated assessment of the prevalence, spectrum and case definition of autoimmune disease. Autoimmun Rev. 2012;11:754-65.

4. Marson A, Housley WJ, Hafler DA. Genetic basis of autoimmunity. J Clin Invest. 2015;125:2234-41.

5. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. N Engl J Med. 2009;360:1989-2003.

6. Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. Annu Rev Pathol. 2011;6:509-37.

7. Denton CP, Khanna D. Systemic sclerosis. Lancet. 2017. Epud ahead of print.

8. Steen VD, Medsger TA. Changes in causes of death in systemic sclerosis, 1972-2002. Ann Rheum Dis. 2007;66:940-4.

9. 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;23:581-90.

10. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol. 1988;15:202-5.

11. LeRoy EC, Medsger TA. Criteria for the classification of early systemic sclerosis. J Rheumatol. 2001;28:1573-6.

12. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, et al. 2013 classification criteria for systemic sclerosis: an American College

of Rheumatology/European League against Rheumatism collaborative initiative. Arthritis Rheum. 2013;65:2737-47.

13. Koenig M, Dieudé M, Senécal JL. Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. Autoimmun Rev. 2008;7:588-93.

14. Okano Y, Steen VD, Medsger TA. Autoantibody reactive with RNA polymerase III in systemic sclerosis. Ann Intern Med. 1993;119:1005-13.

15. Parker JC, Burlingame RW, Webb TT, Bunn CC. Anti-RNA polymerase III antibodies in patients with systemic sclerosis detected by indirect immunofluorescence and ELISA. Rheumatology (Oxford). 2008;47:976-9.

16. Amur S, Parekh A, Mummaneni P. Sex differences and genomics in autoimmune diseases. J Autoimmun. 2012;38:J254-65.

17. Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol. 2012;24:165-70.

18. Elhai M, Avouac J, Walker UA, Matucci-Cerinic M, Riemekasten G, Airò P, et al. A gender gap in primary and secondary heart dysfunctions in systemic sclerosis: a EUSTAR prospective study. Ann Rheum Dis. 2016;75(1):163-9.

19. Ranque B, Mouthon L. Geoepidemiology of systemic sclerosis. Autoimmun Rev. 2010;9:A311-8.

20. Reveille JD. Ethnicity and race and systemic sclerosis: how it affects susceptibility, severity, antibody genetics, and clinical manifestations. Curr Rheumatol Rep. 2003;5:160-7.

21. Arnett FC, Howard RF, Tan F, Moulds JM, Bias WB, Durban E, et al. Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. Arthritis Rheum. 1996;39:1362-70.

22. Chifflot H, Fautrel B, Sordet C, Chatelus E, Sibilia J. Incidence and prevalence of systemic sclerosis: a systematic literature review. Semin Arthritis Rheum. 2008;37:223-35.

23. Arias-Nuñez MC, Llorca J, Vazquez-Rodriguez TR, Gomez-Acebo I, Miranda-Filloy JA, Martin J, et al. Systemic sclerosis in northwestern Spain: a 19-year epidemiologic study. Medicine (Baltimore). 2008;87:272-80.

24. Simeón-Aznar CP, Fonollosa-Plá V, Tolosa-Vilella C, Espinosa-Garriga G, Ramos-Casals M, Campillo-Grau M, et al. Registry of the Spanish network for systemic sclerosis: clinical pattern according to cutaneous subsets and immunological status. Semin Arthritis Rheum. 2012;41:789-800.

25. Trojanowska M. Cellular and molecular aspects of vascular dysfunction in systemic sclerosis. Nat Rev Rheumatol. 2010;6:453-60.

26. Varga J, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. J Clin Invest. 2007;117:557-67.

27. Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Primers. 2015;1:15002.

28. Prescott RJ, Freemont AJ, Jones CJ, Hoyland J, Fielding P. Sequential dermal microvascular and perivascular changes in the development of scleroderma. J Pathol. 1992;166:255-63.

29. Kalogerou A, Gelou E, Mountantonakis S, Settas L, Zafiriou E, Sakkas L. Early T cell activation in the skin from patients with systemic sclerosis. Ann Rheum Dis. 2005;64:1233-5.

30. Assassi S, Mayes MD, Arnett FC, Gourh P, Agarwal SK, McNearney TA, et al. Systemic sclerosis and lupus: points in an interferon-mediated continuum. Arthritis Rheum. 2010;62:589-98.

31. Kim D, Peck A, Santer D, Patole P, Schwartz SM, Molitor JA, et al. Induction of interferon-alpha by scleroderma sera containing autoantibodies to topoisomerase I: association of higher interferon-alpha activity with lung fibrosis. Arthritis Rheum. 2008;58:2163-73.

32. York MR, Nagai T, Mangini AJ, Lemaire R, van Seventer JM, Lafyatis R. A macrophage marker, Siglec-1, is increased on circulating monocytes in patients with systemic sclerosis and induced by type I interferons and toll-like receptor agonists. Arthritis Rheum. 2007;56:1010-20.

33. Yurovsky VV, Wigley FM, Wise RA, White B. Skewing of the CD8+ T-cell repertoire in the lungs of patients with systemic sclerosis. Hum Immunol. 1996;48:84-97.

34. Sakkas LI, Xu B, Artlett CM, Lu S, Jimenez SA, Platsoucas CD. Oligoclonal T cell expansion in the skin of patients with systemic sclerosis. J Immunol. 2002;168:3649-59.

35. Farge D, Henegar C, Carmagnat M, Daneshpouy M, Marjanovic Z, Rabian C, et al. Analysis of immune reconstitution after autologous bone marrow transplantation in systemic sclerosis. Arthritis Rheum. 2005;52:1555-63.

36. O'Reilly S, Hügle T, van Laar JM. T cells in systemic sclerosis: a reappraisal. Rheumatology (Oxford). 2012;51:1540-9.

37. O'Reilly S. Role of interleukin-13 in fibrosis, particularly systemic sclerosis. Biofactors. 2013;39:593-6.

38. Wynn TA. Fibrotic disease and the T(H)1/T(H)2 paradigm. Nat Rev Immunol. 2004;4:583-94.

39. Baraut J, Michel L, Verrecchia F, Farge D. Relationship between cytokine profiles and clinical outcomes in patients with systemic sclerosis. Autoimmun Rev. 2010;10:65-73.

40. Brembilla NC, Montanari E, Truchetet ME, Raschi E, Meroni P, Chizzolini C. Th17 cells favor inflammatory responses while inhibiting type I collagen deposition by dermal fibroblasts: differential effects in healthy and systemic sclerosis fibroblasts. Arthritis Res Ther. 2013;15:R151.

41. Bosello S, De Luca G, Tolusso B, Lama G, Angelucci C, Sica G, et al. B cells in systemic sclerosis: a possible target for therapy. Autoimmun Rev. 2011;10(10):624-30.

42. Khan K, Xu S, Nihtyanova S, Derrett-Smith E, Abraham D, Denton CP, et al. Clinical and pathological significance of interleukin 6 overexpression in systemic sclerosis. Ann Rheum Dis. 2012;71:1235-42.

43. Kitaba S, Murota H, Terao M, Azukizawa H, Terabe F, Shima Y, et al. Blockade of interleukin-6 receptor alleviates disease in mouse model of scleroderma. Am J Pathol. 2012;180:165-76.

44. Yoshizaki A, Yanaba K, Ogawa A, Asano Y, Kadono T, Sato S. Immunization with DNA topoisomerase I and Freund's complete adjuvant induces skin and lung fibrosis and autoimmunity via interleukin-6 signaling. Arthritis Rheum. 2011;63:3575-85.

45. Varga J, Pasche B. Transforming growth factor beta as a therapeutic target in systemic sclerosis. Nat Rev Rheumatol. 2009;5:200-6.

46. Leask A. Scar wars: is TGFbeta the phantom menace in scleroderma? Arthritis Res Ther. 2006;8:213.

47. Manetti M, Guiducci S, Matucci-Cerinic M. The origin of the myofibroblast in fibroproliferative vasculopathy: does the endothelial cell steer the pathophysiology of systemic sclerosis? Arthritis Rheum. 2011;63:2164-7.

48. Gabrielli A, Svegliati S, Moroncini G, Pomponio G, Santillo M, Avvedimento EV. Oxidative stress and the pathogenesis of scleroderma: the Murrell's hypothesis revisited. Semin Immunopathol. 2008;30:329-37.

49. Servettaz A, Goulvestre C, Kavian N, Nicco C, Guilpain P, Chéreau C, et al. Selective oxidation of DNA topoisomerase 1 induces systemic sclerosis in the mouse. J Immunol. 2009;182:5855-64.

50. Sambo P, Baroni SS, Luchetti M, Paroncini P, Dusi S, Orlandini G, et al. Oxidative stress in scleroderma: maintenance of scleroderma fibroblast phenotype by the constitutive up-regulation of reactive oxygen species generation through the NADPH oxidase complex pathway. Arthritis Rheum. 2001;44:2653-64.

51. Servettaz A, Guilpain P, Goulvestre C, Chéreau C, Hercend C, Nicco C, et al. Radical oxygen species production induced by advanced oxidation protein products predicts clinical evolution and response to treatment in systemic sclerosis. Ann Rheum Dis. 2007;66:1202-9.

52. Bramwell B. Diffuse scleroderma: its frequency, its occurrence in stonemasons, its treatment by fibrinolysin, elevation of temperature do to fibrinolysin injections. Edinburgh Med J. 1914;12:387–401.

53. LD E. Scleroderma in goldminers on the Witwatersrand with particular reference to pulmonary manifestations. S Afr Lab Clin Med 1957;3:209–31.

54. Sluis-Cremer GK, Hessel PA, Nizdo EH, Churchill AR, Zeiss EA. Silica, silicosis, and progressive systemic sclerosis. Br J Ind Med. 1985;42:838-43.

55. Silman AJ, Jones S. What is the contribution of occupational environmental factors to the occurrence of scleroderma in men? Ann Rheum Dis. 1992;51:1322-4.

56. Englert H, Small-McMahon J, Davis K, O'Connor H, Chambers P, Brooks P. Male systemic sclerosis and occupational silica exposure-a population-based study. Aust N Z J Med. 2000;30:215-20.

57. Diot E, Lesire V, Guilmot JL, Metzger MD, Pilore R, Rogier S, et al. Systemic sclerosis and occupational risk factors: a case-control study. Occup Environ Med. 2002;59:545-9.

58. Calvert GM, Rice FL, Boiano JM, Sheehy JW, Sanderson WT. Occupational silica exposure and risk of various diseases: an analysis using death certificates from 27 states of the United States. Occup Environ Med. 2003;60:122-9.

59. Maeda M, Nishimura Y, Kumagai N, Hayashi H, Hatayama T, Katoh M, et al. Dysregulation of the immune system caused by silica and asbestos. J Immunotoxicol. 2010;7:268-78.

60. Lee S, Hayashi H, Maeda M, Chen Y, Matsuzaki H, Takei-Kumagai N, et al. Environmental factors producing autoimmune dysregulation--chronic activation of T cells caused by silica exposure. Immunobiology. 2012;217:743-8.

61. Otsuki T, Hayashi H, Nishimura Y, Hyodo F, Maeda M, Kumagai N, et al. Dysregulation of autoimmunity caused by silica exposure and alteration of Fas-mediated apoptosis in T lymphocytes derived from silicosis patients. Int J Immunopathol Pharmacol. 2011;24:11S-6S.

62. Miller FW, Alfredsson L, Costenbader KH, Kamen DL, Nelson LM, Norris JM, 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;39:259-71.

63. Marie I, Gehanno JF, Bubenheim M, Duval-Modeste AB, Joly P, Dominique S, et al. Prospective study to evaluate the association between systemic sclerosis and occupational exposure and review of the literature. Autoimmun Rev. 2014;13:151-6.

64. Marie I, Menard JF, Duval-Modeste AB, Joly P, Dominique S, Bravard P, et al. Association of occupational exposure with features of systemic sclerosis. J Am Acad Dermatol. 2015;72:456-64.

65. Marie I, Gehanno JF. Environmental risk factors of systemic sclerosis. Semin Immunopathol. 2015;37:463-73.

66. Patterson R, Germolec D. Review article toxic oil syndrome: review of immune aspects of the disease. J Immunotoxicol. 2005;2:51-8.

67. Cassinotti P, Burtonboy G, Fopp M, Siegl G. Evidence for persistence of human parvovirus B19 DNA in bone marrow. J Med Virol. 1997;53:229-32.

68. Pandey JP, LeRoy EC. Human cytomegalovirus and the vasculopathies of autoimmune diseases (especially scleroderma), allograft rejection, and coronary restenosis. Arthritis Rheum. 1998;41:10-5.

69. Lunardi C, Bason C, Navone R, Millo E, Damonte G, Corrocher R, et al. Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. Nat Med. 2000;6:1183-6.

70. Farina A, Cirone M, York M, Lenna S, Padilla C, McLaughlin S, et al. Epstein-Barr virus infection induces aberrant TLR activation pathway and fibroblast-myofibroblast conversion in scleroderma. J Invest Dermatol. 2014;134:954-64.

71. Maul GG, Jimenez SA, Riggs E, Ziemnicka-Kotula D. Determination of an epitope of the diffuse systemic sclerosis marker antigen DNA topoisomerase I: sequence similarity with retroviral p30gag protein suggests a possible cause for autoimmunity in systemic sclerosis. Proc Natl Acad Sci U S A. 1989;86:8492-6.

72. Adams Waldorf KM, Nelson JL. Autoimmune disease during pregnancy and the microchimerism legacy of pregnancy. Immunol Invest. 2008;37:631-44.

73. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis Rheum. 2001;44:1359-62.

74. Feghali-Bostwick C, Medsger TA, Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. Arthritis Rheum. 2003;48:1956-63.

75. Assassi S, Arnett FC, Reveille JD, Gourh P, Mayes MD. Clinical, immunologic, and genetic features of familial systemic sclerosis. Arthritis Rheum. 2007;56:2031-7.

76. Bossini-Castillo L, Lopez-Isac E, Martin J. Immunogenetics of systemic sclerosis: Defining heritability, functional variants and shared-autoimmunity pathways. J Autoimmun. 2015;64:53-65.

77. Lee SH, Wray NR, Goddard ME, Visscher PM. Estimating missing heritability for disease from genome-wide association studies. Am J Hum Genet. 2011;88:294-305.

78. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 2010;42:565-9.

79. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76-82.

80. Genomes Project C, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491:56-65.

81. Lewis CM. Genetic association studies: design, analysis and interpretation. Brief Bioinform. 2002;3:146-53.

82. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature. 2001;409:860-921.

83. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931-45.

84. International HapMap C. A haplotype map of the human genome. Nature. 2005;437:1299-320.

85. International HapMap C, Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:851-61.

86. International HapMap C, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. Nature. 2010;467:52-8.

87. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, et al. Complement factor H polymorphism in age-related macular degeneration. Science. 2005;308:385-9.

88. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet Epidemiol. 2008;32:381-5.

89. Marchini J, Howie B. Genotype imputation for genome-wide association studies. Nat Rev Genet. 2010;11:499-511.

90. de Bakker PI, Ferreira MA, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum Mol Genet. 2008;17:R122-8.

91. Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010;11:446-50.

92. Cortes A, Brown MA. Promise and pitfalls of the Immunochip. Arthritis Res Ther. 2011;13:101.

93. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003;31:3812-4.

94. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. Curr Protoc Hum Genet. 2013;Chapter 7:Unit7 20.

95. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57-74.

96. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28:1045-8.

97. Bossini-Castillo L, López-Isac E, Martín J. Immunogenetics of systemic sclerosis: Defining heritability, functional variants and shared-autoimmunity pathways. J Autoimmun. 2015;64:53-65.

98. Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010;42:426-9.

99. Dieudé P, Boileau C, Guedj M, Avouac J, Ruiz B, Hachulla E, et al. Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. Ann Rheum Dis. 2011;70:1695-6.

100. Allanore Y, Saad M, Dieudé P, Avouac J, Distler JH, Amouyel P, et al. Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet. 2011;7:e1002091.

101. Bossini-Castillo L, Martin JE, Broen J, Simeon CP, Beretta L, Gorlova OY, 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;72:602-7.

102. Zhou X, Lee JE, Arnett FC, Xiong M, Park MY, Yoo YK, et al. HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. Arthritis Rheum. 2009;60:3807-14.

103. Bossini-Castillo L, Martin JE, Broen J, Gorlova O, Simeón CP, Beretta 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;21:926-33.

104. Martin JE, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Hum Mol Genet. 2012;21:2825-35.

105. López-Isac E, Bossini-Castillo L, Simeon CP, Egurbide MV, Alegre-Sancho JJ, Callejas JL, et al. A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility. Arthritis Res Ther. 2014;16:R6.

106. Parkes M, Cortes A, van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013;14:661-73.

107. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 2014;94:47-61.

108. Zochling J, Newell F, Charlesworth JC, Leo P, Stankovich J, Cortes A, et al. An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. Arthritis Res Ther. 2014;16:438.

109. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet. 2011;7:e1002004.

110. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS Genet. 2011;7:e1001283.

111. Ellinghaus D, Ellinghaus E, Nair RP, Stuart PE, Esko T, Metspalu A, et al. Combined analysis of genome-wide association studies for Crohn disease and psoriasis identifies seven shared susceptibility loci. Am J Hum Genet. 2012;90:636-47.

112. Li YR, Li J, Zhao SD, Bradfield JP, Mentch FD, Maggadottir SM, et al. Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. Nat Med. 2015;21:1018-27.

113. Martin JE, Assassi S, Diaz-Gallo LM, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum Mol Genet. 2013;22:4021-9.

114. Marquez A, Vidal-Bralo L, Rodriguez-Rodriguez L, Gonzalez-Gay MA, Balsa A, Gonzalez-Alvaro I, et al. A combined large-scale meta-analysis identifies COG6 as a novel shared risk locus for rheumatoid arthritis and systemic lupus erythematosus. Ann Rheum Dis. 2017;76:286-94.

115. Lopez-Isac E, Martin JE, Assassi S, Simeon CP, Carreira P, Ortego-Centeno N, et al. Brief Report: IRF4 Newly Identified as a Common Susceptibility Locus for Systemic Sclerosis and Rheumatoid Arthritis in a Cross-Disease Meta-Analysis of Genome-Wide Association Studies. Arthritis Rheumatol. 2016;68:2338-44.

116. Okada Y, Wu D, Trynka G, Raj T, Terao C, Ikari K, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014;506:376-81.

117. Bentham J, Morris DL, Cunninghame Graham DS, Pinder CL, Tombleson P, Behrens TW, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet. 2015;47:1457-64.

118. Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. Nat Rev Genet. 2016;17:160-74.

119. Evangelou E, Ioannidis JP. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet. 2013;14:379-89.

120. Lopez-Isac E, Bossini-Castillo L, Guerra SG, Denton C, Fonseca C, Assassi S, et al. Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus. Arthritis Rheumatol. 2014;66:3521-3.

121. Lopez-Isac E, Campillo-Davo D, Bossini-Castillo L, Guerra SG, Assassi S, Simeon CP, et al. Influence of TYK2 in systemic sclerosis susceptibility: a new locus in the IL-12 pathway. Ann Rheum Dis. 2016;75:1521-6.

122. Lopez-Isac E, Bossini-Castillo L, Palma AB, Assassi S, Simeon CP, Ortego-Centeno N, et al. Analysis of ATP8B4 F436L missense variant in a large Systemic sclerosis cohort. Arthritis Rheumatol. 2017.

123. Sham PC, Purcell SM. Statistical power and significance testing in large-scale genetic studies. Nat Rev Genet. 2014;15:335-46.

124. Tontonoz P, Spiegelman BM. Fat and beyond: the diverse biology of PPARgamma. Annu Rev Biochem. 2008;77:289-312.

125. Ghosh AK, Bhattacharyya S, Lakos G, Chen SJ, Mori Y, Varga J. Disruption of transforming growth factor beta signaling and profibrotic responses in normal skin fibroblasts by peroxisome proliferator-activated receptor gamma. Arthritis Rheum. 2004;50:1305-18.

126. Burgess HA, Daugherty LE, Thatcher TH, Lakatos HF, Ray DM, Redonnet M, 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:L1146-53.

127. Kulkarni AA, Thatcher TH, Olsen KC, Maggirwar SB, Phipps RP, Sime PJ. PPAR-gamma ligands repress TGFbeta-induced myofibroblast differentiation by targeting the PI3K/Akt pathway: implications for therapy of fibrosis. PLoS One. 2011;6:e15909.

128. Wei J, Zhu H, Komura K, Lord G, Tomcik M, Wang W, et al. A synthetic PPAR-gamma agonist triterpenoid ameliorates experimental fibrosis: PPAR-gamma-independent suppression of fibrotic responses. Ann Rheum Dis. 2014;73:446-54.

129. Wu M, Melichian DS, Chang E, Warner-Blankenship M, Ghosh AK, Varga J. Rosiglitazone abrogates bleomycin-induced scleroderma and blocks profibrotic responses through peroxisome proliferator-activated receptor-gamma. Am J Pathol. 2009;174:519-33.

130. Ruzehaji N, Frantz C, Ponsoye M, Avouac J, Pezet S, Guilbert T, et al. Pan PPAR agonist IVA337 is effective in prevention and treatment of experimental skin fibrosis. Ann Rheum Dis. 2016;75:2175-83.

131. Wei J, Ghosh AK, Sargent JL, Komura K, Wu M, Huang QQ, et al. PPARgamma downregulation by TGFss in fibroblast and impaired expression and function in systemic sclerosis: a novel mechanism for progressive fibrogenesis. PLoS One. 2010;5:e13778.

132. Marangoni RG, Korman BD, Allanore Y, Dieude P, Armstrong LL, Rzhetskaya M, et al. A candidate gene study reveals association between a

variant of the Peroxisome Proliferator-Activated Receptor Gamma (PPAR-gamma) gene and systemic sclerosis. Arthritis Res Ther. 2015;17:128.

133. Glas J, Seiderer J, Markus C, Pfennig S, Wetzke M, Paschos E, et al. Role of PPARG gene variants in inflammatory bowel disease. Inflamm Bowel Dis. 2011;17:1057-8.

134. Zhang ZF, Yang N, Zhao G, Zhu L, Wang LX. Association between the Pro12Ala polymorphism of peroxisome proliferator-activated receptor gamma 2 and inflammatory bowel disease: a meta-analysis. PLoS One. 2012;7:e30551.

135. Butt C, Gladman D, Rahman P. PPAR-gamma gene polymorphisms and psoriatic arthritis. J Rheumatol. 2006;33:1631-3.

136. Consortium GT. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015;348:648-60.

137. Arpino V, Brock M, Gill SE. The role of TIMPs in regulation of extracellular matrix proteolysis. Matrix Biol. 2015;44-46:247-54.

138. Elias GJ, Ioannis M, Theodora P, Dimitrios PP, Despoina P, Kostantinos V, et al. Circulating tissue inhibitor of matrix metalloproteinase-4 (TIMP-4) in systemic sclerosis patients with elevated pulmonary arterial pressure. Mediators Inflamm. 2008;2008:164134.

139. de Bruyn M, Vandooren J, Ugarte-Berzal E, Arijs I, Vermeire S, Opdenakker G. The molecular biology of matrix metalloproteinases and tissue inhibitors of metalloproteinases in inflammatory bowel diseases. Crit Rev Biochem Mol Biol. 2016;51:295-358.

140. Sela-Passwell N, Kikkeri R, Dym O, Rozenberg H, Margalit R, Arad-Yellin R, et al. Antibodies targeting the catalytic zinc complex of activated matrix metalloproteinases show therapeutic potential. Nat Med. 2011;18:143-7.

141. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009;461:218-23.

142. Takahashi N, Akahoshi M, Matsuda A, Ebe K, Inomata N, Obara K, et al. Association of the IL12RB1 promoter polymorphisms with increased risk of

atopic dermatitis and other allergic phenotypes. Hum Mol Genet. 2005;14:3149-59.

143. Cunninghame Graham DS, Morris DL, Bhangale TR, Criswell LA, Syvanen AC, Ronnblom L, et al. Association of NCF2, IKZF1, IRF8, IFIH1, and TYK2 with systemic lupus erythematosus. PLoS Genet. 2011;7:e1002341.

144. Hellquist A, Jarvinen TM, Koskenmies S, Zucchelli M, Orsmark-Pietras C, Berglind L, et al. Evidence for genetic association and interaction between the TYK2 and IRF5 genes in systemic lupus erythematosus. J Rheumatol. 2009;36:1631-8.

145. Sigurdsson S, Nordmark G, Goring HH, Lindroos K, Wiman AC, Sturfelt G, et al. Polymorphisms in the tyrosine kinase 2 and interferon regulatory factor 5 genes are associated with systemic lupus erythematosus. Am J Hum Genet. 2005;76:528-37.

146. Diogo D, Bastarache L, Liao KP, Graham RR, Fulton RS, Greenberg JD, et al. TYK2 protein-coding variants protect against rheumatoid arthritis and autoimmunity, with no evidence of major pleiotropic effects on non-autoimmune complex traits. PLoS One. 2015;10:e0122271.

147. Vignali DA, Kuchroo VK. IL-12 family cytokines: immunological playmakers. Nat Immunol. 2012;13:722-8.

148. Sato S, Hanakawa H, Hasegawa M, Nagaoka T, Hamaguchi Y, Nishijima C, et al. Levels of interleukin 12, a cytokine of type 1 helper T cells, are elevated in sera from patients with systemic sclerosis. J Rheumatol. 2000;27:2838-42.

149. MacDonald KG, Dawson NA, Huang Q, Dunne JV, Levings MK, Broady R. Regulatory T cells produce profibrotic cytokines in the skin of patients with systemic sclerosis. J Allergy Clin Immunol. 2015;135:946- e9.

150. Banning U, Krutmann J, Korholz D. The role of IL-4 and IL-12 in the regulation of collagen synthesis by fibroblasts. Immunol Invest. 2006;35:199-207.

151. Matsushita T, Hasegawa M, Hamaguchi Y, Takehara K, Sato S. Longitudinal analysis of serum cytokine concentrations in systemic sclerosis: association of interleukin 12 elevation with spontaneous regression of skin sclerosis. J Rheumatol. 2006;33:275-84.

152. Komura K, Fujimoto M, Hasegawa M, Ogawa F, Hara T, Muroi E, et al. Increased serum interleukin 23 in patients with systemic sclerosis. J Rheumatol. 2008;35:120-5.

153. Kurasawa K, Hirose K, Sano H, Endo H, Shinkai H, Nawata Y, et al. Increased interleukin-17 production in patients with systemic sclerosis. Arthritis Rheum. 2000;43:2455-63.

154. Truchetet ME, Brembilla NC, Montanari E, Lonati P, Raschi E, Zeni S, et al. Interleukin-17A+ cell counts are increased in systemic sclerosis skin and their number is inversely correlated with the extent of skin involvement. Arthritis Rheum. 2013;65:1347-56.

155. Lei L, Zhao C, Qin F, He ZY, Wang X, Zhong XN. Th17 cells and IL-17 promote the skin and lung inflammation and fibrosis process in a bleomycin-induced murine model of systemic sclerosis. Clin Exp Rheumatol. 2016;34 Suppl 100:14-22.

156. Couturier N, Bucciarelli F, Nurtdinov RN, Debouverie M, Lebrun-Frenay C, Defer G, et al. Tyrosine kinase 2 variant influences T lymphocyte polarization and multiple sclerosis susceptibility. Brain. 2011;134:693-703.

157. Li Z, Gakovic M, Ragimbeau J, Eloranta ML, Ronnblom L, Michel F, et al. Two rare disease-associated Tyk2 variants are catalytically impaired but signaling competent. J Immunol. 2013;190:2335-44.

158. Dendrou CA, Cortes A, Shipman L, Evans HG, Attfield KE, Jostins L, et al. Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. Sci Transl Med. 2016;8:363ra149.

159. Leonardi CL, Kimball AB, Papp KA, Yeilding N, Guzzo C, Wang Y, et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 76-week results from a randomised, double-blind, placebo-controlled trial (PHOENIX 1). Lancet. 2008;371:1665-74.

160. Papp KA, Langley RG, Lebwohl M, Krueger GG, Szapary P, Yeilding N, et al. Efficacy and safety of ustekinumab, a human interleukin-12/23 monoclonal antibody, in patients with psoriasis: 52-week results from a randomised,

double-blind, placebo-controlled trial (PHOENIX 2). Lancet. 2008;371:1675-84.

161. Griffiths CE, Strober BE, van de Kerkhof P, Ho V, Fidelus-Gort R, Yeilding N, et al. Comparison of ustekinumab and etanercept for moderate-to-severe psoriasis. N Engl J Med. 2010;362:118-28.

162. Villarino AV, Kanno Y, O'Shea JJ. Mechanisms and consequences of Jak-STAT signaling in the immune system. Nat Immunol. 2017;18:374-84.

163. Hirahara K, Schwartz D, Gadina M, Kanno Y, O'Shea JJ. Targeting cytokine signaling in autoimmunity: back to the future and beyond. Curr Opin Immunol. 2016;43:89-97.

164. O'Shea JJ, Schwartz DM, Villarino AV, Gadina M, McInnes IB, Laurence A. The JAK-STAT pathway: impact on human disease and therapeutic intervention. Annu Rev Med. 2015;66:311-28.

165. Kreins AY, Ciancanelli MJ, Okada S, Kong XF, Ramirez-Alejo N, Kilic SS, et al. Human TYK2 deficiency: Mycobacterial and viral infections without hyper-IgE syndrome. J Exp Med. 2015;212:1641-62.

166. van de Vosse E, Haverkamp MH, Ramirez-Alejo N, Martinez-Gallo M, Blancas-Galicia L, Metin A, et al. IL-12Rbeta1 deficiency: mutation update and description of the IL12RB1 variation database. Hum Mutat. 2013;34:1329-39.

167. Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, et al. The support of human genetic evidence for approved drug indications. Nat Genet. 2015;47:856-60.

168. 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.

169. Cho JH, Feldman M. Heterogeneity of autoimmune diseases: pathophysiologic insights from genetics and implications for new therapies. Nat Med. 2015;21:730-8.

170. Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends Immunol. 2013;34:22-6.

171. Teruel M, McKinney C, Balsa A, Pascual-Salcedo D, Rodriguez-Rodriguez L, Ortiz AM, et al. Association of CD247 polymorphisms with

rheumatoid arthritis: a replication study and a meta-analysis. PLoS One. 2013;8:e68295.

172. Raychaudhuri S, Thomson BP, Remmers EF, Eyre S, Hinks A, Guiducci C, et al. Genetic variants at CD28, PRDM1 and CD2/CD58 are associated with rheumatoid arthritis risk. Nat Genet. 2009;41:1313-8.

173. Terao C, Kawaguchi T, Dieude P, Varga J, Kuwana M, Hudson M, et al. Transethnic meta-analysis identifies GSDMA and PRDM1 as susceptibility genes to systemic sclerosis. Ann Rheum Dis. 2017.

174. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013;45:1238-43.

175. Lappalainen T, Sammeth M, Friedlander MR, t Hoen PA, Monlong J, Rivas MA, et al. Transcriptome and genome sequencing uncovers functional variation in humans. Nature. 2013;501:506-11.

176. Zhang J, Zahir N, Jiang Q, Miliotis H, Heyraud S, Meng X, et al. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. Nat Genet. 2011;43:902-7.

177. Balada E, Simeon-Aznar CP, Serrano-Acedo S, Martinez-Lostao L, Selva-O'Callaghan A, Fonollosa-Pla V, et al. Lack of association of the PTPN22 gene polymorphism R620W with systemic sclerosis. Clin Exp Rheumatol. 2006;24:321-4.

178. Gourh P, Tan FK, Assassi S, Ahn CW, McNearney TA, Fischbach M, et al. Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. Arthritis Rheum. 2006;54:3945-53.

179. Dieude P, Guedj M, Wipff J, Avouac J, Hachulla E, Diot E, et al. The PTPN22 620W allele confers susceptibility to systemic sclerosis: findings of a large case-control study of European Caucasians and a meta-analysis. Arthritis Rheum. 2008;58:2183-8.

180. Diaz-Gallo LM, Gourh P, Broen J, Simeon C, Fonollosa V, Ortego-Centeno N, et al. Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Ann Rheum Dis. 2011;70:454-62.

181. Wipff J, Allanore Y, Kahan A, Meyer O, Mouthon L, Guillevin L, et al. Lack of association between the protein tyrosine phosphatase non-receptor 22 (PTPN22)*620W allele and systemic sclerosis in the French Caucasian population. Ann Rheum Dis. 2006;65:1230-2.

182. Gutierrez-Achury J, Zorro MM, Ricano-Ponce I, Zhernakova DV, Coeliac Disease Immunochip Consortium RC, Diogo D, et al. Functional implications of disease-specific variants in loci jointly associated with coeliac disease and rheumatoid arthritis. Hum Mol Genet. 2016;25:180-90.

183. Trynka G, Sandor C, Han B, Xu H, Stranger BE, Liu XS, et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet. 2013;45:124-30.

184. Gao L, Emond MJ, Louie T, Cheadle C, Berger AE, Rafaels N, et al. Identification of Rare Variants in ATP8B4 as a Risk Factor for Systemic Sclerosis by Whole-Exome Sequencing. Arthritis Rheumatol. 2016;68:191-200.

185. Hunt KA, Smyth DJ, Balschun T, Ban M, Mistry V, Ahmad T, et al. Rare and functional SIAE variants are not associated with autoimmune disease risk in up to 66,924 individuals of European ancestry. Nat Genet. 2011;44:3-5.

186. Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. Nat Rev Genet. 2016;17:333-51.

187. Hu YJ, Liao P, Johnston HR, Allen AS, Satten GA. Testing Rare-Variant Association without Calling Genotypes Allows for Systematic Differences in Sequencing between Cases and Controls. PLoS Genet. 2016;12:e1006040.

188. Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ, Neale BM. Phenotypic extremes in rare variant study designs. Eur J Hum Genet. 2016;24:924-30.

189. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, et al. Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. Nature. 2013;498:232-5.

190. Rivas MA, Beaudoin M, Gardet A, Stevens C, Sharma Y, Zhang CK, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat Genet. 2011;43:1066-73.

191. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015;43:D447-52.

192. Wu M, Assassi S. The role of type 1 interferon in systemic sclerosis. Front Immunol. 2013;4:266.

193. Honda K, Taniguchi T. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. Nat Rev Immunol. 2006;6:644-58.

194. Catrysse L, Vereecke L, Beyaert R, van Loo G. A20 in inflammation and autoimmunity. Trends Immunol. 2014;35:22-31.

195. Fiorillo E, Orru V, Stanford SM, Liu Y, Salek M, Rapini N, et al. Autoimmune-associated PTPN22 R620W variation reduces phosphorylation of lymphoid phosphatase on an inhibitory tyrosine residue. J Biol Chem. 2010;285:26506-18.

196. Zhong MC, Veillette A. Immunology: Csk keeps LYP on a leash. Nat Chem Biol. 2012;8:412-3.

197. Munger JS, Sheppard D. Cross talk among TGF-β signaling pathways, integrins, and the extracellular matrix. Cold Spring Harb Perspect Biol. 2011;3:a005017.

198. Shiokawa D, Tanuma S. Characterization of human DNase I family endonucleases and activation of DNase gamma during apoptosis. Biochemistry. 2001;40:143-52.

199. Errami Y, Naura AS, Kim H, Ju J, Suzuki Y, El-Bahrawy AH, 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;288:3460-8.

200. Okamoto M, Okamoto N, Yashiro H, Shiokawa D, Sunaga S, Yoshimori A, et al. Involvement of DNase gamma in the resected double-strand DNA

breaks in immunoglobulin genes. Biochem Biophys Res Commun. 2005;327:76-83.

201. Bhattacharya A, Eissa NT. Autophagy and autoimmunity crosstalks. Front Immunol. 2013;4:88.

202. Sollid LM, Pos W, Wucherpfennig KW. Molecular mechanisms for contribution of MHC molecules to autoimmune diseases. Curr Opin Immunol. 2014;31:24-30.

203. Chang X, Wang K. wANNOVAR: annotating genetic variants for personal genomes via the web. J Med Genet. 2012;49:433-6.

204. Ueki M, Takeshita H, Fujihara J, Iida R, Yuasa I, Kato H, 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;407:20-4.

205. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, Ortmann W, et al. Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet. 2008;40:1062-4.

206. Gupta B, Hawkins RD. Epigenomics of autoimmune diseases. Immunol Cell Biol. 2015;93:271-6.

207. Zhang X, Joehanes R, Chen BH, Huan T, Ying S, Munson PJ, et al. Identification of common genetic variants controlling transcript isoform variation in human whole blood. Nat Genet. 2015;47:345-52.

208. Rivera CM, Ren B. Mapping human epigenomes. Cell. 2013;155:39-55.

209. Birnbaum RY, Clowney EJ, Agamy O, Kim MJ, Zhao J, Yamanaka T, et al. Coding exons function as tissue-specific enhancers of nearby genes. Genome Res. 2012;22:1059-68.

210. Spain SL, Barrett JC. Strategies for fine-mapping complex traits. Hum Mol Genet. 2015;24:R111-9.

211. Gorlova O, Martin JE, Rueda B, Koeleman BP, Ying J, Teruel M, 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.

212. Lee JC, Biasci D, Roberts R, Gearry RB, Mansfield JC, Ahmad T, et al. Genome-wide association study identifies distinct genetic contributions to prognosis and susceptibility in Crohn's disease. Nat Genet. 2017;49:262-8.

213. Martin P, McGovern A, Orozco G, Duffus K, Yarwood A, Schoenfelder S, et al. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. Nat Commun. 2015;6:10069.

214. McGovern A, Schoenfelder S, Martin P, Massey J, Duffus K, Plant D, et al. Capture Hi-C identifies a novel causal gene, IL20RA, in the pan-autoimmune genetic susceptibility region 6q23. Genome Biol. 2016;17:212.

"No basta examinar, hay que contemplar: impregnemos de emoción y simpatía las cosas observadas; hagámoslas nuestras, tanto por el corazón como por la inteligencia. Sólo así nos entregarán su secreto. Porque el entusiasmo acrecienta y afina nuestra capacidad perceptiva"

Santiago Ramón y Cajal, *Reglas y consejos sobre investigación científica: los tónicos de la voluntad*