Tesis Doctoral

Revisiting the genetic puzzle of autoimmune diseases

Memoria presentada por la Licenciada en Biología Lina Marcela Diaz Gallo para optar al grado de Doctor Internacional por la Universidad de Granada.

Dirigida por: Javier Martín Ibáñez, Profesor de Investigación del CSIC

ugr | Universidad de Granada

ipbln

Instituto de Parasitología y Biomedicina López-Neyra,

IPBLN-CSIC.

Granada, Junio de 2013.

Editor: Editorial de la Universidad de Granada
Autor: Lina Marcela Diaz Gallo
D.L.: GR 379-2014
ISBN: 978-84-9028-776-7

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

Así mismo, garantizamos, al firmar esta tesis doctoral, que las siete publicaciones presentadas no se han utilizado en la defensa de ninguna otra tesis en España u otro país y que no serán utilizadas con tal propósito.

Granada, 11 de Junio de 2013

Director de la Tesis

Fdo.:

Profesor Javier Martín Ibánez

Doctoranda

Fdo.:

Lina Marcela Diaz Gallo

…"¿De qué le sirve a uno ganar el mundo entero si se pierde o se destruye a sí mismo?..Lucas 9:25

…"What good is it for someone to gain the whole world, and yet lose of forfeit their very self?"… Luke9:25

In Memory of Juan Manuel Gallo Morales

CONTENTS

ABBREVIATIONS

ACA	Anti-Centromere Antibody
ACCP	Anti-Citrullinated Protein Antibody
AIDs	Autoimmune Diseases
AITD	Autoimmune Thyroid Diseases
ANA	Antinuclear Antibodies,
ANCA	Antineutrophil Cytoplasmic Antibodies
ARA	Anti-RNA Polymerase III
ATA	Anti-Topoisomerase Antibody
BAK1	BCL2-Antagonist/Killer 1
BCR	B cell receptor
CD	Crohn's Disease
CD24	CD24 molecule
CD247	CD247 molecule
CNEs	Conserved Non-Coding Elements
DC	Dendritic Cells
dcSSc	Diffuse Cutaneous Scleroderma
DZ	Dizygotic Twins
EBV	Epstein-Barr virus
FOXP3	Forkhead box P3
GAB	Intestinal Goblet Cells
GVS	Genetic Variation Score
GWAS	Genome-wide Association Studies
HapMap	Haplotype Map
HLA	Human Leukocyte Antigen
HLA	Human Leukocyte Antigen
HW	Hardy-Weinberg Equilibrium
IBD	Inflammatory Bowel Disease
IFN	Interferon Cytokine Signals
IFN- γ	Interferon- γ
IgG	Immunoglobulin G

IL	Interleukin
IL2-IL21	Interleukin 2 and interleukin 21 locus
IRAK1	Interleukin-1 receptor-associated kinase 1
lcSSc	Limited Cutaneous Scleroderma
LD	Linkage Disequilibrium
LD	Linkage Disequilibrium
LYP	Lymphoid Tyrosine Phosphatase
MECP2	Methyl CpG binding Protein 2
MHC	Major Histocompatibily Complex
MS	Multiple Sclerosis
MZ	Monozygotic Twins
NCBI	National Center fo Biotecnology Information
NK	Natural Killer
NMDA	N-Methyl-d-Aspartate
NOD2	Nucleotide-binding Oligomerization Domain containing 2
OMIM	Online Mendelian Inheritance in Man
OR	Odds Ratio
PAB	Exocrine Pancreas
PADI4	Peptidyl Arginine Deiminase, type IV
PAH	Pulmonary Arterial Hypertension
PBMCs	Peripheral Bone Marrow Cells
PTPN22	Protein tyrosine phosphatase non-receptor type 22
RA	Rheumatoid Arthritis
RF	Rheumatoid Factor
SLE	Systemic Lupus Erythematosus
SNP	Single Nucleotid Polymorphism
SSc	Systemic Sclerosis
STAT4	Signal transducer and activator transcription 4
T1D	Type 1 Diabetes
TCR	T cell receptor
Th	T-helper
Tregs	Regulatory T cells
TULA	T cell Ubiquitin Ligand

UBASH3a	Ubiquitin associated and Src-homology 3 domain containing A
UC	Ulcerative Colitis
UTR	Untranslate Region
UVB	Ultraviolet B

RESUMEN

Actualmente hay más de 80 enfermedades autoinmunes (AIDs, del inglés *Autoimmune Diseases*; OMIM #109100) reconocidas, que en conjunto afectan a entre un 3 y un 10% de la población general. La mayoría de estas enfermedades son crónicas e incurables, con frecuencia llevan a una discapacidad sustancial y, muchas generan complicaciones que amenazan la vida. Las AIDs constituyen un problema de salud pública que además de causar un sufrimiento individual importante, tienen un alto costo social [1]. Estas patologías se caracterizan por el desarrollo de una respuesta inmune por parte del huésped en contra de sus propias células y tejidos. Aunque su etiología no se conoce completamente, se debe a una compleja interacción entre factores genéticos y ambientales [2-5]. Los avances tecnológicos y esfuerzos de cooperación internacional han permitido aumentar la comprensión de los factores genéticos implicados en las AIDs; sin embargo, se estima que únicamente se conoce entre un 5 al 20% del componente hereditario de estas enfermedades [6-8]. Teniendo esto en cuenta, podemos ser llamados "*los cazadores de la heredabilidad perdida de las AIDs*", ya que parte de nuestra pasión es tratar de entender y contribuir al conocimiento de ese 80% del componente genético de las AIDs que está aún por descubrir. Hay diferentes maneras de alcanzar dicho objetivo y una de ellas se aplica en esta Tesis: el estudio del componente genético común de las AIDs.

Hoy por hoy, el estudio del componente genético de las AIDs está basado en la secuenciación del genoma humano, publicado por primera vez hace más de una década [9-11]. Este avance ha proporcionado los pilares donde ensamblar la información sobre estructuras y funciones biológicas; como la variación genética, la distancia de recombinación, el desequilibrio de ligamiento y su relación con enfermedades hereditarias, entre otros. La secuenciación del genoma humano ha proporcionado asimismo, un entorno para conectar modelos de la biología de los sistemas con la fisiología de los seres humanos; así como también un amplio catálogo de información genómica. A raíz de la secuenciación del genoma humano, nuestro trabajo tiene una plataforma de aproximación sistemática para descubrir variaciones genéticas que subyacen en las enfermedades humanas. La identificación de los factores de susceptibilidad genética funciona como una brújula y un mapa que indican el camino para descubrir rutas celulares y mecanismos fisiopatológicos de las AIDs. Esta

aproximación sistemática ha dado como resultado la identificación de cientos de *loci* asociados con una o más AIDs [2, 12]. Se ha demostrado que cerca de la mitad (44%) de los *loci* identificados mediante estudios de asociación del genoma completo (GWAS, del inglés *Genome-Wide Association Studies*) en una enfermedad de origen autoinmune particular, influyen en el riesgo de, al menos dos AIDs [13]. Basándonos en este tipo de evidencia, evaluamos polimorfismos genéticos que habían sido previamente asociados con otras AIDs, en grupos de casos y controles de origen Europea en las siguientes enfermedades: esclerosis sistémica (SSc, del inglés S*ystemic Sclerosis*), enfermedad inflamatoria intestinal (IBD, del inglés *Inflammatory Bowel Disease*), lupus eritematoso sistémico (SLE, del inglés *Systemic Lupus Erythematosus*) y artritis reumatoide (RA, del inglés *Rheumatoid Arthritis*). Los polimorfismos estudiados están localizados en los *loci PTPN22, STAT4, IL2-IL21, CD24, BAK1* y *UBASH3a*. Por la metodología empleada y los hallazgos encontrados, esta Tesis representa una contribución importante al conocimiento de los factores genéticos comunes de las AIDs.

Hemos demostrado por primera vez, la relación de algunas de estas variantes genéticas en patologías de origen autoinmune en las que no habían sido descritas previamente, tales como asociaciones de polimorfismos genéticos en los genes *PTPN22* y *CD24* con IBD, *IL2-IL21* con SSc y *UBASH3a* con SLE. Además, hemos confirmado mediante estudios de replicación y meta-análisis observaciones interesantes, como el efecto protector del polimorfismo de nucleótido simple (SNP, del inglés *single nucleotide polymorphism*) rs2476601 del gen *PTPN22* en enfermedad de Crohn (uno de los dos mayores subtipos de IBD) y el riesgo modesto que confiere a SSc. Asimismo, la asociación de la variante genética rs7574865 del gen *STAT4* con colitis ulcerosa (el otro subtipo más frecuente de IBD). Aparte de las conclusiones específicas obtenidas de nuestros resultados, también observamos algunas características importantes del estudio del componente genético de las AIDs, entre ellas: i) El alelo asociado de un polimorfismo dado puede conferir riesgo a un grupo de AIDs, mientras confiere protección a otro grupo de patologías de origen autoinmune. ii) La fuerza del efecto de un factor genético asociado a varias AIDs varía para cada enfermedad, siendo más fuerte para algunas patologías y menor o más leve para otras. iii) Pueden existir varios polimorfismos asociados dentro de un mismo *loci* de susceptibilidad a AIDs y parecen ser específicos para cada enfermedad. iv) La asociación significativa de una variante genética con una patología de origen autoinmune funciona como una baliza que señala

las posibles implicaciones del gen o los genes localizados en, o cerca de, los *loci* donde se encuentra la variación genética, pero no determina la implicación funcional del mismo o de los genes allí localizados. v) Los SNPs asociados con AIDs parecen ser diferentes en las distintas poblaciones humanas. vi) No es sorprendente que la mayoría de los polimorfismos genéticos asociados comúnmente a AIDs están localizados en regiones intrónicas o intergénicas del genoma humano.

Los *loci* asociados de manera común a diferentes AIDs remarcan rutas inmunológicas implicadas en la etiopatogénesis de estas enfermedades y nos indican un punto de partida para estudiar sus mecanismos específicos. El conocimiento acumulado hasta el momento de la genética de las AIDs representa un buen punto de partida; sin embargo, el progreso en los distintos niveles en este campo está estrechamente ligado a nuestro avance en la comprensión del funcionamiento del genoma humano.

SUMMARY

There are more than 80 recognised autoimmune diseases, AIDs (OMIM #109100) that together affects from 3 to 10 % of the general population. Most of them are chronic and incurable, often leading to substantial disability and many of them may have life-threatening complications. AIDs constitute a major health problem that besides causing important individual suffering, have high social cost [1]. These pathologies are characterized by an immune response from the host against its own cells and tissues. They result from complex interactions between genetic and environmental factors [2-5]. Technological advance and coordinated international efforts have increased the understanding of the genetic factors of AIDs; but it is estimated that only 5-20% of the genetic background of AIDs is known [6-8]. Keeping this in mind, we can be called the "*missing AIDs heritability hunters*" because our passion is to try to understand and contribute to the 80% of the genetic component of AIDs that is left to be discovered. There are different ways to try to reach that goal and one of those is implemented in this thesis: the study of the common genetic component of AIDs.

Now a days, the study of the AIDs genetic background lays on the sequencing of the human genome, first published more than a decade ago [9-11]; because it has provided the comprehensive scaffold to assemble fragmentary information into landscapes of biological structure and functions, like genetic variation, recombinational distance, linkage disequilibrium and association to inherited diseases, among others. The human genome sequencing has also provided a framework to connect the biology of model systems to the physiology of human beings; as well as provide a comprehensive catalogue of genomic information. In the aftermath of the human genome sequencing, our work has systematic approaches to discover the genetic variants underlying a disease. The identification of the susceptibility genetic variants is a compass and a map showing the way to discover cellular pathways and physiologic mechanisms of AIDs. This systematic approach has resulted in the identification of hundreds of *loci* associated with one or more AIDs [2, 12]. It has been shown that nearly half (44%) of *loci* identified in genome-wide association studies (GWAS) of an individual disease influence the risk of at least two AIDs [13]. This thesis has an important contribution to such evidence. We evaluated genetic risk polymorphisms that have been previously associated in other AIDs, in European descendent case-control groups of Systemic

Sclerosis (SSc), inflammatory bowel disease (IBD), Systemic Lupus Erythematosus (SLE) and Rheumatoid arthritis (RA). The studied polymorphisms were located in the *loci PTPN22*, *STAT4*, *IL2-IL21*, *CD24, BAK1* and *UBASH3a*.

Our results report for first time associations of these known susceptibility AIDs *loci* with diseases that has not been described before; such as the association of genetic variants in *PTPN22* and *CD24* with IBD, *IL2-IL21* with SSc and *UBASH3a* with SLE. Moreover, we confirmed through replication and meta-analysis studies, interesting observations like the association with an opposite effect of the rs2476601 *PTPN22* SNP in Crohn diseases, the modest association of this SNP with SSc and the rs7574865 *STAT4* variant influence in ulcerative colitis (UC), a subphenotype of IBD. Besides the specific conclusions get from our results we observed some important features in the study of the genetic component of the AIDs, among them: i) the associated allele of a polymorphism can confer risk to a group of AIDs while is associated with a protective effect against others autoimmune pathologies. ii) The effect's strength of a common associated AIDs genetic factor variety for each disease, being stronger for some diseases and lower for others. These propose a hierarchy role for the common genetic factors that can be different for each AID and its subphenotypes. iii) The associated polymorphisms of a given AIDs susceptibility *loci* seem to be specific for each disease. iv) A significant association of a SNP works as a red flag pointing the possible implication of the gene(s) located in or proxy to the *loci* where the SNP is situated, but do not determined functional implication of the polymorphisms or genes. v) The associated SNPs to AIDs vary across different descendant populations. vi) It is not surprising that most of the common AIDs associated polymorphisms exist in intronic or intergenic regions of the human genome.

The common AIDs associated *loci* highlight the immunologic pathways implicated in the etiopathogenesis of these diseases and give us a start point for study their specific mechanisms. The understanding of the genetic background of AIDs cumulated until now it's a good beginning. The advance in different levels of this field is directly related with our understanding of the function of the human genome.

INTRODUCTION

Autoimmune diseases: an overview

Autoimmune diseases, (AIDs) (OMIM #109100) are a category of human disorders that occur in to 3-10% of the general population and are characterized by an immune response from the host against its own cells and tissues. They result from complex interactions between genetic and environmental factors. [2, 3, 5, 13-15].

Paul Ehrlich suggested the concept of autoimmunity around 1900. He said that a functional immune system must have "*horror autotoxicus*", the unwillingness of the organism to endanger itself by the formation of toxic autoantibodies; not that antibodies against self cannot be formed but that they are prevented "*by certain contrivances*" from exerting any destructive action [16]. Only fifty or sixty years after Ehrlich's idea, the autoimmune diseases (AIDs) were considered. These occurred due to the accumulation of evidence that studied the basic immunology, the pathogenesis of paroxysmal cold hemoglobinuria or the reports of experimental encephalitis, sympathetic ophthalmia, hemolytic anemia or thyroiditis (among others) [16, 17]. Then, the difference between "physiological" and "pathological" autoimmunity was more sharply focuses; in other words, it was recognized that failure of self tolerance is the fundamental cause of AIDs [18].

Self tolerance can be grouped in central and peripheral. In central tolerance, the immature T and B lymphocytes that recognize self antigens are controlled by apoptosis in the generative lymphoid organs, the thymus and bone marrow, respectively. In peripheral tolerance, the mature self-reactive lymphocytes are killed or shut off in peripheral tissues through different mechanisms; such apoptotic cell death, functional unresponsiveness (called anergy), and suppression by regulatory T cells, among others. AIDs develop when there is a control tolerance malfunction, which leads to the escape of self-reactive lymphocytes and their activation [19, 20]. A complex interaction between inheritance component and environmental triggers is thought as the cause of self tolerance's failure (Figure 1). Nevertheless, the exact mechanisms of that complex interaction are not completely understood [2, 19-21]. AIDs can be classified into organ-

specific or systemic according if the aberrant immune reaction against self cells or tissues is restricted to a specific organ, or affects multiple organs, respectively.

Three types of evidence can be marshalled to establish that a human disease is actually of autoimmune origin: direct proof, indirect evidence and circumstantial evidence [22]. The direct proof is when the disease is reproduced in a normal recipient by direct transfer of autoantibody. For example, the transplacental transmission of pathognomonic autoantibody from an afflicted mother to the fetus in neonatal myasthenia gravis [23] or congenital heart block [24]. But the transplacental transmission of autoantibody does not always trigger an AID in the newborn, indicating that other factors are important for the establishment of the AID. The indirect evidence is based on the isolation of autoantibodies or self-reactive T cells from the organs which represent the major target of autoimmune attack using experimental models, ideally followed by the reproduction of the AID in experimental animal models [22]. Nevertheless, the experimental evidence is difficult to carry out and many times the animal models do not resemble the human disease [25]. Different animal species, even different strains of the same species, vary greatly in their susceptibility to experimental AID animal models, a fact that could be extrapolated to humans because the incidence and prevalence of AID vary between human populations [1]. Last, the circumstantial is the combination of certain descriptive markers such a: positive family history of the same or others AIDs, presence of other AID in the same patient, presence of infiltrating mononuclear cells in the affected organ or tissue, statistical association with a particular mayor histocompatibily complex (MHC) allele or haplotype or aberrant expression of human leukocyte antigen (HLA) molecule, high serum levels of IgG autoantibodies, deposition of antigen-antibody complex in the affected organ or tissue, improvement of the symptoms with the use of immune suppressive drugs [22]. The degree of clustering of diseases in families can be estimated of the ratio (λ) of the risk for relatives of patients with a diseases and the population prevalence of that disease. There is familial aggregation in AIDs (familial autoimmunity) with λ values that ranging from 1.5 to 100 [26-28]. Although, familial autoimmunity has an important genetic base, this is not the fully responsible because environmental factors can also cause familial clustering. The simultaneous presence of two or more AIDs in the same individual was described as polyautoimmunity or multiple autoimmune syndrome [29], and occur in excess than expected by chance; this phenomenon support the genetic background and early life

programming due to the environmental exposures as the fundament for susceptibility to AIDs [30, 31]. Although these three theoretical criterions to identify AIDs seem to be obvious, their application is not straightforward. We know clearly that **the AIDs aetiology is based on a complex interaction between the genetic component of the individuals and their environment; and that the individual effects of these genetic and environmental factors are moderate, increasing the risk by 10-50% (Figure 1).**

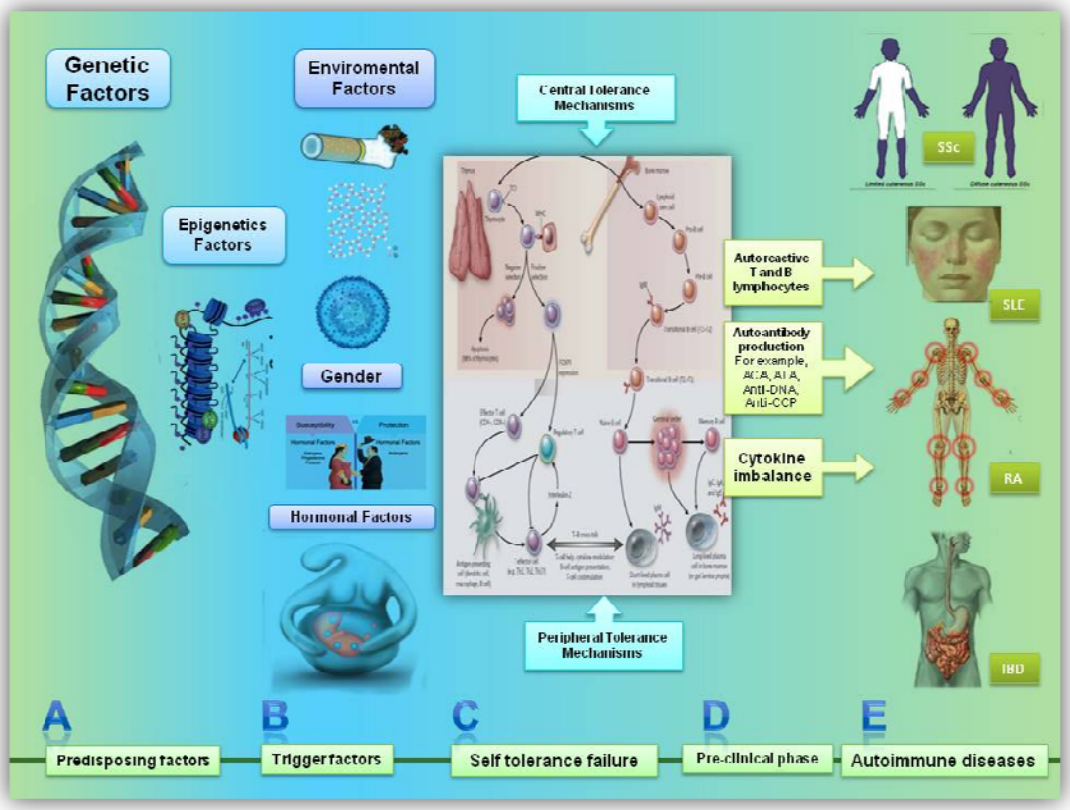


Figure 1. AIDs aetiology is based on a complex interaction between the genetic component of the individuals and their environment. (A) Multilocus genetic variants are associated with higher risk to develop a variety of AIDs [2, 5]. Epigenetic mechanism regulates gene expression and are sensible to external stimuli; connecting genetic with environmental factors. Epigenetic modifications include DNA methylation, histone modifications and nucleosome positioning. There is evidence that have related epigenetic process with AIDs. For example, T cells from AIDs patients have an overexpression of methylation-regulated genes, implying T cell autoreactivity [32]. **(B)** Epidemiologic and experimental evidence have shown that, chemical factors such smoking or exposure to crystalline silica; physical factors, such sunlight and biologic agents, such as viruses, bacteria or parasites infection-causing are important trigger of AIDs [15]. Most AIDs are more prevalent in women than in men. Different lines of evidence have suggested some factors that are the basis of the gender bias in AIDs [33, 34]. Usually, women have a stronger humoral and cellular immune response compared

to men and this fact is related with the sexual hormones. Estrogens, androgens and prolactin are hormones that affect both innate and adaptive immune system and they have been related with AIDs. A hormonal imbalance can result in lower immune-suppressive androgens and higher immune-enhancer estrogens. For example, during pregnancy and postpartum the manifestations of certain AIDs such severity, are altered [35]. Environmental and hormonal factors can interact, adding more complexity to the aetiology of AIDs. A recent study identified a direct interaction between sex hormones and microbial exposures and show that microbiome manipulations can provoke testosterone-dependent protection from autoimmune in a genetically high-risk rodent model [36]. **(C)** The genetic predisposition, epigenetic mechanisms and trigger factors such environmental and hormonal imbalance, influence the self tolerance mechanisms. These complex interactions lead to a failure probably in both central and peripheral check point of autoreactivity that leads to develop of AIDs [19, 20]. In the central tolerance mechanisms for example, more than 98% of developing thymocytes die from apoptosis in the thymus, because of excessive reactivity to self-peptides bound to MHC molecules, followed by positive selection for functionally competent effector T cells (CD4+ and CD8+) that are released into the periphery. A similar process occurs in the bone marrow, where self reactivity B cells are dramatically reduced. These process as the production of regulatory T cells (Tregs) are under genetic control, for instance the expression regulatory factor *FOXP3* is part of the production control of Tregs, and its absence leads to severe autoimmunity [2]. Self-reactive cells that escape from central control might never encounter the self-antigen they recognize there for exists in a stage of ignorance. If the encounter is produce the self-reactive cells will be regulated by peripheral tolerance mechanisms such anergy, a non-pathogenic expression profile of cytokines and chemokines receptors or apoptosis. Also the activated self-reactive cells are controlled by Tregs [20]. Genetic alterations in any of those mechanisms could direct the appearance of AIDs, for example the interleukin-2 (*IL2*) pathway may influence the efficiency of Treg regulation [2, 20]. **(D)** Tolerance mechanisms failure produce a profile of increase levels of T and B self-reactive cells, autoantibodies production and an imbalance in the cytokine profile that could constitute the pre-clinical stage of AIDs. **(E)** Although the genetic factors, environmental factors, epigenetic profiles and immunologic mechanisms seems to be similar the outcome of AIDs has a broadly list of different phenotypes. In example the AIDs studied in this Thesis; Systemic Sclerosis (SSc) that affects the connective tissues by extensive fibrosis [37, 38]. Systemic Lupus Erythematosus (SLE) characterized by production of antinuclear autoantibodies, immune-complex deposition, and subsequent multi-organ damage [39]. Rheumatoid Arthritis (RA) a systemic AID in which there are synovial inflammation and hyperplasia, cartilage and bone destruction, systemic features such as cardiovascular, pulmonary, psychological and skeletal disorders [40, 41]. Inflammatory Bowel Disease (IBD) encompasses both chronic relapsing inflammatory disorders of the gastrointestinal tract, Crohn's disease (CD) and Ulcerative colitis (UC) [42, 43].

The nature of different AIDs seems to have the same root, highlighting that these diseases have specific but also common environmental and genetic factors [2-6, 15, 20, 21, 44-49]. **The present Thesis is focus in the common genetic factors of AIDs.**

Thanks to a long period of work through national and international collaborations our laboratory has available an important collection of genomic DNA samples from patients suffering of different AIDs and healthy donors. Those individuals are from European descendant populations, mainly from Spain. Actually, the main studied AID in our group is the Systemic Sclerosis (SSc), taking advantage of this and the collection of samples of other AIDs such Rheumatoid Arthritis (RA), Systemic Lupus Erythematosus (SLE) and Inflammatory Bowel Diseases (IBD) we studied different immune-related loci in these three systemic and one organ-specific AIDs. The next sections of the introduction are mainly focus on these diseases, having in mind the main focus of this Thesis, common genetic factors in AIDs.

The term "environmental factors" is frequently used to describe all those susceptibility factors leading to AIDs that are not explicable on the bases of an identifiable genetic marker. Environment could be taken to refer to those factors external to the individual; for example, factors associated with diet or airborne exposures. Those environmental factors could be also internal to the subject with a multilevel interaction between external factors, biologic and genetic basis, for example, infections (**Figure 1**) [50].

The environmental factors can be grouped in chemical, physical and biological agents [15]. **i) Chemical factors** such silica, pesticides, solvents, nail polish, hairdressing and smoking have been related through epidemiologic studies with AIDs [51] [52] [53] [54]. Freshly fractured particulate silica (crystalline silica or quartz) is released typically in mining but also in the so-called dusty trades, including sandblasting, rock drilling, sand factory work, granite cutting, construction work, brick laying, tilling, and cement work. Animal studies have shown that silica can exacerbate autoimmunity in a genetic susceptible autoimmune strains [51]. Moreover, epidemiologic studies have shown that silica is an important environmental factor for SSc . Pesticides have been suggested as a risk factor for AIDs but there is insufficient evidence. Studies on the induction and/or exacerbation of autoimmunity in experimental animals by trichloroethylene, a kind of solvent, have demonstrated an accelerated autoimmune response including increased autoantibodies, T-cell activation and inflammatory cytokines [52]. Although more epidemiology studies that evaluated the solvents and cosmetic products influence on develop of AIDs, there are some indications of a relation between working with nail polish and the risk to Systemic Lupus Erythematosus (SLE). Also hairdressing

chemicals could be connected with Rheumatoid Arthritis (RA) [15]. Smoking contributes to the development of RA [53]. Higher risk to develop RA is associated with greater exposure and several studies have shown that smoking has a modulator role in the immune system; for example, altering T-cells anergy, depleting immune reactions against pulmonary infections, altering normal cytokines production, triggering oxidative stress, among others effects [54]. There have been fewer investigations of smoking in relation to SLE, Systemic Sclerosis (SSc) or Inflammatory Bowel Disease (IBD). Suggestive results indicated that smoking could be a risk factor to SLE. On the other hand, the relationship between cigarette smoking and IBD is complex in that it is likely that current smoking contributes to the development of Crohn's disease (CD) but is protective for the development of and Ulcerative colitis (UC) [15, 43]

ii) The sunlight is the **physical factor** that has been more clearly involved in AIDs, base on epidemiological studies. Several studies have found inverse associations between ecologic measures of ultraviolet B (UVB) irradiance and/or latitude and incidence of RA. Mine while evidence that sunlight might be a risk factor for SLE is not so strong; this disease has been associated with outdoor work in the year before diagnosis, particularly among those who are prone to sunburn and among persons who had serious sunburns before age 20 (Reviewed in [15]).

iii) Biologic agents such as viruses, bacteria or parasites infection-causing has been suspected as an important trigger of AIDs since the beginning of the identification of these entities. Nevertheless, there is not a conclusive proof. There is evidence that the parvovirus B19 is associated with the development of RA, and the authors of the same study observed a suggestive result for a similar but not so strong relationship with SLE [55]. The Epstein-Barr virus (EBV) might be involve in the initiation of RA and SLE [56-59]. Other studies relate a significant low percentage of antibodies for hepatitis B virus in SLE patients compare with healthy controls [60]; or the eradication of *Helicobacter pylori* infection from patients with Raynaud's phenomenon (one of the principal symptoms of SSc) and the completely disappearance or improve of the symptoms of Raynaud's attacks [61]. But none of those studies have been replicated or prove the direct relationship between the infection and AIDs. Unavoidably, the possibility of a relation between infections and AIDs is related with the hypothesis of molecular mimicry, where a self and microbial peptide antigen is strictly related to autoimmune pathology [3, 62, 63]. Although there is strong and punctual evidence that

support molecular mimicry between EBV and multiple sclerosis (MS), as part of the pathogenic mechanisms [64], this mimicry mechanism remains largely hypothetical for RA, SLE, SSc and IBD. As it is mention in the previous section, the relationship between microbes and IBD is complex and relevant for the development of this disease. Interesting, a very recent study identified a direct interaction between sex hormones and microbial exposures and show that microbiome manipulations can provoke testosterone-dependent protection from autoimmune in a genetically high-risk rodent model [36]. This study of Markle *et al.* (2013) linked in a model two considers factors of AIDs, the microbiome and the gender bias.

Different lines of evidence have suggested some factors that are the basis of the gender bias in AIDs (Figure 2) [33, 34]. Usually, women have a stronger humoral and cellular immune response compared to men and this fact is related with the sexual hormones. Estrogens, androgens and prolactin are hormones that affect both innate and adaptative immune system and they have been related with AIDs. A hormonal imbalance can result in lower immune-suppressive androgens and higher immune-enhancer estrogens. Divers studies in RA, SLE and SSc patients suggest this (reviewed in [34]). For example, during pregnancy and postpartum the RA diseases lessen while the behaviour of SLE is opposite [35]. The sexual chromosomes are another important factor in the gender bias of AIDs. Since females have two X chromosomes and males have only one, genes located on this chromosome might differ in their expression level between both genders. To avoid these differences in the expression levels, one X chromosome in females is inactivated. Though it has been reported that X chromosome is not complete inactivated, resulting in overexpression of 10-15% of X-linked genes which can contribute to AIDs [65]. For instance, polymorphisms of the X-linked *loci* that contain the genes *IRAK1* and *MECP2* have been associated with SLE, RA and SSc [66]. Moreover, our preliminary observations indicated that there is an overexpression of *MECP2* in SSc female patients compared with healthy women [67]. On the other hand, the X chromosome inactivation could be skewed. This phenomenon is not pathogenic itself but has been related with AIDs. Some studies have observed higher skewed in RA and SSc female patients than in healthy women [68-70]. Finally, the harbouring of hematopoietic stem cells from fetal to maternal or the other way around (fetal microchimerism and maternal microchimerism, respectively) has been related with AIDs. The clearest evidence of microchimerism has been described in SSc, some

evidence suggest relation of this phenomenon with RA but not with SLE [71-74]. Despite of the described evidence, gender bias is still one of the many puzzling features in AIDs.

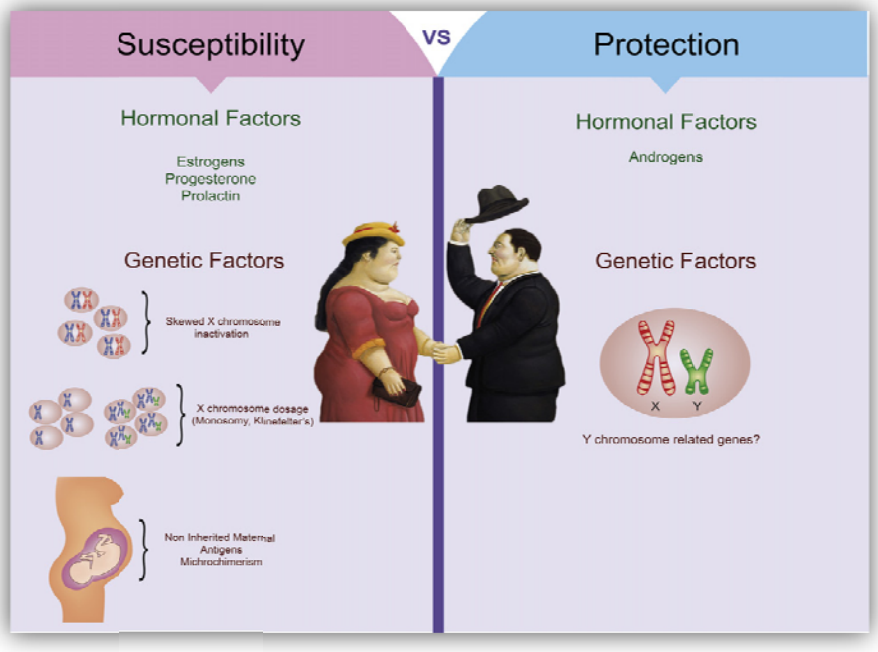


Figure 2. Representation of the possible factors involved in the gender bias of AIDs. Different lines of evidence have suggested some factors that are the basis of the gender bias in AIDs. Such a sexual hormones, sexual chromosomes, skewed of X chromosome inactivation and fetal and maternal microchimerism in pregnancy. Copyright © Quintero 2012 *et al.*[34]

Genetic factors in autoimmune diseases

Genetics is the study of single genes and their effects. Models of inheritance have been established for thousands of conditions by mutations in single genes. The point mutations, a change in a single DNA base in the sequence or single nucleotide polymorphism (SNP), are the most common. The SNP is merely a marker of biologic diversity that happens to correlate with health because of its relation to the genetic factor that could be the actual cause. The greatest contribution of genetics in human health until now is the revelation of mechanisms of common, complex diseases [75].

Humans have known for millennia that heredity affects health. However, it seems to be the normal slow process in science, many years had passed before the Mendel's contribution that explains how the heredity affects phenotype were understood. Only one century ago, thanks to the work of people like Archibald Garrod the Mendel's contribution was applied to human trails and health [7, 75]. Before the human genome sequencing era, the genomic of AIDs knowledge was based on animal models and family-based linkage studies leading to recognized around 40 genes associated with these pathologies by 1994 [27]. The long haplotype structure of the human genome due to the LD, and its systematic mapping by the International Haplotype Map (HapMap) Project [76], allowed single nucleotide polymorphisms (SNPs), to be used as markers for common haplotypes. Then the SNPs could be genotyped using TaqMan® 5' allelic discrimination technology in the candidate gene approach; or chip technology in the GWAS hypothesis free (Figure 3) [12, 77]. After the outstanding GWAS approach there are more than 200 *loci* that have been associated with one or more AIDs, according with one of the most recent reviews in the field, published one year and a half ago [2].

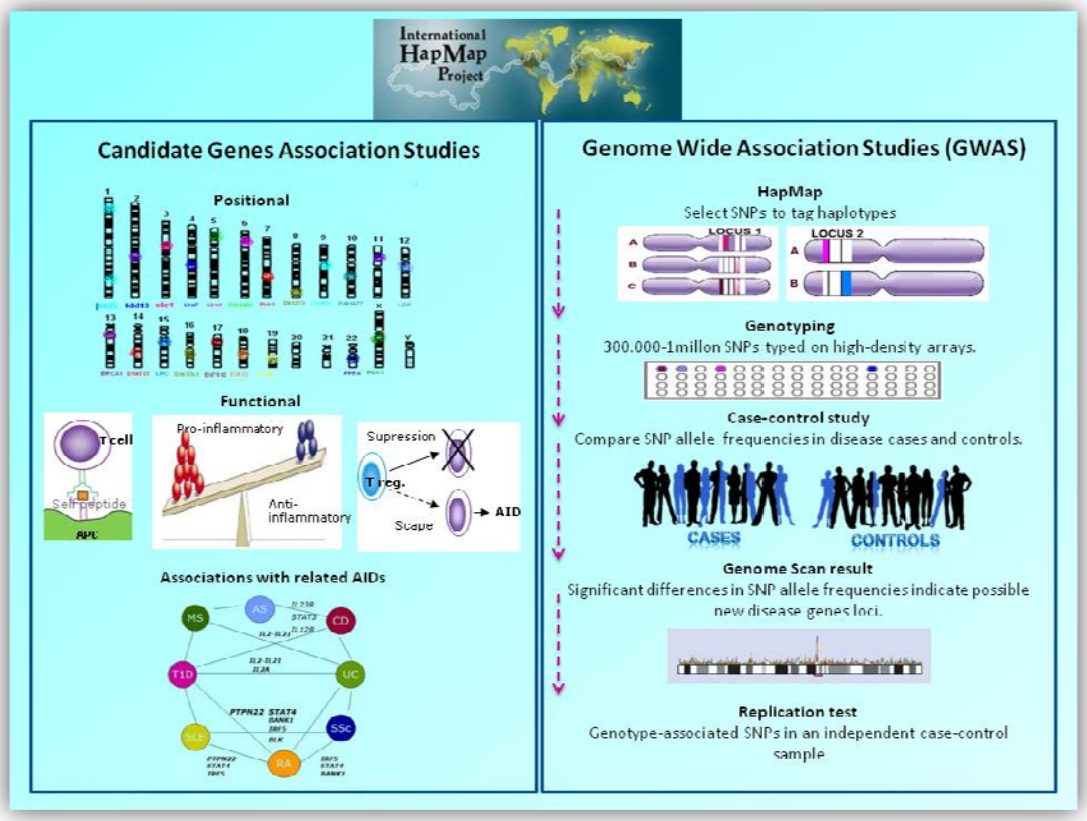


Figure 3. Principal strategies apply in the study of the genetic component of AIDs. The base of both candidate gene association studies and genome wide association studies is the same, the comparison of the allele frequency of the studied polymorphism between individuals with the diseases and healthy donors. In the **candidate genes association studies** predetermine genetic variants are study. They are selected using previous observations such positional signals obtain from family-based linkage studies, functional implications of the genes and/or evidence of previous associations with related diseases. By contrast, the **GWAS** are hypothesis free because multiple genetic variants can be screen all along the human genome simultaneously. This is based on the long haplotype structure of the human genome due to the LD that has been systematically characterized by the HapMap project [76]. After the determination of the higher significant signals a replication in independent case-control groups is performed to refine the associated variants.

Common genetic susceptibility in autoimmune diseases

The concept of a shared origin for AIDs has at times been called; "mosaic of autoimmune manifestations" [78], "kaleidoscope of autoimmunity"[45, 49] or "autoimmune tautology" [44]. Basically those concepts illustrate that AIDs share a number of factors, such as clinical manifestations, immunopathological mechanisms, familial autoimmunity, appearance of polyautoimmunity, and genetic risk factors. Understanding what is common and different in the genetic background is a clue towards figuring out the puzzle of AIDs.

It has been repetitively mentioned before that the main genetic region linked to AIDs is the MHC *loci* (Figure 4) [2, 3, 5, 27, 79, 80]. The MHC region is located in the 6p21 chromosomic region, it extends ~7.6Mb, it has one of the highest genes densities of all the human genome and around 40% of its genes have immunological related functions. The MHC is divided into class I (*HLA-A, -B, -C*), class II (*HLA-DP, -DQ, -DR*) and class III (non-antigen processing and presentation genes). Among its main characteristics is the high LD that represents a big challenge in the AIDs genetic susceptibility research. The main conclusions from the studies of the MHC in AIDs are [79-83]: i) The MHC associations with AIDs result from complex, multilocus effects that span the entire region. Although more of the *loci* associated with AIDs point similar *loci* (i.e. *HLA-DR*), the alleles associated to each disease seems to be specific. For example the most shared diseases susceptibility allele arise from *HLA-DR4* except for SSc and UC, but the high resolution alleles of this group are specific for each disease (i.e. *HLA-DRB1*01:01* in RA and *HLA-DRB1*01:03* in CD). ii) The divers associations of the MHC in one AID can be classified in a hierarchy of association. For example the top association for SLE is *HLA-DRB1*03:01* (marked with high LD by the SNP rs1269852 that is located between *TNXB* and *CREBL1* genes); but there are number of secondary signals that clearly could have biological relevance like *NOTCH4* in class III. iii) The characterization of the LD patterns between the highly polymorphic *HLA* genes and >7,500 common SNPs and deletion-insertion polymorphisms of the MHC region across different populations [84], has allowed refined MHC association studies in RA, SLE, SSc and IBD [79, 80, 82, 83]. Since SNPs outside of the *HLA* genes are informative about *HLA* types. Additionally, alleles in the MHC that are likely to have undergone positive selective pressure have provided good candidates to test AIDs

associations. iv) The MHC associated alleles for the same AIDs vary across the populations, due to difference in the haplotypic structure between human populations. Likewise most of the genetic association studies, the MHC related ones have been performed mainly in Caucasian populations. It is necessary that a more comprehensive approach to the study of the MHC in AIDs is undertaken in huge sample size from different populations. Also it is necessary, the identification of the causal alleles distinguishing the clade associated with each disease by means of use dense SNP genotyping and by direct re-sequencing of the corresponding *locus*.

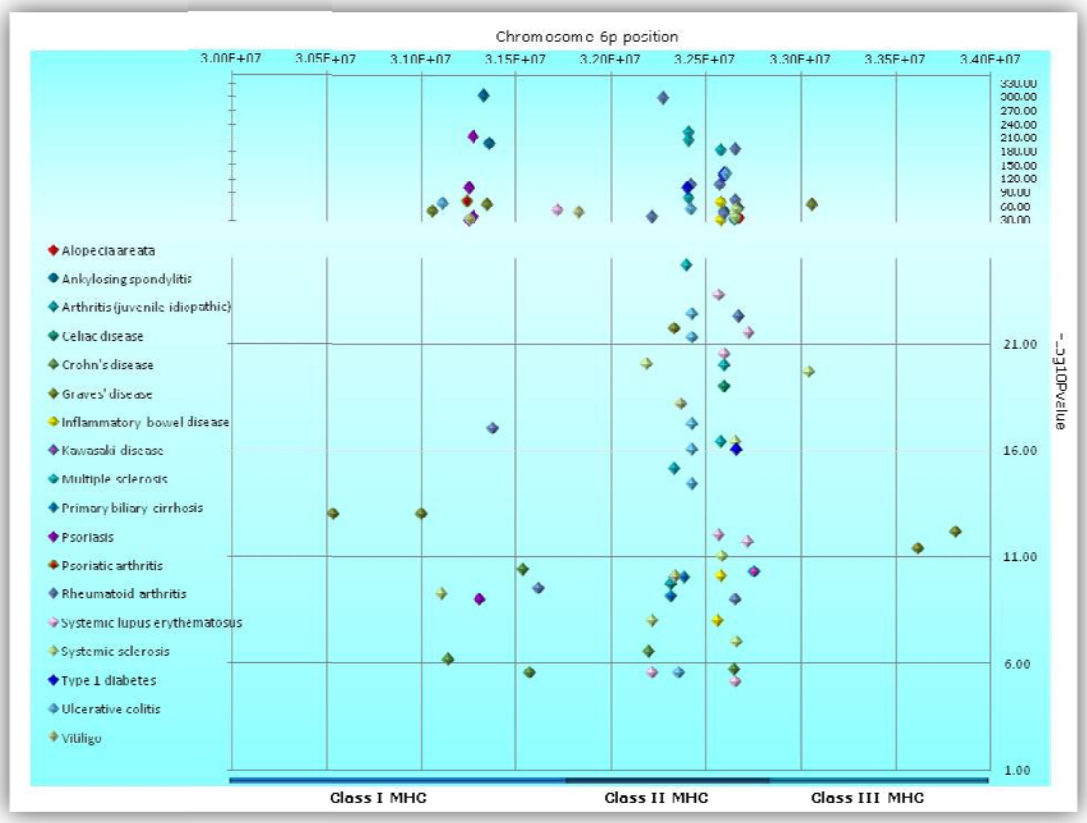


Figure 4. Genetic variants in the MHC associate to several AIDs. The main genetic region linked to AIDs is the MHC. The MHC region is located in the 6p21 chromosomic region, it extends ~7.6Mb, it has one of the highest genes densities of all the human genome and around 40% of its genes have immunological-related functions. The MHC is divided into class I, that contains genes such as *HLA-A, -B, -C*; class II, where the genes *HLA-DP, -DQ, -DR* are located and class III, where non-antigen processing and presentation genes are situated. This plot represents the GWAS level association signals reported until August of 2011 for divers AIDs [85]. Most of the signals are located in the class II followed by the class I, highlighting the previous associations described through different methodologies including case control association studies between *HLA* molecules and AIDs. The plot also shows how the related signals are specific for each AID.

Moreover the stunning GWA approach in AIDs and their combination called pan-meta-analysis [86], are increasing exponentially the number of SNPs associated to inflammatory pathologies [7]. One of the first pan-meta-analysis, combined CD, UC and type 1 diabetes and showed that the same polymorphisms located in the *loci MHC*, *IL27*, *IL18RAP*, *IL10* and *PTPN22* confer risk for type 1 diabetes while have a protection effect against CD and UC [87]. A similar study combining GWAS of CD and celiac diseases found new associations with polymorphisms in the *TAGAP* and *PUS10 loci* [88]. The combine analysis between RA and celiac disease GWAS, reported new genetic variants located in or close to the *DDX6*, *CD247*, *UBE2L3* and *UBASH3a loci* [89]. The pan-meta-analysis in SLE and SSc described for the first time the association of a variant in the *KIAA0319L locus* to AIDs [86].

Complete reviews compiling the non-MHC genes associated to RA, SLE, SSc and IBD have been published [90-94]; along a side with the reviews of the shared genes involved in AIDs [2, 5, 46, 47, 95]. Moreover different tools are available online to explore the genetic associations with AIDs, mainly based in the results from GWAS. Among them are the GWAS catalogue [85], ImmunoBase [96], and the 1000 Genome project browser [97]. A summary of those *loci* for RA, SLE, SSc and IBD are showed in the Figure 5, which shows different examples of *locus* associated with more than one AIDs. Bellow we will focus in the *loci* studied in this thesis.

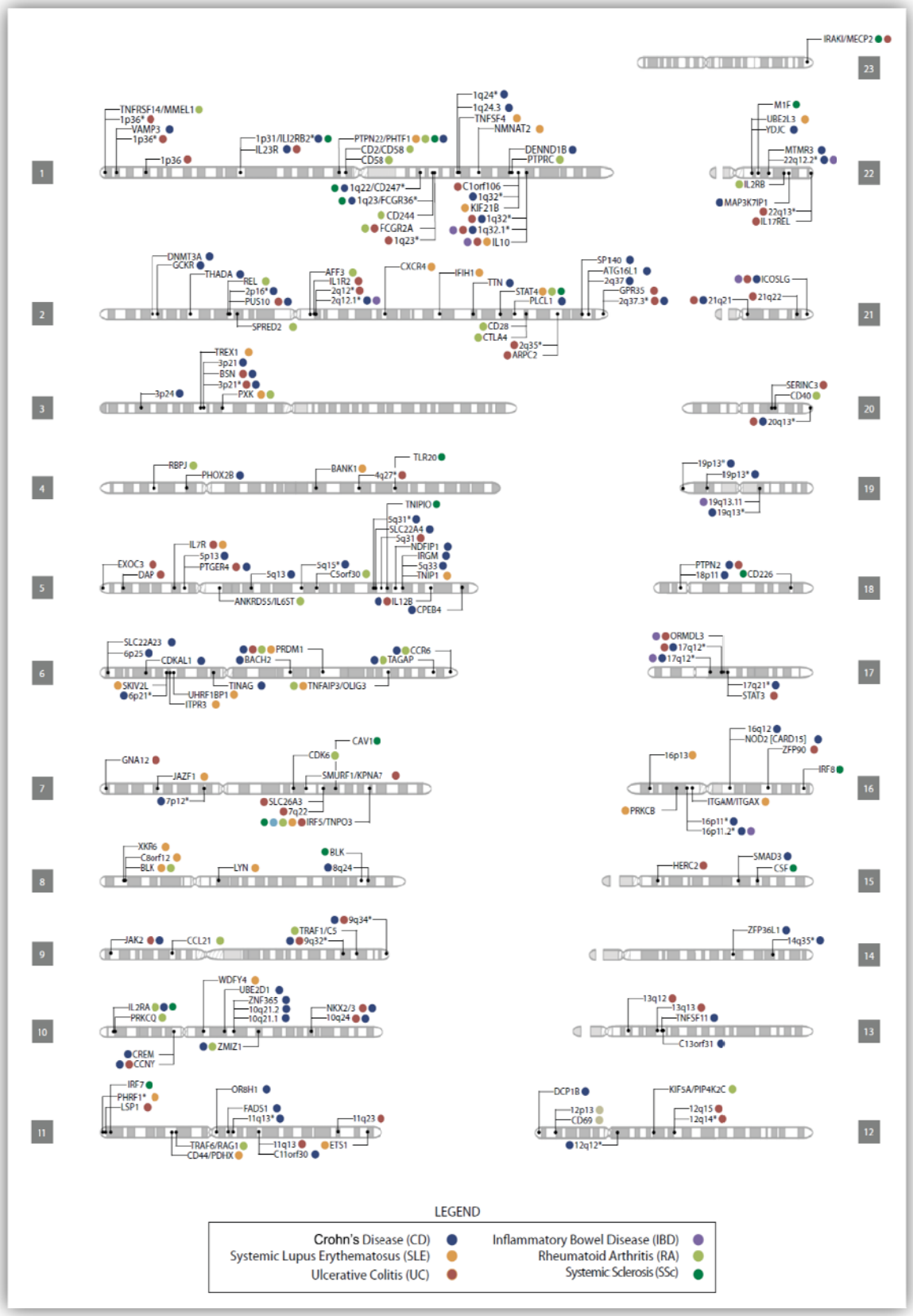

1
TNFRSF14/MMEL1
1p36*
VAMP3
1p36*
1p36
1p31/IL12RB2*
IL23R
PTPN22/PHTF1
CD2/CD58
CD58
1q22/CD247*
1q23/FCGR36*
CD244
FCGR2A
1q23*
1q24*
1q24.3
TNFSF4
NMNAT2
DENND1B
PTPRC
C1orf106
1q32*
KIF21B
1q32*
1q32.1*
IL10
2
DNMT3A
GCKR
THADA
REL
2p16*
PUS10
SPRED2
AFF3
IL1R2
2q12*
2q12.1*
CXCR4
IFIH1
TTN
STAT4
PLCL1
CD28
CTLA4
2q35*
ARPC2
SP140
ATG16L1
2q37
GPR35
2q37.3*
3
TREX1
3p21
BSN
3p21*
PXK
3p24
4
RBPJ
PHOX2B
BANK1
4q27*
TLR20
5
TNIP10
5q31*
SLC22A4
5q31
NDFIP1
IRGM
5q33
TNIP1
IL7R
5p13
PTGER4
EXOC3
DAP
5q13
5q15*
C5orf30
ANKRD55/IL6ST
IL12B
CPEB4
6
SLC22A23
6p25
CDKAL1
SKIV2L
6p21*
UHRF1BP1
ITPR3
PRDM1
BACH2
TINAG
CCR6
TAGAP
TNFAIP3/OLIG3
7
CAV1
GNA12
CDK6
JAZF1
7p12*
SMURF1/KPNA7
SLC26A3
7q22
IRF5/TNPO3
8
XKR6
C8orf12
BLK
LYN
BLK
8q24
9
9q34*
JAK2
CCL21
TRAF1/C5
9q32*
10
WDFY4
UBE2D1
IL2RA
PRKCQ
ZNF365
10q21.2
10q21.1
NKX2/3
10q24
ZMIZ1
CREM
CCNY
11
IRF7
PHRF1*
LSP1
OR8H1
FADS1
11q13*
11q23
TRAF6/RAG1
CD44/PDHX
11q13
C11orf30
ETS1
12
DCP1B
12p13
CD69
12q12*
KIF5A/PIP4K2C
12q15
12q14*
13
13q12
13q13
TNFSF11
C13orf31
14
ZFP36L1
14q35*
15
HERC2
SMAD3
CSF
16
16q12
NOD2 [CARD15]
ZFP90
16p13
IRF8
PRKCB
ITGAM/ITGAX
16p11*
16p11.2*
17
ORMDL3
17q12*
17q12*
17q21*
STAT3
18
PTPN2
18p11
CD226
19
19p13*
19p13*
19q13.11
19q13*
20
SERINC3
CD40
20q13*
21
ICOSLG
21q21
21q22
22
M1F
UBE2L3
YDJC
MTMR3
22q12.2*
IL2RB
MAP3K7IP1
22q13*
IL17REL
23
IRAK1/MECP2
LEGEND
Crohn's Disease (CD)
Systemic Lupus Erythematosus (SLE)
Ulcerative Colitis (UC)
Inflammatory Bowel Disease (IBD)
Rheumatoid Arthritis (RA)
Systemic Sclerosis (SSc)

Figure 5. Human karyogram indicating *loci* associated with the AIDs studied in this thesis. The figure present the most representative *locus* associated with RA, SLE, CD, UC and this two last together referred as IBD. Comparatively the study of the SSc genetic component is recent and it is one of the main focuses in our research group. For that reason most of the *loci* significantly associated to SSc are represented. Modified from Lessard *et al.*[47].

Protein tyrosine phosphatase non-receptor type 22 (PTPN22)

PTPN22 encodes the lymphoid tyrosine phosphatase (LYP), that is a critical regulator of signalling in T and B lymphocytes, and is expressed exclusively in hematopoietic cells. In the T cell receptor (TCR) signalling, LYP acts as a negative regulator. Major substrates of LYP in T cells are pY residues in the activation motif of tyrosine kinases involved in mediating early TCR signaling, such as leukocyte especific protein tyrosine kinase (Lck), FYN oncogene related to SRC, FGR, YES (Fyn), and ζ chain-associated protein tyrosine kinase 70 (ZAP70) [98-100]. The precise mechanisms by which LYP acts in the signalling of the B cell receptor (BCR) are not yet established, but its known that this tyrosine phosphatase protein affects the modulation of the B cell populations and their tolerance checkpoints [101, 102]. Little is known about the role of LYP in other immune cells subpopulations, but it is known that natural killer (NK) and dendritic cells (DC) highly express LYP [96].

The non-synonymous SNP rs2476601 (C1858T, R620W) located in the exon 14 of the *PTPN22* gene, is the leading example of a genetic variant that confers risk of developing AIDs. It is the strongest association for these diseases outside the MHC. Using a CGAS this variant was first associated with type 1 diabetes [103], and subsequently with multiple AIDs in Caucasian populations, among them RA and SLE [104-107]. It was considered that the association was related to diseases characterized by humoral component where autoantibodies are prominent. However, later on the variant was associated in with CD [108, 109]. The association of the rs2476601 with RA and SLE has been also observed in different GWAS [110-112], mine while for SSc. the variant did not exceed the GWAS p value association threshold ($p<10^{-8}$) [113]. During the exploration of other possible *PTPN22* genetics variants associated to RA, two coding SNPs were discovered in the catalytic domain of the protein [114]. Among them, the SNP rs33996649 (G788A, R263Q) in the exon 10 of the *PTPN22* gene, was associated with SLE [115]. The minor allele of this variant exhibited a protection role

for the development of SLE and a similar results have been observed in RA [116]. There is not GWAS information for this variant, because one of the inclusion criteria for this approach is a minor allele frequency upper than 5%. The frequency for the A allele of rs33996649 is around 2% for European ancestry populations [97].

Signal transducer and activator transcription 4 (STAT4)

STAT4 is a critical regulator in the inflammatory process. It transducers interleukin-12 (IL-12), interleukin-23 (IL-23), and type 1 interferon cytokine signals (IFN) in T cells and monocytes, leading to T-helper type 1 (Th1) and T-helper type 17 (Th17) differentiation, monocytes activation, and interferon- γ (IFN- γ) production [117]. *STAT4* is located in the long arm of the chromosome 2. This region was first associated with RA and SLE in genome wide linkage scans in Caucasians families [118-120]. Case-control studies in the same diseases that implemented tagging selection and fine mapping strategies, identified a significant association with the SNP rs7574865 in the third intron of *STAT4* [121, 122]. A correlation between the homozygous for the risk allele of rs7574865 (TT) and higher expression of *STAT4* has observed, indicating a possible functional implication of the genetic variant [122]. Different studies replicated the association of the SNP rs7574865 with RA and SLE mainly in Caucasians but also in Asian and South Latin American populations [123, 124]. The influence of this *STAT4* variant was subsequently revealed in other AIDs such Sjögren's syndrome [125, 126] and lcSSc [127]. Also, the rs7574865 polymorphism was associated with IBD in a small Spanish case-control cohort [128]. The association of the rs7574865 with RA, SLE, SSc and celiac diseases has been detected in several GWAS [89, 129-133].

Interleukin 2 and interleukin 21 *locus* (IL2-IL21)

The first associations with this *locus* to AIDs were observed through GWAS approach in grave's disease, T1D [134, 135] and celiac disease [136]. Almost simultaneously a case-control study in SLE found associated two polymorphisms (rs907715 and rs2221903) of the *IL21* gene [137]. The hit association of the celiac disease study (rs6822844) was evaluated a T1D and RA candidate gene association studies. The results showed that the minor allele of the SNP confers decrease risk for both diseases

[138]. These studies were followed by similar results in other cohorts of RA [139, 140], UC [141, 142], CD [143, 144] and psoriasis [145]. A transethnic fine mapping of this *locus* in SLE restricted the association to two variants: rs907715 and rs6835457 [146]. Reports of the associated *IL2IL21* genetic variants in SSc were not performed until then.

The extended *IL2-IL21 locus* (~480kb) contains also the genes *KIAAA1109* and *ADAD1*. However, the biological role and the location of the rs6822844 polymorphism suggest that the *IL2* and *IL21* genes could be functional implicated in the AIDs susceptibility. IL2 has a paradoxical function in the immune system; it is determinant in the first steps of the T cell proliferation process but its absence produce lethal autoimmunity in mice. The suggestion that *IL2* could be involved in the production the regulatory T cells can explain in part the paradox [147, 148]. On the other hand, *IL21* is predominantly produced by activated CD4+ T cells and natural killer (NK) cells. *IL21* has pleiotropic effects in both innate and adaptive immune response. For example, it can enhance the differentiation of Th17 cells, which are present in inflamed tissues of patients with RA or CD. *IL21* also promote the B cell differentiation, activation or death in the humoral immune response; this feature makes it a good candidate as a therapeutic target [149, 150].

CD24

The small glycosylphosphatidylinositol-anchored cell surface protein, encode by the *CD24* gene, is expressed in cells that participate in the immunopathogenic mechanism of AIDs like macrophagues, T and B cells [96, 151]. Indeed, this protein is involved in the costimulatory activity of the clonal expansion of T cells and it is a check point in the homeostatic proliferation of T cells [152]. *CD24* gene (GeneID 100133941) is located in the *locus* 6q21, and pseudogenes have been anotated in chromosomes 15 and Y, probably due to a segmental duplication. The cDNA of this gene has particular features. Around 10% of its ~2 kb of mRNA, is the protein coding portion; the leading 90% constitute a long 3´untranslate region (UTR) [153, 154]. Two functional polymorphisms of *CD24* have been associated with multiple sclerosis, RA, SLE, autoimmune thyroid diseases and giant cell arteritis [151, 155-161]. A SNP in the exon 2 (rs8734, T226C), that generates a substitution of A by V in the amino acid 57 [161]. The other genetic variant associated with a decrease risk to AIDs is a dinucleotid

deletion at position 1527~1528 (rs3838646, P1527del) from the translation start site that modulates the *CD24* mRNA stability [151, 161].

BCL2-Antagonist/Killer 1 (BAK1)

Immune cell homeostasis is regulated by apoptosis. This is an important process for tolerance because T and B cells expressing antigen receptors that recognize self-antigens with high avidity are eliminated by apoptosis. Moreover, activated T and B lymphocytes are eliminated through apoptosis, after the successful elimination of a pathogen. The purpose of such elimination is to make space for subsequent immune response and to limit collateral damage to healthy tissues. Defects in apoptosis may allow the escape of potential dangerous cells, that could affect the development of AIDs [162]. BCL-2 family member proteins are important in the regulation of the apoptosis pathway; among them *BAK1* plays a proapoptotic role [163]. *BAK1* is overexpressed in the autoimmune lesions of patients with different autoimmune pathologies [164-167]. All together indicate that *BAK1* was well qualified for CGAS. In fact *BAK1* polymorphisms (rs513349, rs561276 and rs5745582) were associated with RA, SLE and Sjögren's syndrome in a Colombian population [168]. However, other studies have not been testing these results.

Ubiquitin associated and Src-homology 3 domain containing A (UBASH3a)

UBASH3a encodes a protein that facilitates T cell apoptosis through the caspasas activation mechanism; it belongs to the T cell ubiquitin ligand (TULA) proteins family that are involved in the TRC suppression signalling. It is expressed exclusively in lymphoid cells and it is implicated in the signalling regulation mediated by protein tyrosine kinases. *UBASH3a* is located in the chromosome region 21q22, contains 15 exons and spans 40kb [169-171]. Different variants of the *UBAHS3a* have been associated with RA, T1D, celiac diseases and vitiligo in GWAS [89, 172-174].

Rheumatoid Arthritis (RA)

RA (MIM#180300)[175] is characterized by synovial inflammation and hyperplasia, production of rheumatoid factor (RF) and anti-citrullinated protein antibody (ACPA) autoantibodies, cartilage and bone destruction, and systemic features, including cardiovascular, pulmonary, psychological and skeletal disorders [40, 41]. Like in most of the AIDs the frequency of affected women is significantly higher than men, with RA two to three times more common among women [34]. RA is defined by criteria that has been useful in harmonizing clinical trials and clinical practice but that are not based on what is known about its etiopathogenesis [176]. One of the seven criteria is the presence of RF, which is not specific for RA but is rather a general consequence of immune activation in the context of immune complex formation. There is not experimental evidence that show the proarthriogenic effect of RF. Comparatively, ACPA autoantibodies are highly specific for RA. They can be found in 60% of RA patients while only in around 2% of the general population and are very rare in other autoimmune conditions [177]. The occurrence of antibodies to citrullinated proteins is seen several years before the onset of the disease, and very few patients develop ACPA after onset of their symptoms [178]. Another important characteristic of RA is that the presence of ACPA is closely linked to the presence of MHC class II alleles that predispose for this disease.

In the early 1970's different studies showed that RA patients tend to share the same HLA genes. Subsequently, serologic studies identified an increased proportion of RA patients positive for HLA-DR*w*4 allele, in comparison to healthy controls; demonstrating that the HLA region is a genetic susceptibility factor to RA (reviewed in [90]). Better characterization of the HLA *locus* allowed determining the different risk alleles in the *HLA-DRB1* gene that encode a conserved amino acid sequence, which led to the "shared epitope" hypothesis [179]. A sequence of five amino acids located in the antigen binding-groove of the HLA protein, and encoded by share epitope alleles are associated mostly with the ACPA-positive subset of RA [53, 180]. It seems that this feature influences the affinity of binding to the citrullinated peptides, and to modulate T-cell responses, nevertheless the precise biological implications for the share epitope are not yet clear. Moreover genetic associations with ACPA-negative RA complicate the shared epitope theory.

Concordance rates of between 15 to 30% among monozygotic twins (MZ) and 5 % among dizygotic twins (DZ) support the genetic component of RA. This suggests that the genetic contribution to this disease approaches 60% [181]. The prevalence of this diseases range between 0.5 to 1.0% in European and North-America populations. On the other hand the American Indian populations have the highest recorded occurrence of RA, with a prevalence of 5.8% for Pima Indians and 6.8% for the Chippewa Indians. By contrast, some studies in South Africa and Nigeria fail to find any RA cases with sample sizes of 500 and 2000 adults, respectively. Also, studies in Southeast Asia, including China and Japan, have shown very low occurrence of RA, of around 0.2-0.3% [50]. Familial aggregation, studies of candidate genes and GWAS sustain clearly a considerable RA genetic component [90].

Table 1. General features of the autoimmune diseases (AIDs) studied in the present thesis.

Phenotype	Abbreviation	Main affection	Auto antibodies	~ % Monozygotic (MZ) twins concordance	% dizygotic (DZ) twins concordance	Ancestry group with higher prevalence	Women: men ratio	Environmental risk factors***	Reference of clinical criteria of classification
Rheumatoid arthritis	**RA**	Synovial inflammation and hyperplasia.	RF, ACPA	15 - 30	5	Native American Indians	3:1	Smoking, crystalline silica, pesticides, hairdressing chemicals, UVB (protective), parvovirus B19, Epstein-Barr virus	[182]
Systemic Lupus Erythematosus	**SLE**	Systemic affection, characterized by production of antinuclear autoantibodies, immune-complex deposition, and subsequent organ damage.	ANAs, anti-dsDNA, anti-Sm, anti-U1 RNP, anti-Ro/SSA, anti-La/SSB, anti-NMDA	24	2	Afro descendants	9:1	Crystalline silica, pesticides, nail polish, sun light, parvovirus B19, Epstein-Barr virus, hepatitis B virus	[183]
Systemic Sclerosis*	**SSc**	Affection of the connective tissue; involved first vascular alterations and extensive fibrosis and autoantibodies against various cellular antigens.	ACA, ATA, ARA	4.2	5.6	Choctaw Native Americans and Afro descendants	3:1 - 8:1	Crystalline silica, occupational exposure to solvents (trichloroethylene), *Helicobacter pylori* infection	[184, 185]

Chron's disease**	**CD**	Chronic relapsing inflammatory disorders of the gastrointestinal tract. The affection is presented in a non-continuous fashion; most commonly the terminal ileum or the perianal region is involved.	--	50	0-3.8	Europe and North America	1:1	Smoking, microflora (complex interaction with the host)	[186]
Ulcerative colitis**	**UC**	Chronic relapsing inflammatory disorders of the gastrointestinal tract. The inflammation is limited to colon and its manifested in a continuous fashion.	--	20	0-4.5	Europe and North America	1:1	Smoking (protective), microflora (complex interaction with the host)	[186]

*The main subtypes of systemic sclerosis are limited cutaneous scleroderma (lcSSc) and diffuse cutaneous scleroderma (dcSSc).

* *There is not a clear consensus related with the specific autoantibodies in IBD (CD and UC).

***Environmental risk factors well establish or suggested by epidemiological studies.

Systemic Lupus Erythematosus (SLE)

SLE (MIM#152700) is considered as prototypic AID because of the production of antinuclear autoantibodies, immune-complex deposition, and subsequent organ damage. The term "lupus erythematosus" was introduced by 19th-century physicians to describe skin lesions and 100 years later it was discovered that the diseases is systemic [39]. The rate between women and men is nine to one, affecting women of childbearing age. Although SLE is not as prevalent as some other chronic diseases, for example hypertension; the overall impact of SLE, measured as health-related quality of life, is significantly worse and affects all health domains at an earlier age in comparison to patients with some other common chronic diseases [34, 187]. The prevalence of SLE ranges from 0.02% in North Europeans people 0.2% among Afro descendent populations. In the United States, people from Africa, Hispanic, or Asian ancestry, as compared with those of other ethnic groups, tend to have an increase prevalence of SLE and greater involvement of organs [188]. The diverse clinical manifestations of SLE present a challenge to the clinician. They range from rash and arthritis through anemia and thrombocytopenia to serositis, nephritis, seizures, and psychosis. A patient with one of these clinical problems, especially in female patients between 15 to 50 years of age, should be part of the SLE differential diagnosis. For the formal diagnosis of SLE, the appearance of at least four of the 11 clinical criteria defined is required [183].

Some of the most recognized autoantibodies in SLE are against: i) the nucleosome (antinuclear antibodies, ANAs) specifically directed to double stranded DNA (anti-dsDNA) and histones. ii) The spliceosome, directed to Sm (nuclear particles consisting of several different polypeptides, anti-Sm) and U1 component of the small nuclear ribonucleoproteins (anti-U1 RNP). iii) Cytoplasmic ribonucleoproteins (anti-Ro/SSA and anti-La/SSB); and iv) the N-methyl-d-aspartate (NMDA) receptor (anti-NMDA). Anti-dsDNA antibodies are present in around 70% of SLE patients but in less than 0.5% of healthy individuals or patients with other AIDs. There are relations between the presences of certain autoantibody specificities in SLE and determined clinical manifestations. For example, anti-dsDNA antibodies are related with nephritis, the anti-NMDA antibodies are related with brain disease and as mention before, the presence of anti-Ro/SSa and anti-La/SSB are associated with fetal loss. From the extensive spectrum of autoantibody against self-antigen in SLE, some of them have been shown to contribute to diseases-related tissue injury [39, 189].

The λ in SLE ranges from 5.8 to 29. Interestingly there is also familiar aggregation of other AIDS in SLE patients, for example there is a λ between 2.02 to 3.3 of RA in SLE patients [26]. Other reports showed that the concordance for MZ is 24% and 2% for DZ [190]. These epidemiologic studies together with the whole-genome scans of families, spontaneous mouse models of SLE, classical candidate gene association studies (CGAS) and GWAS have lead the identification of several SLE susceptibility *loci* [91, 92].

In rare cases, SLE may be associated with recessive genetic modes that exhibit very strong effect sizes. Those are alleles in the complement component genes, *C2, C4A, C4B* y *C1q* (revised in [191]). It is not surprising that the first genetic association in SLE would be the MHC region. Although the history of these *loci* in SLE and others AIDs is not the same as the RA shared epitope hypothesis, the association with the MHC *loci* and AIDs is a constant. The most consistent HLA associations with SLE reside with the class II alleles, *HLA-DRB1*03:01* and *HLA-DRB1*15:01*; moreover the genes *C2*, *C4A* and *C4B* are localized in the class II region of the MHC, showing that the signal from this region is attribute to diverse genetic variants [79, 81, 192].

Systemic Sclerosis (SSc)

SSc (MIM%181750) or scleroderma, affects the connective tissue; and in general its pathogenic processes involves first vascular alterations followed by extensive fibrosis and autoantibodies against various cellular antigens. Although the characteristic attribute is skin thickening, hence the term scleroderma (Greek; "skleros" and "dermos" that mean thickening and skin, respectively); SSc can affect any organ system, including pulmonary, renal, cardiac and gastrointestinal systems [37, 38]. It has the highest mortality of any rheumatic condition [38]. The general SSc classification criterion includes patients, who present Raynaud's phenomenon, sclerodactyly alone, esophageal dysmotility, telangiectasia, calcinosis and scleroderma-related autoantibodies among others [184, 185]. There is considerable variation in the classification of SSc, particularly in the early stages of the diseases. However, the classification criteria are applied to those in whom the diagnosis has already been made, so its relevance is that it can be used by researchers to ensure uniformity in diseases cohorts. There is ongoing work from large international interactions that will help to mould the existing criterion and new proposals allowing better, more accurate comparisons in the near future [184, 185, 193]

The three most frequent autoantibodies in SSc are usually mutually exclusive and important diagnosis markers: i) anti-centromere antibody (ACA), ii) anti-topoisomerase antibody (ATA) and iii) anti-RNA polymerase III (ARA). There are two major subgroups by which SSc is classified, limited cutaneous scleroderma (lcSSc) and diffuse cutaneous scleroderma (dcSSc); they are determined by the difference in the extent of the skin involvement. There is a positive correlation between these subtypes and the presence of the most common autoantibodies. ACA is positive for 50 to 70% of lcSSc patients; ATA is present in 30-40 % of dcSSc patients, ARA is strongly associated with dcSSc [194]. Moreover, strong associations between the type of autoantibody and the organ complications and survival in SSc patients are indicated. ACA are considerably more frequent with SSc-associated pulmonary arterial hypertension (PAH); the presence of ATA is a strong predictor of the development of pulmonary fibrosis and digital ulcers, also ATA levels have been shown to correlate positively with skin score, disease severity and disease activity in SSc. Patients in whom ATA disappear during follow-up have

been reported to have milder disease and better survival. The presence of ARA is a strongest predictor of renal crisis in SSc [37, 38, 178].

The SSc prevalence varies geographically and ethnically, from 0.005-0.3%; also the female to male ratio ranges from 3:1 to 8:1. The highest rate of SSc has been described in Choctaw Native Americans and Afro descendants while the lowest is seen n Japanese people [195, 196]. There are fewer studies of twins and familial aggregation in SSc than RA, SLE or IBD. Nevertheless, the available studies suggest that the concordance for SSc is similar between MZ (4.2%) and DZ (5.6%) [197]; other studies have evaluated the familial aggregation of other AIDs in SSc more than the familial aggregation of SSc itself [198-200]. The studies agreed that two of the most frequent AIDs in relatives of patients with SSc are autoimmune thyroid diseases (AITD) and RA. One of those research also evaluated the SSc familial aggregation in a French population observing a λ of 3.5 [200]. Interesting, SSc probands with ACA and lcSSc subtype were more likely to report familial autoimmunity [198]. Together, these data point that the genetic component of SSc seems to be minor compare with SLE, RA or IBD. These estimates are, however, based on a small number of twins and only on Caucasian populations. This leaves a question mark about the calculations of the percentage of the heritable component in AIDs. Even thou the contribution of heritage is less in SSc than other AIDs, there is still a contribution to described and understand. Indeed, the first significant advances in the study of the genetic of SSc have been stepped recently [94]; for example, there are 11 GWAS in SLE and four in SSc registered in the NHGRI GWAS by now [82, 85, 113, 129, 201]. Classical genetic association studies and GWAS have contributed significantly to the identification of the genetic *locus* related with SSc [94].

Inflammatory Bowel Diseases (IBD)

IBD (MIM#266600) encompass both chronic relapsing inflammatory disorders of the gastrointestinal tract, Crohn's disease (CD) and Ulcerative colitis (UC). It is consider that an inappropriate and continuing inflammatory response to commensal microbes in a genetically susceptibility host is the start point of IBD [42, 43]. For IBD the contribution of autoimmune mechanisms is questioned, but the overlap of genetic associations that have been identifying during the last decade suggests a share immune pathogenesis with AIDs [2, 5]. Moreover, some IBD patient's exhibit extraintestinal autoimmune affectations such as inflammation of the skin, eyes and joins. Also the successful immunosuppressive therapy and the presence of a variety of autoantibodies including antineutrophil cytoplasmic antibodies (ANCA), antibodies against exocrine pancreas (PAB) or intestinal goblet cells (GAB) support the autoimmune background of IBD[202, 203].

The range of IBD is more clearly related with geography than ethnicity. The highest incidence and prevalence of both CD and UC, has been identify in Europe and North America. However increase of IBD's incidence and prevalence from other areas of the world such Africa, Asia and Latin America has been reported. The prevalence for CD ranges from 0.026 to 0.19% and for UC from 0.037 to 0.25%; and the peak onset is in persons 15 to 10 years of age [204]. CD involves any part of gastrointestinal tract; most commonly the terminal ileum or the perianal region, in a non-continuous fashion and, is commonly associated with complications such as strictures, abscesses and fistulas. The microscopic features of CD include thickened submucosa, transmural inflammation, fissuring ulceration and non-caseating granulomas. By contrast, UC is characterized by inflammation that is limited to colon: it begins in the rectum, spreads proximally in a continuous fashion and frequently involves the periappendiceal region. UC shows superficial inflammation changes limited to the mucosa and submucosa with cryptitis and crypt abscesses. IBD diagnosis is based upon the coevaluation of clinical findings, endoscopic, radiological, and histological with the main goal of excluding other conditions with similar presentations and defining the extent and severity of inflammation. The criteria classification for CD takes in account the age at diagnosis, location and behavior. On the other hand, the extension of the affection and the severity are the main factors for the classification in UC [186, 205]. IBD differentiates from RA,

SLE and SSc in the gender bias. The proportion of male and female affected by CD or UC is almost 50:50 [42, 43].

The higher concordance rate among MZ than among DZ twin pairs strongly pointed to a genetic influence of IBD, which seems to be more pronounce with regard to CD (around 50%) than to UC (around 20%) [206, 207]. Among complex diseases, genome association studies have been highly successful in IBD, identifying 163 genetic risk *loci*, as well as the implication of previously unexpected mechanisms, such as autophagy, in its pathogenesis. Other successful result of the GWAS in IBD is the finding that around 110 *loci* are shared between CD and UC, and some others have been associated with different inflammatory diseases. This indicated that CD and UC engage common pathways and maybe part of a mechanistic continuum [43, 93, 208]. Several independent genetic association studies in both CD and UC have shown evidence of linkage to the MHC, suggesting that this region may exert a greater effect in susceptibility to UC rather than CD [48, 81]. Compared with RA, SLE and SSc, the record of published GWAS can be attributed to IBD, with 24 registered GWAS in the NHGRI catalogue [85]. This remark the necessity of increasing the sample size, perform replications and combine genetic association studies (i.e. meta-analysis) to reach similar IBD successful with the previous mentioned diseases.

There are increasing studies showing that the host-microbe interactions shape the genetic architecture of IBD. The gut microflora is a community that has co-evolved with the host and confers beneficial effects, including helping to metabolize nutrients, modulate immune responses and defend against pathogens. The dysregulation of that normal homeostatic relationship is a central factor in the development of IBD. The microbiome can be perceived as a conserved functional entity in which clustering the abundance of genes in certain categories in a species-independent fashion shows high host interindividual similarity. Differences in the abundance of both bacterial species and functional gene categories (such as bacterial motility, sugar and iron metabolism) can differentiate patients with IBD from healthy individuals, demonstrating IBD-related changes in gut microbial ecology [93]. A recent study showed that *loci* associated with IBD, as a group, are subject to both types of selective pressures; directional (consistently favouring one allele over the course of human history) and balancing (favouring the retention of both alleles within populations) selection. That means that both infections organisms and host-microbe co-evolution have influence the genetic

susceptibility to IBD. The authors of the study observed, for example, a considerable overlap between susceptibility *loci* for IBD and mycobacterial infection [208]. All these mentioned characteristics of IBD prompt out this diseased as role model of the missing link between infectious diseases and AIDs.

OBJECTIVES

The main aim of this thesis was to evaluate genetic markers, which have been previously associated with other AIDs, in diseases that have not been studied for such genetic variants. Other objective was to integrate the observed results with the previous findings and to distinguish similarities and differences between the diseases. An additional purpose was to try to detected patrons or characteristics in the share AID genetic variants.

The specific objectives were:

- To re-evaluate the influence of the rs2476601 *PTPN22* polymorphism in SSc and IBD in Caucasian populations.
- To determine whether the novel functional rs33996649 *PTPN22* variant was associated with SSc and IBD in Caucasian populations.
- To re-evaluate the association of the rs7574865 *STAT4* polymorphisms with IBD in Spanish population.
- To study for the first time the influence in SSc of the previous associated *IL2-IL21 locus* polymorphisms.
- To determine whether genetic variants of the *UBASH3a* gene were associated with SLE in Caucasian populations.
- To find out if two polymorphisms (rs8734 and rs3838646) of *CD24* gene influence the susceptibility to IBD in a Spanish population.
- To assess if the *BAK1* variants associated in AIDs in Latin American populations had the same effect in RA and SLE from a Spanish population.

PUBLICATIONS

Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis. Annals of Rheumatic Diseases, 2011

Extended report

Analysis of the influence of *PTPN22* gene polymorphisms in systemic sclerosis

LM Diaz-Gallo,[1] P Gourh,[2] J Broen,[3] C Simeon,[4] V Fonollosa,[4] N Ortego-Centeno,[5] S Agarwal,[2] MC Vonk,[3] M Coenen,[6] G Riemekasten,[7] N Hunzelmann,[8] R Hesselstrand,[9] FK Tan,[2] JD Reveille,[2] S Assassi,[2] FJ García-Hernandez,[10] P Carreira,[11] MT Camps,[12] A Fernandez-Nebro,[13] P Garcia de la Peña,[14] T Nearney,[15] D Hilda,[16] MA González-Gay,[17] P Airo,[18] L Beretta,[19] R Scorza,[19] A Herrick,[20] J Worthington,[20] A Pros,[21] I Gómez-Gracia,[22] L Trapiella,[23] G Espinosa,[24] I Castellvi,[25] T Witte,[26] F de Keyser,[27] M Vanthuyne,[28] MD Mayes,[2] TRDJ Radstake,[3] FC Arnett,[2] J Martin,[1] B Rueda[1]

► Additional data (supplementary figure 1) are published online only. To view these articles please visit the journal online (http://ard.bmj.com)

For numbered affiliations see end of article

Correspondence to
Dr Blanca Rueda, Instituto de Parasitología y Biomedicina 'López-Neyra', Consejo Superior deInvestigaciones Científicas, Parque Tecnológico Ciencias de la Salud, Avenida delConocimiento s/n, 18100-Armilla, Granada, Spain; blarume@ugr.es

Accepted 3 October 2010
Published Online First 3 December 2010

ABSTRACT

Objective Two functional single nucleotide polymorphisms (SNP) in the *PTPN22* gene (rs24746601 and rs33996649) have been associated with autoimmunity. The aim of this study was to investigate the role of the R263Q SNP for the first time and to re-evaluate the role of the R620W SNP in the genetic predisposition to systemic sclerosis (SSc) susceptibility and clinical phenotypes.

Methods 3422 SSc patients (2020 with limited cutaneous SSc and 1208 with diffuse cutaneous SSc) and 3638 healthy controls of Caucasian ancestry from an initial case–control set of Spain and seven additional independent replication cohorts were included in our study. Both rs33996649 and rs2476601 PTPN22 polymorphisms were genotyped by TaqMan allelic discrimination assay. A meta-analysis was performed to test the overall effect of these *PTPN22* polymorphisms in SSc.

Results The meta-analysis revealed evidence of association of the rs2476601 T allele with SSc susceptibility ($p_{FDRcorrected}$=0.03 pooled, OR 1.15, 95% CI 1.03 to 1.28). In addition, the rs2476601 T allele was significantly associated with anticentromere-positive status ($p_{FDRcorrected}$=0.02 pooled, OR 1.22, 95% CI 1.05 to 1.42). Although the rs33996649 A allele was significantly associated with SSc in the Spanish population ($p_{FDRcorrected}$=0.04, OR 0.58, 95% CI 0.36 to 0.92), this association was not confirmed in the meta-analysis (p=0.36 pooled, OR 0.89, 95% CI 0.72 to 1.1).

Conclusion The study suggests that the *PTPN22* R620W polymorphism influences SSc genetic susceptibility but the novel R263Q genetic variant does not. These data strengthen evidence that the R620W mutation is a common risk factor in autoimmune diseases.

Systemic sclerosis (SSc) is a complex disease with an autoimmune origin in which extensive fibrosis, vascular alterations and autoantibodies against various cellular antigens are among the principal features.[1] There are two major subgroups in the actual classification of SSc: limited cutaneous (lcSSc) and diffuse cutaneous (dcSSc).[2] In lcSSc, fibrosis is mainly restricted to the hands, arms and face. Anticentromere antibodies (ACA) occur in 50–90% of lcSSc patients. Conversely, dcSSc is a rapidly progressing disorder that affects a large area of skin and compromises one or more internal organs. Antitopoisomerase I antibodies (ATA) are more frequently associated with this form of SSc.[1 2]

SSc occurs in genetically predisposed individuals who have encountered specific environmental factors and/or other stochastic factors.[1–3] Similar to other autoimmune disorders, the most consistent and reproducible genetic association with SSc corresponds to the major histocompatibility complex.[3] Genes encoding molecules involved in immune function have also recently been associated with susceptibility to SSc, such as *IRF5*, *STAT4* genes and the *C8orf13-BLK* region.[4–9] In spite of these findings, the complete genetic background of SSc, the nature of its genetic determinants and how they contribute to SSc susceptibility and clinical manifestations are still poorly understood.[1 3]

The protein tyrosine phosphatase non-receptor 22 (*PTPN22*) gene encodes the protein tyrosine phosphatase lymphoid tyrosine phosphatase, which is a critical gatekeeper of T-cell receptor (TCR) signalling. In T cells, lymphoid tyrosine phosphatase potently inhibits signalling through dephosphorylation of several substrates, including the Src family kinases Lck and Fyn, as well as ZAP-70 and TCRzeta.[10–12] Interestingly, *PTPN22* has emerged as an important genetic risk factor for human autoimmunity. In particular, two missense single nucleotide polymorphisms (SNP) are associated with autoimmune disorders. The R620W (C1858T, rs2476601) polymorphism in *PTPN22* exon 14 was first associated with type 1 diabetes[13] and subsequently with other autoimmune disorders such as rheumatoid arthritis (RA)[14 15] and systemic lupus erythematosus (SLE)[16] and others (reviewed in Lee *et al*).[17] Interestingly, the role of the R620W polymorphism in SSc has also been investigated and shows a trend of association.[18–21] Another polymorphism in *PTPN22* that is associated with autoimmunity is R263Q (G788A; rs33996649) in exon 10, which alters an amino acid in the catalytic domain of the enzyme. The R263Q polymorphism is a protective factor to

SLE.[22] Both polymorphisms seem to have functional relevance in the immune response.[13 22–26]

In this study, we evaluated the role of the *PTPN22* R263Q polymorphism in SSc for the first time and re-evaluated the influence of the R620W polymorphism in the genetic background of SSc and its clinical phenotypes.

MATERIALS AND METHODS

Patients

A total of 3422 SSc patients and 3638 controls was included in this study. First, we analysed an initial case–control set of 636 SSc patients (370 with lcSSc and 182 with dcSSc) and 1128 healthy controls of Spanish Caucasian ancestry. In addition, seven independent replication cohorts were analysed (Belgium 120 lcSSc, 58 dcSSc and 256 controls; England 344 lcSSc, 128 dcSSc and 373 controls; Germany 164 lcSSc, 128 dcSSc and 288 controls; Italy 292 lcSSc, 115 dcSSc and 371 controls; The Netherlands 131 lcSSc, 41 dcSSc and 277 controls; USA 607 lcSSc, 388 dcSSc and 693 controls; and Sweden 270 lcSSc, 191 dcSSc and 280 controls).

All of the patients fulfilled the 1980 American College of Rheumatology (ACR) classification criteria for SSc.[27] In addition, patients were classified as having limited or diffuse SSc. When patients with SSc have cutaneous involvement distal to the elbows and knees, they fulfil definitions for limited scleroderma.[2] Those SSc patients with cutaneous changes proximal to the elbows and knees were classified as having diffuse SSc.[28] In addition, the following clinical data were collected to ascertain the clinical SSc phenotype: age, gender, disease duration, the presence of SSc-specific autoantibodies and the presence of ACA and ATA (anti-Scl70). The methods used to determine the autoantibodies were the same in all contribution centres and have been described previously.[20] Lung involvement was assessed according to international guidelines.[29] Pulmonary fibrosis was assessed by a CT scan. Restrictive syndrome and diffusion capacity of the lungs was defined as a forced vital capacity of less than 75% of the predicted value and a diffusion capacity for carbon monoxide of less than 75% of the predicted value (based on age, sex, height and ethnic origin). All of the populations studied and their recruiting centres have been reported and described previously.[6 8 9 20 30] The main clinical features of the SSc patients from all of the analysed case sets are summarised in table 1.

Clinical records about the co-occurrence of other autoimmune disorders in SSc patients were not available for all study cohorts. Therefore, to follow a homogeneous inclusion criteria in all case–control sets patients having a concomitant autoimmune condition were not excluded from the analyses. The control population consisted of unrelated, healthy individuals recruited in the same geographical region as the SSc patients and was matched by age, sex and ethnicity with the SSc patients groups. The study was approved by the local ethical committees from all of the participating centres, and written informed consent was obtained from both patients and controls.

PTPN22 genotyping

DNA from patients and controls was obtained using standard methods. Samples were genotyped for the rs33996649 *PTPN22* polymorphism using TaqMan 5′ allelic discrimination assay technology, designed by Custom TaqMan SNP Genotyping Assays (Applied Biosystems, Foster City, California, USA). The rs2476601 *PTPN22* polymorphism was genotyped using a predesigned SNP genotyping assay provided by Applied Biosystems (part number: C__16021387_20). The PCR were performed as described previously.[15]

All samples were genotyped in the same centre to avoid inconsistencies. To verify genotyping accuracy, randomly selected samples were genotyped twice and showed 99% identical genotypes; the call rate was higher than 90% for all studied populations.

Statistical analysis

We tested Hardy–Weinberg equilibrium (HWE) for each case–control set using the program FINETI (http://ihg.gsf.de/cgi-bin/hw/hwa2.pl). Significance was calculated with 2×2 contingency tables and Fisher's exact test to obtain p values, OR and 95% CI using PLINK V.1.07 (http://pngu.mgh.harvard.edu/purcell/plink/). p Values less than 0.05 were considered statistically significant. Multiple testing was corrected by false discovery rate control (p_{FDR}). Linkage disequilibrium measurements (r^2) between rs33996649 and rs2476601 were estimated by the expectation-maximisation algorithm using HAPLOVIEW version 4.1 (Broad Institute of MIT and Harvard).

A search of the literature was made using Medline citations to identify available articles in which the association of the rs2476601 *PTPN22* (C1858T, R620W) polymorphism with SSc disease had been examined. The medical subject heading terms and text words used were 'protein tyrosine phosphatase', 'PTPN22', 'scleroderma' and 'SSc'. A previously published study was included in the meta-analysis if (1) it was published by October 2009, (2) it was original data (independent among studies), (3) it provided enough data to calculate the OR and (4) it included SSc patients diagnosed by the 1980 ACR classification criteria for SSc.[27] An article was excluded if (1) it contained overlapping data, (2) the number of null and wild genotypes could not be ascertained and (3) the patients and controls included were related. We obtained (via personal communication) the frequencies of the R620W polymorphism in the SSc subtypes and autoantibodies if they were not provided in the selected manuscripts. Dieudé *et al*[19]

Table 1 Main clinical features of SSc patients from the Spanish and the seven replication cohorts

	SSc patients population							
	Spain	Belgium	England	Germany	Italy	Netherlands	USA	Sweden
Female (%)	89	77.6	83.5	87	95	72	88.4	78
Limited phenotype (%)	58.2	63.4	72.6	47.2	70.8	68.9	39.2	78.3
ACA positivity (%)	41.8	21.69	28.1	37	42	22.6	28.3	27.4
Anti-Scl70 positivity (%)	19.8	12.5	11.9	26.8	25.3	24.7	16.5	17.7
Pulmonary fibrosis on CT scan (%)	32	–	43	36	36.2	40	–	45
Low FVC (<75% predicted) (%)	28	–	30	18	18	24	–	35
Low DLCO (<75% predicted) (%)	45	–	11.4	50	50	20	–	24

Values are expressed in %
ACA, anticentromere antibody; DLCO, diffusion capacity for carbon monoxide; FVC, forced vital capacity; SSc, systemic sclerosis.

thus provided a more complete dataset for their French cohort, and Gourh *et al*[20] provided genotypic frequency distribution in the subtypes of SSc.

The analysis of the combined data from all populations was performed using Stats Direct version 2.6.6 (StatsDirect Ltd, Cheshire, UK) for global SSc, the clinical subgroups of the disease (lcSSc and dcSSc) and the autoantibody classification (ACA and ATA). Homogeneity of the OR among cohorts was calculated using Breslow–Day and Cochran's *Q* test methods. Higgins' test (I^2) was used to determine if the percentage of total variation across the studies is due to heterogeneity rather than chance (I^2 <25% low, I^2 50% moderate and I^2 >75% high).[31] The pooled OR were calculated under a fixed-effects model (Mantel–Haenszel meta-analysis if I^2 <25% or I^2 50%) or random effects model (DerSimonian–Laird if I^2 >75%).[32] The estimation of the statistical power of the study was performed using the CaTS-Power Calculator (Andrew Skol and Gonçalo Abecasi, 2006).

RESULTS

The *PTPN22* R263Q polymorphism is associated with SSc in the Spanish population

First, we conducted an association study in a case–controls set of Spanish Caucasian ancestry. The distribution of the allelic frequencies of the two studied polymorphisms, R263Q and R620W, was in HWE in both Spanish SSc patients and controls (table 2). The allele frequencies for the R620W variant in the Spanish population (minor allelic frequency (MAF) 0.07) were very similar to that reported previously in Caucasian populations, including the international HapMap project (MAF 0.1) (http://www.hapmap.org) and previous studies (MAF 0.05–0.1).[13–17]

Table 2 Genotype and allele frequencies for rs33996649 and rs2476601 PTPN22 polymorphisms in healthy controls and patients with SSc from eight countries

SNP	Population		GG	GA	AA	G	A	p Value*	OR	(95% CI)
rs33996649	Spain	SSc (n=599)	0.960	0.040	0	0.979	0.0211			
		Controls (n=1128)	0.932	0.068	0	0.966	0.0341	0.04¶	0.58	0.36 to 0.92
	Belgium	SSc (n=187)	0.952	0.048	0	0.976	0.0241			
		Controls (n=236)	0.949	0.051	0	0.975	0.0254	0.9	0.95	0.39 to 2.27
	England	SSc (n=477)	0.956	0.042	0.002	0.977	0.0231			
		Controls (n=382)	0.969	0.031	0	0.984	0.0157	0.28	1.48	0.73 to 3.01
	Germany	SSc (n=354)	0.966	0.034	0	0.983	0.0169			
		Controls (n=285)	0.958	0.042	0	0.979	0.0211	0.59	0.8	0.36 to 1.8
	Italy	SSc (n=419)	0.945	0.055	0	0.973	0.0274			
		Controls (n=371)	0.938	0.062	0	0.969	0.0310	0.68	0.88	0.49 to 1.59
	Netherlands	SSc (n=185)	0.959	0.041	0	0.979	0.0207			
		Controls (n=263)	0.951	0.049	0	0.975	0.0247	0.72	0.83	0.32 to 2.22
	USA	SSc (n=1050)	0.947	0.053	0	0.973	0.0267			
		Controls (n=693)	0.958	0.042	0	0.979	0.0209	0.28	1.28	0.81to 2.02
	Sweden	SSc (n=191)	0.974	0.026	0	0.987	0.0131			
		Controls (n=280)	0.946	0.054	0	0.973	0.0268	0.15	0.48	0.17 to1.34
	Pooled†	SSc(n=3422)	0.954	0.045	0	0.977	0.0230			
		Controls (n=3638)	0.947	0.053	0	0.973	0.0530			
						Fixed model		0.36	0.89	0.72 to 1.12
SNP	**Population**		**CC**	**CT**	**TT**	**C**	**T**	**p Value***	**OR**	**(95% CI)**
rs2476601	Spain	SSc (n=636)	0.857	0.137	0.006	0.925	0.075			
		Controls (n=1128)	0.857	0.136	0.007	0.925	0.075	0.98	0.99	0.77 to 1.29
	Belgium	SSc (n=189)	0.788	0.212	0.000	0.894	0.106			
		Controls (n=256)	0.859	0.129	0.012	0.927	0.073	0.08	1.51	0.95 to 2.42
	England	SSc (n=463)	0.793	0.201	0.006	0.893	0.107			
		Controls (n=373)	0.853	0.147	0.000	0.926	0.074	0.04¶	1.5	1.07 to 2.12
	Germany	SSc (n=343)	0.802	0.175	0.023	0.889	0.111			
		Controls (n=288)	0.774	0.212	0.014	0.880	0.120	0.62	0.92	0.65 to 1.3
	Italy	SSc (n=383)	0.890	0.104	0.005	0.943	0.057			
		Controls (n=371)	0.935	0.059	0.005	0.965	0.035	0.08¶	1.68	1.02 to 2.76
	Netherlands	SSc (n=190)	0.805	0.189	0.005	0.900	0.100			
		Controls (n=277)	0.819	0.177	0.004	0.908	0.092	0.68	1.1	0.7 to 1.71
	Sweden	SSc (n=175)	0.789	0.200	0.011	0.889	0.111			
		Controls (n=279)	0.789	0.190	0.022	0.884	0.116	0.82	0.95	0.62 to 1.45
	Dieudé *et al*[19]‡	SSc (n=1018)	0.820	0.172	0.008	0.906	0.094			
		Controls (n=1004)	0.828	0.163	0.009	0.909	0.091	0.73	1.04	0.84 to 1.28
	Gourh *et al*[20]‡	SSc (n=666)	0.787	0.203	0.011	0.888	0.112			
		Controls (n=430)	0.844	0.147	0.009	0.917	0.083	0.06¶	1.39	1.03 to 1.87
	Pooled§	SSc (n=4063)	0.819	0.173	0.009	0.905	0.095			
		Controls (n=4406)	0.843	0.148	0.008	0.917	0.083			
						Fixed model		0.03¶	1.15	1.03 to 1.28

*p Value for the minor allele.
†I^2=17.6%, Breslow–Day p=0.281 Q=8.49 p=0.29; The statistical power for this pooled analysis was 97%.
‡The authors of those papers provided by personal communication the actualisation and specific data of their works.
§I^2=33.8%, Breslow–Day p=0.145 Q=12.08 p=0.15. The statistical power for this pooled analysis was 100%.
¶False discovery rate correction p value.
SNP, single nucleotide polymorphism; SSc, systemic sclerosis.

After comparing the genotypic and allelic frequencies of the R263Q and R620W *PTPN22* genetic variants between SSc Spanish patients and healthy controls, we observed that the R263Q A allele was associated with SSc (p=0.02 ($p_{FDRcorrected}$=0.04), OR 0.58, 95% CI 0.36 to 0.92). However, we did not observe any significant difference for the R620W polymorphism in this population (p=0.98, OR 0.99, 95% CI 0.77 to 1.29) (table 2).

In addition, we performed an analysis stratifying the patients according to their clinical outcome, that is, dcSSc, lcSSc, ACA or ATA-positive patients. Interestingly, we observed that the frequency of the A allele of the R236Q polymorphism was higher in healthy controls (3%) compared with lcSSc Spanish patients (2%) (p=0.02 ($p_{FDRcorrected}$=0.04), OR 0.49, 95% CI 0.27 to 0.91) (table 3). However, no significant association was found between R620W and the subtypes of the disease or between the R263Q or R620W *PTPN22* polymorphisms and ACA and ATA-positive subsets of SSc (tables 3 and 4).

A replication study and meta-analysis showed that the R620W polymorphism is associated with SSc and ACA-positive patients

In view of the interesting findings observed in the Spanish population, we conducted a large replication study including seven independent populations with Caucasian ancestry. All analysed control populations were in HWE for both *PTPN22* R263Q and R620W genetic variants. As previously reported,[22] no linkage disequilibrium between the *PTPN22* R263Q and R620W genetic variants in any population was observed (r^2<0.03 for all studied populations).

None of the seven replication cohorts showed a significant association of the R263Q *PTPN22* polymorphism with SSc susceptibility (table 2), with the disease subtypes (lcSSc and dcSSc) (table 3), or with ACA and ATA status (table 4). Neither did the meta-analysis confirm the significant association of the R263Q genetic variant and SSc, observed in the Spanish population (p=0.36, pooled OR 0.89, 95% CI 0.72 to 1.12) (figure 1A and table 2).

The replication study of the *PTPN22* R620W polymorphism showed a slightly increased frequency of the T allele in SSc patients compared with controls in most cohorts (table 2). However, this difference was statistically significant after FDR correction only in the English population (p=0.02 ($p_{FDRcorrected}$=0.04), OR 1.50, 95% CI 1.07 to 2.12) (table 2, figure 1B). In this regard, the frequency of the T allele was significantly increased in lcSSc English patients (11%) compared with healthy controls (7%) (p=0.005 ($p_{FDRcorrected}$=0.01), OR 1.67, 95% CI 1.17 to 2.40) (table 3). To evaluate the overall effect of the classic *PTPN22* R620W genetic variant in SSc susceptibility, we performed a meta-analysis including the seven case–control sets analysed in this study (from Spain, Belgium, England, Germany, Italy, The Netherlands and Sweden), together with previously published reports. Four studies that analysed the *PTPN22* R620W polymorphism in SSc were identified through a Medline search (published in October 2009),[18–21] but two of these studies were excluded due to overlapping data,[21 18] The authors of the selected manuscripts were contacted to obtain more detailed data information (see Materials and methods).

Heterogeneity was not observed, and the inconsistency was low in the meta-analysis of the R620W *PTPN22* polymorphism and SSc (Q=12.08, p=0.15, I^2=33.8%). Accordingly, in this pooled analysis, the T allele was significantly associated with susceptibility to develop SSc (p=0.010 ($p_{FDRcorrected}$=0.03), OR 1.15, 95% CI 1.03 to 1.28) (figure 1B and table 2). Moreover, the meta-analysis of the T allele of *PTPN22* R620W and the lcSSc form showed a significant association under the fixed effects model (p=0.02, OR 1.15, 95% CI 1.02 to 1.31). However, heterogeneity was detected (Q=19.84, p=0.01, I^2=59.8%), and the final result was based on the random effects model that showed no significant association between the variant allele and the lcSSc subtype (p=0.12, OR 1.18, 95% CI 0.96 to 1.44) (figure2A and table 3). Neither the meta-analysis for the R620W polymorphism nor for the dcSSc form showed significant association (p=0.10, pooled OR 1.15, 95% CI 0.97 to 1.35) (table 3). However, the meta-analysis performed after stratifying SSc patients according to autoantibody status showed that the T allele of the *PTPN22* R620W polymorphism was a risk factor for the ACA-positive subset of SSc (p=0.01 ($p_{FDRcorrected}$=0.02), pooled OR 1.22, 95% CI 1.05 to 1.42) (figure 2B and table 4). In this line, we performed a meta-analysis for the *PTPN22* R620W polymorphism between the ACA-positive versus ACA-negative patients in the available data cohorts and we observed the same direction of the OR but the results did not reach significant difference (p=0.16, pooled OR 1.15, 95% CI 0.95 to 1.39, see supplementary figure 1, available online only).

DISCUSSION

A loss-of-function *PTPN22* variant (R263Q, G788A, rs33996649), which is less effective in reducing TCR signalling, was recently associated with SLE.[22] Recent findings also demonstrate an overlap in the genetic factors involved in both SLE and SSc, with consistent associations of genetic variants in the *STAT4*, *IRF5* and *BANK1* genes.[6–9 20 30 33 34] This fact prompted us to analyse the possible influence of the R263Q *PTPN22* mutation in the genetic component of SSc and its clinical manifestations. We studied eight Caucasian cohorts and performed a pooled analysis. Although we observed a significant association with the 788A allele in SSc and its lcSSc subtype in the Spanish cohort, this was not confirmed in the pooled analysis that included seven Caucasian ancestry replication cohorts. Considering the statistical power reached by the pooled analysis (97%), we suggest that the R263Q *PTPN22* polymorphism does not influence the genetic background of SSc. In spite of the lack of significant association, the pooled OR shows a protective trend (0.87) in the meta-analysis for the R263Q variant and SSc, the same direction observed for the association of this polymorphism in SLE.[22] In order to determine with greater accuracy if this polymorphism is implicated in the clinical outcome of SSc or the ACA and ATA autoantibody subsets (the statistical power for these analyses range between 42% and 81%), further studies are needed.

Conversely, in the present study, we wanted to re-evaluate the role of the *PTPN22* R620W (C1858T, rs24746601) polymorphism in SSc genetic susceptibility by taking advantage of our large Caucasian cohort and performing a meta-analysis, which is a robust tool for resolving contradictory results and increasing the statistical power in genetic association studies.[35–37] Our meta-analysis had high statistical power (100% for the global SSc disease, 100% for SSc in the lcSSc clinical subgroup, 99% for dcSSc, 99% for SSc ACA autoantibody status and 97% for ATA) and showed evidence of association of the 1858T allele with SSc, the ACA subset, and (to a lesser extent) the lcSSc form. Our data confirm and extend previous reports showing a trend of association of the R620W *PTPN22* variant with SSc.[19 20] For instance, Gourh *et al*[20] observed an increased frequency of the T allele in SSc patients compared with controls in a Caucasian population (p=0.13, OR 1.25). Similarly, Dieudé *et al*[19] conducted a meta-analysis including three Caucasian cohorts, and they

Extended report

Table 3 Distribution of PTPN22 rs33996649 and rs2476601 polymorphisms in SSc subtypes and healthy controls

SNP	Population		GG	GA	AA	G	A	p Value*	OR	(95% CI)
rs33996649	Spain	Limited (n=351)	0.966	0.034	0	0.983	0.0171	0.04††	0.49	0.27 to 0.91
		Diffuse (n=169)	0.947	0.053	0	0.973	0.0266	0.47	0.77	0.38 to 1.56
		Controls (n=1128)	0.932	0.068	0	0.966	0.0341			
	Belgium	Limited (n=119)	0.950	0.050	0	0.975	0.0252	0.99	0.99	0.37 to 2.68
		Diffuse (n=58)	0.931	0.052	0	0.957	0.0259	0.96	1.04	0.29 to 3.73
		Controls (n=236)	0.949	0.051	0	0.975	0.0254			
	England	Limited (n=344)	0.962	0.035	0.003	0.980	0.0203	0.51	1.30	0.60 to 2.83
		Diffuse (n=128)	0.945	0.055	0	0.973	0.0273	0.23	1.76	0.69 to 4.52
		Controls (n=382)	0.969	0.031	0	0.984	0.0157			
	Germany	Limited (n=164)	0.970	0.030	0	0.985	0.0152	0.54	0.72	0.25 to 2.06
		Diffuse (n=128)	0.977	0.023	0	0.992	0.0078	0.35	0.55	0.15 to 1.97
		Controls (n=285)	0.958	0.042	0	0.979	0.0211			
	Netherlands	Limited (n=109)	0.954	0.046	0	0.977	0.0229	0.89	0.93	0.33 to 2.63
		Diffuse (n=31)	0.968	0.032	0	0.984	0.0161	0.67	0.65	0.08 to 5.03
		Controls (n=263)	0.951	0.049	0	0.975	0.0247			
	Italy	Limited (n=292)	0.952	0.062	0	0.976	0.0240	0.44	0.77	0.39 to 1.51
		Diffuse (n=115)	0.939	0.061	0	0.970	0.0304	0.97	0.98	0.42 to 2.32
		Controls (n=371)	0.938	0.062	0	0.969	0.0310			
	USA	Limited (n=607)	0.956	0.044	0	0.978	0.0222	0.82	1.06	0.63 to 1.81
		Diffuse (n=388)	0.938	0.062	0	0.969	0.0309	0.15	1.49	0.86 to 2.58
		Controls (n=693)	0.958	0.042	0	0.979	0.0209			
	Sweden	Limited (n=270)	0.493	0.015	0	0.984	0.0160	0.27	0.54	0.18 to 1.64
		Diffuse (n=191)	0.974	0.026	0	0.987	0.0131	0.28	0.34	0.04 to 2.60
		Controls (n=280)	0.946	0.054	0	0.973	0.0268			
	Pooled†‡	Limited (n=2020)	0.959	0.040	0	0.980	0.0200	0.13	0.81	0.62 to 1.05
		Diffuse (n=1208)	0.951	0.049	0	0.975	0.0250	0.99	0.99	0.73 to 1.33
		Controls (n=3638)	0.947	0.053	0	0.973	0.0530			
SNP	**Population**		**CC**	**CT**	**TT**	**C**	**T**	**p Value***	**OR**	**(95% CI)**
rs2476601	Spain	Limited (n=370)	0.889	0.108	0.003	0.943	0.0568	0.09	0.74	0.52 to 1.05
		Diffuse (n=182)	0.808	0.181	0.011	0.898	0.1016	0.08	1.40	0.96 to 2.03
		Controls (n=1128)	0.857	0.136	0.007	0.925	0.0749			
	Belgium	Limited (n=120)	0.775	0.225	0.000	0.888	0.1125	0.07	1.62	0.96 to 2.73
		Diffuse (n=58)	0.845	0.155	0.000	0.922	0.0776	0.85	1.08	0.50 to 2.30
		Controls (n=256)	0.859	0.129	0.012	0.927	0.0725			
	England	Limited (n=336)	0.774	0.217	0.009	0.882	0.1176	0.01††	1.67	1.17 to 2.40
		Diffuse (n=122)	0.852	0.148	0.000	0.926	0.0738	1.00	1.00	0.58 to 1.74
		Controls (n=373)	0.853	0.147	0.000	0.926	0.0737			
	Germany	Limited (n=162)	0.772	0.185	0.043	0.864	0.1358	0.49	1.16	0.77 to 1.73
		Diffuse (n=120)	0.817	0.175	0.008	0.904	0.0958	0.32	0.78	0.47 to 1.28
		Controls (n=288)	0.774	0.212	0.014	0.880	0.1198			
	Netherlands	Limited (n=131)	0.794	0.206	0.000	0.897	0.1031	0.62	1.13	0.69 to 1.85
		Diffuse (n=41)	0.854	0.122	0.024	0.915	0.0854	0.84	0.92	0.40 to 2.10
		Controls (n=277)	0.819	0.177	0.004	0.908	0.0921			
	Italy	Limited (n=271)	0.886	0.111	0.000	0.941	0.0590	0.09††	1.73	1.02 to 2.94
		Diffuse (n=102)	0.892	0.098	0.010	0.941	0.0588	0.13	1.72	0.85 to 3.47
		Controls (n=371)	0.935	0.059	0.005	0.965	0.0350			
	Sweden	Limited (n=137)	0.803	0.190	0.007	0.898	0.1022	0.54	0.86	0.54 to 1.38
		Diffuse (n=175)	0.789	0.200	0.011	0.889	0.1114	0.48	1.28	0.64 to 2.56
		Controls (n=279)	0.789	0.190	0.022	0.884	0.1165			
	Dieudé *et al*[19]§	Limited (n=641)	0.811	0.181	0.008	0.902	0.0983	0.46	1.10	0.86 to 1.39
		Diffuse (n=315)	0.838	0.156	0.006	0.916	0.0841	0.62	0.93	0.68 to 1.28
		Controls (n=1004)	0.828	0.163	0.009	0.909	0.0906			
	Gourh *et al*[20]§	Limited (n=378)	0.778	0.209	0.013	0.882	0.1177	0.04††	1.48	1.07 to 2.05
		Diffuse (n=254)	0.791	0.201	0.008	0.892	0.1083	0.11	1.35	0.94 to 1.96
		Controls (n=430)	0.844	0.147	0.009	0.917	0.0826			
	Pooled¶**	Limited (n=2546)	0.815	0.176	0.009	0.903	0.0970	0.12	1.18	0.96 to 1.44
		Diffuse (n=1459)	0.820	0.173	0.007	0.906	0.0940	0.28	1.09	0.94 to 1.26
		Controls (n=4406)	0.843	0.148	0.008	0.917	0.0830			

*p Value for the minor allele
†Heterogeneity for systemic sclerosis (SSc) limited analysis $I^2=0.0\%$, Breslow–Day p=0.57 Q=5.59 p=0.58.
‡Heterogeneity for SSc diffused analysis $I^2=9.1\%$, Breslow–Day p=0.33 Q=7.70 p=0.36. The statistical power for this pooled analysis was 81%.
§The authors of those papers provided by personal communication the actualisation and specific data of their works.
¶Heterogeneity for SSc limited analysis $I^2=59.79\%$ Breslow–Day p=0.0102 Q=19.84 p=0.0109. The values showed for the limited form of the SSc meta-analysis correspond to the random effects model. The statistical power for this pooled analysis was 100%.
**Heterogeneity for SSc diffused analysis $I^2=0\%$, Breslow–Day p=0.43 Q=7.91 p=0.44. The statistical power for this pooled analysis was 99%.
††False discovery rate correction p value.
SNP, single nucleotide polymorphism.

Table 4 Distribution of PTPN22 genetic variants according with scleroderma-specific autoantibody status

SNP	Population		GG	GA	AA	G	A	p Value*	OR	(95% CI)
rs33996649	Spain	ACA (n=253)	0.964	0.036	0	0.982	0.0178	0.06	0.51	0.26 to 1.03
		ATA (n=112)	0.964	0.036	0	0.982	0.0179	0.19	0.52	0.19 to 1.42
		Controls (n=1128)	0.932	0.068	0	0.966	0.0341			
	Belgium	ACA (n=40)	0.950	0.050	0	0.975	0.0250	0.98	0.98	0.22 to 4.48
		ATA (n=32)	0.938	0.063	0	0.969	0.0313	0.78	1.24	0.27 to 5.66
		Controls (n=236)	0.949	0.051	0	0.975	0.0254			
	England	ACA (n=133)	0.962	0.038	0	0.981	0.0188	0.73	1.20	0.42 to 3.44
		ATA (n=56)	0.982	0.018	0	0.991	0.0089	0.58	0.56	0.07 to 4.38
		Controls (n=382)	0.969	0.031	0	0.984	0.0157			
	Germany	ACA (n=127)	0.953	0.047	0	0.976	0.0236	0.82	1.13	0.42 to 3.03
		ATA (n=97)	0.979	0.021	0	0.990	0.0103	0.34	0.48	0.11 to 2.18
		Controls (n=285)	0.958	0.042	0	0.979	0.0211			
	Netherlands	ACA (n=37)	0.946	0.054	0	0.973	0.0270	0.91	1.10	0.24 to 4.96
		ATA (n=32)	0.938	0.063	0	0.969	0.0313	0.58	0.79	0.35 to 1.81
		Controls (n=263)	0.951	0.049	0	0.975	0.0247			
	Italy	ACA (n=177)	0.944	0.056	0	0.972	0.0282	0.80	0.91	0.43 to 1.93
		ATA (n=114)	0.939	0.061	0	0.969	0.0307	0.98	0.99	0.42 to 2.34
		Controls (n=371)	0.938	0.062	0	0.969	0.0310			
	USA	ACA (n=296)	0.956	0.044	0	0.978	0.0220	0.18	1.39	0.86 to 2.23
		ATA (n=174)	0.954	0.046	0	0.977	0.0230	0.15	1.49	0.86 to 2.58
		Controls (n=693)	0.958	0.042	0	0.979	0.0209			
	Sweden	ACA (n=51)	0.980	0.020	0	0.990	0.0098	0.30	0.36	0.05 to 2.75
		ATA (n=33)	0.939	0.061	0	0.970	0.0303	0.87	1.14	0.25 to 5.08
		Controls (n=280)	0.946	0.054	0	0.973	0.0268			
	Pooled†‡	ACA (n=1114)	0.957	0.043	0	0.978	0.0225	0.46	0.87	0.63 to 1.21
		ATA (n=650)	0.957	0.043	0	0.978	0.0215	0.48	0.85	0.56 to 1.27
		Controls (n=3638)	0.947	0.053	0	0.973	0.0265			
SNP	**Population**		**CC**	**CT**	**TT**	**C**	**T**	**p Value***	**OR**	**(95% CI)**
rs2476601	Spain	ACA (n=266)	0.853	0.139	0.008	0.923	0.0771	0.87	1.03	0.72 to 1.47
		ATA (n=126)	0.881	0.111	0.008	0.937	0.0635	0.51	0.84	0.49 to 1.42
		Controls (n=1128)	0.857	0.136	0.007	0.925	0.0749			
	Belgium	ACA (n=41)	0.805	0.195	0.000	0.902	0.0976	0.43	1.38	0.62 to 3.08
		ATA (n=32)	0.750	0.250	0.000	0.875	0.1250	0.14	1.83	0.81 to 4.12
		Controls (n=255)	0.863	0.129	0.008	0.927	0.0725			
	England	ACA (n=130)	0.777	0.223	0.000	0.888	0.1115	0.06	1.58	0.98 to 2.53
		ATA (n=55)	0.891	0.091	0.018	0.936	0.0636	0.70	0.85	0.38 to 1.93
		Controls (n=373)	0.853	0.147	0.000	0.926	0.0737			
	Germany	ACA (n=127)	0.787	0.157	0.055	0.866	0.1339	0.57	1.14	0.73 to 1.76
		ATA (n=92)	0.804	0.196	0.000	0.902	0.0978	0.42	0.80	0.46 to 1.38
		Controls (n=288)	0.774	0.212	0.014	0.880	0.1198			
	Netherlands	ACA (n=43)	0.791	0.209	0.000	0.895	0.1047	0.71	1.15	0.55 to 2.44
		ATA (n=47)	0.851	0.149	0.000	0.926	0.0745	0.58	0.79	0.35 to 1.81
		Controls (n=277)	0.819	0.177	0.004	0.908	0.0921			
	Italy	ACA (n=161)	0.888	0.106	0.006	0.941	0.0590	0.07	1.73	0.94 to 3.17
		ATA (n=97)	0.907	0.093	0.000	0.954	0.0464	0.46	1.34	0.62 to 2.91
		Controls (n=371)	0.935	0.059	0.005	0.965	0.0350			
	Sweden	ACA (n=48)	0.792	0.188	0.021	0.885	0.1146	0.96	0.98	0.50 to 1.94
		ATA (n=31)	0.613	0.355	0.032	0.790	0.2097	0.08††	2.01	1.04 to 3.91
		Controls (n=279)	0.789	0.190	0.022	0.884	0.1165			
	Dieudé *et al*[19]§	ACA (n=372)	0.812	0.183	0.005	0.903	0.0968	0.62	1.08	0.81 to 1.44
		ATA (n=247)	0.834	0.158	0.008	0.913	0.0870	0.80	0.97	0.69 to 1.37
		Controls (n=1004)	0.828	0.163	0.009	0.909	0.0906			
	Gourh *et al*[20]§	ACA (n=188)	0.761	0.229	0.011	0.875	0.1250	0.04††	1.60	1.08 to 2.35
		ATA (n=107)	0.710	0.271	0.019	0.846	0.1542	0.005††	2.05	1.32 to 3.19
		Controls (n=430)	0.844	0.147	0.009	0.917	0.0826			
	Pooled¶**	ACA (n=1376)	0.815	0.174	0.011	0.902	0.0981	0.02††	1.22	1.05 to 1.42
		ATA (n=834)	0.824	0.168	0.008	0.908	0.0923	0.26	1.17	0.89 to 1.55
		Controls (n=4126)	0.847	0.145	0.007	0.920	0.0800			

*p Value for the minor allele.
†Heterogeneity for systemic sclerosis (SSc) patients with anticentromere antibody (ACA)-positive and healthy controls analysis $I^2=0\%$, Breslow–Day p=0.61 Q=5.01 p=0.66. The statistical power for this pooled analysis was 69%.
‡Heterogeneity for SSc patients with antitopoisomerase antibody (ATA)-positive and healthy controls analysis $I^2=0.0\%$, Breslow–Day p=0.89 Q=2.82 p=0.90. The statistical power for this pooled analysis was 42%.
‡§These data were provided by personal communication by the authors, because they did not showed the complete frequencies in the previous work.
¶Heterogeneity for SSc patients with ACA-positive and healthy controls analysis $I^2=0\%$, Breslow–Day p=0.60 Q=6.36 p=0.61. The statistical power for this pooled analysis was 99%.
**Heterogeneity for SSc patients with ATA-positive and healthy controls analysis $I^2=49.9\%$, Breslow–Day p=0.04 Q=15.96 p=0.04. The values showed for the ATA-positive status correspond to the random effects model. The statistical power for this pooled analysis was 90%, in this case the values correspond to the random effect model.
††False discovery rate correction p value.
SNP, single nucleotide polymorphism.

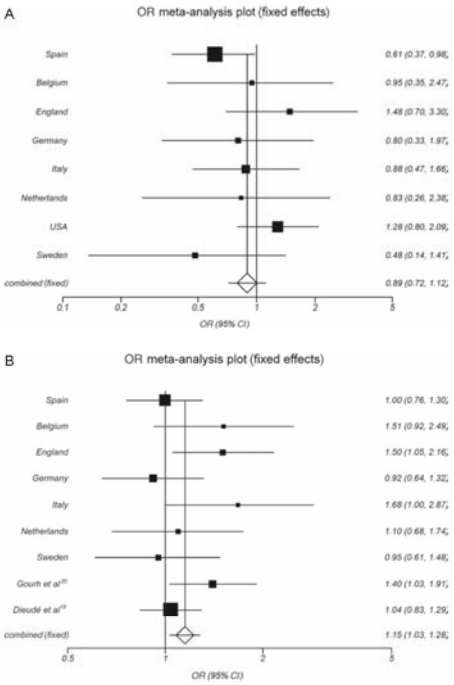


Figure 1 (A) Forest plot for the meta-analysis of the *PTPN22* R263Q (G788A; rs33996649) polymorphism in systemic sclerosis in eight Caucasian cohorts. (B) Forest plot of the *PTPN22* R620W (C1858T; rs2476601) polymorphism and SSc in seven Caucasian cohorts and two previous studies.

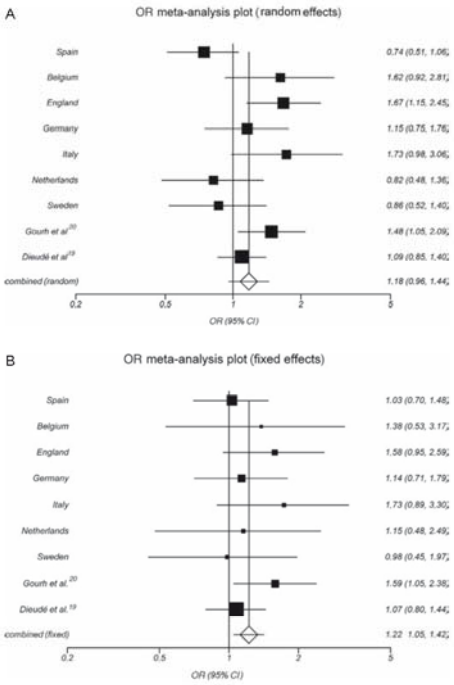


Figure 2 (A) Forest plot for the meta-analysis of the *PTPN22* the R620W (C1858T; rs2476601) polymorphism and limited cutaneous systemic sclerosis subtype, in seven Caucasian cohorts and previous studies. (B) Forest plot for the meta-analysis of the *PTPN22* the R620W (C1858T; rs2476601) polymorphism and anticentromere antibody-positive autoantibody status of systemic sclerosis, in seven Caucasian cohorts and previous studies.

observed a trend of association between the T allele and SSc (p=0.12, OR 1.18). Data from our more powerful meta-analysis showed a statistically significant association of the T allele with SSc ($P_{FRDcorrected}$=0.03, OR 1.15), which highlights the need for large cohorts and a meta-analysis approach to detect minor associations in genetic studies.[35–37] The major discrepancies between our study and previous reports are related to the association of the R620W *PTPN22* variant with autoantibody status. Gourh *et al*[20] showed a significant association between the 1858 T allele and both ACA and ATA-positive status. However, Dieudé *et al*[19] observed a weaker effect (OR 1.08) between the CT/TT genotypes and ATA but not ACA-positive status. In our large study, we only observed an association between the 1858T allele and ACA-positive status. Interestingly, a meta-analysis for this allele between the ACA-positive versus ACA-negative status confirmed this tendency and showed that the OR maintains the risk direction (OR 1.15) but due to the statistical power (75%) we could not detect a significant difference (see supplementary figure 1, available online only). The discrepancies between these results could be explained partly by the clinical heterogeneity of the disease between populations.[38] In addition, it is now clear that the R620W polymorphism displays a wide range of allele frequencies in normal Caucasian populations.[37]

The observed effect magnitude of the 1858T allele on genetic susceptibility to SSc (OR 1.15) seems to be weaker than that of other autoimmune diseases, such as SLE and RA, indicating that the *PTPN22* gene contributes to a lesser extent to SSc genetic susceptibility. However, the specific immunological mechanisms of each disease may explain such results.[37 39]

Some limitations could be attributed to our study, as patients with SSc complicated by other autoimmune diseases could not be excluded from the analysis. Therefore, given that the R620W polymorphism is associated with multiple autoimmune phenotypes, it can be argued that our findings may result from the genetic background of those SSc patients presenting with another autoimmune disease associated with the *PTPN22* gene. Although this possibility cannot be completely discounted, this seems not to be the case, because the most frequent co-autoimmune disease reported in SSc patients, Sjögren's syndrome, is not associated with R620W in Caucasian populations.[17 40] In this regard, it is worth mentioning that Sjögren's syndrome frequently presents concomitantly with other autoimmune diseases. This association is well described for RA or SLE in which the *PTPN22* gene is an important genetic factor.[41]

Conversely, SLE and SSc fit within the same spectrum of interferon-mediated diseases. A subset of SSc patients shows a 'lupus-like' high interferon-inducible gene expression pattern,[38] and recently Kariuki and Niewold[24] demonstrated skewing of serum cytokine profiles in SLE patients carrying the 1858T risk allele towards high serum IFN-α. This implies that the 1858T allele could be a heritable risk factor for SSc through the interferon pathway. On the other hand, other functional studies have shown that primary T cells from patients with autoimmunity (type 1 diabetes and RA) carrying the W620 allele exhibited reduced IL-2 response to TCR engagement.[42 43] IL-2 is known as one of the molecules that shapes immune responses and tolerance.[44] All this together points out the pathways by which the R620W *PTPN22* variant influence autoimmunity, but further functional genetics studies are needed to solve this completely. In conclusion, our results suggest that the R263Q *PTPN22* variant is not associated with SSc, in contrast to the R620W polymorphism that is a known susceptibility factor for SSc and the ACA-positive subset. Our results indicate that compared with the predisposition conferred by the R620W variant in autoimmunity, the protective effect of R263Q appears to be weaker. One possible explanation for these observations is that the loss-of-function effect on TCR signalling due to the R263Q variant may be more easily compensated for than a gain-of-function effect caused by the R620W variant.[22]

Acknowledgments The authors would like to thank Sofia Vargas and Sonia Garcia for their excellent technical assistance, the patients and control donors for their essential collaboration. The authors also thank Dieudé *et al* and Gourh *et al*, for kindly providing an update of their results to do a more complete analysis in the present study.

Funding This work was supported by grants SAF2009-11110, Junta de Andalucía, grants CTS-4977 and CTS-180 and by the RETICS Program, RD08/0075 (RIER) from the Instituto de Salud Carlos III (ISCIII). BR was supported by the I3P CSIC programme funded by the 'Fondo Social Europeo'. LMDG was supported by COLFUTURO and the 'Ayudas Predoctorales de Formación en Investigación en Salud (PFIS – FI09/00544)' from the Instituto de Salud Carlos III. The USA contributors were supported by NIH/NIAMS grants P50AR054144, N01-AR-0-2251 and NIH/NCRR 3UL1RR024148.

Competing interests None.

Patient consent Obtained.

Ethics approval This study was conducted with the approval of each ethics committee of the participating hospitals.

Provenance and peer review Not commissioned; externally peer reviewed.

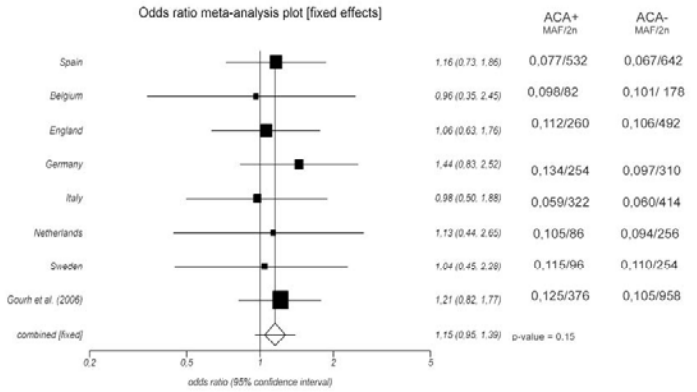
Author affiliations [1]Instituto de Parasitología y Biomedicina Lopez-Neyra CSIC, Granada, Spain
[2]Department of Internal Medicine, Division of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center at Houston Medical School, Houston, USA
[3]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[4]Servicio de Medicina Interna, Hospital Valle de Hebron, Barcelona, Spain
[5]Servicio de Medicina Interna, Hospital Clinico Universitario, Granada, Spain
[6]Departmentof Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[7]Department of Rheumatology and Clinical Immunology, Charité University Hospital, Berlin, Germany
[8]Department of Dermatology, University of Cologne, Germany
[9]Department of Rheumatology, Lund University Hospital, Lund, Sweden
[10]Servicio de Medicina Interna, Hospital Virgen del Rocio, Sevilla, Spain
[11]Servicio de Reumatología, Hospital 12 de Octubre, Madrid, Spain
[12]Servicio de Medicina Interna, Hospital Carlos Haya, Málaga, Spain
[13]Servicio de Reumatología, Hospital Carlos Haya, Málaga, Spain
[14]Servicio de Reumatología, Hospital Ramon y Cajal, Madrid, Spain
[15]University of Texas Medical Branch at Galveston, Galveston, Texas, USA
[16]The University of Texas Health Science Center, San Antonio, Texas, USA
[17]Servicio de Reumatología, Hospital Marqués de Valdecilla, Santander, Spain
[18]Servizio di Reumatologia ed Immunologia Clinica Spedali Civili, Brescia, Italy
[19]Referral Center for Systemic Autoimmune Diseases, University of Milan, Milan, Italy
[20]Rheumatic Diseases Centre, University of Manchester, Salford Royal NHS Foundation Trust, UK
[21]Servicio de Reumatología, Hospital del Mar, Barcelona, Spain
[22]Servicio de Reumatología, Hospital Reina Sofía, Córdoba, Spain
[23]Servicio de Medicina Interna, Hospital Universitario Central de Asturias, Oviedo, Spain
[24]Servicio de Medicina Interna, Hospital Clinico, Barcelona, Spain
[25]Servicio de Reumatología, Hospital de Sant Pau, Barcelona, Spain
[26]Hannover Medical School, Hannover, Germany
[27] Department of Rheumatology, University of Ghent, Belgium
[28]Department of Rheumatology, Université Catholique de Louvain, Belgium

REFERENCES

1. **Gabrielli A,** Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009;**360**:1989–2003.
2. **LeRoy EC,** Medsger TA, Jr. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001;**28**:1573–6.
3. **Agarwal SK,** Tan FK, Arnett FC. Genetics and genomic studies in scleroderma (systemic sclerosis). *Rheum Dis Clin North Am* 2008;**34**:17–40, v.
4. **Dieudé P,** Guedj M, Wipff J, *et al*. Association between the IRF5 rs2004640 functional polymorphism and systemic sclerosis: a new perspective for pulmonary fibrosis. *Arthritis Rheum* 2009;**60**:225–33.
5. **Ito I,** Kawaguchi Y, Kawasaki A, *et al*. Association of a functional polymorphism in the IRF5 region with systemic sclerosis in a Japanese population. *Arthritis Rheum* 2009;**60**:1845–50.
6. **Rueda B,** Broen J, Simeon C, *et al*. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. *Hum Mol Genet* 2009;**18**:2071–7.
7. **Dieudé P,** Guedj M, Wipff J, *et al*. STAT4 is a genetic risk factor for systemic sclerosis having additive effects with IRF5 on disease susceptibility and related pulmonary fibrosis. *Arthritis Rheum* 2009;**60**:2472–9.
8. **Gourh P,** Agarwal SK, Martin E, *et al*. Association of the C8orf13-BLK region with systemic sclerosis in North-American and European populations. *J Autoimmun* 2010;**34**:155–62.
9. **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.
10. **Cloutier JF,** Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* 1999;**189**:111–21.
11. **Wu J,** Katrekar A, Honigberg LA, *et al*. Identification of substrates of human protein-tyrosine phosphatase PTPN22. *J Biol Chem* 2006;**281**:11002–10.
12. **Mustelin T,** Rahmouni S, Bottini N, *et al*. Role of protein tyrosine phosphatases in T cell activation. *Immunol Rev* 2003;**191**:139–47.
13. **Bottini N,** Musumeci L, Alonso A, *et al*. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 2004;**36**:337–8.
14. **Begovich AB,** Carlton VE, Honigberg LA, *et al*. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330–7.
15. **Orozco G,** Sánchez E, González-Gay MA, *et al*. Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus. *Arthritis Rheum* 2005;**52**:219–24.
16. **Kyogoku C,** Langefeld CD, Ortmann WA, *et al*. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 2004;**75**:504–7.
17. **Lee YH,** Rho YH, Choi SJ, *et al*. The PTPN22 C1858T functional polymorphism and autoimmune diseases – a meta-analysis. *Rheumatology (Oxford)* 2007;**46**:49–56.
18. **Balada E,** Simeón-Aznar CP, Serrano-Acedo S, *et al*. Lack of association of the PTPN22 gene polymorphism R620W with systemic sclerosis. *Clin Exp Rheumatol* 2006;**24**:321–4.
19. **Dieudé P,** Guedj M, Wipff J, *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.
20. **Gourh P,** Tan FK, Assassi S, *et al*. Association of the PTPN22 R620W polymorphism with anti-topoisomerase I- and anticentromere antibody-positive systemic sclerosis. *Arthritis Rheum* 2006;**54**:3945–53.
21. **Wipff J,** Allanore Y, Kahan A, *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.
22. **Orrú V,** Tsai SJ, Rueda B, *et al*. A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. *Hum Mol Genet* 2009;**18**:569–79.
23. **Chuang WY,** Ströbel P, Belharazem D, *et al*. The PTPN22 gain-of-function + 1858T(+) genotypes correlate with low IL-2 expression in thymomas and predispose to myasthenia gravis. *Genes Immun* 2009;**10**:667–72.
24. **Kariuki SN,** Niewold TB. Genetic regulation of serum cytokines in systemic lupus erythematosus. *Transl Res* 2010;**155**:109–17.

25. **Stanford SM,** Mustelin TM, Bottini N. Lymphoid tyrosine phosphatase and autoimmunity: human genetics rediscovers tyrosine phosphatases. *Semin Immunopathol* 2010;**32**:127–36.
26. **Zikherman J,** Hermiston M, Steiner D, *et al.* PTPN22 deficiency cooperates with the CD45.E613R allele to break tolerance on a non-autoimmune background. *J Immunol* 2009;**182**:4093–106.
27. 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.
28. **LeRoy EC,** Black C, Fleischmajer R, *et al.* Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;**15**:202–5.
29. **Matucci-Cerinic M,** D'Angelo S, Denton CP, *et al.* Assessment of lung involvement. *Clin Exp Rheumatol* 2003;**21**(3 Suppl 29):S19–23.
30. **Rueda B,** Gourh P, Broen J, *et al.* BANK1 functional variants are associated with susceptibility to diffuse systemic sclerosis in Caucasians. *Ann Rheum Dis* 2010;**69**:700–5.
31. **Higgins JP,** Thompson SG, Deeks JJ, *et al.* Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**:557–60.
32. **Huedo-Medina TB,** Sánchez-Meca J, Marín-Martínez F, *et al.* Assessing heterogeneity in meta-analysis: Q statistic or I^2 index? *Psychol Methods* 2006;**11**:193–206.
33. **Assassi S,** Mayes MD, Arnett FC, *et al.* Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum* 2010;**62**:589–98.
34. **Dieudé P,** Wipff J, Guedj M, *et al.* BANK1 is a genetic risk factor for diffuse cutaneous systemic sclerosis and has additive effects with IRF5 and STAT4. *Arthritis Rheum* 2009;**60**:3447–54.
35. **Attia J,** Ioannidis JP, Thakkinstian A, *et al.* How to use an article about genetic association: B: are the results of the study valid? *JAMA* 2009;**301**:191–7.
36. **Cardon LR,** Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;**2**:91–9.
37. **Gregersen PK,** Olsson LM. Recent advances in the genetics of autoimmune disease. *Annu Rev Immunol* 2009;**27**:363–91.
38. **Walker UA,** Tyndall A, Czirják L, *et al.* Geographical variation of disease manifestations in systemic sclerosis: a report from the EULAR Scleroderma Trials and Research (EUSTAR) group database. *Ann Rheum Dis* 2009;**68**:856–62.
39. **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.
40. **Avouac J,** Airò P, Dieudé P, *et al.* Associated autoimmune diseases in systemic sclerosis define a subset of patients with milder disease: results from 2 large cohorts of European Caucasian patients. *J Rheumatol* 2010;**37**:608–14.
41. **Theander E,** Jacobsson LT. Relationship of Sjögren's syndrome to other connective tissue and autoimmune disorders. *Rheum Dis Clin North Am* 2008;**34**:935–47, viii–ix.
42. **Aarnisalo J,** Treszl A, Svec P, *et al.* Reduced CD4+ T cell activation in children with type 1 diabetes carrying the PTPN22/Lyp 620Trp variant. *J Autoimmun* 2008;**31**:13–21.
43. **Rieck M,** Arechiga A, Onengut-Gumuscu S, *et al.* Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol* 2007;**179**:4704–10.
44. **Malek TR,** Castro I. Interleukin-2 receptor signaling: at the interface between tolerance and immunity. *Immunity* 2010;**33**:153–65.

SUPPLEMENTAL MATERIAL



Supplementary Figure S1. Forest plot for the meta-analysis of the *PTPN22* the R620W (C1858T; rs2476601) polymorphism and ACA-positive versus ACA-negative status patients, in eight Caucasian cohorts. Heterogeneity for this analysis I2=0,0%, Breslow-Day P=0,98 Q=1,41 p=0,98.

Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis. Inflammatory Bowel Diseases, 2011

ORIGINAL ARTICLE

Differential Association of Two *PTPN22* Coding Variants with Crohn's Disease and Ulcerative Colitis

Lina-Marcela Diaz-Gallo, Msc,[1] Laura Espino-Paisán, MSc,[2] Karin Fransen, Msc,[3,4] María Gómez-García, MD, PhD,[5] Suzanne van Sommeren, MD,[3] Carlos Cardeña, MD,[6] Luis Rodrigo, MD, PhD,[7] Juan Luis Mendoza, MD, PhD,[8] Carlos Taxonera, MD, PhD,[8] Antonio Nieto, MD, PhD,[9] Guillermo Alcain, MD,[10] Ignacio Cueto, MD,[10] Miguel A. López-Nevot, MD, PhD,[11] Nunzio Bottini, MD,[12] Murray L. Barclay, MD,[13] J. Bart Crusius, PhD,[14] Adriaan A. van Bodegraven, MD, PhD,[15] Cisca Wijmenga, PhD,[4] Cyriel Y. Ponsioen, MD,[16] Richard B. Gearry, MB, ChB, PhD,[17] Rebecca L. Roberts, PhD,[18] Rinse K. Weersma, MD,[3] Elena Urcelay, MD, PhD,[2] Tony R. Merriman, PhD,[18] Behrooz Z. Alizadeh, MD, PhD,[19]* and Javier Martin, MD, PhD[1]*

Background: The *PTPN22* gene is an important risk factor for human autoimmunity. The aim of this study was to evaluate for the first time the role of the R263Q *PTPN22* polymorphism in ulcerative colitis (UC) and Crohn's disease (CD), and to reevaluate the association of the R620W *PTPN22* polymorphism with both diseases.

Methods: A total of 1677 UC patients, 1903 CD patients, and 3111 healthy controls from an initial case–control set of Spanish Caucasian ancestry and two independent sample sets of European ancestry (Dutch and New Zealand) were included in the study. Genotyping was performed using TaqMan SNP assays for the R263Q (*rs33996649*) and R620W (*rs2476601*) *PTPN22* polymorphisms. Meta-analysis was performed on 6977 CD patients, 5695 UC patients, and 9254 controls to test the overall effect of the minor allele of R620W and R263Q polymorphisms.

Results: The *PTPN22* 263Q loss-of-function variant showed initial evidence of association with UC in the Spanish cohort ($P = 0.026$, odds ratio [OR] = 0.61, 95% confidence interval [CI]: 0.39–0.95), which was confirmed in the meta-analysis ($P = 0.013$ pooled, OR = 0.69, 95% CI: 0.51–0.93). In contrast, the 263Q allele showed no association with CD ($P = 0.22$ pooled, OR = 1.16, 95% CI: 0.91–1.47). We found in the pooled analysis that the PTPN22 620W gain-of-function variant was associated with reduced risk of CD ($P = 7.4E\text{-}06$ pooled OR = 0.81, 95% CI: 0.75–0.89) but not of UC ($P = 0.88$ pooled, OR = 0.98, 95% CI: 0.85–1.15).

Conclusions: Our data suggest that two autoimmunity-associated polymorphisms of the *PTPN22* gene are differentially associated with CD and UC. The R263Q polymorphism only associated with UC, whereas the R620W was significantly associated with only CD.

(Inflamm Bowel Dis 2011;17:2287–2294)

Key Words: protein tyrosine phosphatase, nonreceptor type 22 (PTPN22) gene, inflammatory bowel disease (IBD), ulcerative colitis (UC), Crohn's disease (CD)

Additional supporting information may be found in the online version of this article.

Received for publication December 1, 2010; Accepted December 8, 2010.

From the [1]Instituto de Parasitología y Biomedicina "López-Neyra," CSIC, Granada, Spain, [2]Department of Clinical Immunology, Hospital Clínico S. Carlos, Madrid, Spain, [3]Department of Gastroenterology and Hepatology, University Medical Center Groningen, Groningen, The Netherlands, [4]Department of Genetics, University Medical Center Groningen, The Netherlands, [5]Department of Gastroenterology, Hospital Virgen de las Nieves, Granada, Spain, [6]Department of Gastroenterology, Hospital Clínico San Cecilio, Granada, Spain, [7]Department of Gastroenterology, Hospital Central de Asturias, Oviedo, Spain, [8]Department of Gastroenterology, Hospital Universitario S. Carlos, Madrid, Spain, [9]Department of Immunology, Hospital Puerta del Mar, Cádiz, Spain, [10]Department of Gastroenterology, Hospital Virgen de la Victoria, Málaga, Spain, [11]Department of Immunology, Hospital Virgen de las Nieves, Granada, Spain, [12]Division of Cell Biology, La Jolla Institute for Allergy and Immunology, La Jolla, CA, USA, [13]Department of Medicine, University of Otago, Christchurch, New Zealand, [14]Laboratory of Immunogenetics, Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands, [15]Department of Gastroenterology, VU University Medical Center, Amsterdam, The Netherlands, [16]Department of Gastroenterology and Hepatology, Academic Medical Center Amsterdam, The Netherlands, [17]Department of Medicine, University of Otago, Christchurch, New Zealand, [18]Biochemistry Department, University of Otago, Dunedin, New Zealand, [19]Unit of Genetic Epidemiology & Bioinformatics, Department of Epidemiology, University Medical Center Groningen, Groningen, The Netherlands.

Supported by Plan Nacional de I+D, grant: SAF06-00398), Junta de Andalucía, grant: CTS-1180 and FIS PI08/1676. Health Research Council (HRC) of NZ, grant: 08/075C. R.L.R. is the recipient of a Sir Charles Hercus Health Research Fellowship (HRC). Banco Nacional de ADN (University of Salamanca, Spain) supplied part of the control DNA samples. L.M.D.G. was supported by the "Ayudas Predoctorales de Formación en Investigación en Salud (PFIS – FI09/00544)" and COLFUTURO. B.Z.A. was supported by grants from the Netherlands Organization for Health Research and Development (ZonMw, grant number 016.096.121).

*These authors contributed equally to this work.

Reprints: Javier Martin, MD, PhD, Instituto de Parasitología y Biomedicina "López-Neyra," Consejo Superior de Investigaciones Científicas, Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18100-Armilla (Granada), Spain (e-mail: martin@ipb.csic.es)

DOI 10.1002/ibd.21630

Published online 1 February 2011 in Wiley Online Library (wileyonlinelibrary.com).

Crohn's disease (CD) and ulcerative colitis (UC) are the main types of inflammatory bowel disease (IBD). They are relapsing and chronic inflammatory disorders that result from the complex interaction of genetic, immune, and environmental factors. It is estimated that the current number of loci associated with IBD only explain 10%–20% of the genetic risk attributed to UC and CD. Thus, additional genetic contributions clearly remain to be discovered.[1–4]

The protein tyrosine phosphatase nonreceptor 22 (*PTPN22*) gene encodes the gatekeeper of T-cell receptor (TCR) signaling, protein tyrosine phosphatase (PTP, also known as LYP), and as such is a compelling candidate risk factor for IBD. In T cells, LYP (lymphoid tyrosine phosphatase) potently inhibits signaling through dephosphorylation of several substrates, including the Src-family kinases Lck and Fyn, as well as ZAP-70 and TCRzeta. Moreover, *PTPN22* has emerged as an important genetic risk factor for human autoimmunity.[5–8] Specifically, two missense single nucleotide polymorphisms (SNPs), both with functional influence,[6,8–12] have been associated with autoimmune diseases. The R620W (1858C>T, rs2476601) polymorphism in exon 14 of *PTPN22* was first associated with type 1 diabetes (T1D), and subsequently with autoimmune disorders such as rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), IBD, and other autoimmune diseases.[13–16] The R620W variation disrupts the interaction between Lck and LYP, leading to reduced phosphorylation of LYP, which ultimately contributes to gain-of-function inhibition of T-cell signaling.[17] The Q minor allele of R263Q (788G>A, rs33996649) in exon 10, within the catalytic domain of the enzyme, is a loss-of-function mutation that confers protection against development of SLE and RA.[12,18]

In this study we sought first to determine whether the newly described amino acid substitution, R263Q (788G>A, rs33996649) is associated with altered susceptibility to CD and UC and, second, to reevaluate the influence of the R620W (1858C>T, rs2476601) polymorphism on these diseases by conducting a case–control study and meta-analysis.

MATERIALS AND METHODS

Case–Control Study

Study Population

A total of 1903 CD patients, 1677 UC patients, and 3111 healthy controls from an initial case–control set of Spanish Caucasian ancestry (699 CD patients, 658 UC patients, and 1685 healthy controls) and two independent sample sets of European ancestry from The Netherlands (694 CD patients, 548 UC patients, and 863 healthy controls) and New Zealand (510 CD patients, 471 UC patients, and 563 healthy controls) were included in the case–control study. All IBD patients were diagnosed according to standard clinical, endoscopic, radiologic, and histopathologic criteria.[19 21] Control individuals were matched by Caucasian origin, age, and gender. Written informed consent was obtained from all participants. The study was approved by the Ethics Committee of the Spanish and Dutch hospitals, and by the Upper (cases) and Lower (controls) South Regional Ethics Committees of New Zealand.

PTPN22 Genotyping

DNA from patients and controls was obtained using standard extraction methods. Samples were genotyped for SNP rs33996649 using a Custom TaqMan SNP Genotyping Assay (Applied Biosystems, Foster City, CA). The primer sequences were: forward 5′ TTTGAACTAATGAAGGCCTCT GTGT 3′ and reverse 5′ ATTCCTGAGAACTTCAGTGTTTT CAGT 3′. The specific minor groove binder probe sequences were 5′ TTGATCCGGGAAATG 3′ and 5′ TTGATCCAGGA AATG 3′. The samples were genotyped for rs2476601 polymorphism via TaqMan 5′ allelic discrimination assay using a predesigned probe (Part number: C__16021387_20; Applied Biosystems). To verify the genotyping consistency 10% of samples from each studied cohort were genotyped twice. The concordance between original and repeat genotypes was 99%. The genotype call rate was >90% for all studied populations.

Data Analysis

Deviation from Hardy–Weinberg equilibrium (HWE) was tested by standard chi-square analysis. The differences in genotype distribution and allele frequency among cases and controls were calculated by contingency tables and when necessary by Fisher's exact test. An association was considered statistically significant if $P < 0.05$. Linkage disequilibrium (LD) measurements (r^2) between rs33996649 and rs2476601 were estimated by the expectation-maximization algorithm using HAPLOVIEW v. 4.1 (© Broad Institute of MIT and Harvard 2008, Cambridge, MA). Case–control association analysis was performed using PLINK (v. 1.07) (http://pngu.mgh.harvard.edu/purcell/plink/) to estimate odds ratios (OR) and 95% confidence intervals (CI).[22] To test for associations of the *PTPN22* polymorphisms with clinical features, a univariate analysis using χ^2 or Fisher's exact test was applied. The Montreal Classification[19] criteria were used to determine the clinical variables. We compare each variable with the healthy controls and within cases (see Supporting Information Tables 1–4). Multiple testing was corrected by false discovery rate control (p_{FDR}). Analysis was conducted using PLINK (v. 1.07) and Stats Direct (v. 2.6.6 http://www.statsdirect.com) softwares.

Meta-analysis

Study Selection and Data Extraction

To estimate the common effect of the *PTPN22* R620W polymorphism on IBD we conducted a search on MEDLINE and PUBMED electronic databases up to April 2010 to identify available articles in which this polymorphism was genotyped in patients with CD or UC and healthy controls. The search strategy included Medical Subject Heading (MeSH) terms and text words as follows: "Inflammatory Bowel Disease" [MeSH] OR

TABLE 1. Genotype and Allele Frequencies for the R263Q *PTPN22* (rs33996649) Polymorphism in Healthy Controls and IBD Patients from Three Different Populations

Population		GG	%	GA	%	AA	%	Allele G	%	Allele A	%	*P*-value	OR	(95 % CI)	
Spanish	CD patients ($n = 699$)	640	91.6	59	8.4	0	0.0	1339	95.8	59	4.2	0.073	1.34	0.97	1.85
	UC patients ($n = 658$)	632	96.0	26	4.0	0	0.0	1290	98.0	26	2.0	0.026	0.61	0.39	0.95
	Controls ($n = 1685$)	1580	93.8	103	6.1	2	0.1	3263	96.8	107	3.2				
Dutch	CD patients ($n = 694$)	658	94.8	36	5.2	0	0.0	1352	97.4	36	2.6	0.98	0.99	0.64	1.55
	UC patients ($n = 548$)	523	95.4	25	4.6	0	0.0	1071	97.7	25	2.3	0.58	0.87	0.53	1.43
	Controls ($n = 863$)	818	94.8	45	5.2	0	0.0	1681	97.4	45	2.6				
New Zealand	CD patients ($n = 510$)	490	96.1	20	3.9	0	0.0	1000	98.0	20	2.0	0.87	0.95	0.52	1.74
	UC patients ($n = 471$)	459	97.5	12	2.5	0	0.0	930	98.7	12	1.3	0.17	0.61	0.30	1.24
	Controls ($n = 559$)	536	95.9	23	4.1	0	0.0	1095	97.9	23	2.1				
Pooled	CD patients ($n = 1903$)	1788	94.0	115	6.0	0	0.0	3691	97.0	115	3.0	0.22[a]	1.16	0.91	1.47
	UC patients ($n = 1677$)	1614	96.2	63	3.8	0	0.0	3291	98.1	63	1.9	0.013[b]	0.69	0.51	0.93
	Controls ($n = 3107$)	2934	94.4	171	5.5	2	0.1	6039	97.2	175	2.8				

CD, Crohn's disease'. UC, ulcerative colitis. *P*-value for the minor allele.
[a]Meta-analysis calculated through the fixed effects model. Breslow-Day $P = 0.44$.
[b]Meta-analysis calculated through the fixed effects model. Breslow-Day $P = 0.54$.

"Crohn's Disease"[MeSH]) OR "Colitis, Ulcerative" [MeSH] AND "PTPN22 protein, human "[Substance Name] OR PTPN22. References in the studies were reviewed to identify additional studies not indexed by MEDLINE.

Studies for the meta-analysis were selected if they met the following conditions: 1) diagnosis and phenotype was established by means of the Vienna or Montreal Classifications[19–21]; 2) data were collected in Caucasian populations; 3) the study had a case–control design; 4) the SNPs genotyped were rs2476601 or rs6679677 (both are in complete linkage disequilibrium in Caucasian populations, http://www.hapmap.org); 5) the study supplied enough information to calculate the OR, or the authors provided the data by personal communication (the authors of articles which did not show complete data were contacted by email); 6) the study provided original data (independent of other studies included in the meta-analysis); and 7) the article was published in a peer-reviewed journal as a full article, not as an abstract or similar type of summary.

Our systematic review of the literature identified 28 potential studies for the meta-analysis of R620W in IBD.[13,16,23–47] A total of 15 studies were not included in our analysis.[13,27,28,31–36,38,39,41,44–46] Five of these were not case–control studies[31,34,35,38,41] and three did not genotype rs2476601 or rs6679677.[27,28,36] Another five did not supply enough information to calculate the OR.[13,32,44–46] One included some samples of our Spanish cohort[33] and another was carried out only on patients with ileal CD.[39]

Data Analysis

The analysis of the combined data from all populations was performed using Stats Direct software, v. 2.6.6. The summarized ORs and CIs were obtained by means of both the random (DerSimonian-Laird) and the fixed (Mantel-Haenszel meta-analysis) effect models. The heterogeneity of ORs among cohorts was calculated using Breslow-Day test. The statistical power of the R263Q and R620W meta-analysis was 97%, 99% for CD, and 96%, 99% for UC, respectively (assuming a $P = 0.01$; disease prevalence of 0.1% and allele frequency of 5%; done using CaTS software http://www.sph.umich.edu/csg/abecasis/CaTS/index.html).

RESULTS

R263Q Polymorphism of PTPN22 Is Associated with Reduced Risk of UC

First we conducted an association study in a case–control set of Spanish Caucasian ancestry. The distribution of the allelic frequencies of the two polymorphisms, R263Q and R620W (Tables 1, 2) were in HWE both in patients and controls. As previously reported,[12,18] no LD between the *PTPN22* R263Q and R620W genetic variants was observed in any population ($r^2 < 0.03$ for each studied population).

We observed that the 263Q allele was significantly associated with UC ($P = 0.026$, OR = 0.61, 95% CI: 0.39–0.95) but not with CD ($P = 0.07$, OR = 1.34, 95% CI: 0.97–1.85) (Table 1).

We then conducted a follow-up study in two independent Caucasian populations. The case–control analysis in the Dutch and New Zealand cohorts did not show significant association with the R263Q polymorphism in either the CD (Dutch: $P = 0.98$, OR = 0.99 95%, CI:0.64–1.55, New Zealand: $P = 0.87$, OR = 0.95, 95% CI: 0.52–1.74) or the UC sample sets (Dutch: $P = 0.58$, OR = 0.87, 95% CI: 0.53–1.43, New Zealand: $P = 0.17$, OR = 0.61, 95% CI: 0.30–1.24) (Table 1).

TABLE 2. Genotype and Allele Frequencies for R620W PTPN22 (rs2476601) Polymorphism in Healthy Controls and IBD Patients from 14 Different Populations

Population		CC	%	CT	%	TT	%	Allele C	%	Allele T	%	*P*-value	OR	(95 % CI)
Spanish	CD patients ($n = 699$)	626	89.6	69	9.9	4	0.6	1321	94.5	77	5.5	0.11	0.81	0.62 1.1
	UC patients ($n = 658$)	571	86.8	81	12.3	6	0.9	1223	92.9	93	7.1	0.68	1.05	0.82 1.35
	Controls ($n = 1685$)	1467	87.1	209	12.4	9	0.5	3143	93.3	227	6.7			
Dutch	CD patients ($n = 672$)	575	85.6	94	14.0	3	0.4	1244	92.6	100	7.4	0.036	0.76	0.58 0.98
	UC patients ($n = 539$)	468	86.8	67	12.4	4	0.7	1003	93.0	75	7.0	0.015	0.7	0.52 0.93
	Controls ($n = 834$)	683	81.9	142	17.0	9	1.1	1508	90.4	160	9.6			
New Zealand	CD patients ($n = 477$)	414	86.8	60	12.6	3	0.6	888	93.1	66	6.9	0.014	0.67	0.49 0.92
	UC patients ($n = 448$)	366	81.7	76	17.0	6	1.3	808	90.2	88	9.8	0.93	0.99	0.73 1.32
	Controls ($n = 563$)	454	80.6	106	18.8	3	0.5	1014	90.1	112	9.9			
Anderson et al. (2009) British	UC patients ($n = 2471$)	2024	81.9	425	17.2	22	0.9	4473	90.5	469	9.5	0.74	0.98	0.86 1.12
	Controls ($n = 2483$)	2025	81.6	435	17.5	23	0.9	4485	90.3	481	9.7			
De Jager et al (2006) Canadian	CD patients ($n = 249$)	225	90.0	23	9.6	1	0.4	468	94.8	30	6.0	0.33	1.33	0.71 2.50
	Controls ($n = 207$)	191	92.3	16	7.7	0	0.0	398	95.9	16	3.9			
Duerr et al. (2006) Caucasian European	CD patients ($n = 541$)	473	87.4	68	12.6	0	0.0	1014	93.7	68	6.3	0.003	0.63	0.46 0.86
	Controls ($n = 541$)	441	81.5	95	17.6	5	0.9	977	90.3	105	9.7			
Hradsky et al (2008) Czech	CD patients ($n = 345$)	275	79.7	66	19.1	4	1.2	616	89.3	74	10.7	0.92	1.02	0.74 1.39
	Controls ($n = 501$)	398	79.4	100	20.0	3	0.6	896	89.4	106	10.6			
Latiano et al. (2007)Italian	CD patients ($n = 301$)	283	94.0	18	6.0	0	0.0	584	97.0	18	3.0	0.31	0.73	0.39 1.37
	UC patients ($n = 306$)	278	90.8	28	9.2	0	0.0	584	95.4	28	4.6	0.70	1.10	0.63 1.96
	Controls ($n = 256$)	235	91.8	21	8.2	0	0.0	491	95.9	21	4.1			
Morgan et al. (2010) New Zealand	CD patients ($n = 315$)	260	82.5	52	16.5	3	1.0	572	90.8	58	9.2	0.33	0.85	0.60 1.19
	Controls ($n = 472$)	379	80.3	85	18.0	8	1.7	843	89.3	101	10.7			
Prescott et al. (2005) British	CD patients ($n = 294$)	254	86.4	37	12.6	3	1.0	545	92.7	43	7.3	0.46	0.86	0.58 1.29
	UC patients ($n = 220$)	192	86.9	26	12.2	2	0.9	410	92.8	30	6.8	0.38	0.83	0.53 1.30
	Controls ($n = 374$)	312	83.4	61	16.3	1	0.3	685	91.6	63	8.4			
Silverberg et al. (2009) Caucasian European	UC patients ($n = 1052$)	852	81.0	189	18.0	11	1.0	1893	90.0	211	10.0	0.008	1.27	1.06 1.50
	Controls ($n = 2571$)	2171	84.4	383	14.9	17	0.7	4725	91.9	417	8.1			
Van Oene et al. (2005) Canadian	CD patients ($n = 455$)	389	85.5	63	13.8	3	0.7	841	92.4	69	7.6	0.55	0.91	0.66 1.25
	Controls ($n = 603$)	508	84.2	90	14.9	5	0.8	1106	91.7	100	8.3			
Wagenleiter et al (2005) German	CD patients ($n = 146$)	122	83.6	23	15.8	1	0.7	267	91.4	25	8.6	0.390	0.82	0.49 1.34
	Controls ($n = 254$)	204	80.3	47	18.5	3	1.2	455	89.6	53	10.4			
WTCC (2007) Caucasian	CD patients ($n = 2005$)	1703	84.9	291	14.5	11	0.5	3697	92.2	313	7.8	0.001	0.79	0.69 0.91
	Controls ($n = 3004$)	2447	81.5	533	17.7	24	0.8	5427	90.3	581	9.7			
Pooled	CD patients ($n = 6977$)	6013	86.2	925	13.3	39	0.6	12951	92.8	1003	7.2	7.4E-06a	0.81	0.75 0.89
	Controls ($n = 9254$)	7718	83.4	1467	15.9	69	0.7	16903	91.3	1605	8.7			
	UC patients ($n = 5695$)	4751	83.4	893	15.7	51	0.9	10395	91.3	995	8.7	0.88[b]	0.98	0.85 1.15
	Controls ($n = 8766$)	7347	83.8	1357	15.5	62	0.7	16051	91.6	1481	8.4			

CD, Crohn's disease. UC, ulcerative colitis. *P*-value for the minor allele
[a]Meta-analysis calculated through the fixed effects model. Breslow-Day $P = 0.18$.
[b]Meta-analysis calculated through the random effects model. Breslow-Day $P = 0.03$.

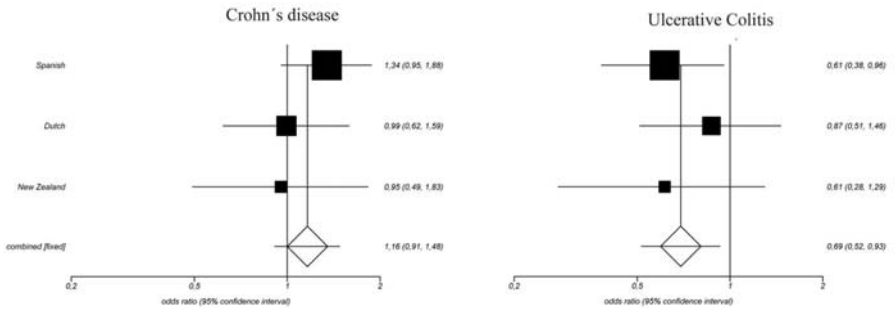


FIGURE 1. Forest plots for the meta-analyses of the *PTPN22* R263Q (G788A; rs33996649) polymorphism in CD and UC. The analyses correspond to the frequency of the minor (A) allele in the three Caucasian IBD sample sets.

Our combined analysis of the three studied Caucasian sample sets did not reveal a significant association between the R263Q polymorphism and CD (P = 0.22 pooled, OR = 1.16, 95% CI: 0.91–1.47) but it did strengthen the initial association observed in UC in the Spanish sample set (P = 0.013 pooled, OR = 0.69, 95% CI: 0.51–0.93) (Table 1; Fig. 1), suggesting that the 263Q variant of the *PTPN22* gene may reduce the risk of UC.

620W Allele of PTPN22 Is Associated with Reduced Risk of CD

In order to reevaluate the role of the R620W polymorphism of the *PTPN22* gene on IBD, we conducted a case–control study in the three Caucasian cohorts. We did not observe a significant difference in genotype or in the minor allele frequency (MAF) between CD patients and healthy controls in the Spanish sample set (P = 0.11, OR = 0.81, 95% CI: 0.62–1.1). In contrast, we observed that the R620W variant was associated with reduced risk of CD in the Dutch sample set (P = 0.036, OR = 0.76, 95% CI: 0.58–0.98) and in the New Zealand sample set (P = 0.014, OR = 0.67, 95% CI: 0.49–0.92) (Table 2). For the UC analysis, we did not observe a significant difference in either the Spanish or the New Zealand sample sets for the R620W polymorphism (Spanish: P = 0.68, OR = 1.05, 95% CI: 0.82–1.35, New Zealand: P = 0.93, OR = 0.99, 95% CI: 0.73–1.32). However, the 620W allele was associated with a reduced risk of UC in the Dutch sample set (P = 0.015, OR = 0.70, 95% CI = 0.52–0.93) (Table 2).

We performed a meta-analysis to reevaluate the role of the R620W polymorphism in IBD. From the remaining 13 studies, three studies fulfilled inclusion criteria for meta-analysis of the R620W *PTPN22* polymorphism in UC,[24,30,37] and Silverberg et al[40] provided the minor allele frequencies of R620W in their initial cohort by personal communication. In CD, eight studies fulfilled inclusion criteria for meta-analysis of the R620W *PTPN22* polymorphism,[23,25,29,30,37,42,43,47] and Duerr et al[26] provided the minor allele frequencies of R620W in their initial cohort by personal communication.

A strong association between the 620W variant and CD was demonstrated (P = 7.4E-06 pooled, OR = 0.81, 95% CI: 0.75–0.89) (Table 2; Fig. 2). This confirms the association of the reduced risk observed between this allele and CD in our initial case–control study in the Dutch and New Zealand sample sets and in the previous meta-analysis reported by Barrett et al.[13] In contrast, no association was observed between the 620W allele and UC (P = 0.88 pooled, OR = 0.98, 95% CI: 0.85–1.15) (Table 2; Fig. 2).

620W Allele of PTPN22 Is Associated with Reduced Risk of Ileal Location in CD

We evaluated the possible associations of the R263Q and R620W variants of *PTPN22* with the clinical phenotypes of UC and CD (Supplementary Tables 1–4). Meta-analysis revealed the 620W variant was significantly associated with reduced risk of ileal location of CD when compared to healthy controls ($P_{FDRcorrected}$ = 9E-03) pooled OR = 0.64, 95% CI = 0.49-0.84, Supplementary Table 2). We observed no significant association of the R263Q polymorphism with CD or UC clinical manifestations.

DISCUSSION

This article reports for the first time the role of the newly identified R263Q polymorphism of *PTPN22* in IBD. In addition, we performed a case–control study in Spanish, Dutch, and New Zealand populations and a meta-analysis to assess the role of the R620W *PTPN22* polymorphism with CD and UC. Our results indicate that there is a differential association of the R263Q and R620W polymorphisms

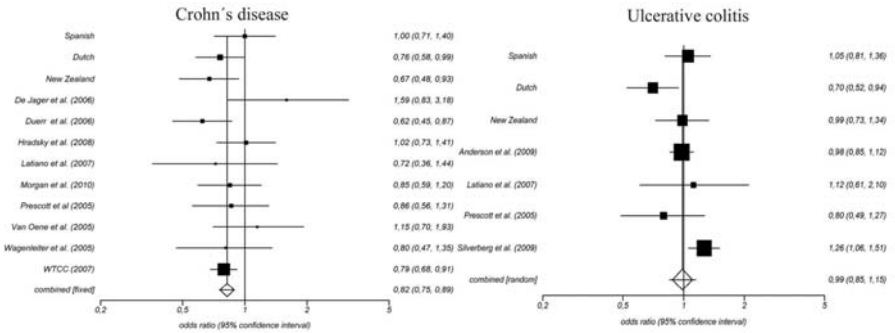


FIGURE 2. Forest plots for the meta-analyses of the *PTPN22* the R620W (C1858T; rs2476601) polymorphism in CD and UC. The analyses correspond to the frequency of the minor (T) allele in 12 Caucasian IBD sample sets.

with IBD. On the one hand, the PTPN22 263Q loss-of-function variant is a protective factor for UC, with no relationship to CD; on the other hand, the 620W gain-of-function variant confers protection against CD, while showing no association with UC. The effect size observed between the R263Q polymorphism and UC (0.69) is similar to that reported for SLE (i.e., 0.63) by Orru et al,[12,18] suggesting that this polymorphism could be another common genetic component in autoimmunity. In addition, we confirmed in the Dutch and New Zealand CD cohorts, together with a combined analysis, the previously reported protective role of the 620W allele in CD but not in UC.[12,15,23,26,32,39,45–47] Thus, there is support for the hypothesis that both outcomes of IBD have a partially different genetic component. On the other hand, we have reported evidence of a reduced risk factor of the 620W allele in the ileal location of CD. Nevertheless, these result should be taken cautiously, since we observed no significant difference when comparing the ileal location against colonic/ileocolonic location of the disease. This may be an artifact of low statistical power of these stratified analyses (i.e., 50%–65% power). Replication studies are needed to confirm this new finding. Increased emphasis has been placed in the recent years on predictive biomarkers to predict the onset or future course of disease.[48] In this regard, the present report supports the idea that subtle genetic differences combined with assessment of the pattern of critical mediators (i.e., presence of autoantibodies) may be useful for tracing progression of the disease.

To determine the immunological implications of the differential association of R263Q and R620W *PTPN22* polymorphisms with CD and UC, functional approaches are required. Nevertheless, there is strong evidence to suggest that the 263Q allele is a loss-of-function variant which is less effective in reducing TCR signaling than 263R.[12] This supports the hypothesis that positive modulation of the TCR helps in reestablishing tolerance in at least a subset of autoimmune patients.[6,49] This functional evidence, together with the significant association that we observed with UC, suggests that TCR signaling is more important in this disease than in CD. Actually, autoantibodies are more often detected in UC than in CD patients. It is estimated that 60%–70% of UC patients are positive for atypical antineutrophilic cytoplasmic antibodies, whereas only few CD patients present autoantibodies (atypical antineutrophilic cytoplasmic antibodies 5%–25%, pancreatic autoantibodies 27%–37%, and thrombophilia-associated antibodies 3%–37%).[50]

The present study confirms that the 620W allele is associated with a reduced risk of developing CD, in contrast to the increasing risk that this genetic variant confers to other autoimmune diseases such as T1D, SLE, and RA.[6,14–16] Several authors have shown that 620W *PTPN22* is a gain-of-function variant that reduces TCR signaling leading to decreased elimination of potentially autoreactive T cells and/or decreased production of natural regulatory T cells (Treg) (reviewed[6]). This could explain the loss of tolerance that takes place in autoimmune diseases like T1D, SLE, and RA, but not the protective role 620W allele appears to confer against CD. A possible explanation could be that IBD may represent an inappropriate immune response to the commensal microbiota in a genetically predisposed host,[3] mimicking an infection process. This hypothesis is supported by the fact that the 620W allele confers protection towards some highly prevalent infectious diseases.[6] Previous studies have reported a significant protective role of the 620W allele in tuberculosis (TB).[51,52] Moreover, the R263Q polymorphism has been associated with increasing risk to develop TB,[52] the opposite of the reported

associations with SLE[12] and RA[18] and UC in the present study. Our findings suggest that many of the genetic loci involved in autoimmunity may be under balanced selection due to antagonistic pleiotropic effects. Genetic variants such as R620W and R263Q with opposite effects in different diseases may facilitate the maintenance of common susceptibility alleles in human populations.[6,46,53] Moreover, our results also support the idea that CD and UC differ in some genetic risk factors, thereby suggesting the involvement of different immunological mechanisms with a related nature.[24,45,46,54,55]

ACKNOWLEDGMENTS

We thank Sofia Vargas and Sonia Garcia for excellent technical assistance. We thank all the donors, patients, and controls.

REFERENCES

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. *Lancet*. 2007;369:1627–1640.
2. Budarf ML, Labbe C, David G, et al. GWA studies: rewriting the story of IBD. *Trends Genet*. 2009;25:137–146.
3. Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. *Annu Rev Immunol*. 2010;28:573–621.
4. Manolio TA, Collins FS, Cox NJ, et al. Finding the missing heritability of complex diseases. *Nature*. 2009;461:747–753.
5. Cloutier JF, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med*. 1999;189:111–121.
6. Stanford SM, Mustelin TM, Bottini N. Lymphoid tyrosine phosphatase and autoimmunity: human genetics rediscovers tyrosine phosphatases. *Semin Immunopathol*. 2010;32:127–36.
7. Wu J, Katrekar A, Honigberg LA, et al. Identification of substrates of human protein-tyrosine phosphatase PTPN22. *J Biol Chem*. 2006;281:11002–11010.
8. Yu X, Sun JP, He Y, et al. Structure, inhibitor, and regulatory mechanism of Lyp, a lymphoid-specific tyrosine phosphatase implicated in autoimmune diseases. *Proc Natl Acad Sci U S A*. 2007;104:19767–19772.
9. Arechiga AF, Habib T, He Y, et al. Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. *J Immunol*. 2009;182:3343–3347.
10. Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet*. 2004;36:337–338.
11. Liu Y, Stanford SM, Jog SP, et al. Regulation of lymphoid tyrosine phosphatase activity: inhibition of the catalytic domain by the proximal interdomain. *Biochemistry*. 2009;48:7525–7532.
12. Orru V, Tsai SJ, Rueda B, et al. A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus. *Hum Mol Genet*. 2009;18:569–579.
13. Barrett JC, Hansoul S, Nicolae DL, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*. 2008;40:955–962.
14. Bottini N, Vang T, Cucca F, et al. Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol*. 2006;18:207–213.
15. Gregersen PK, Lee HS, Batliwalla F, et al. PTPN22: setting thresholds for autoimmunity. *Semin Immunol*. 2006;18:214–223.
16. Lee YH, Rho YH, Choi SJ, et al. The PTPN22 C1858T functional polymorphism and autoimmune diseases—a meta-analysis. *Rheumatology (Oxford)*. 2007;46:49–56.
17. Fiorillo E, Orru V, Stanford SM, et al. Autoimmune-associated PTPN22 R620W variation reduces phosphorylation of lymphoid phosphatase on an inhibitory tyrosine residue. *J Biol Chem*. 2010;285:26506–25618.
18. Rodriguez-Rodriguez L, Wan Taib WR, Topless R, et al. The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in Caucasian case-control samples. *Arthritis Rheum*. 2010, Nov 15. [Epub ahead of print].
19. Gasche C, Scholmerich J, Brynskov J, et al. A simple classification of Crohn's disease: report of the Working Party for the World Congresses of Gastroenterology, Vienna 1998. *Inflamm Bowel Dis*. 2000;6:8–15.
20. Satsangi J, Silverberg MS, Vermeire S, et al. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut*. 2006;55:749–753.
21. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol*. 2005;19(suppl A):5–36.
22. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–575.
23. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447:661–678.
24. Anderson CA, Massey DC, Barrett JC, et al. Investigation of Crohn's disease risk loci in ulcerative colitis further defines their molecular relationship. *Gastroenterology*. 2009;136:523–529 e523.
25. De Jager PL, Sawcer S, Waliszewska A, et al. Evaluating the role of the 620W allele of protein tyrosine phosphatase PTPN22 in Crohn's disease and multiple sclerosis. *Eur J Hum Genet*. 2006;14:317–321.
26. Duerr RH, Taylor KD, Brant SR, et al. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science*. 2006;314:1461–1463.
27. Franke A, Balschun T, Karlsen TH, et al. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet*. 2008;40:713–715.
28. Hampe J, Franke A, Rosenstiel P, et al. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet*. 2007;39:207–211.
29. Hradsky O, Lenicek M, Dusatkova P, et al. Variants of CARD15, TNFA and PTPN22 and susceptibility to Crohn's disease in the Czech population: high frequency of the CARD15 1007fs. *Tissue Antigens*. 2008;71:538–547.
30. Latiano A, Palmieri O, Valvano MR, et al. Evaluating the role of the genetic variations of PTPN22, NFKB1, and FcGRIIIA genes in inflammatory bowel disease: a meta-analysis. *Inflamm Bowel Dis*. 2007;13:1212–1219.
31. Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. *Hum Mol Genet*. 2008;17:R116–121.
32. Libioulle C, Louis E, Hansoul S, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p13.1 and modulates expression of PTGER4. *PLoS Genet*. 2007;3:e58.
33. Martin MC, Oliver J, Urcelay E, et al. The functional genetic variation in the PTPN22 gene has a negligible effect on the susceptibility to develop inflammatory bowel disease. *Tissue Antigens*. 2005;66:314–317.
34. Massey DC, Parkes M. Genome-wide association scanning highlights two autophagy genes, ATG16L1 and IRGM, as being significantly associated with Crohn's disease. *Autophagy*. 2007;3:649–651.
35. Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet*. 2008;9:9–14.
36. Parkes M, Barrett JC, Prescott NJ, et al. Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to Crohn's disease susceptibility. *Nat Genet*. 2007;39:830–832.
37. Prescott NJ, Fisher SA, Onnie C, et al. A general autoimmunity gene (PTPN22) is not associated with inflammatory bowel disease in a British population. *Tissue Antigens*. 2005;66:318–320.
38. Raelson JV, Little RD, Ruether A, et al. Genome-wide association study for Crohn's disease in the Quebec Founder Population identifies multiple validated disease loci. *Proc Natl Acad Sci U S A*. 2007;104:14747–14752.
39. Rioux JD, Xavier RJ, Taylor KD, et al. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet*. 2007;39:596–604.
40. Silverberg MS, Cho JH, Rioux JD, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. *Nat Genet*. 2009;41:216–220.

41. Torkamani A, Topol EJ, Schork NJ. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics*. 2008;92:265–272.
42. van Oene M, Wintle RF, Liu X, et al. Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. *Arthritis Rheum*. 2005;52:1993–1998.
43. Wagenleiter SE, Klein W, Griga T, et al. A case-control study of tyrosine phosphatase (PTPN22) confirms the lack of association with Crohn's disease. *Int J Immunogenet*. 2005;32:323–324.
44. Yamazaki K, McGovern D, Ragoussis J, et al. Single nucleotide polymorphisms in TNFSF15 confer susceptibility to Crohn's disease. *Hum Mol Genet*. 2005;14:3499–3506.
45. McGovern DP, Gardet A, Torkvist L, et al. Genome-wide association identifies multiple ulcerative colitis susceptibility loci. *Nat Genet*. 2010;42:332–337.
46. Wang K, Baldassano R, Zhang H, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. *Hum Mol Genet*. 2010;19:2059–67.
47. Morgan AR, Han DY, Huebner C, et al. PTPN2 but not PTPN22 is associated with Crohn's disease in a New Zealand population. *Tissue Antigens*. 2010;76:119–125.
48. Rose NR. Predictors of autoimmune disease: autoantibodies and beyond. *Autoimmunity*. 2008;41:419–428.
49. Chatenoud L. CD3-specific antibodies as promising tools to aim at immune tolerance in the clinic. *Int Rev Immunol*. 2006;25:215–233.
50. Behr MA, Divangahi M, Lalande JD. What's in a name? The (mis)labelling of Crohn's as an autoimmune disease. *Lancet*. 2010;376:202–203.
51. Gomez LM, Anaya JM, Martin J. Genetic influence of PTPN22 R620W polymorphism in tuberculosis. *Hum Immunol*. 2005;66:1242–1247.
52. Lamsyah H, Rueda B, Baassi L, et al. Association of PTPN22 gene functional variants with development of pulmonary tuberculosis in Moroccan population. *Tissue Antigens*. 2009;74:228–232.
53. Dean M, Carrington M, O'Brien SJ. Balanced polymorphism selected by genetic versus infectious human disease. *Annu Rev Genomics Hum Genet*. 2002;3:263–292.
54. Brant SR. Exposed: the genetic underpinnings of ulcerative colitis relative to Crohn's disease. *Gastroenterology*. 2009;136:396–399.
55. Diaz-Gallo LM, Palomino-Morales RJ, Gomez-Garcia M, et al. STAT4 gene influences genetic predisposition to ulcerative colitis but not Crohn's disease in the Spanish population: a replication study. *Hum Immunol*. 2010;71:515–519.

SUPPLEMENTAL TABLES

Supplementary Table 1. Allele frequencies distribution of the R263Q polymorphism in three Caucasian cohorts according to the clinical classification variables of CD[18-20] and healthy controls.

Population	Clinical Variable (N)[b]	2N	G	%	A	%	p value	OR	95% CI	
Spain	**Diagnosis Age A (248)**									
	<16 (A1)	66	64	96.97	2	3.03	0.94	1.39	0.39	5
	17 - 40 (A2)	332	317	95.48	15	4.52	0.19	1.52	0.88	2.62
	>40 (A3)	98	95	96.94	3	3.06	0.95	1.26	0.43	3.72
	Disease Location L (482)									
	Ileal (L1)	416	395	94.95	21	5.05	0.41[c]	1.68	1.04	2.7
	Colonic (L2)	178	171	96.07	7	3.93	0.58	1.41	0.66	2.9
	Ileocolonic (L3)	370	356	96.22	14	3.78	0.53	1.27	0.73	2.22
	Upper GI Tract (L4)	40	40	100.00	0	0.00	-	-	-	-
	Disease Behavior B (340)	680	651	95.74	29	4.26				
	Perforating (B3)	198	191	96.46	7	3.54	0.78	1.26	0.59	2.68
	Stricturing (B2)	90	88	97.78	2	2.22	0.61	1.02	0.29	3.63
	Inflammatory (B1)	392	372	94.90	20	5.10	0.29[c]	1.76	1.1	2.85
	Controls (1685)	3370	3263	96.82	107	3.18				
Dutch	**Diagnosis Age A (452)**									
	<16 (A1)	152	147	96.71	5	3.29	0.61	1.48	0.6	3.65
	17 - 40 (A2)	584	573	98.12	11	1.88	0.33	0.76	0.39	1.47
	>40 (A3)	168	164	97.62	4	2.38	0.86	1.1	0.42	2.96
	Disease Location L (425)									
	Ileal (L1)	226	220	97.35	6	2.65	0.97	1.16	0.5	2.67
	Colonic (L2)	206	203	98.54	3	1.46	0.32	0.72	0.24	2.15
	Ileocolonic (L3)	418	409	97.85	9	2.15	0.6	0.89	0.44	1.81
	Upper GI Tract (L4)	92	91	98.91	1	1.09	0.37	0.79	0.15	4.1
	Disease Behavior B (446)									
	Perforating (B3)	194	189	97.42	5	2.58	0.98	1.15	0.47	2.83
	Stricturing (B2)	372	364	97.85	8	2.15	0.61	0.9	0.43	1.89
	Inflammatory (B1)	326	319	97.85	7	2.15	0.63	0.91	0.42	1.99
	Controls (863)	1726	1681	97.39	45	2.61				
New Zealand[a]	**Diagnosis Age A (497)**									
	<16 (A1)	112	111	99.11	1	0.89	0.39	0.42	0.06	3.19
	17 - 40 (A2) + >40 (A3)	882	862	97.73	20	2.27	0.75	1.1	0.61	2.02
	Disease Location L(493)									
	Ileal (L1)	576	568	98.61	8	1.39	0.33	0.72	0.33	1.59
	Colonic (L2)	410	397	96.83	13	3.17	0.2	1.61	0.81	3.17
	Disease Behavior B (497)									
	Inflammatory (B1)	564	549	97.34	15	2.66	0.43	1.33	0.69	2.54
	(B2+B3)	430	424	98.60	6	1.40	0.38	0.75	0.31	1.81
	Controls (559)	1118	1095	97.94	23	2.06				

[a]Only showed the available data. [b]The meta-analysis calculated through the fixed effects model for each clinical variable did not show significant associations. [c]FDR correction p-value (based on nine comparisons)

Supplementary Table 2. Allelic frequencies of the R620W polymorphism in three Caucasian cohorts according to the clinical classification of CD[18-20] and healthy controls.

Population	Clinical Variable (N)	2N	C	%	T	%	p value	OR	95% CI	
Spain	**Diagnosis Age A (248)**									
	<16 (A1)	66	60	90.9	6	9.1	0.45	1.58	0.69	3.59
	17 - 40 (A2)	332	316	95.2	16	4.8	0.17	0.74	0.44	1.24
	>40 (A3)	98	95	96.9	3	3.1	0.15	0.57	0.2	1.68
	Disease Location L (482)									
	Ileal (L1)	416	394	94.7	22	5.3	0.26	0.8	0.51	1.25
	Colonic (L2)	178	170	95.5	8	4.5	0.24	0.73	0.36	1.46
	Ileocolonic (L3)	370	349	94.3	21	5.7	0.44	0.87	0.55	1.37
	Upper GI Tract (L4)	40	39	97.5	1	2.5	0.29	0.69	0.13	3.54
	Disease Behavior B (340)									
	Perforating (B3)	198	187	94.4	11	5.6	0.52	0.88	0.48	1.62
	Stricturing (B2)	90	84	93.3	6	6.7	0.98	1.14	0.51	2.55
	Inflammatory (B1)	392	373	95.2	19	4.8	0.15	0.74	0.46	1.19
	Controls (1685)	3370	3143	93.3	227	6.7				
Dutch	**Diagnosis Age A (697)**									
	<16 (A1)	654	620	94.8	34	5.2	5.4E-03[c]	0.53	0.36	0.77
	17 - 40 (A2)	578	551	95.3	27	4.7	1.8E-03[c]	0.48	0.31	0.72
	>40 (A3)	162	146	90.1	16	9.9	0.91	1.1	0.63	1.85
	Disease Location L (417)									
	Ileal (L1)	220	211	95.9	9	4.1	0.06[c]	0.44	0.23	0.86
	Colonic (L2)	206	194	94.2	12	5.8	0.08	0.62	0.34	1.13
	Ileocolonic (L3)	408	374	91.7	34	8.3	0.43	0.87	0.59	1.29
	Upper GI Tract (L4)	84	80	95.2	4	4.8	0.13	0.58	0.22	1.52
	Disease Behavior B (435)									
	Perforating (B3)	186	175	94.1	11	5.9	0.1	0.64	0.34	1.18
	Stricturing (B2)	364	343	94.2	21	5.8	0.18[c]	0.6	0.38	0.95
	Inflammatory (B1)	320	294	91.9	26	8.1	0.41	0.86	0.56	1.32
	Controls (834)	1668	1508	90.4	160	9.6				
New Zealand[a]	**Diagnosis Age A (465)**									
	<16 (A1)	106	96	90.6	10	9.4	0.87	0.94	0.48	1.86
	17 - 40 (A2)	-	-	-	-	-	-	-	-	-
	>40 (A3)	-	-	-	-	-	-	-	-	-
	Disease Location L(461)									
	Ileal (L1)	542	504	93.0	38	7.0	0.44[c]	0.68	0.47	1
	Colonic (L2)	380	351	92.4	29	7.6	0.18	0.75	0.49	1.45
	Ileocolonic (L3)	-	-	-	-	-	-	-	-	-
	Upper GI Tract (L4)	-	-	-	-	-	-	-	-	-
	Disease Behavior B (465)									
	Inflammatory (B1)	528	492	93.2	36	6.8	0.36[c]	0.66	0.45	0.98
	(B2+B3)	402	371	92.3	31	7.7	0.19	1.32	0.87	2
	Controls (563)	1126	1014	90.1	112	9.9				
Pooled[b,c]	**Diagnosis Age**									
	<16 (A1)	826	776	93.9	50	6.1	0.038[c]	0.64	0.47	0.88

Disease Location									
Ileal (L1)	1178	1109	94.1	69	5.9	9E-03[c]	0.64	0.49	0.84

[a]Only showed the available data. [b]Meta-analysis calculated through the fixed effects model and only showed the most relevant results. [C]FDR correction p-value (based on nine comparisons). We did not observe significant associations after the pooled analysis between the different subsets of the diseases (i.e. A1 vs. A2+A3).

Supplementary Table 3. Allele frequencies distribution of the R263Q polymorphism in three Caucasian cohorts according to the clinical classification variables of UC[18-20] and healthy controls.

Population	Clinical Variable (N)[b]	2N	G	%	A	%	p value	OR	95% CI	
Spain	**Diagnosis Age A (212)**									
	<16 (A1)	12	12	100	0	0	-	-	-	-
	17 - 40 (A2)	232	229	98,71	3	1,29	0,11	0,53	0,18	1,54
	>40 (A3)	180	176	97,78	4	2,22	0,47	0,85	0,33	2,22
	Disease Extension E (388)									
	Ulcerative Proctitis (E1)	64	63	98,44	1	1,56	0,46	0,94	0,19	4,282
	Left-side UC (E2)	398	391	98,24	7	1,76	0,12	0,62	0,29	1,3
	Extensive UC (E3)	314	306	97,45	8	2,55	0,54	0,89	0,44	1,8
	Controls (1685)	3370	3263	96,82	107	3,18				
Dutch	**Diagnosis Age A (406)**									
	<16 (A1)	76	74	97,37	2	2,63	0,99	1,46	0,4	5,34
	17 - 40 (A2)	546	535	97,99	11	2,01	0,44	0,82	0,43	1,57
	>40 (A3)	200	194	97	6	3	0,74	1,31	0,57	3,03
	Disease Extension E (364)									
	Ulcerative Proctitis (E1)	126	125	99,21	1	0,79	0,21	0,58	0,11	2,98
	Left-side UC (E2)	204	196	96,08	8	3,92	0,28	1,67	0,79	3,53
	Extensive UC (E3)	398	390	97,99	8	2,01	0,49	0,84	0,4	1,77
	Controls (834)	1726	1681	97,39	45	2,61				
New Zealand[a]	**Diagnosis Age A (481)**									
	<16 (A1)	52	52	100	0	0	-	-	-	-
	17 - 40 (A2) + >40 (A3)	910	899	98,79	11	1,21	0,14	0,61	0,29	1,24
	Disease Extension E (476)									
	Ulcerative Proctitis (E1) + Left-side UC (E2)	384	380	98,96	4	1,04	0,2	0,5	0,17	1,46
	Extensive UC (E3)	568	561	98,77	7	1,23	0,23	0,65	0,28	1,49
	Controls (559)	1118	1095	97,94	23	2,06				

[a]Only showed the available data. [b]The meta-analysis calculated through the fixed effects model for each clinical variable did not show significant associations. [C]FDR correction p-value (based on nine comparisons)

Supplementary Table 4. Allele frequencies distribution of the R620W polymorphism in three Caucasian cohorts according to the clinical classification variables of UC[18-20] and healthy controls.

Population	Clinical Variable (N)[b]	2N	C	%	T	%	p value	OR	95% CI	
Spain	**Diagnosis Age A (212)**									
	<16 (A1)	12	10	83,33	2	0,33	0,17	3,8	0,94	15
	17 - 40 (A2)	232	217	93,53	15	0,13	0,87	1,01	0,59	1,7
	>40 (A3)	180	163	90,56	17	0,19	0,16	1,5	0,91	2,52
	Disease Extension E (388)									
	Ulcerative Proctitis (E1)	64	62	96,88	2	0,06	0,25	0,66	0,18	2,34
	Left-side UC (E2)	398	372	93,47	26	0,13	0,88	1	0,66	1,51
	Extensive UC (E3)	314	296	94,27	18	0,11	0,49	0,88	0,54	1,44
	Controls (1685)	3370	3143		227					
Dutch	**Diagnosis Age A (406)**									
	<16 (A1)	76	68	89,47	8	0,21	0,79	1,2	0,59	2,54
	17 - 40 (A2)	540	511	94,63	29	0,11	0.018[c]	0,55	0,37	0,82
	>40 (A3)	196	186	94,90	10	0,10	0.351[c]	0,55	0,29	1,05
	Disease Extension E (359)									
	Ulcerative Proctitis (E1)	122	111	90,98	11	0,18	0,83	1	0,54	1,88
	Left-side UC (E2)	200	188	94,00	12	0,12	0,097	0,64	0,36	1,17
	Extensive UC (E3)	396	377	95,20	19	0,10	0.018[c]	0,5	0,31	0,8
	Controls (834)	1668	1508		160					
New Zealand[a]	**Diagnosis Age A (431)**									
	<16 (A1)	50	48	96,00	2	0,08	0,16	0,38	0,09	1,57
	17 - 40 (A2) + >40 (A3)	812	729	89,78	83	0,20	0,84	1,03	0,77	1,39
	Disease Extension E (426)									
	Ulcerative Proctitis (E1) + Left-side UC (E2)	532	476	89,47	56	0,21	0,52	1,15	0,75	1,78
	Extensive UC (E3)	320	292	91,25	28	0,18	0,52	0,89	0,58	1,37
	Controls (563)	1126	1014		112					

[a]Only showed the available data. [b]The meta-analysis calculated through the fixed effects model for each clinical variable did not show significant associations. [c]FDR correction p-value (based on nine comparisons)

STAT4 gene influences genetic predisposition to ulcerative colitis but not Crohn's disease in the Spanish population: a replication study. Human Immunology, 2010

Human Immunology 71 (2010) 515–519

Contents lists available at ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

STAT4 gene influences genetic predisposition to ulcerative colitis but not Crohn's disease in the Spanish population: A replication study

Lina Marcela Diaz-Gallo [a], Rogelio J. Palomino-Morales [a], María Gómez-García [b], Carlos Cardeña [c], Luis Rodrigo [d], Antonio Nieto [e], Guillermo Alcain [f], Ignacio Cueto [f], Miguel A. López-Nevot [g], Javier Martin [a,*]

[a] Instituto de Parasitología y Biomedicina "López-Neyra", CSIC, Granada, Spain
[b] Servicio de Digestivo, Hospital Universitario Virgen de las Nieves, Granada, Spain
[c] Servicio de Digestivo, Hospital Clínico San Cecilio, Granada, Spain
[d] Servicio de Digestivo, Hospital Universitario Central de Asturias, Oviedo, Spain
[e] Servicio de Inmunología, Hospital Puerta del Mar, Cádiz, Spain
[f] Servicio de Digestivo, Hospital Virgen de la Victoria, Málaga, Spain
[g] Servicio de Inmunología, Hospital Universitario Virgen de las Nieves, Granada, Spain

ARTICLE INFO

Article history:
Received 23 September 2009
Accepted 1 February 2010
Available online 7 March 2010

Keywords:
Signal transducer and activator of transcription 4 (*STAT4*) gene
Inflammatory bowel disease
Crohn's disease
Ulcerative colitis
Genetic association study

ABSTRACT

Recently, the signal transducer and activator of transcription 4 (*STAT4*) gene has been associated with multiple autoimmune diseases. Interestingly, a recent work showed that the T allele of the rs7574865 *STAT4* SNP was associated with inflammatory bowel disease (IBD) in a Spanish population. The aim of the present study was to reevaluate the role of the *STAT4* rs7574865 polymorphism on IBD. The present case-control study included 498 Crohn's disease (CD) patients, 402 ulcerative colitis (UC) patients, and 1296 healthy matched controls. Genotyping was performed using a PCR system with a pre-developed TaqMan allelic discrimination assay for the rs7574865 *STAT4* SNP. Moreover, a meta-analysis was performed with the previous work in a Spanish population and the current study, including a final sample size of 1574 IBD patients (820 with CD and 754 with UC) and 2012 healthy controls. No evidence of association was found for the current case-control study (CD: $p = 0.23$, OR = 0.9, 95% CI = 0.75–1.1; UC: $p = 0.17$, OR = 1.14, 95% CI = 0.95–1.38). However, the meta-analysis showed that the *STAT4* rs7574865 T allele was significantly associated with susceptibility to UC ($p = 0.012$ pooled; OR = 1.20, 95% CI = 1.04–1.39) but not CD ($p = 0.71$ pooled; OR = 0.93, 95% CI = 0.65–1.34). Our data suggest that the rs7574865 *STAT4* SNP is a genetic susceptibility variant for UC but not CD in the Spanish population.

1. Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the main types of inflammatory bowel disease (IBD). They are relapsing and chronic inflammatory disorders that result from the complex interaction of genetic, immune, and environmental factors [1–4]. Despite the fact that there is an increasingly list of genetic factors associated with IBD, it is estimated that the current number of loci associated with *IBD* represents only a small fraction of the genetic risk. Thus, additional genetic contributions clearly remain to be discovered [3,5–12].

The signal transducer and activator of transcription 4 (*STAT4*) is an interesting candidate gene that may play a role in the genetic background of IBD for two principal reasons. First, *STAT4* is an important player in directing T helper cells toward the Th1 and Th17 lineages, both of which are implicated in autoimmune processes, including IBD pathogenesis [13–19]. Second, recent findings suggest that the *STAT4* gene may be a common genetic factor in autoimmunity [20,21] Thus, the *STAT4* gene was first associated with a genetic predisposition to systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), with a single nucleotide polymorphism (SNP) of *STAT4* intron 3 (rs7574865) showing the strongest effect [22–26]. Subsequent works have found similar results with different autoimmune diseases, such as type 1 diabetes, Sjögren's syndrome, and systemic sclerosis [27–29]. Interestingly, a recent work showed that the T allele of the *STAT4* rs7574865 was associated with IBD in a Spanish population [30]. However, the statistical power of this study was low (70% for each disease), and the significance of the association quite modest (p=0.03 for each disease [CD and UC]). On this basis, we aimed to re-evaluate the possible role of the *STAT4* gene in IBD susceptibility through an independent replication study and meta-analysis.

* Corresponding author.
E-mail address: martin@ipb.csic.es (J. Martin).

0198-8859/10/$32.00 - see front matter © 2010 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2010.02.005

2. Subjects and methods

2.1. Study population

A total of 900 unrelated Spanish patients with IBD (498 with CD and 402 with UC) and 1296 gender-, age-, and ethnicity-matched healthy controls were included in the case-control study. The patients were recruited from Hospital Universitario Virgen de la Nieves (Granada), Hospital Clínico San Cecilio (Granada), Hospital Virgen de la Victoria (Málaga) and Hospital Universitario Central de Asturias (Oviedo). IBD was diagnosed according to standard clinical, endoscopic, radiologic, and histopathologic criteria [31,32]. Written informed consent was obtained from all participants. The study was approved by the ethics committeea of the hospitals.

2.2. STAT4 genotyping

DNA from patients and controls was obtained using standard methods. Samples were genotyped for the rs7574865 polymorphism via Taqman 5′ allelic discrimination assay technology using a predesigned SNP genotyping assay provided by Applied Biosystems (Part number: C__29882391_10, Foster City, CA, USA) as described previously [24]. All samples were genotyped in the same center to avoid genotyping inconsistencies and to verify the genotyping consistency; random samples were genotyped twice, showing 99% identical genotypes [24].

2.3. Data analysis

Deviation from Hardy-Weinberg equilibrium (HWE) was tested by χ^2 analysis using the program FINETI (http://ihg2.helmholtz-muenchen.de/cgi-bin/hw/hwa1.pl). The differences in genotype distribution and allele frequency among cases and controls were calculated by 2x2 contingency tables and Fisher's exact test. The p values, odds ratios (OR) and 95% confidence intervals (CI) were calculated using Statcalc software (Epi Info 2002; Centers for Disease Control and Prevention, Atlanta, GA). To evaluate possible associations with the genetic polymorphism and clinical features (Table 1), a univariate analysis using χ^2 or Fisher's exact test was used. Values of $p < 0.05$ were considered statistically significant. Analyses were conducted using the SPSS software package v.15.0 (SPSS Inc., Chicago, IL).

The analysis of the combined data from all populations was performed using StatsDirect software, V2.6.6 (StatsDirect. http://www.statsdirect.com. England: StatsDirect 2008.). The summarized OR and CIs were obtained by means of both the random (DerSimonian-Laird) and the fixed (Mantel–Haenszel meta-analysis) effect models. The random effect model was chosen because it assumes that there is a distribution of true effect sizes rather than one true effect and thus assigns a more balanced weight to each study [33]. The homogeneity of ORs among cohorts was calculated using Cochran's Q test [34]. The estimation of the power of the study was performed using Quanto v0.5 software (Department of Preventive Medicine University of Southern California, Los Angeles, CA).

3. Results

To re-evaluate the association of the *STAT4* rs7574865 polymorphism in an independent Spanish cohort, 498 CD patients, 402 UC patients and 1296 healthy controls were genotyped for this SNP. Clinical features of the patients are shown in Table 1. The genotype and allele distributions for the STAT4 rs7574865 genetic variant in the studied Spanish population are represented in Table 2. The frequencies of the *STAT4* rs7574865 polymorphism were in Hardy–Weinberg equilibrium in both healthy controls and IBD patients. The minor allele frequency (MAF) of this SNP in the control group did not significantly deviate from the frequency in the European data from either the International Hapmap Project (MAF rs7574865, 20.8%; http://www.hapmap.org) or the healthy controls from previously reported works [22,24–29].

Table 1
Clinical features of patients with IBD

	CD	UC
Total number	498	402
General characteristics		
Male (%)	54.3	55.2
Age, y, (mean ± SD)	42.1 ± 13.9	46.6 ± 15.9
Age at onset, y (mean ± SD)	31.9 ± 13.9	28.8 ± 22.6
Smoking habits (%)		
Never	46.6	63.3
Ex/current	53.4	36.7
Extra intestinal manifestations (%)	37.8	41.7
Surgery (%)	47.9	9.1
Montreal criteria [28,29]		
Diagnosis age (%)		
<16 (A1)	9.9	–
17–40 (A2)	68.8	–
>40 (A3)	21.3	–
Disease location (%)		
Ileal (L1)	40.3	–
Colonic (L2)	17.6	–
Ileocolonic (L3)	39.4	–
Upper GI tract (L4)	2.6	–
Disease behavior (%)		
Inflammatory (B1)	38.9	–
Structuring (B2)	22.2	–
Perforating (B3)	38.9	–
Disease extension (%)		
Ulcerative proctitis (E1)	–	16.8
Left-side UC (E2)	–	39.3
Extensive UC (E3)	–	43.9
Disease severity (%)		
Clinical remission (S0)	–	27.7
Mild UC (S1)	–	40.7
Moderate UC (S2)	–	30.3
Severe UC (S3)	–	1.3

Abbreviations: CD, Crohn's disease; UC, ulcerative colitis; GI, gastrointestinal.

The *STAT4* rs7574865 T allele frequency showed a slight increase in UC patients (23%) compared with healthy controls (20.8%). However, there were no statistically significant differences between UC patients and controls ($p = 0.17$; OR = 1.14 95%, CI = 0.95–1.38). Similarly, the MAF of CD patients (19%) was not significantly differed with respect to that of healthy controls ($p = 0.23$; OR = 0.9 95%, CI = 0.75–1.1) (Table 2).

To improve the statistical power and to determine the common effect size of the *STAT4* polymorphism on CD and UC, Spanish patients, a meta-analysis was performed with the current work and the study reported by Martinez et al. [30]. Heterogeneity was observed in the meta-analysis for the *STAT4* genetic variant and CD ($p = 0.013$, $Q = 6.21$, df = 1). This was mainly due to the difference in the T allele frequency between the patients of each study (*i.e.*, 23% in the Martinez et al. study and 19% in the current work). Therefore, only the random-effects model for the meta-analysis in CD patients, which displayed no statistically significant difference between groups, should be considered ($p = 0.71$, pooled OR = 1.07, 95% CI = 0.74–1.54) (Fig. 1).

In contrast, heterogeneity was not observed in the UC meta-analysis ($p = 0.42$, $Q = 0.64$, df = 1). The combined analysis under the fixed and random-effect models showed evidence for an association of the *STAT4* rs7574865 T allele with UC ($p = 0.013$, pooled OR = 1.20, 95% CI = 1.04–1.39) (Fig. 2).

To address whether the *STAT4* variant rs7574865 affects clinical manifestations of both CD and UC, a univariate analysis was performed to compare its frequency in terms of gender, age at onset, extraintestinal manifestations, surgery, and the Montreal classification criteria. However, no association was found between the

Table 2
Genotype and allele frequencies of the rs7574865 *STAT4* polymorphism in healthy controls and CD and UC, Spanish patients

STAT4 rs7574865	Controls $n = 1296$ (%)	CD $n = 498$ (%)	p Value	OR (95% CI)	UC $n = 402$ (%)	p Value	OR (95% CI)
GG	813 (62.7)	320 (64.3)	0.55	1.07 (0.86–1.32)	241 (59.9)	0.32	0.89 (0.70–1.12)
GT	428 (33)	167 (33.5)	0.84	1.03 (0.82–1.28)	137 (34.1)	0.69	1.05 (0.83–1.33)
TT	55 (4.2)	11 (2.2)	0.04	0.55 (0.29–1.04)	24 (6)	0.15	1.46 (0.89–2.40)
G	2054 (79.2)	807 (81)	0.23	1.12 (0.93–1.34)	619 (77)	0.17	0.87 (0.72–1.06)
T	538 (20.8)	189 (19)	0.23	0.90 (0.75–1.08)	185 (23)	0.17	1.14 (0.95–1.38)

Abbreviations: CD, Crohn's disease; UC, ulcerative colitis; CI, confidence interval; OR, odds ratio.

STAT4 rs7574865 genetic variant and the phenotype of either IBD disease (data not shown).

4. Discussion

Confidence in genetic association studies is partly supported by their reproducibility among populations obtained via replication studies [35,36]. Meta-analysis provides a robust tool for resolving contradictory results and increasing the statistical power in genetic association studies [35,36]. In the present study, we used both approaches to evaluate the role of the rs7574865 polymorphism of the *STAT4* gene in IBD in Spanish patients.

According to the results of the meta-analysis, the rs7574865 T allele may be considered a genetic risk factor in the development of UC However, the results of both the present case-control study and the meta-analysis do not replicate the association between the *STAT4* rs7574865 genetic variant and CD reported by Martinez et al. The lack of concordance in the results from these independent association studies is probably due to their low statistical power; when considered individually, the statistical power of these two studies did not rise above than 72% for each disease. In contrast, the power of the meta-analysis is 98% for CD and 97% for UC. This power supports the findings of (1) the lack of association between the rs7574865 *STAT4* polymorphism and CD and (2) the identification of the T allele as a risk factor for UC. Indeed, CD and UC share many characteristics. As some clinical features also distinguish them, however, they likely share some genetic susceptibility loci but differ at others [5]. Along with this line (and excluding the MHC region), a number of genetic loci (*e.g.*, polymorphisms in the *ATG16L1*, *IRGM*, *MIF*, *IL23R*, and *IL12B* genes) associated with both diseases have been described. Meanwhile, the *NOD2* gene, which is strongly associated with CD, is not a risk factor for UC susceptibility. However, the *ECM1*, *ARPC2*, and *IL-10* genes as well as regions on chromosomes 1p36, 2q24–37, and 12q15 are mainly associated with UC [3,5–10] These differences may be caused by pleiotropic effects of specific genes on a common polygenic and multifactor background [20,21,37]. The distinct influence of the *STAT4* gene on IBD risk observed in our study provides an example of the differences in genetic background between these two intestinal inflammatory disorders.

The *STAT4* gene is located on the 2q32.3 region of the human genome (http://www.hapmap.org), and the genome-wide linkage study performed by Barmada et al. [10] showed a linkage with the 2q24–q37 (D2S1776/D2S1391) locus only for UC. The WTCCC study genotyped 40 SNPs present in the 2q31–q32 region for CD, showing that none of these SNPs was significantly associated with disease susceptibility [38]. The lack of association between these 40 SNPs in the 2q31–q32 locus and CD strengthens our suggestion of the genetic influence of the *STAT4* variant in UC only. In other genetic association studies, polymorphisms in the *ECM1*, *ARPC2*, *IL-10*, and MYO9B genes and 1p36, 2q24–q37 (D2S1776/D2S1391) and 12q15 genome regions have been associated with UC but not CD [3,5,6,8–10,39].

Accumulating evidence indicates that *STAT4* is a pathogenic factor in IBD, both in murine models [40,41] and in human beings [18]. Mice that constitutively express *Stat4* develop chronic transmural colitis [40], and increased levels of constitutive *STAT4* were observed in mucosal cells from UC patients [18]. It has been suggested that the CD presents a Th1/Th17 cytokines response in contrast to the Th2 type response in UC [2,42,43]. However to understand the influence of the rs7574865 *STAT4* variant in the different immunologic pathways in IBD, further functional studies are needed. Nevertheless, a difference between the amount of phosphorylated STAT4 present in cells of affected tissues from CD and UC patients was described [17,18,44]. Based on this, we spec-

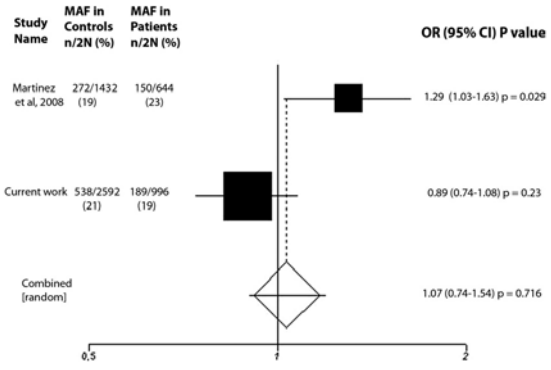


Fig. 1. Forest plot for meta-analysis of the T allele frequency (rs7574865 *STAT4* polymorphism) in Spanish CD patients. Common effect size, odds ratios, with 95% confidence intervals were calculated by the DerSimonian-Laird method, under the random effect model. Filled squares represent the studies in relation to their weights.

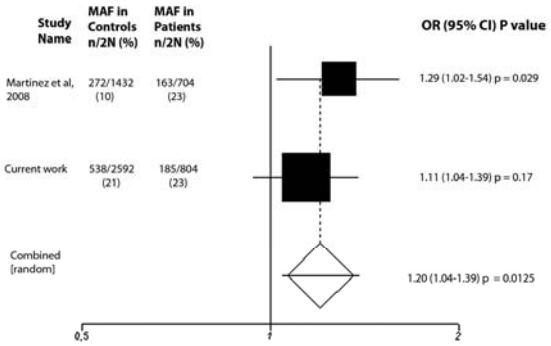


Fig. 2. Forest plot for meta-analysis of the T allele frequency (rs7574865 *STAT4* polymorphism) in Spanish UC patients. Common effect size, odds ratios, with 95% confidence intervals were calculated by the DerSimonian-Laird method, under the random effect model. Filled squares represent the studies in relation to their weights.

ulate that the T allele of the rs7574865 polymorphism could indirectly influence the mechanisms of STAT4 phosphorylation and, in turn, different profile of active cytokines in UC and CD. Moreover, the rs7574865 *STAT4* polymorphism could influence the *STAT4* transcription rate by disrupting a transcription factor binding site or a binding site for modified histone proteins [19,22–24]. A correlation between the *STAT4* expression levels in peripheral blood mononuclear cells from SLE patients and the SNP rs7574865 has previously been demonstrated [24]. Together, these ideas suggest that the *STAT4* gene may be involved in dysregulation of the immune response in UC patients. Further genetic association studies in other populations with different genetic backgrounds that use different functional approaches are needed to fully clarify the role of *STAT4* in the genetics of IBD.

Acknowledgments

We thank Sofia Vargas for excellent technical assistance, and all of the patients and control donors for their essential collaboration.

This work was supported by grants SAF2006-00398 and Junta de Andalucía, grants: CTS-1180. Diaz-Gallo LM was supported by COLFUTURO and the "Ayudas Predoctorales de Formación en Investigación en Salud (PFIS - FI09/00544)" from the Instituto de Salud Carlos III.

References

[1] Baumgart DC, Carding SR. Inflammatory bowel disease: Cause and immunobiology. Lancet 2007;369:1627.
[2] Bouma G, Strober W. The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 2003;3:521.
[3] Budarf ML, Labbe C, David G, Rioux JD. GWA studies: Rewriting the story of IBD. Trends Genet 2009;25:137.
[4] Danese S, Scaldaferri F, Papa A, Pola R, Sans M, Gasbarrini G, et al. Platelets: New players in the mucosal scenario of inflammatory bowel disease. Eur Rev Med Pharmacol Sci 2004;8:193.
[5] Silverberg MS, Cho JH, Rioux JD, McGovern DP, Wu J, Annese V, et al. Ulcerative colitis-risk loci on chromosomes 1p36 and 12q15 found by genome-wide association study. Nat Genet 2009;41:216.
[6] Marquez A, Varade J, Robledo G, Martinez A, Mendoza JL, Taxonera C, et al. Specific association of a CLEC16A/KIAA0350 polymorphism with NOD2/CARD15(−) Crohn's disease patients. Eur J Hum Genet 2009.
[7] Oliver J, Rueda B, Lopez-Nevot MA, Gomez-Garcia M, Martin J. Replication of an association between IL23R gene polymorphism with inflammatory bowel disease. Clin Gastroenterol Hepatol 2007;5:977.
[8] Oliver J, Marquez A, Gomez-Garcia M, Martinez A, Mendoza JL, Vilchez JR, et al. Association of the macrophage migration inhibitory factor gene polymorphisms with inflammatory bowel disease. Gut 2007;56:150.
[9] Palomino-Morales RJ, Oliver J, Gomez-Garcia M, Lopez-Nevot MA, Rodrigo L, Nieto A, et al.. Association of ATG16L1 and IRGM genes polymorphisms with inflammatory bowel disease: A meta-analysis approach. Genes Immun 2009;10:356.
[10] Barmada MM, Brant SR, Nicolae DL, Achkar JP, Panhuysen CI, Bayless TM, et al. A genome scan in 260 inflammatory bowel disease-affected relative pairs. Inflamm Bowel Dis 2004;10:513.
[11] Wang K, Zhang H, Kugathasan S, Annese V, Bradfield JP, Russell RK, et al. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn disease. Am J Hum Genet 2009;84:399.
[12] Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, et al. Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. Nat Genet 2008;40:955.
[13] Kaplan MH: STAT4: A critical regulator of inflammation in vivo. Immunol Res 2005;31:231.
[14] Watford WT, Hissong BD, Bream JH, Kanno Y, Muul L, O'Shea JJ. Signaling by *IL-12 and IL-23 and the immunoregulatory roles of STAT4. Immunol Rev* 2004;202:139.
[15] Bettelli E, Korn T, Kuchroo VK. Th17: The third member of the effector T cell trilogy. Curr Opin Immunol 2007;19:652.
[16] Mathur AN, Chang HC, Zisoulis DG, Stritesky GL, Yu Q, O'Malley JT, et al. Stat3 and Stat4 direct development of IL-17-secreting T_h cells. J Immunol 2007;178:4901.
[17] Mudter J, Weigmann B, Bartsch B, Kiesslich R, Strand D, Galle PR, et al. Activation pattern of signal transducers and activators of transcription (STAT) factors in inflammatory bowel diseases. Am J Gastroenterol 2005;100:64.
[18] Pang YH, Zheng CQ, Yang XZ, Zhang WJ. Increased expression and activation of IL-12-induced Stat4 signaling in the mucosa of ulcerative colitis patients. Cell Immunol 2007;248:115.
[19] Korman BD, Kastner DL, Gregersen PK, Remmers EF. STAT4: Genetics, mechanisms, and implications for autoimmunity. Curr Allergy Asthma Rep 2008;8:398.
[20] Gregersen PK, Olsson LM. Recent advances in the genetics of autoimmune disease. Annu Rev Immunol 2009;27:363.
[21] Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nat Rev Genet 2009;10:43.
[22] Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 2007;357:977.
[23] Orozco G, Alizadeh BZ, Delgado-Vega AM, Gonzalez-Gay MA, Balsa A, Pascual-Salcedo D, et al. Association of STAT4 with rheumatoid arthritis: A replication study in three European populations. Arthritis Rheum 2008;58:1974.
[24] Abelson AK, Delgado-Vega AM, Kozyrev SV, Sanchez E, Velazquez-Cruz R, Eriksson N, et al. STAT4 associates with SLE through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk. Ann Rheum Dis 2008;68(11):1746.
[25] Palomino-Morales RJ, Rojas-Villarraga A, Gonzalez CI, Ramirez G, Anaya JM, Martin J. STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians. Genes Immun 2008;9:379.
[26] Sigurdsson S, Nordmark G, Garnier S, Grundberg E, Kwan T, Nilsson O, et al. A risk haplotype of STAT4 for systemic lupus erythematosus is over-expressed, correlates with anti-dsDNA and shows additive effects with two risk alleles of IRF5. Hum Mol Genet 2008;17:2868.
[27] Zervou MI, Mamoulakis D, Panierakis C, Boumpas DT, Goulielmos GN. STAT4: A risk factor for type 1 diabetes? Hum Immunol 2008;69:647.
[28] Korman BD, Alba MI, Le JM, Alevizos I, Smith JA, Nikolov NP, et al. Variant form of STAT4 is associated with primary Sjogren's syndrome. Genes Immun 2008;9:267.

[29] Rueda B, Broen J, Simeon C, Hesselstrand R, Diaz B, Suarez H, et al. The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. Hum Mol Genet 2009;18:2071.
[30] Martinez A, Varade J, Marquez A, Cenit MC, Espino L, Perdigones N, et al. Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases. Arthritis Rheum 2008;58:2598.
[31] Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, Brant SR, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a working party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19(Suppl A):5.
[32] Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: Controversies, consensus, and implications. Gut 2006;55:749.
[33] Huedo-Medina TB, Sanchez-Meca J, Marin-Martinez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? Psychol Methods 2006; 11:193.
[34] Gavaghan DJ, Moore RA, McQuay HJ. An evaluation of homogeneity tests in meta-analyses in pain using simulations of individual patient data. Pain 2000;85: 415.
[35] Attia J, Ioannidis JP, Thakkinstian A, McEvoy M, Scott RJ, Minelli C, et al. How to use an article about genetic association: B: Are the results of the study valid? J Am Med Assoc 2009;301:191.
[36] Zintzaras E, Lau J. Synthesis of genetic association studies for pertinent gene-disease associations requires appropriate methodological and statistical approaches. J Clin Epidemiol 2008;61:634.
[37] Anaya JM, Gomez L, Castiblanco J. Is there a common genetic basis for autoimmune diseases? Clin Dev Immunol 2006;13:185.
[38] Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661.
[39] Nunez C, Oliver J, Mendoza JL, Gomez-Garcia M, Pinero A, Taxonera C, et al. MYO9B polymorphisms in patients with inflammatory bowel disease. Gut 2007;56:1321.
[40] Wirtz S, Finotto S, Kanzler S, Lohse AW, Blessing M, Lehr HA, et al. Cutting edge: Chronic intestinal inflammation in STAT-4 transgenic mice: Characterization of disease and adoptive transfer by TNF-plus IFN-gamma-producing CD4$^+$ T cells that respond to bacterial antigens. J Immunol 1999;162:1884.
[41] O'Malley JT, Eri RD, Stritesky GL, Mathur AN, Chang HC, Hogenesch H, et al. STAT4 isoforms differentially regulate T_h1 cytokine production and the severity of inflammatory bowel disease. J Immunol 2008;181:5062.
[42] Abraham C, Cho JH. IL-23 and autoimmunity: New insights into the pathogenesis of inflammatory bowel disease. Annu Rev Med 2009;60:97.
[43] Brand S. Crohn's disease: T_h1, Th17 or both? The change of a paradigm: New immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn's disease. Gut 2009;58:1152.
[44] Parrello T, Monteleone G, Cucchiara S, Monteleone I, Sebkova L, Doldo P, et al. Up-regulation of the IL-12 receptor beta 2 chain in Crohn's disease. J Immunol 2000;165:7234.

Implication of IL-2/IL-21 region in systemic sclerosis genetic susceptibility. Annals of Rheumatic Diseases, 2012

EXTENDED REPORT

Implication of *IL-2/IL-21* region in systemic sclerosis genetic susceptibility

Lina-Marcela Diaz-Gallo,[1] Carmen P Simeon,[2] Jasper C Broen,[3] Norberto Ortego-Centeno,[4] Lorenzo Beretta,[5] Madelon C Vonk,[3] Patricia E Carreira,[6] Sofia Vargas,[1] José Andrés Román-Ivorra,[7] Miguel A González-Gay,[8] Carlos Tolosa,[9] Francisco Javier López-Longo,[10] Gerard Espinosa,[11] Esther F Vicente,[12] Roger Hesselstrand,[13] Gabriela Riemekasten,[14] Torsten Witte,[15] Jörg H W Distler,[16] Alexandre E Voskuyl,[17] Annemie J Schuerwegh,[18] Paul G Shiels,[19] Annika Nordin,[20] Leonid Padyukov,[20] Anna-Maria Hoffmann-Vold,[21] Raffaella Scorza,[5] Claudio Lunardi,[22] Paolo Airo,[23] Jacob M van Laar,[24] Nicolas Hunzelmann,[25] Birgit S Gathof,[26] Alexander Kreuter,[27] Ariane Herrick,[28] Jane Worthington,[28] Christopher P Denton,[29] Xiaodong Zhou,[30] Frank C Arnett,[30] Carmen Fonseca,[29] Bobby PC Koeleman,[31] Shervin Assasi,[30] Timothy R D J Radstake,[32] Maureen D Mayes,[30] Javier Martín,[1] The Spanish Scleroderma Group

► Additional supplementary data are published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/annrheumdis-2012-202357).

For numbered affiliations see end of article.

Correspondence to
Lina-Marcela Diaz-Gallo, Cellular Biology and Immunology Department, Instituto de Parasitología y Biomedicina López-Neyra (IPBLN-CSIC), Parque Tecnológico Ciencias de la Salud, Avenida del Conocimiento s/n 18100-Armilla, Granada 18100, Spain; lina.diaz@ipb.csic.es

Accepted 21 October 2012
Published Online First 21 November 2012

ABSTRACT

Objective The interleukin 2 (*IL-2*) and interleukin 21 (*IL-21*) locus at chromosome 4q27 has been associated with several autoimmune diseases, and both genes are related to immune system functions. The aim of this study was to evaluate the role of the *IL-2/IL-21* locus in systemic sclerosis (SSc).

Patients and methods The case control study included 4493 SSc Caucasian patients and 5856 healthy controls from eight Caucasian populations (Spain, Germany, The Netherlands, USA, Italy, Sweden, UK and Norway). Four single nucleotide polymorphisms (rs2069762, rs6822844, rs6835457 and rs907715) were genotyped using TaqMan allelic discrimination assays.

Results We observed evidence of association of the rs6822844 and rs907715 variants with global SSc (p_c=6.6E-4 and p_c=7.2E-3, respectively). Similar statistically significant associations were observed for the limited cutaneous form of the disease. The conditional regression analysis suggested that the most likely genetic variation responsible for the association was the rs6822844 polymorphism. Consistently, the rs2069762A-rs6822844T-rs6835457G-rs907715T allelic combination showed evidence of association with SSc and limited cutaneous SSc subtype (p_c=1.7E-03 and p_c=8E-4, respectively).

Conclusions These results suggested that the *IL-2/IL-21* locus influences the genetic susceptibility to SSc. Moreover, this study provided further support for the *IL-2/IL-21* locus as a common genetic factor in autoimmune diseases.

INTRODUCTION

Interleukin 2 (*IL-2*) and interleukin 21 (*IL-21*) are equally attractive biological candidates that may influence the pathogenesis of autoimmune diseases. Both are cytokines involved in the proliferation of T and B lymphocytes and different immunological activation pathways.[1] Moreover, the *IL-2* and *IL-21* genes cover a region of approximately 200 kb that maps in the 4q27 locus. *IL-2* has an important role in the maintenance of immune system homeostasis and self-tolerance. This cytokine has two paradoxical roles: promoting T cell proliferation and terminating T cell responses. Moreover, *IL-2* facilitates the production of immunoglobulins through B cells and induces the differentiation and proliferation of natural killer cells.[1 2] *IL-21* is a potent immunomodulatory cytokine with pleiotropic effects on both innate and adaptive immune responses. These actions include the following positive effects: enhanced proliferation of lymphoid cells, increased cytotoxicity of CD8 T cells and natural killer cells, and differentiation of B cells into plasma cells. *IL-21* is also produced by T helper 17 (Th17) cells and is a critical regulator of Th17 development.[1 3] Genetic association studies have demonstrated that several *IL-2/IL-21* polymorphisms influence the risk for autoimmune diseases (AIDs). The first evidence of this association was found in type 1 diabetes, Graves' disease, coeliac diseases and rheumatoid arthritis.[4–7] These results have been confirmed through replication studies in different populations and extended to other autoimmune diseases, such as inflammatory bowel diseases, giant cell arthritis, psoriasis and systemic lupus erythematosus (SLE).[8–17]

Systemic sclerosis (SSc) is a chronic fibrotic autoimmune disease in which patients are commonly classified into the following two major subgroups that are related to the specific autoantibodies against several nuclear and/or nucleolar antigens: (i) limited cutaneous SSc (lcSSc), which is related to the positive status of anticentromere autoantibodies (ACA) and (ii) diffuse cutaneous (dcSSc), which is related to the positive status of antitopoisomerase

autoantibodies (ATA).[18 22] More than 40 susceptibility loci to SSc have been identified during the last 10 years. Half of these variants need to be replicated in different populations and many of these variants are shared among different AIDs, especially SLE.[22 25] In this regard, one single nucleotide polymorphism (SNP) of the *IL-2* gene was proposed as risk factor to lcSSc subtype,[26] but this association has not been confirmed by other studies. Moreover the *IL-21* gene has been implicated as a potential driver of AIDs and recently a fine-mapping in SLE demonstrated that variants of the *IL-2/IL-21* region are implicated in the genetic susceptibility to SLE.[12 16] Thus, the aim of this study was to evaluate the influence of the *IL-2/IL-21* region in SSc genetic susceptibility.

PATIENTS AND METHODS

Subjects

This case-control association study was comprised of 4493 SSc patients and 5896 controls of Caucasian ancestry. The discover cohort included the Spanish group, which consisted of 1176 SSc patients and 1721 healthy controls. The follow-up phase consisted of the following subjects: 609 SSc cases and 426 controls from Germany, 365 SSc cases and 784 controls from the Netherlands, 916 SSc cases and 884 controls from USA, 595 SSc cases and 1107 controls from Italy, 225 SSc cases and 273 controls from Sweden, 374 SSc cases and 436 controls from the UK and 102 SSc cases and 278 controls from Norway. There was an overlapping of 1726 SSc and 2578 controls with the previous GWAS in SSc.[25] The patients fulfilled the 1980 American College of Rheumatology classification criteria for SSc[27] or the criteria proposed for early SSc.[21] In addition, the patients were classified as having lcSSc or dcSSc as described by LeRoy *et al*.[21] The following clinical data were collected for the ascertainment of the clinical phenotype of the SSc patients: age, gender and presence of SSc-specific autoantibodies (Ab; ACA and ATA). The control population consisted of unrelated healthy individuals recruited in the same geographical regions as the SSc patients, and they were matched by age, sex and ethnicity with the SSc patient groups. The study was approved by local ethical committees from all the participating centres. Both patients and controls were included in the study after written informed consent was obtained.

SNP Selection and genotyping

Four SNPs of the *IL-2/IL-21* region were selected for this study. The rs2069762 SNP was selected because it has been suggested to be a genetic factor of lcSSc subtype susceptibility by a study in a small Italian cohort.[26] SSc and SLE share some immunogenetic pathways; thus, the rs6822844, rs6835457 and rs907715 *IL-2/IL-21* polymorphisms were studied because they are the most associated variants in a recent fine-mapping of the region in SLE.[12]

DNA from the patients and the controls were extracted from peripheral white blood cells following standard procedures. The samples were genotyped for the rs2069762, rs6822844, rs6835457 and rs907715 *IL-2/IL-21* region polymorphisms using predesigned SNP genotyping assays from Applied Biosystems (Assay IDs: C__15859930_10, C__28983601_10, C___1597475_10 and C___8949748_10, respectively). TaqMan SNP genotyping was performed using a 7900HT Real-Time PCR system from Applied Biosystems following the manufacturer's suggestions (Foster City, California, USA). In all the cohorts, the genotyping success rate was greater than 95%, and randomly selected samples were genotyped twice to verify the genotyping accuracy. Ninety-nine per cent of the genotypes were identical.

Statistical analysis

The Hardy-Weinberg equilibrium was tested for all the SNPs in all the studied populations. Significance was calculated using 2×2 contingency tables and Fisher's exact test or the χ^2 test when necessary to obtain p values, ORs and 95% CIs using PLINK (V.1.07) software (http://pngu.mgh.harvard.edu/purcell/plink/).[28] The p values less than 0.05 were considered to be statistically significant. The Bonferroni correction was applied to the significant p values and referred in the text as p_c ($p_{corrected}$). Cochran-Mantel-Haenszel meta-analysis was performed to control the differences among populations as implemented by the PLINK software. In addition, the Breslow-Day test (BD test) and the Higgins' test (I^2) were performed using the PLINK software in each meta-analysis. The random-effects model was checked in the significant BD P_{values} analysis. The dependency of the association between each SNP and every studied genetic variant was determined by a conditional logistic regression analysis (considering the different cohorts as covariates) using the PLINK software. Linkage disequilibrium (LD) patterns between the four studied SNPs were estimated by the expectation-maximisation algorithm using HAPLOVIEW (V.4.2; Broad Institute of MIT and Harvard) and PLINK software. To evaluate the allelic combination difference between cases and controls, the conditional haplotype-based associations test was applied using the PLINK software.[29] The statistical power of the combined analysis was between 91% and 99% for all the SNPs, allowing for the detection of associations with an OR equal to 1.2 at a 5% significance level and the lowest minor allelic frequency, according to the Power Calculator for Genetic Studies 2006 software, which uses the methods described by Skol *et al*.[30]

RESULTS

The cases and controls of the eight Caucasian populations were in Hardy-Weinberg equilibrium at a 5% significance level. Additionally, the minor allelic frequencies of the four studied SNPs were similar to those reported by the HapMap project for the Utah residents with ancestry from northern and western Europe (CEU) population (http://hapmap.ncbi.nlm.nih.gov/). The LD structure of the eight cohorts is shown in the supplemental material (see online supplementary figure S1).

First, an association study was conducted in a Spanish case-control set, and a significant association was observed between the rs907715 SNPs minor allele and the global SSc (p_c=0.03, OR=0.85 95% CI 0.8 to 0.9) and the lcSSc subtype (p_c=0.04, OR=0.83 95% CI 0.7 to 0.9). A trend of association was observed between the minor allele of the rs6822844 SNP and the global SSc (p_{value}=0.04, OR=0.84 95% CI 0.7 to 1) and lcSSc subtype (p_{value}=0.04, OR=0.79 95% CI 0.7 to 0.9). Also a trend of association was detected between the minor allele of rs6835457 and lcSSc subtype in this population (p_{value}=0.03, OR=0.87 95% CI 0.8 to 1). In contrast, no association was observed with the rs2069762 SNP (p_{value}=0.8 for both SSc and lcSSc) (see online supplementary tables S1–S3). Based on these observations, we decided to evaluate other Caucasian cohorts and to perform a meta-analysis.

Table 1 shows the meta-analysis results for the *IL-2/IL-21* SNPs, the global SSc, the main SSc subtypes, the ACA and the ATA antibodies positive status. The combined analysis showed that the minor allele frequencies of the rs6822844 and rs907715 SNPs were significantly higher in controls than in SSc (p_c=6.6E-04 OR=0.86 95% CI 0.79 to 0.93 and p_c=7.2E-3 OR=0.91 95% CI 0.85 to 0.96, respectively)

Table 1 Genotype and minor allele frequencies of meta-analysis of four *IL-2/IL-21* SNPs located in SSc patients and healthy controls from European and US populations

			Genotype, N (%)				Allele test		
SNP	1/2	Subgroup (N)	1/1	1/2	2/2	MAF (%)	p Value*	p_c†	OR (CI 95%)‡
rs2069762	C/A	Controls (n=5482)	510 (9.30)	2266 (41.34)	2706 (49.36)	29.97			
		SSc (n=4281)	429 (10.02)	1778 (41.53)	2074 (48.45)	30.79	0.08	NA	1.06 (0.99 to 1.13)
		lcSSc (n=2897)	295 (10.18)	1203 (41.53)	1399 (48.29)	30.95	0.09	NA	1.06 (0.99 to 1.14)
		dcSSc (n=1384)	134 (9.68)	575 (41.55)	675 (48.77)	30.46	0.31	NA	1.05 (0.96 to 1.15)
		ACA+ (n=1736)	170 (9.79)	730 (42.05)	836 (48.16)	30.82	0.25	NA	1.05 (0.97 to 1.14)
		ATA+ (n=1031)	94 (9.12)	428 (41.51)	509 (49.37)	29.87	0.98	NA	1.00 (0.90 to 1.11)
rs6822844	T/G	Controls (n=5792)	149 (2.57)	1475 (25.47)	4168 (71.96)	15.31			
		SSc (n=4407)**	98 (2.22)	996 (22.60)	3313 (75.18)	13.52	1.7E-04	6.6E-04	0.86 (0.79 to 0.93)
		lcSSc (n=2977)***	67 (2.25)	659 (22.14)	2251 (75.61)	13.32	1.5E-04	6.0E-04	0.84 (0.76 to 0.92)
		dcSSc (n=1430)	31 (2.17)	337 (23.57)	1062 (74.27)	13.95	0.06	NA	0.89 (0.79 to 1)
		ACA+ (n=1763)	38 (2.16)	395 (22.40)	1330 (75.44)	13.36	0.01	0.06	0.87 (0.78 to 0.97)
		ATA+ (n=1074)	29 (2.70)	257 (23.93)	788 (73.37)	14.66	0.67	NA	0.97 (0.85 to 1.11)
rs6835457	G/A	Controls (n=5720)	668 (11.68)	2507 (43.83)	2545 (44.49)	33.59			
		SSc (n=4392)****	445 (10.13)	1908 (43.44)	2039 (46.43)	31.85	0.013	0.05	0.93 (0.87 to 0.98)
		lcSSc (n=2965)*****	312 (10.52)	1255 (42.33)	1398 (47.15)	31.69	0.014	0.06	0.92 (0.86 to 0.98)
		dcSSc (n=1427)	133 (9.32)	653 (45.76)	641 (44.92)	32.20	0.28	NA	0.95 (0.87 to 1.04)
		ACA+ (n=1765)	186 (10.54)	756 (42.83)	823 (46.63)	31.95	0.12	NA	0.94 (0.86 to 1.02)
		ATA+ (n=1064)	113 (10.62)	481 (45.21)	470 (44.17)	33.22	0.99	NA	1.00 (0.90 to 1.10)
rs907715	T/C	Controls (n=5644)	670 (11.87)	2491 (44.14)	2483 (43.99)	33.94			
		SSc (n=4341)******	437 (10.07)	1883 (43.38)	2021 (46.56)	31.76	1.8E-03	7.2E-03	0.91 (0.85 to 0.96)
		lcSSc (n=2929)*******	307 (10.48)	1236 (42.20)	1386 (47.32)	31.58	2.7E-03	0.01	0.90 (0.84 to 0.96)
		dcSSc (n=1412)	130 (9.21)	647 (45.82)	635 (44.97)	32.12	0.14	NA	0.93 (0.85 to 1.02)
		ACA+ (n=1744)	180 (10.32)	754 (43.23)	810 (46.44)	31.94	0.05	NA	0.92 (0.85 to 1)
		ATA+ (n=1056)	109 (10.32)	475 (44.98)	472 (44.70)	32.81	0.48	NA	0.96 (0.87 to 1.07)

*All p values have been calculated for the allelic model.
**Breslow-Day p_{value}=0.29. Higgins' test (I^2)=17.3%. Random-effects model p_{value}=8.8E-04 p_c=3.5E-3 Random-effects OR=0.86.
***Breslow-Day p_{value}=0.16. I^2=33.9%. Random-effects model p_{value}=4.1E-03. Random-effects OR=0.84.
****Breslow-Day p_{value}=0.06. I^2=48.6%. Random-effects model p_{value}=0.1. Random-effects OR estimate=0.93.
*****Breslow-Day p_{value}=0.09. I^2=43.4%. Random-effects model p_{value}=0.11. Random-effects OR estimate=0.92.
******Breslow-Day p_{value}=0.02. I^2=58%. Random-effects model p_{value}=0.08. Random-effects OR estimate=0.91.
*******Breslow-Day p_{value}=0.09. I^2=43.7%. Random-effects model p_{value}=0.05. Random-effects OR estimate=0.91.
†If it is applicable, Bonferroni correction is shown.
‡OR for the minor allele.
ACA, anticentromere autoantibodies; ATA, antitopoisomerase autoantibodies; dcSSc, diffuse cutaneous SSc; NA, not applicable; SNP, single nucleotide polymorphisms; SSc, systemic sclerosis.

and lcSSc patients (p_c=6E-4 OR=0.84 95% CI 0.76 to 0.92 and p_c=0.01 OR=0.9 95% CI 0.84 to 0.96, respectively). A trend of association was observed in the meta-analysis for the rs6822844 and rs6835457 variants and ACA positive status (p_{value}=0.01 OR=0.87 95% CI 0.78 to 0.97 and p_{value}=0.05 OR=0.92 95% CI 0.85 to 1, respectively). The rs6835457 SNP also had a trend of association with global SSc and lcSSc (p_{value}=0.01 OR=0.93 95% CI 0.87 to 0.98 and

Table 2 Conditional logistic regression analysis for the *IL-2/IL-21* SNPs located in SSc considering the eight European and US populations as covariate

						r2 with rs6822844							
Group of analysis	SNP	MAF Cases	MAF Controls	p value of each SNP conditioned by rs6822844	p value of rs6822844 conditioned by each SNP	Spain	Germany	The Netherlands	USA	Italy	Sweden	UK	Norway
SSc													
	rs2069762	0.31	0.30	0.69	1.30E-03	0.06	0.06	0.09	0.07	0.07	0.11	0.06	0.09
	rs6835457	0.32	0.34	0.43	0.024	0.25	0.37	0.38	0.36	0.28	0.48	0.37	0.49
	rs907715	0.32	0.34	0.19	0.026	0.26	0.37	0.39	0.36	0.29	0.39	0.37	0.49
lcSSc													
	rs2069762	0.31	0.30	0.69	9.07E-04	–	–	–	–	–	–	–	–
	rs6835457	0.32	0.34	0.53	0.014	–	–	–	–	–	–	–	–
	rs907715	0.32	0.34	0.3	0.015	–	–	–	–	–	–	–	–
ACA+													
	rs2069762	0.31	0.30	0.64	0.015	–	–	–	–	–	–	–	–
	rs6835457	0.32	0.34	0.81	0.061	–	–	–	–	–	–	–	–
	rs907715	0.32	0.34	0.56	0.063	–	–	–	–	–	–	–	–

ACA, anticentromere autoantibodies; lcSSC, limited cutaneous SSc; MAF, minor allelic frequencies; SNP, single nucleotide polymorphisms; SSc, systemic sclerosis.

p_{value}=0.01 OR=0.92 95% CI 0.86 to 0.98, respectively). We did not detect any significant association between the rs6835457 or rs2069762 SNPs and the global SSc diseases or its different phenotypes (for detailed information see supplementary tables S1 through S3). It is worth noting that the minor I^2 percentage was observed in the meta-analysis for the rs6822844 SNP with SSc (17.3%) and lcSSc (33.9%), suggesting that the variation between the populations is moderate. Moreover, these analyses were the only ones that remained significant in the random-effect model (rs6822844 and SSc p_c=3.5E-3, rs6822844 and lcSSc p_c=0.016).

A conditional logistic regression analysis was used to identify which SNP could be the causal SNP for the observed associations between the studied polymorphisms. The association of each SNP was evaluated using the populations as covariates, and the association was conditioned to the rs6822844 SNP because the lowest p_{value} and strongest effect (OR) were observed in this locus. Pairwise conditional analysis showed that the association of the rs907715 SNP was explained by the rs6822844 effect, because only the latter SNP remained significant after conditioned to each other (rs907715 conditioned p_{value}=0.19; rs6822844 conditioned p_{value}=0.026). Moreover, the rs2069762 and rs6835457 SNPs exhibited significance only when conditioned to the rs6822844 SNP. These results suggested that the rs6822844 signal could explain the association observed in the *IL-2/IL-21* locus (table 2).

Finally, the results of the conditional haplotype-based association testing are shown in table 3. The allelic combination formed by the rs2069762 major allele and the rs6822844, rs6835457 and rs907715 minor alleles was significantly increased in the controls compared with the global SSc (p_c=1.7E-3, OR=0.89 95% CI 0.81 to 0.98), the lcSSc subtype (p_c=8E-4, OR=0.86 95% CI 0.77 to 0.96) and the ACA positive status (p_c=2.7E-2, OR=0.86 95% CI 0.75 to 0.98). Interestingly, the OR observed for this analysis was not different from the one observed in the allelic test. Moreover, the significant effect of the omnibus analyses for SSc, lcSSc and ACA positive status disappeared when they were controlled by the rs6822844 SNP (p_{values} of the likelihood ratio test were: p_{value}=0.66 for global SSc, p_{value}=0.74 for lcSSc and p_{value}=0.93 for ACA+).

DISCUSSION

Our study suggests for the first time the influence of the rs6822844 polymorphism of the *IL-2/IL-21* region in susceptibility to SSc. This variant also influences the lcSSc subtype of the diseases and probably the ACA positive status due to the trend of association observed between the rs6822844 polymorphism and this phenotype. Although, our study had sufficient statistical power for both dcSSc and ATA analysis (95% and 91%, respectively), we observed that there were not significant associations between the four *IL-2/IL-21* SNPs and dcSSc or ATA positive status. The ORs exhibited the same direction as the significant associations with SSc and lcSSc, suggesting that an increment in the sample size with future studies could show a significant relation between the *IL-2/IL-21* SNPs and dcSSc or ATA. Interestingly, the rs6822844 variant was associated in the same OR direction as that observed in SLE. The minor allele of this variant is more frequent in healthy donors than in SSc patients, lcSSc subtype subjects and SLE patients.[12 16] The logistic regression and the allelic combination analyses support that the rs6822844 SNP association was responsible for the observed associations. The rs2069762A-rs6822844T-rs6835457G-rs907715T allelic combination was associated as a protective factor to SSc, lcSSc subtype and ACA positive status, which is the same effect observed for the T allele of the rs6822844. Importantly, the ORs observed for this allelic combination were not different from the ORs observed for the rs6822844 SNP analysis. These observations were slightly different from the results of the SLE study performed by Hughes *et al*[12] where the observed association between *IL-2/IL-21* region and SLE could be explained by the rs6835457 and rs907715 SNPs. Together, these results support the idea that the common genetic factors in autoimmune diseases may be associated at a regional level but differ in the specific SNPs associated with each disease, including the magnitude and direction of the association.[31 32] Although, the logistic regression test and the allelic combination analyses conditioned by the rs6822844 SNP suggest that this variant is responsible for the association observed in the region; we cannot totally discard a slight role of the rs6835457 and rs907715 polymorphisms in SSc due to the moderate LD between them and the r6822844 SNP.

The rs6822844 and rs6835457 SNPs are located in the flanking 3′-untranslated region of *IL-21*, and the rs907715 polymorphism is located in intron 3 of the *IL-21* gene. In contrast, the rs2069762 SNP is located in the flanking 5′-untranslated region of *IL-2*, which did not exhibit significant association with SSc, the subtypes of the disease or the antibodies' status. The rs2069762 minor allele has been previously associated with the lcSSc subtype.[26] Our study has a considerably larger sample size than the previous study; therefore the previously reported

Table 3 Conditional haplotype-based association analysis of four *IL-2/IL-21* SNPs located according to diseases, lcSSc diseases subtype and ACA status and considering the eight European and US populations as covariate

	Frequency					Frequency				Frequency			
Allelic combination†	Controls	SSc	OR (CI 95%)	P*	Pc‡	lcSSc	OR (CI 95%)	P*	Pc‡	ACA+	OR (CI 95%)	P*	Pc‡
AGGT	0.182	0.181	-ref-	–**	–	0.181	-ref-	–***	–	0.183	-ref-	–****	–
ATGT	0.152	0.133	0.89 (0.81 to 0.98)	4.12E-04	1.65E-03	0.131	0.86 (0.77 to 0.96)	2.01E-04	8.04E-04	0.132	0.86 (0.75 to 0.98)	6.81E-03	2.72E-02
CGAC	0.298	0.306	1.04 (0.96 to 1.12)	0.18	NA	0.308	1.03 (0.94 to 1.13)	0.2	NA	0.308	1.02 (0.92 to 1.14)	0.27	NA
AGAC	0.369	0.379	1.04 (0.96 to 1.13)	0.1	NA	0.38	1.03 (0.94 to 1.13)	0.12	NA	0.377	1.02 (0.91 to 1.13)	0.4	NA

†The order of the SNPs is rs2069762, rs6822844, rs6835457, rs907715.
*p_{value} of the likelihood ratio test. Based on WHAP method.[29]
‡If it is applicable, Bonferroni correction is shown. Not applicable (NA).
**Omnibus test X^2=14.5 (df=3); p value=2.35E-03; Pc=9.4E-03.
***Omnibus test X^2=14.5 (df=3); p value=2.32E-03; Pc=9.28E-03.
****Omnibus test X^2=8.92 (df=3); p value=0.03; Pc=0.12.
ACA, anticentromere autoantibodies; lcSSc, limited cutaneous SSc; SNP, single nucleotide polymorphisms; SSc, systemic sclerosis.

significant association for rs2069762 might stem from type 1 statistical error. This fact together with the location of the associated SNP suggests a highlighted role of the *IL-21* cytokine. By examining the expression and regulation of *IL-21* and the *IL-21* receptor (*IL-21R*) in patients with SSc, a previous study demonstrated an upregulation of *IL-21R* in epidermis samples.[33] However, a recent study has demonstrated that the scleroderma burden in allogeneic haemopoietic stem cell transplantations is driven by Th17 induction via *IL-21* and *IL-23* signalling.[34] Together, these results suggest that *IL-21/IL-21R* signalling has a pathogenic function in SSc.

The role of *IL-2* and *IL-21* in the immune system makes these genes plausible candidates for the genetic component of autoimmune diseases.[1 35 36] Our results increase the evidence that have showed that the rs6822844 is significantly associated with multiple autoimmune diseases.[10 13] According to the HapMap project for the CEU population (http://hapmap.ncbi.nlm.nih.gov/), the rs682284 polymorphism tags seven other variants located along the *IL-2/IL-21* region (rs13132245, rs13122573, rs4459999, rs13151961, rs13140464, rs6814280 and rs2069778), but clear evidence that connects any of these variants with the *IL-2* and/or *IL-21* regulation is lacking. Together, all point out these genetic variants as good candidates for functional studies in SSc pathogenesis and in other autoimmune diseases.

Although, the combined analyses of the rs6822844 polymorphism did not show heterogeneity (BD p_{value}=0.29. I^2=17.3%) between the eight European populations, a weak point of our study is that we did not have enough data available to control the association by principal component. Furthermore, as we mentioned before, an increment in the sample size for the stratified analysis could define in an accurate way the role of the studied variants in different clinical manifestations of SSc as their influence in the presence of coautoimmunity. Consequently, it is necessary to replicate the actual observation.

To conclude, consistent with previous studies on autoimmune diseases, the *IL-2/IL-21* region is a susceptibility genetic factor for SSc and its lcSSc subtype. The rs6822844 polymorphism confers the best association signal for SSc. It is also worth mentioning that this study shows the importance of the study of different populations and broad collaboration to find the missing heritability for relatively rare diseases like SSc.

Author affiliations
[1]Cellular Biology and Immunology Department, Instituto de Parasitología y Biomedicina López-Neyra, IPBLN-CSIC, Granada, Spain
[2]Department of Internal Medicine, Hospital Vall d'Hebron, Barcelona, Spain
[3]Department of Rheumatology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands
[4]Department of Internal Medicine, Hospital Clínico Universitario San Cecilio, Granada, Spain
[5]Department of Allergy, Clinical Immunology and Rheumatology, IRCCS Fondazione Policlinico-Mangiagalli-Regina Elena & University of Milan, Milan, Italy
[6]Department of Rheumatology, Hospital Universitario 12 de Octubre, Madrid, Spain
[7]Department of Rheumatology, Hospital Universitario y Politécnico La Fe, Valencia, Spain
[8]Department of Rheumatology, Hospital Marqués de Valdecilla, IFIMAV, Santander, Spain
[9]Department of Internal Medicine, Hospital Parc Taulí, Sabadell, Spain
[10]Department of Rheumatology, Hospital General Universitario Gregorio Marañon, Madrid, Spain
[11]Department of Systemic Autoimmune Diseases, Hospital Clinic de Barcelona, Barcelona, Spain
[12]Department of Rheumatology, Hospital Universitario La Princesa, Madrid, Spain
[13]Department of Rheumatology, Lund University Hospital, Lund, Sweden
[14]Department of Rheumatology and Clinical Immunology, Charité Universitätsmedizin and German Rheumatism Research Centre, a Leibniz institute. Berlin, Germany
[15]Department of Clinical Immunology and Rheumatology, Hannover Medical School, Hannover, Germany
[16]Department of Internal Medicine 3, 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]Institute of Cancer Sciences, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow, Scotland
[20]Rheumatology Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden
[21]Department of Rheumatology, Rikshospitalet, Oslo University Hospital, Oslo, Norway
[22]Department of Medicine, Università degli Studi di Verona, Verona, Italy
[23]UO Reumatologia ed Immunologia Clinica, Spedali Civili, Brescia, Italy
[24]Musculoskeletal Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle, UK
[25]Department of Dermatology, University of Cologne, Cologne, Germany
[26]Division of Transfusion Medicine, University Hospital of Cologne, Cologne, Germany
[27]Department of Dermatology, Allergology, and Venereology, Ruhr University of Bochum, Bochum, Germany
[28]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester Academic Health Science Centre, Manchester, UK
[29]Centre for Rheumatology, Royal Free and University College Medical School, London, UK
[30]Department of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center at Houston, Houston, Texas, USA
[31]Section Complex Genetics, Department of Medical Genetics, University Medical Center Utrecht, Utrecht, The Netherlands
[32]Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands

Acknowledgements We thank 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). We are also thankful to EUSTAR (The EULAR Scleroderma Trials and Research group) and the German Network of Systemic Sclerosis for the facilitation of this project.

Spanish Scleroderma Group Jose Luis Callejas and Raquel Ríos, Unidad de Enfermedades Sistémicas Autoinmunes, Servicio de Medicina Interna, Hospital Clínico Universitario San Cecilio, Granada; Nuria Navarrete, Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada; Rosa García Portales, Servicio de Reumatología, Hospital Virgen de la Victoria, Málaga; María Teresa Camps, Servicio de Medicina Interna, Hospital Carlos Haya, Málaga; Antonio Fernández-Nebro, Servicio de Reumatología, Hospital Carlos Haya, Málaga; María F. González-Escribano, Servicio de Inmunología, Hospital Virgen del Rocío, Sevilla; Julio Sánchez-Román, Francisco J García-Hernández and Mª Jesús Castillo, Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla; Mª Ángeles Aguirre and Inmaculada Gómez-Gracia, Servicio de Reumatología, Hospital Reina Sofía, Córdoba; Benjamín Fernández-Gutiérrez and Luis Rodríguez-Rodríguez, Servicio de Reumatología, Hospital Clínico San Carlos, Madrid; José Luis Andreu, Servicio de Reumatología, Hospital Puerta del Hierro, Madrid; Paloma García de la Peña, Servicio de Reumatología, Hospital Madrid Norte Sanchinarro, Madrid; Lina Martínez, Servicio de Reumatología, Hospital General Universitario Gregorio Marañón, Madrid; María Ángeles Robles, Natividad Oreiro, Servicio de Reumatología, INIBIC-Hospital Universitario A Coruña, La Coruña; Vicente Fonollosa, Servicio de Medicina Interna, Hospital Valle de Hebrón, Barcelona; Anna Pros, Servicio de Reumatología, Hospital Del Mar, Barcelona; Mónica Rodríguez Carballeira, Servicio de Medicina Interna, Hospital Universitari Mútua Terrasa, Barcelona; Francisco Javier Narváez, Servicio de Reumatología, Hospital Universitari de Bellvitge, Barcelona; Bernardino Díaz, Luis Tropiella and María Gallego, Servicio de Medicina Interna, Hospital Central de Asturias, Oviedo; María del Carmen Freire and Inés Vaqueiro, Unidad de Trombosis y Vasculitis, Servicio de Medicina Interna, Hospital Xeral-Complexo Hospitalario Universitario de Vigo, Vigo; María Victoria Egurbide, Servicio de Medicina Interna, Hospital de Cruces, Barakaldo; Luis Sáez-Comet, Unidad de Enfermedades Autoinmunes Sistémicas, Servicio de Medicina Interna, Hospital Universitario Miguel Servet, Zaragoza; Federico Díaz and Vanesa Hernández, Servicio de Reumatología, Hospital Universitario de Canarias, Tenerife; Emma Beltrán, Hospital General Universitario, Valencia Spain.

Contributors All the authors listed participated in all or at list one of these activities: Conception and design, acquisition of data or analysis and interpretation of data. Drafting the article or revising it critically for important intellectual content. Final approval of the version published.

Funding This work was supported by the following grants: LMDG was funded by the 'Ayudas Predoctorales de Formación en Investigación en Salud (PFIS-FI09/00544)' from the Instituto de Salud Carlos III. JM was funded by GEN-FER from the Spanish

Society of Rheumatology, SAF2009-11110 from the Spanish Ministry of Science, CTS-4977 from Junta de Andalucía, Spain, in part by Redes Temáticas de Investigación Cooperativa Sanitaria Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), Spain and by Fondo Europeo de Desarrollo Regional (FEDER). TRDJR was funded by the VIDI laureate from the Dutch Association of Research (NWO) and Dutch Arthritis Foundation (National Reumafonds). JM and TRDJR were sponsored by the Orphan Disease Program grant from the European League Against Rheumatism (EULAR). BPCK is supported by the Dutch Diabetes Research Foundation (grant 2008.40.001) and the Dutch Arthritis Foundation (Reumafonds, grant NR 09-1-408). TW was granted by DFG WI 1031/6.1 and DFG KFO 250 TP03. NOC was funded by PI-0590-2010, Consejería de Salud, Junta de Andalucía, Spain.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 and by the NIH-NIAMS Center of Research Translation (CORT) in SSc (P50AR054144), K23AR061436, and the Department of Defense Congressionally Directed Medical Research Programs (W81XWH-07-01-0111), (NIH/NIAID) 1U01AI09090.

Competing interests None.

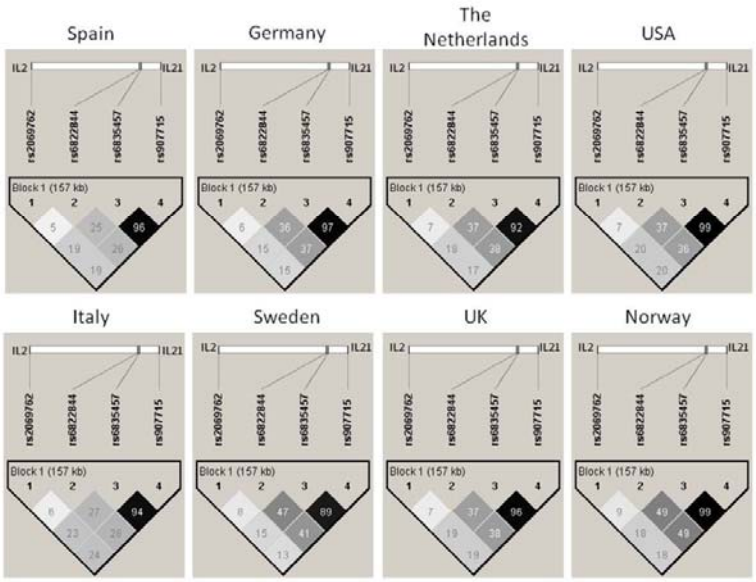
Ethics approval The local ethical committees in accordance with the tenets of the Declaration of Helsinki.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1. **McGuire HM,** Vogelzang A, Hill N, *et al*. Loss of parity between IL-2 and IL-21 in the NOD Idd3 locus. *Proc Natl Acad Sci USA* 2009;**106**:19438–43.
2. **Yamanouchi J,** Rainbow D, Serra P, *et al*. Interleukin-2 gene variation impairs regulatory T cell function and causes autoimmunity. *Nat Genet* 2007;**39**:329–37.
3. **Bubier JA,** Sproule TJ, Foreman O, *et al*. A critical role for IL-21 receptor signaling in the pathogenesis of systemic lupus erythematosus in BXSB-Yaa mice. *Proc Natl Acad Sci USA* 2009;**106**:1518–23.
4. **Wellcome Trust Case Control Consortium.** Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;**447**:661–78.
5. **Todd JA,** Walker NM, Cooper JD, *et al*. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007;**39**:857–64.
6. **van Heel DA,** Franke L, Hunt KA, *et al*. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat Genet* 2007;**39**:827–9.
7. **Zhernakova A,** Alizadeh BZ, Bevova M, *et al*. Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases. *Am J Hum Genet* 2007;**81**:1284–8.
8. **Barton A,** Eyre S, Ke X, *et al*. Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes. *Hum Mol Genet* 2009;**18**:2518–22.
9. **Festen EA,** Goyette P, Scott R, *et al*. Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis. *Gut* 2009;**58**:799–804.
10. **Hollis-Moffatt JE,** Chen-Xu M, Topless R, *et al*. Only one independent genetic association with rheumatoid arthritis within the KIAA1109-TENR-IL2-IL21 locus in Caucasian sample sets: confirmation of association of rs6822844 with rheumatoid arthritis at a genome-wide level of significance. *Arthritis Res Ther* 2010;**12**:R116.
11. **Hollis-Moffatt JE,** Gearry RB, Barclay ML, *et al*. Consolidation of evidence for association of the KIAA1109-TENR-IL2-IL21 rs6822844 variant with Crohn's disease. *Am J Gastroenterol* 2010;**105**:1204–5.
12. **Hughes T,** Kim-Howard X, Kelly JA, *et al*. Fine-mapping and transethnic genotyping establish IL2/IL21 genetic association with lupus and localize this genetic effect to IL21. *Arthritis Rheum* 2011;**63**:1689–97.
13. **Maiti AK,** Kim-Howard X, Viswanathan P, *et al*. Confirmation of an association between rs6822844 at the IL2-IL21 region and multiple autoimmune diseases: evidence of a general susceptibility locus. *Arthritis Rheum* 2010;**62**:323–9.
14. **Marquez A,** Orozco G, Martínez A, *et al*. Novel association of the interleukin 2-interleukin 21 region with inflammatory bowel disease. *Am J Gastroenterol* 2009;**104**:1968–75.
15. **Rodriguez-Rodriguez L,** Castaneda S, Vazquez-Rodriguez TR, *et al*. Role of the rs6822844 gene polymorphism at the IL2-IL21 region in biopsy-proven giant cell arteritis. *Clin Exp Rheumatol* 2011;**29**(1 Suppl 64):S12–16.
16. **Sawalha AH,** Kaufman KM, Kelly JA, *et al*. Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus. *Ann Rheum Dis* 2008;**67**:458–61.
17. **Warren RB,** Smith RL, Flynn E, *et al*. A systematic investigation of confirmed autoimmune loci in early-onset psoriasis reveals an association with IL2/IL21. *Br J Dermatol* 2011;**164**:660–4.
18. **Gabrielli A,** Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009;**360**:1989–2003.
19. **Katsumoto TR,** Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol* 2011;**6**:509–37.
20. **Koenig M,** Dieude M, Senecal JL. Predictive value of antinuclear autoantibodies: the lessons of the systemic sclerosis autoantibodies. *Autoimmun Rev* 2008;**7**:588–93.
21. **LeRoy EC,** Black C, Fleischmajer R, *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. *J Rheumatol* 1988;**15**:202–5.
22. **Martin JE,** Bossini-Castillo L, Martin J. Unraveling the genetic component of systemic sclerosis. *Hum Genet* 2012;**131**:1023–37.
23. **Assassi S,** Mayes MD, Arnett FC, *et al*. Systemic sclerosis and lupus: points in an interferon-mediated continuum. *Arthritis Rheum* 2010;**62**:589–98.
24. **Cannona FD,** Gutala R, Simeon CP, *et al*. Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis. *Ann Rheum Dis* 2011;**71**:114–19.
25. **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.
26. **Mattuzzi S,** Barbi S, Carletto A, *et al*. Association of polymorphisms in the IL1B and IL2 genes with susceptibility and severity of systemic sclerosis. *J Rheumatol* 2007;**34**:997–1004.
27. **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.
28. **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.
29. **Purcell S,** Daly MJ, Sham PC. WHAP: haplotype-based association analysis. *Bioinformatics* 2007;**23**:255–6.
30. **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.
31. **Cho JH,** Gregersen PK. Genomics and the multifactorial nature of human autoimmune disease. *N Engl J Med* 2011;**365**:1612–23.
32. **Sirota M,** Schaub MA, Batzoglou S, *et al*. Autoimmune disease classification by inverse association with SNP alleles. *PLoS Genet* 2009;**5**:e1000792.
33. **Distler JH,** Jungel A, Kowal-Bielecka O, *et al*. Expression of interleukin-21 receptor in epidermis from patients with systemic sclerosis. *Arthritis Rheum* 2005;**52**:856–64.
34. **Hill GR,** Olver SD, Kuns RD, *et al*. Stem cell mobilization with G-CSF induces type 17 differentiation and promotes scleroderma. *Blood* 2010;**116**:819–28.
35. **Crispin JC,** Tsokos GC. Transcriptional regulation of IL-2 in health and autoimmunity. *Autoimmun Rev* 2009;**8**:190–5.
36. **Ettinger R,** Kuchen S, Lipsky PE. Interleukin 21 as a target of intervention in autoimmune disease. *Ann Rheum Dis* 2008;**67**(Suppl 3):iii83–6.

Supplementary Figure 1. LD plot based on the r^2 values for the eight European and US American descendent populations.



Supplementary Table 1. Genotype and minor allele frequencies of meta-analysis of four *IL2-IL21* SNPs located in Caucasian SSc patients and healthy controls from Europe and USA.

SNP	Alleles	Population	Genotype, N (%)			MAF (%)	Allele test		
			1/1	1/2	2/2		*P*-value[a]	*Pc*[b]	OR [CI 95%][c]
rs2069762	C/A	**Spain**							
		Controls (n=1716)	157 (9.15)	698 (40.69)	861 (50.15)	29.5			
		SSc (n=1149)	108 (9.40)	468 (40.73)	573 (49.87)	29.77	0.83	NA	1.01 [0.90-1.14]
		Germany							
		Controls (n=411)	26 (6.33)	164 (39.90)	221 (53.77)	26.28			
		SSc (n=591)	58 (9.81)	238 (40.27)	295 (49.92)	29.95	0.07	NA	1.20 [0.98-1.46]
		The Netherlands							
		Controls (n=656)	56 (8.54)	266 (40.55)	334 (50.91)	28.81			
		SSc (n=342)	30 (8.77)	133 (38.89)	179 (52.34)	28.22	0.78	NA	0.97 [0.79-1.19]
		USA							
		Controls (n=864)	84 (9.72)	358 (41.44)	422 (48.84)	30.44			
		SSc (n=916)	87 (9.50)	379 (41.38)	450 (49.13)	30.19	0.87	NA	0.99 [0.86-1.14]
		Italy							
		Controls (n=1024)	131 (12.79)	453 (44.24)	440 (42.97)	34.91			
		SSc (n=573)	83 (14.49)	261 (45.55)	229 (39.97)	37.26	0.18	NA	1.11 [0.95-1.29]
		Sweden							
		Controls (n=262)	20 (7.63)	97 (37.02)	145 (55.34)	26.15			
		SSc (n=195)	13 (6.67)	77 (39.49)	105 (53.85)	26.41	0.93	NA	1.01 [0.75-1.37]
		UK							
		Controls (n=371)	31 (8.36)	145 (39.08)	195 (52.56)	27.9			
		SSc (n=415)	39 (9.40)	177 (42.65)	199 (47.95)	30.72	0.22	NA	1.15 [0.92-1.43]
		Norway							
		Controls (n=277)	14 (5.05)	125 (45.13)	138 (49.82)	27.62			
		SSc (n=100)	11 (11.00)	45 (45.00)	44 (44.00)	33.5	0.12	NA	1.32 [0.93-1.87]
		Combined							
		Controls (n=5432)	510 (9.30)	2266 (41.34)	2706 (49.36)	29.97			
		SSc (n=4281)	429 (10.02)	1778 (41.53)	2074 (48.45)	30.79	**0.08**	**NA**	**1.06 [0.99-1.13]**
rs6822844	T/G	**Spain**							
		Controls (n=1721)	34 (1.98)	351 (20.40)	1336 (77.63)	12.17			
		SSc (n=1176)	10 (0.85)	225 (19.13)	941 (80.02)	10.42	0.04	0.16	0.84 [0.71-0.99]
		Germany							
		Controls (n=423)	9 (2.13)	109 (25.77)	305 (72.10)	15.01			
		SSc (n=605)	12 (1.98)	139 (22.98)	454 (75.04)	13.47	0.32	NA	0.88 [0.69-1.13]
		The Netherlands							
		Controls (n=734)	27 (3.68)	194 (26.43)	513 (69.89)	16.89			
		SSc (n=365)	16 (4.38)	97 (26.58)	252 (69.04)	17.67	0.65	NA	1.06 [0.84-1.33]
		USA							
		Controls (n=884)	22 (2.49)	243 (27.49)	619 (70.02)	16.23			
		SSc (n=916)	14 (1.53)	244 (26.64)	658 (71.83)	14.85	0.25	NA	0.90 [0.75-1.08]
		Italy							
		Controls (n=1107)	22 (1.99)	269 (24.30)	816 (73.71)	14.14			

		SSc (n=595)	14 (2.35)	103 (17.31)	478 (80.34)	11.01	0.01	0.04	0.75 [0.60-0.93]
		Sweden							
		Controls (n=273)	7 (2.56)	92 (33.70)	174 (63.74)	19.41			
		SSc (n=225)	13 (5.78)	61 (27.11)	151 (67.11)	19.33	0.97	NA	0.99 [0.73-1.36]
		UK							
		Controls (n=372)	15 (4.03)	123 (33.06)	234 (62.90)	20.56			
		SSc (n=424)	14 (3.30)	101 (23.82)	309 (72.88)	15.21	0.01	0.02	0.69 [0.54-0.90]
		Norway							
		Controls (n=273)	13 (4.68)	94 (33.81)	171 (61.51)	21.58			
		SSc (n=101)	5 (4.95)	26 (25.74)	70 (69.31)	17.82	0.26	NA	0.79 [0.52-1.19]
		Combined							
		Controls (n=5792)	149 (2.57)	1475 (25.47)	4168 (71.96)	15.31			
		SSc (n=4407)	98 (2.22)	996 (22.60)	3313 (75.18)	13.52	**1.66E-04**	**6.33E-04**	**0.86 [0.79-0.93]**
		Spain							
		Controls (n=1695)	206 (12.15)	721 (42.54)	768 (45.31)	33.42			
		SSc (n=1171)	105 (8.97)	514 (43.89)	552 (47.14)	30.91	0.05	0.18	0.89 [0.80-1.0]
		Germany							
		Controls (n=428)	33 (7.75)	185 (43.43)	208 (48.83)	29.46			
		SSc (n=609)	67 (11.00)	247 (40.56)	295 (48.44)	31.28	0.38	NA	1.09 [0.90-1.32]
		The Netherlands							
		Controls (n=717)	87 (12.13)	284 (39.61)	346 (48.26)	31.94			
		SSc (n=358)	37 (10.34)	162 (45.25)	159 (44.41)	32.96	0.63	NA	1.05 [0.87-1.27]
		USA							
		Controls (n=884)	88 (9.95)	405 (45.81)	391 (44.23)	32.86			
		SSc (n=915)	96 (10.49)	418 (45.68)	401 (43.83)	33.33	0.76	NA	1.02 [0.89-1.17]
rs6835457	**G/A**	**Italy**							
		Controls (n=1076)	143 (13.29)	479 (44.52)	454 (42.19)	35.55			
		SSc (n=587)	59 (10.05)	237 (40.37)	291 (49.57)	30.24	1.96E-03	7.33E-03	0.79 [0.67-0.92]
		Sweden							
		Controls (n=271)	30 (11.07)	121 (44.65)	120 (44.28)	33.39			
		SSc (n=220)	25 (11.36)	94 (42.73)	101 (45.91)	32.73	8.25E-01	NA	0.97 [0.74-1.27]
		UK							
		Controls (n=374)	46 (12.30)	182 (48.66)	146 (39.04)	36.63			
		SSc (n=430)	48 (11.16)	193 (44.88)	189 (43.95)	33.6	2.05E-01	NA	0.88 [0.71-1.08]
		Norway							
		Controls (n=277)	35 (12.64)	130 (46.93)	112 (40.43)	36.1			
		SSc (n=102)	8 (7.84)	43 (42.16)	51 (50.00)	28.92	0.06	NA	0.72 [0.51-1.02]
		Combined							
		Controls (n=5720)	668 (11.68)	2507 (43.83)	2545 (44.49)	33.59			
		SSc (n=4392)	445 (10.13)	1908 (43.44)	2039 (46.43)	31.85	**0.01**	**0.05**	**0.93 [0.87-0.98]**
		Spain							
		Controls (n=1661)	208 (12.52)	723 (43.53)	730 (43.95)	34.29			
		SSc (n=1144)	103 (9.00)	499 (43.62)	542 (47.38)	30.81	6.49E-03	0.03	0.85 [0.76-0.96]
		Germany							
		Controls (n=407)	32 (7.86)	176 (43.24)	199 (48.89)	29.48			

		SSc (n=598)	65 (10.87)	244 (40.80)	289 (48.33)	31.27	0.39	NA	1.09 [0.90-1.3]
		The Netherlands							
		Controls (n=709)	85 (11.99)	283 (39.92)	341 (48.10)	31.95			
		SSc (n=351)	35 (9.97)	159 (45.30)	157 (44.73)	32.62	0.75	NA	1.03 [0.85-1.25]
		USA							
		Controls (n=882)	88 (9.98)	401 (45.46)	393 (44.56)	32.71			
		SSc (n=913)	97 (10.62)	416 (45.56)	400 (43.81)	33.41	0.66	NA	1.03 [0.90-1.19]
rs907715	**T/C**	**Italy**							
		Controls (n=1063)	146 (13.73)	476 (44.78)	441 (41.49)	36.12			
		SSc (n=584)	52 (8.90)	244 (41.78)	288 (49.32)	29.79	2.41E-04	9.53E-04	0.75 [0.64-0.87]
		Sweden							
		Controls (n=270)	29 (10.74)	119 (44.07)	122 (45.19)	32.78			
		SSc (n=216)	24 (11.11)	89 (41.20)	103 (47.69)	31.71	0.72	NA	0.95 [0.73-1.25]
		UK							
		Controls (n=374)	47 (12.57)	183 (48.93)	144 (38.50)	37.03			
		SSc (n=436)	53 (12.16)	189 (43.35)	194 (44.50)	33.83	0.18	NA	0.87 [0.71-1.07]
		Norway							
		Controls (n=273)	35 (12.59)	130 (46.76)	113 (40.65)	35.97			
		SSc (n=99)	8 (8.08)	43 (43.43)	48 (48.48)	29.8	0.12	NA	0.76 [0.53-1.07]
		Combined							
		Controls (n=5644)	670 (11.87)	2491 (44.14)	2483 (43.99)	33.94			
		SSc (n=4341)	437 (10.07)	1883 (43.38)	2021 (46.56)	31.76	**1.79E-03**	**7.16E-03**	**0.91 [0.85-0.96]**

[a]All P-values have been calculated for the allelic model. [b]If it is applicable, Bonferroni correction is shown [c]Odds ratio for the minor allele.

Supplementary Table 2. Genotype and minor allele frequencies of meta-analysis of four *IL2-IL21* SNPs located in Caucasian SSc subtypes and healthy controls from Europe and the USA.

SNP	Alleles	Population	Genotype, N (%)			MAF (%)	Allele test		
			1/1	1/2	2/2		*P*-value[a]	*Pc*[b]	OR [CI 95%][c]
		Spain							
		Controls (n=1716)	157 (9.15)	698(40.69)	861 (50.15)	29.5			
		lcSSc (n=782)	74 (9.46)	320(40.92)	388 (49.62)	29.92	0.76	NA	1.02 [0.89-1.17]
		dcSSc (n=367)	34 (9.26)	148(40.33)	185 (50.41)	29.43	0.97	NA	1.0 [0.84-1.19]
		Germany							
		Controls (n=411)	26 (6.33)	164(39.90)	221 (53.77)	26.28			
		lcSSc (n=348)	38 (10.92)	143(41.09)	167 (47.99)	31.47	0.03	0.10	1.29 [1.03-1.61]
		dcSSc (n=243)	20 (8.23)	95 (39.09)	128 (52.67)	27.78	0.55	NA	1.08 [0.84-1.39]
		The Netherlands							
		Controls (n=656)	56 (8.54)	266(40.55)	334 (50.91)	28.81			
		lcSSc (n=234)	21 (8.97)	86 (36.75)	127 (54.27)	27.35	0.55	NA	0.93 [0.73-1.18]
		dcSSc (n=108)	9 (8.33)	47 (43.52)	52 (48.15)	30.09	0.70	NA	1.06 [0.78-1.46]
		USA							
		Controls (n=864)	84 (9.72)	358(41.44)	422 (48.84)	30.44			
		lcSSc (n=595)	48 (8.07)	254(42.69)	293 (49.24)	29.41	0.55	NA	0.95 [0.81-1.12]
		dcSSc (n=321)	39 (12.15)	125(38.94)	157 (48.91)	31.62	0.58	NA	1.06 [0.87-1.29]
rs2069762	C/A	**Italy**							
		Controls (n=1024)	131 (12.79)	453(44.24)	440 (42.97)	34.91			
		lcSSc (n=418)	68 (16.27)	186(44.50)	164 (39.23)	38.52	0.07	NA	1.17 [0.99-1.38]
		dcSSc (n=155)	15 (9.68)	75 (48.39)	65 (41.94)	33.87	0.72	NA	0.95 [0.74-1.23]
		Sweden							
		Controls (n=262)	20 (7.63)	97 (37.02)	145 (55.34)	26.15			
		lcSSc (n=149)	10 (6.71)	56 (37.58)	83 (55.70)	25.5	0.84	NA	0.97 [0.70-1.34]
		dcSSc (n=46)	3 (6.52)	21 (45.65)	22 (47.83)	29.35	0.52	NA	1.17 [0.72-1.91]
		UK							
		Controls (n=371)	31 (8.36)	145(39.08)	195 (52.56)	27.9			
		lcSSc (n=305)	30 (9.84)	125(40.98)	150 (49.18)	30.33	0.33	NA	1.13 [0.89-1.42]
		dcSSc (n=110)	9 (8.18)	52 (47.27)	49 (44.55)	31.82	0.26	NA	1.21 [0.87-1.67]
		Norway							
		Controls (n=277)	14 (5.05)	125(45.13)	138 (49.82)	27.62			
		lcSSc (n=66)	6 (9.09)	33 (50.00)	27 (40.91)	34.09	0.14	NA	1.36 [0.90-2.03]
		dcSSc (n=34)	5 (14.71)	12 (35.29)	17 (50.00)	32.35	0.41	NA	1.25 [0.73-2.15]
		Combined							
		Controls (n=5482)	510 (9.30)	2266 (41.34)	2706 (49.36)	29.97			
		lcSSc (n=2897)	295 (10.18)	1203 (41.53)	1399 (48.29)	30.95	**0.09**	**NA**	**1.06 [0.99-1.14]**
		dcSSc (n=1384)	134 (9.68)	575(41.55)	675 (48.77)	30.46	**0.31**	**NA**	**1.05 [0.96-1.15]**
		Spain							
		Controls (n=1721)	34 (1.98)	351(20.40)	1336 (77.63)	12.17			
		lcSSc (n=805)	7 (0.87)	145(18.01)	653 (81.12)	9.876	0.02	0.13	0.79 [0.65-0.96]
		dcSSc (n=371)	3 (0.81)	80 (21.56)	288 (77.63)	11.59	0.66	NA	0.95 [0.74-1.21]
		Germany							
		Controls (n=423)	9 (2.13)	109(25.77)	305 (72.10)	15.01			
		lcSSc (n=357)	5 (1.40)	81 (22.69)	271 (75.91)	12.75	0.20	NA	0.83 [0.62-1.11]
		dcSSc (n=248)	7 (2.82)	58 (23.39)	183 (73.79)	14.52	0.81	NA	0.96 [0.70-1.32]
		The Netherlands							
		Controls (n=734)	27 (3.68)	194(26.43)	513 (69.89)	16.89			
		lcSSc (n=248)	12 (4.84)	58 (23.39)	178 (71.77)	16.53	0.85	NA	0.97 [0.74-1.28]
		dcSSc (n=117)	4 (3.42)	39 (33.33)	74 (63.25)	20.09	0.23	NA	1.24 [0.87-1.75]
		USA							
		Controls (n=884)	22 (2.49)	243(27.49)	619 (70.02)	16.23			
		lcSSc (n=595)	8 (1.34)	159(26.72)	428 (71.93)	14.71	0.26	NA	0.89 [0.73-1.09]
		dcSSc (n=321)	6 (1.87)	85 (26.48)	230 (71.65)	15.11	0.51	NA	0.92 [0.72-1.18]
rs6822844	G/T	**Italy**							
		Controls (n=1107)	22 (1.99)	269(24.30)	816 (73.71)	14.14			
		lcSSc (n=431)	8 (1.86)	77 (17.87)	346 (80.28)	10.79	0.01	0.05	0.73 [0.57-0.94]

		dcSSc (n=164)	6 (3.66)	26 (15.85)	132 (80.49)	11.59	0.21	NA	0.80 [0.56-1.14]
		Sweden							
		Controls (n=273)	7 (2.56)	92 (33.70)	174 (63.74)	19.41			
		lcSSc (n=164)	13 (7.93)	50 (30.49)	101 (61.59)	23.17	0.19	NA	1.25 [0.90-1.75]
		dcSSc (n=61)	0 (0.00)	11 (18.03)	50 (81.97)	9.016	0.01	0.03	0.41 [0.21-0.79]
		UK							
		Controls (n=372)	15 (4.03)	123(33.06)	234 (62.90)	20.56			
		lcSSc (n=311)	11 (3.54)	74 (23.79)	226 (72.67)	15.43	0.01	0.06	0.71 [0.53-0.93]
		dcSSc (n=113)	3 (2.65)	27 (23.89)	83 (73.45)	14.6	0.05	NA	0.66 [0.44-1.00]
		Norway							
		Controls (n=278)	13 (4.68)	94 (33.81)	171 (61.51)	21.58			
		lcSSc (n=66)	3 (4.55)	15 (22.73)	48 (72.73)	15.91	0.15	NA	0.69 [0.41-1.14]
		dcSSc (n=35)	2 (5.71)	11 (31.43)	22 (62.86)	21.43	0.98	NA	0.99 [0.54-1.82]
		Combined							
		Controls (n=5792)	149 (2.57)	1475 (25.47)	4168 (71.96)	15.31			
		lcSSc (n=2977)	67 (2.25)	659(22.14)	2251 (75.61)	13.32	**1.5E-04**	**6.0E-04**	**0.84 [0.76-0.92]**
		dcSSc (n=1430)	31 (2.17)	337(23.57)	1062 (74.27)	13.95	**0.06**	**NA**	**0.89 [0.79-1.0]**
		Spain							
		Controls (n=1695)	206 (12.15)	721(42.54)	768 (45.31)	33.42			
		lcSSc (n=800)	74 (9.25)	338(42.25)	388 (48.50)	30.38	0.03	0.07	0.87 [0.76-0.99]
		dcSSc (n=371)	31 (8.36)	176(47.44)	164 (44.20)	32.08	0.48	NA	0.94 [0.79-1.12]
		Germany							
		Controls (n=426)	33 (7.75)	185 (43.43)	208 (48.83)	29.46			
		lcSSc (n=359)	39 (10.86)	139(38.72)	181 (50.42)	30.22	0.74	NA	1.04 [0.83-1.29]
		dcSSc (n=250)	28 (11.20)	108(43.20)	114 (45.60)	32.8	0.20	NA	1.17 [0.92-1.48]
		The Netherlands							
		Controls (n=717)	87 (12.13)	284(39.61)	346 (48.26)	31.94			
		lcSSc (n=243)	29 (11.93)	106(43.62)	108 (44.44)	33.74	0.46	NA	1.09 [0.87-1.35]
		dcSSc (n=115)	8 (6.96)	56 (48.70)	51 (44.35)	31.3	0.85	NA	0.97 [0.72-1.31]
		USA							
		Controls (n=884)	88 (9.95)	405(45.81)	391 (44.23)	32.86			
		lcSSc (n=595)	60 (10.08)	276(46.39)	259 (43.53)	33.28	0.81	NA	1.02 [0.87-1.19]
		dcSSc (n=320)	36 (11.25)	142(44.38)	142 (44.38)	33.44	0.79	NA	1.03 [0.85-1.24]
rs6835457	**G/A**	**Italy**							
		Controls (n=1076)	143 (13.29)	479(44.52)	454 (42.19)	35.55			
		lcSSc (n=428)	45 (10.51)	167(39.02)	216 (50.47)	30.02	3.9E-03	0.02	0.78 [0.66-0.92]
		dcSSc (n=159)	14 (8.81)	70 (44.03)	75 (47.17)	30.82	0.10	NA	0.81 [0.63-1.04]
		Sweden							
		Controls (n=271)	30 (11.07)	121(44.65)	120 (44.28)	33.39			
		lcSSc (n=162)	21 (12.96)	69 (42.59)	72 (44.44)	34.26	0.79	NA	1.04 [0.78-1.39]
		dcSSc (n=58)	4 (6.90)	25 (43.10)	29 (50.00)	28.45	0.30	NA	0.79 [0.51-1.23]
		UK							
		Controls (n=374)	46 (12.30)	182(48.66)	146 (39.04)	36.63			
		lcSSc (n=311)	38 (12.22)	135(43.41)	138 (44.37)	33.92	0.30	NA	0.89 [0.71-1.11]
		dcSSc (n=119)	10 (8.40)	58 (48.74)	51 (42.86)	32.77	0.28	NA	0.84 [0.62-1.15]
		Norway							
		Controls (n=277)	35 (12.64)	130(46.93)	112 (40.43)	36.1			
		lcSSc (n=67)	6 (8.96)	25 (37.31)	36 (53.73)	27.61	0.06	NA	0.68 [0.45-1.02]
		dcSSc (n=35)	2 (5.71)	18 (51.43)	15 (42.86)	31.43	0.44	NA	0.81 [0.48-1.38]
		Combined							
		Controls (n=5720)	668 (11.68)	2507 (43.83)	2545 (44.49)	33.59			
		lcSSc (n=2965)	312 (10.52)	1255 (42.33)	1398 (47.15)	31.69	**0.01**	**0.06**	**0.92 [0.86-0.98]**
		dcSSc (n=1427)	133 (9.32)	653(45.76)	641 (44.92)	32.2	**0.28**	**NA**	**0.95 [0.87-1.04]**
		Spain							
		Controls (n=1661)	208 (12.52)	723(43.53)	730 (43.95)	34.29			
		lcSSc (n=780)	73 (9.36)	327(41.92)	380 (48.72)	30.32	0.01	0.02	0.83 [0.73-0.95]
		dcSSc (n=364)	30 (8.24)	172(47.25)	162 (44.51)	31.87	0.21	NA	0.90 [0.76-1.06]
		Germany							
		Controls (n=407)	32 (7.86)	176(43.24)	199 (48.89)	29.48			
		lcSSc (n=352)	37 (10.51)	137(38.92)	178 (50.57)	29.97	0.84	NA	1.02 [0.82-1.28]
		dcSSc (n=246)	28 (11.38)	107(43.50)	111 (45.12)	33.13	0.17	NA	1.19 [0.93-1.51]

		The Netherlands							
		Controls (n=709)	85 (11.99)	283(39.92)	341 (48.10)	31.95			
		lcSSc (n=236)	28 (11.86)	100(42.37)	108 (45.76)	33.05	0.66	NA	1.05 [0.84-1.31]
		dcSSc (n=115)	7 (6.09)	59 (51.30)	49 (42.61)	31.74	0.95	NA	0.99 [0.73-1.34]
		USA							
		Controls (n=882)	88 (9.98)	401 (45.46)	393 (44.56)	32.71			
		lcSSc (n=593)	60 (10.12)	274(46.21)	259 (43.68)	33.22	0.77	NA	1.02 [0.88-1.20]
		dcSSc (n=320)	37 (11.56)	142(44.38)	141 (44.06)	33.75	0.63	NA	1.05 [0.87-1.27]
rs907715	**T/C**	**Italy**							
		Controls (n=1063)	146 (13.73)	476(44.78)	441 (41.49)	36.12			
		lcSSc (n=426)	42 (9.86)	173(40.61)	211 (49.53)	30.16	2.0E-03	0.01	0.76 [0.64-0.91]
		dcSSc (n=158)	10 (6.33)	71 (44.94)	77 (48.73)	28.8	0.01	0.04	0.72 [0.55-0.93]
		Sweden							
		Controls (n=270)	29 (10.74)	119(44.07)	122 (45.19)	32.78			
		lcSSc (n=160)	19 (11.88)	67 (41.88)	74 (46.25)	32.81	0.99	NA	1.00 [0.75-1.35]
		dcSSc (n=56)	5 (8.93)	22 (39.29)	29 (51.79)	28.57	0.39	NA	0.82 [0.52-1.28]
		UK							
		Controls (n=374)	47 (12.57)	183(48.93)	144 (38.50)	37.03			
		lcSSc (n=317)	42 (13.25)	133(41.96)	142 (44.79)	34.23	0.28	NA	0.88 [0.71-1.10]
		dcSSc (n=119)	11 (9.24)	56 (47.06)	52 (43.70)	32.77	0.23	NA	0.83 [0.61-1.13]
		Norway							
		Controls (n=278)	35 (12.59)	130(46.76)	113 (40.65)	35.97			
		SSc (n=99)	8 (8.08)	43 (43.43)	48 (48.48)	29.8	0.12	NA	0.76 [0.53-1.07]
		lcSSc (n=65)	6 (9.23)	25 (38.46)	34 (52.31)	28.46	0.11	NA	0.71 [0.47-1.08]
		Combined							
		Controls (n=5644)	670 (11.87)	2491 (44.14)	2483 (43.99)	33.94			
		lcSSc (n=2929)	307 (10.48)	1236 (42.20)	1386 (47.32)	31.58	**2.73E-03**	**0.01**	**0.90 [0.84-0.96]**
		dcSSc (n=1412)	130 (9.21)	647(45.82)	635 (44.97)	32.12	**0.14**	**NA**	**0.93 [0.85-1.02]**

[a]All P-values have been calculated for the allelic model. [b]If it is applicable, Bonferroni correction is shown. [c]Odds ratio for the minor allele.

Supplementary Table 3. Genotype and minor allele frequencies of meta-analysis of four *IL2-IL21* SNPs located in Caucasian SSc specific autoantibody and healthy controls from Europe and the USA.

SNP	Alleles	Population	Genotype, N (%)			MAF (%)	Allele test		
			1/1	1/2	2/2		*P*-value[a]	*Pc*[b]	OR [CI 95%][c]
rs2069762	C/A	**Spain**							
		Controls (n=1716)	157 (9.15)	698 (40.69)	861 (50.15)	29.5			
		ACA+ (n=580)	56 (9.66)	223 (38.45)	301 (51.90)	28.88	0.69	NA	0.97 [0.84-1.13]
		ATA+ (n=279)	22 (7.89)	114 (40.86)	143 (51.25)	28.32	0.57	NA	0.94 [0.77-1.15]
		Germany							
		Controls (n=411)	26 (6.33)	164 (39.90)	221 (53.77)	26.28			
		ACA+ (n=234)	23 (9.83)	99 (42.31)	112 (47.86)	30.98	0.07	NA	1.26 [0.98-1.62]
		ATA+ (n=183)	17 (9.29)	68 (37.16)	98 (53.55)	27.87	0.57	NA	1.08 [0.82-1.43]
		The Netherlands							
		Controls (n=656)	56 (8.54)	266 (40.55)	334 (50.91)	28.81			
		ACA+ (n=86)	7 (8.14)	35 (40.70)	44 (51.16)	28.49	0.93	NA	0.98 [0.69-1.40]
		ATA+ (n=91)	9 (9.89)	36 (39.56)	46 (50.55)	29.67	0.81	NA	1.04 [0.74-1.46]
		USA							
		Controls (n=864)	84 (9.72)	358 (41.44)	422 (48.84)	30.44			
		ACA+ (n=285)	20 (7.02)	127 (44.56)	138 (48.42)	29.3	0.61	NA	0.95 [0.77-1.17]
		ATA+ (n=164)	13 (7.93)	70 (42.68)	81 (49.39)	29.27	0.67	NA	0.95 [0.73-1.23]
		Italy							
		Controls (n=1024)	131 (12.79)	453 (44.24)	440 (42.97)	34.91			
		ACA+ (n=288)	41 (14.24)	140 (48.61)	107 (37.15)	38.54	0.11	NA	1.17 [0.97-1.42]
		ATA+ (n=203)	27 (13.30)	91 (44.83)	85 (41.87)	35.71	0.76	NA	1.04 [0.83-1.29]
		Sweden							
		Controls (n=262)	20 (7.63)	97 (37.02)	145 (55.34)	26.15			
		ACA+ (n=57)	5 (8.77)	23 (40.35)	29 (50.88)	28.95	0.5397	NA	1.15 [0.73-1.80]
		ATA+ (n=35)	0 (0.00)	16 (45.71)	19 (54.29)	22.86	0.5546	NA	0.84 [0.46-1.51]
		UK							
		Controls (n=371)	31 (8.36)	145 (39.08)	195 (52.56)	27.9			
		ACA+ (n=152)	12 (7.89)	55 (36.18)	85 (55.92)	25.99	0.53	NA	0.91 [0.67-1.23]
		ATA+ (n=63)	5 (7.94)	28 (44.44)	30 (47.62)	30.16	0.60	NA	1.12 [0.74-1.69]
		Norway							
		Controls (n=277)	14 (5.05)	125 (45.13)	138 (49.82)	27.62			
		ACA+ (n=54)	6 (11.11)	28 (51.85)	20 (37.04)	37.04	0.05	NA	1.54 [1.00-2.38]
		ATA+ (n=13)	1 (7.69)	5 (38.46)	7 (53.85)	26.92	0.94	NA	0.97 [0.40-2.34]
		Combined							
		Controls (n=5482)	510 (9.30)	2266 (41.34)	2706 (49.36)	29.97			
		ACA+ (n=1736)	170 (9.79)	730 (42.05)	836 (48.16)	30.82	**0.25**	**NA**	**1.05 [0.97-1.14]**
		ATA+ (n=1031)	94 (9.12)	428 (41.51)	509 (49.37)	29.87	**0.98**	**NA**	**1.00 [0.90-1.11]**
		Spain							
		Controls (n=1721)	34 (1.98)	351 (20.40)	1336 (77.63)	12.17			
		ACA+ (n=590)	5 (0.85)	115 (19.49)	470 (79.66)	10.59	0.1461	NA	0.85 [0.69-1.06]

		ATA+ (n=286)	4 (1.40)	74 (25.87)	208 (72.73)	14.34	0.1473	NA	1.21 [0.94-1.56]
		Germany							
		Controls (n=423)	9 (2.13)	109 (25.77)	305 (72.10)	15.01			
		ACA+ (n=238)	5 (2.10)	55 (23.53)	177 (74.37)	13.87	0.57	NA	0.91 [0.66-1.26]
		ATA+ (n=188)	4 (2.13)	49 (26.06)	135 (71.81)	15.16	0.95	NA	1.01 [0.72-1.42]
		The Netherlands							
		Controls (n=734)	27 (3.68)	194 (26.43)	513 (69.89)	16.89			
		ACA+ (n=94)	6 (6.38)	24 (25.53)	64 (68.09)	19.15	0.44	NA	1.17 [0.79-1.72]
		ATA+ (n=101)	4 (3.96)	20 (19.80)	77 (76.24)	13.86	0.28	NA	0.79 [0.52-1.21]
		USA							
		Controls (n=884)	22 (2.49)	243 (27.49)	619 (70.02)	16.23			
		ACA+ (n=285)	6 (2.11)	77 (27.02)	202 (70.88)	15.61	0.73	NA	0.95 [0.74-1.24]
		ATA+ (n=165)	5 (3.03)	46 (27.88)	114 (69.09)	16.97	0.74	NA	1.06 [0.77-1.44]
rs6822844	**G/T**	**Italy**							
		Controls (n=1107)	22 (1.99)	269 (24.30)	816 (73.71)	14.14			
		ACA+ (n=289)	3 (1.04)	53 (18.34)	233 (80.62)	10.21	0.01	0.05	0.69 [0.51-0.93]
		ATA+ (n=216)	6 (2.78)	38 (17.59)	172 (79.63)	11.57	0.16	NA	0.80 [0.58-1.09]
		Sweden							
		Controls (n=273)	7 (2.56)	92 (33.70)	174 (63.74)	19.41			
		ACA+ (n=62)	6 (9.68)	18 (29.03)	38 (61.29)	24.19	0.23	NA	1.33 [0.83-2.10]
		ATA+ (n=2040)	2 (0.10)	11 (0.54)	2027 (99.36)	18.75	0.89	NA	0.96 [0.53-1.75]
		UK							
		Controls (n=372)	15 (4.03)	123 (33.06)	234 (62.90)	20.56			
		ACA+ (n=152)	5 (3.29)	40 (26.32)	107 (70.39)	16.45	0.13	NA	0.76 [0.54-1.08]
		ATA+ (n=64)	3 (4.69)	13 (20.31)	48 (75.00)	14.84	0.13	NA	0.67 [0.40-1.13]
		Norway							
		Controls (n=278)	13 (4.68)	94 (33.81)	171 (61.51)	21.58			
		ACA+ (n=53)	2 (3.77)	12 (22.64)	39 (73.58)	15.09	0.13	NA	0.65 [0.37-1.14]
		ATA+ (n=14)	1 (7.14)	6 (42.86)	7 (50.00)	28.57	0.38	NA	1.45 [0.62-3.38]
		Combined							
		Controls (n=5792)	149 (2.57)	1475 (25.47)	4168 (71.96)	15.31			
		ACA+ (n=1763)	38 (2.16)	395 (22.40)	1330 (75.44)	13.36	**0.015**	**0.06**	**0.87 [0.78-0.97]**
		ATA+ (n=1074)	29 (2.70)	257 (23.93)	788 (73.37)	14.66	**0.67**	**NA**	**0.97 [0.85-1.11]**
		Spain							
		Controls (n=1695)	206 (12.15)	721 (42.54)	768 (45.31)	33.42			
		ACA+ (n=589)	58 (9.85)	263 (44.65)	268 (45.50)	32.17	0.43	0.00	0.94 [0.82-1.09]
		ATA+ (n=283)	33 (11.66)	124 (43.82)	126 (44.52)	33.57	0.95	0.00	1.0 [0.83-1.22]
		Germany							
		Controls (n=426)	33 (7.75)	185 (43.43)	208 (48.83)	29.46			
		ACA+ (n=240)	23 (9.58)	87 (36.25)	130 (54.17)	27.71	0.50	NA	0.92 [0.72-1.18]
		ATA+ (n=190)	20 (10.53)	83 (43.68)	87 (45.79)	32.37	0.31	NA	1.15 [0.88-1.49]
		The Netherlands							
		Controls (n=717)	87 (12.13)	284 (39.61)	346 (48.26)	31.94			
		ACA+ (n=92)	15 (16.30)	40 (43.48)	37 (40.22)	38.04	0.10	NA	1.31 [0.95-1.80]
		ATA+ (n=97)	7 (7.22)	42 (43.30)	48 (49.48)	28.87	0.39	NA	0.86 [0.62-1.20]

		USA							
		Controls (n=884)	88 (9.95)	405 (45.81)	391 (44.23)	32.86			
		ACA+ (n=285)	27 (9.47)	139 (48.77)	119 (41.75)	33.86	0.66	NA	1.05 [0.86-1.28]
		ATA+ (n=165)	21 (12.73)	84 (50.91)	60 (36.36)	38.18	0.06	NA	1.26 [0.99-1.61]
rs6835457	**G/A**	**Italy**							
		Controls (n=1076)	143 (13.29)	479 (44.52)	454 (42.19)	35.55			
		ACA+ (n=288)	32 (11.11)	114 (39.58)	142 (49.31)	30.9	0.04	0.07	0.81 [0.67-0.99]
		ATA+ (n=211)	18 (8.53)	95 (45.02)	98 (46.45)	31.04	0.08	NA	0.82 [0.65-1.02]
		Sweden							
		Controls (n=271)	30 (11.07)	121 (44.65)	120 (44.28)	33.39			
		ACA+ (n=59)	7 (11.86)	26 (44.07)	26 (44.07)	33.9	0.92	NA	1.02 [0.67-1.56]
		ATA+ (n=40)	5 (12.50)	15 (37.50)	20 (50.00)	31.25	0.70	NA	0.91 [0.55-1.50]
		UK							
		Controls (n=374)	46 (12.30)	182 (48.66)	146 (39.04)	36.63			
		ACA+ (n=158)	20 (12.66)	67 (42.41)	71 (44.94)	33.86	0.39	NA	0.89 [0.67-1.17]
		ATA+ (n=64)	8 (12.50)	29 (45.31)	27 (42.19)	35.16	0.75	NA	0.94 [0.63-1.39]
		Norway							
		Controls (n=277)	35 (12.64)	130 (46.93)	112 (40.43)	36.1			
		ACA+ (n=54)	4 (7.41)	20 (37.04)	30 (55.56)	25.93	0.04	0.17	0.62 [0.39-0.99]
		ATA+ (n=14)	1 (7.14)	9 (64.29)	4 (28.57)	39.29	0.73	NA	1.15 [0.53-2.49]
		Combined							
		Controls (n=5720)	668 (11.68)	2507 (43.83)	2545 (44.49)	33.59			
		ACA+ (n=1765)	186 (10.54)	756 (42.83)	823 (46.63)	31.95	**0.12**	**NA**	**0.94 [0.86-1.02]**
		ATA+ (n=1064)	113 (10.62)	481 (45.21)	470 (44.17)	33.22	**0.99**	**NA**	**1.0 [0.90-1.10]**
		Spain							
		Controls (n=1661)	208 (12.52)	723 (43.53)	730 (43.95)	34.29			
		ACA+ (n=575)	57 (9.91)	257 (44.70)	261 (45.39)	32.26	0.21	NA	0.91 [0.79-1.05]
		ATA+ (n=278)	32 (11.51)	121 (43.53)	125 (44.96)	33.27	0.64	NA	0.96 [0.79-1.16]
		Germany							
		Controls (n=407)	32 (7.86)	176 (43.24)	199 (48.89)	29.48			
		ACA+ (n=237)	21 (8.86)	85 (35.86)	131 (55.27)	26.79	0.30	NA	0.88 [0.68-1.13]
		ATA+ (n=186)	18 (9.68)	83 (44.62)	85 (45.70)	31.99	0.38	NA	1.13 [0.86-1.47]
		The Netherlands							
		Controls (n=709)	85 (11.99)	283 (39.92)	341 (48.10)	31.95			
		ACA+ (n=86)	13 (15.12)	39 (45.35)	34 (39.53)	37.79	0.12	NA	1.29 [0.93-1.80]
		ATA+ (n=98)	8 (8.16)	39 (39.80)	51 (52.04)	28.06	0.27	NA	0.83 [0.60-1.16]
		USA							
		Controls (n=882)	88 (9.98)	401 (45.46)	393 (44.56)	32.71			
		ACA+ (n=285)	27 (9.47)	139 (48.77)	119 (41.75)	33.86	0.61	NA	1.05 [0.86-1.29]
		ATA+ (n=164)	21 (12.80)	83 (50.61)	60 (36.59)	38.11	0.06	NA	1.27 [0.99-1.62]
rs907715	**T/C**	**Italy**							
		Controls (n=1063)	146 (13.73)	476 (44.78)	441 (41.49)	36.12			
		ACA+ (n=290)	29 (10.00)	121 (41.72)	140 (48.28)	30.86	0.019	0.07	0.79 [0.65-0.96]
		ATA+ (n=211)	14 (6.64)	100 (47.39)	97 (45.97)	30.33	0.02	0.09	0.77 [0.61-0.96]
		Sweden							

Controls (n=270)	29 (10.74)	119 (44.07)	122 (45.19)	32.78			
ACA+ (n=59)	6 (10.17)	27 (45.76)	26 (44.07)	33.05	0.95	NA	1.01 [0.66-1.55]
ATA+ (n=36)	5 (13.89)	11 (30.56)	20 (55.56)	29.17	0.54	NA	0.84 [0.49-1.45]
UK							
Controls (n=374)	47 (12.57)	183 (48.93)	144 (38.50)	37.03			
ACA+ (n=159)	23 (14.47)	66 (41.51)	70 (44.03)	35.22	0.57	NA	0.92 [0.70-1.22]
ATA+ (n=69)	10 (14.49)	29 (42.03)	30 (43.48)	35.51	0.73	NA	0.94 [0.64-1.37]
Norway							
Controls (n=278)	35 (12.59)	130 (46.76)	113 (40.65)	35.97			
ACA+ (n=53)	4 (7.55)	20 (37.74)	29 (54.72)	26.42	0.06	NA	0.64 [0.40-1.02]
ATA+ (n=14)	1 (7.14)	9 (64.29)	4 (28.57)	39.29	0.72	NA	1.15 [0.53-2.51]
Combined							
Controls (n=5644)	670 (11.87)	2491 (44.14)	2483 (43.99)	33.94			
ACA+ (n=1744)	180 (10.32)	754 (43.23)	810 (46.44)	31.94	**0.05**	**NA**	**0.92 [0.85-1.00]**
ATA+ (n=1056)	109 (10.32)	475 (44.98)	472 (44.70)	32.81	**0.48**	**NA**	**0.96 [0.87-1.07]**

[a]All P-values have been calculated for the allelic model. [b]If it is applicable, Bonferroni correction is shown. [c]Odds ratio for the minor allele.

Analysis of the influence of two CD24 genetic variants in Crohn's disease and ulcerative colitis. Human Immunology, 2011

Human Immunology 72 (2011) 969-972

Contents lists available at SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/humimm

Analysis of the influence of two *CD24* genetic variants in Crohn's disease and ulcerative colitis

Lina-Marcela Diaz-Gallo [a,*], Luz María Medrano [b], María Gómez-García [c], Carlos Cardeña [d], Luis Rodrigo [e], Juan Luis Mendoza [f], Carlos Taxonera [f], Antonio Nieto [g], Guillermo Alcain [h], Ignacio Cueto [h], Miguel A. López-Nevot [i], Elena Urcelay [b], Javier Martin [a]

[a] Instituto de Parasitología y Biomedicina "López-Neyra," Consejo Superior de Investigaciones Científicas (CSIC), Granada, Spain
[b] Department of Clinical Immunology, Hospital Clínico S. Carlos, Madrid, Spain
[c] Department of Gastroenterology, Hospital Virgen de las Nieves, Granada, Spain
[d] Department of Gastroenterology, Hospital Clínico San Cecilio, Granada, Spain
[e] Department of Gastroenterology, Hospital Central de, Asturias, Oviedo, Spain
[f] Department of Gastroenterology, Hospital Universitario S. Carlos, Madrid, Spain
[g] Department of Immunology, Hospital Puerta del Mar, Cádiz, Spain
[h] Department of Gastroenterology, Hospital Virgen de la Victoria, Málaga, Spain
[i] Department of Immunology, Hospital Virgen de las Nieves, Granada, Spain

ARTICLE INFO

Article history:
Received 19 January 2011
Accepted 25 May 2011
Available online 17 June 2011

Keywords:
CD24 gene
Crohn's disease
Ulcerative colitis

ABSTRACT

The aim of this study was to evaluate the possible implication of *CD24* gene in the genetic predisposition to inflammatory bowel disease (IBD). Our study population consisted of 1321 female Spanish individuals (369 Crohn's disease [CD] patients, 323 ulcerative colitis [UC] patients, and 629 healthy matched controls). Two putative functional polymorphisms, a C to T coding polymorphism (rs8734) and a TG deletion in the 3′ untranslated region (rs3838646), were used as *CD24* genetic markers and genotyped using a Taqman 5′ allelic discrimination assay. The "del" allele of the dinucleotide deletion was associated with an increased risk of CD (odds ratio = 1.61, 95% confidence interval = 1.17–2.21, p_{FDR} = 6.4E-03) but not with UC. Moreover, this allele was significant associated with the age of CD diagnosis between 17 and 40 years, the ileocolonic location, and the inflammatory behavior of CD. We observed no significant differences between the allelic or genotypic frequencies of the A57V polymorphism in our studied IBD cohort. Our results suggest that the rs3838646 *CD24* polymorphism is part of the genetic background of CD.

1. Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are the main clinical phenotypes of inflammatory bowel disease (IBD). Both are relapsing and chronic inflammatory disorders that result from the complex interaction of genetic, immune, and environmental factors. There is an increasingly long list of genetic factors associated with IBD. However, it is estimated that the current number of loci associated with IBD represents only a small fraction of the genetic risk [1–3]. Thus, additional genetic contributions clearly remain to be discovered.

The *CD24* gene encodes a small (ranging between 20 and 70 amino acids), heavily glycosylated protein that is attached to the cell membrane by a glyco-phosphotidyl-inistol [GPI] anchor and is expressed in a broad range of cell types [4]. *CD24*, referred to as a B-cell differentiation marker at first, plays crucial roles in lymphocyte maturation [5,6,7], neuronal development [8], and tissue renewal homeostasis under physiologic conditions [9]. This molecule has been proposed as a genetic checkpoint in T-cell homoeostasis and pathogenesis of autoimmune diseases [10]. Moreover, the *CD24* gene has been recently added to the list of putative intestinal stem cell markers [8,11] and it has been demonstrated that is up-regulated in regenerative mucosa in both UC and CD [12]. By contrast, two functional genetic polymorphisms, a C to T coding polymorphism (rs8734) [13] and a TG deletion in the 3′ untranslated region (rs3838646) [14], have been associated as susceptibility factors for multiple sclerosis (MS), rheumatoid arthritis (RA), systemic lupus erythematosus (SLE), and giant cell arthritis (GCA) [13–19]. The aim of this study was to investigate the potential implication of the rs8734 and rs3838646 *CD24* gene functional variants in the genetic susceptibility to IBD.

2. Subjects and methods

Our study included 692 IBD female Spanish patients (371 with CD and 323 with UC) meeting the standard clinical, endoscopic,

* Corresponding author.
E-mail address: lina.diaz@ipb.csic.es (L.-M. Diaz-Gallo).
E.U. and J.M. contribute equally to this work.

0198-8859/11/$32.00 - see front matter © 2011 American Society for Histocompatibility and Immunogenetics. Published by Elsevier Inc. All rights reserved.
doi:10.1016/j.humimm.2011.05.028

Table 1
Allele and genotype frequencies of the rs3838646 *CD24* variant in Crohn's disease female patients and healthy controls

	CD24 genotype (%)			Allele frequency (%)		*p* Value (allele)	OR (95% CI)
	TG/TG	TG/del	del/del	TG	del		
Controls [629]	547 (87)	76 (12.1)	6 (1)	1170 (93)	88 (7)		
Crohn's disease [371]	301 (81.1)	60 (16.2)	10 (2.7)	662 (89.2)	80 (10.8)	**6.4E-03**[a]	1.61 (1.17–2.21)
Diagnosis age A [268]							
<16 (A1)	17 (73.9)	5 (21.7)	1 (4.3)	39 (84.7)	7 (15.2)	0.04	2.39 (1.04–5.49)
17–40 (A2)	144 (75)	40 (20.8)	8 (4.2)	328 (85.4)	56 (14.6)	**6.62E-05**[b]	2.27 (1.59–3.24)
>40 (A3)	48 (90.6)	5 (9.4)	0(0)	101 (95.3)	5 (4.7)	0.37	0.66 (0.26–1.66)
Disease location L [263]							
Ileal (L1)	86 (80.4)	17 (15.9)	4 (3.7)	189 (88.3)	25 (11.7)	0.017	1.75 (1.1–2.81)
Colonic (L2)	46 (88.5)	5 (9.6)	1 (1.9)	97 (93.3)	7 (6.7)	0.92	0.96 (0.43–2.13)
Ileocolonic (L3)	76 (73.1)	24 (23.1)	4 (3.8)	176 (84.6)	32 (15.4)	**4.78E-04**[c]	2.42 (1.57–3.73)
Disease behavior B [264]							
Perforating (B3)	54 (80.6)	11 (16.4)	2 (3)	119 (88.8)	15 (11.2)	0.08	1.68 (0.94–2.99)
Structuring (B2)	46 (80.7)	9 (15.8)	2 (3.5)	101 (88.6)	13 (11.4)	0.08	1.71 (0.92–3.17)
Inflammatory (B1)	107 (76.4)	28 (20)	5 (3.6)	242 (86.4)	38 (13.6)	**3.13E-03**[d]	2.1 (1.39–3.13)

[a] *p* Value corrected by FDR correction.
[b] *p* Value corected by Bonferroni correction. When A2 was analyzed againsnt A1 + A3, *p* value = 0.016; OR = 2.1 95% CI = 1.14–3.88.
[c] *p* Value corrected by Bonferroni correction. When L3 was analyzed against L1+L2. p = 0.046, OR = 1.66, 95% CI = 1.0–2.72.
[d] Value corrected by Bonferroni correction. When B1 was analyzed against B2+B3, *p* = 0.29, OR = 1.31, 95% CI = 0.8–2.13.

radiologic, and histopathologic criteria [20], and 629 matched female healthy control subjects. Demographic and clinical characteristics of the subjects have been previously described [21]. Written informed consent was obtained from all participants. The ethics committees of all participating Spanish hospitals approved the study.

DNA was obtained from peripheral blood using standard methods. Samples were genotyped for the *CD24* rs8734 and rs3838646 genetic variants using Taqman 5′ allelic discrimination assays as previously described [16]. The primers used in this study showed a 99% alignment identity in both the *CD24* gene (chromosome 6, RefSeq. NM_013230) and a pseudogene located at chromosome Y (RefSeq. NG_006012)[22]. Hence, only females were included in the study to avoid nonspecific amplifications. Deviation from Hardy–Weinberg equilibrium (HWE) was tested by standard χ^2 analysis. The differences in genotype distribution and allele frequency among cases and controls were calculated by contingency tables and when necessary by Fisher's exact test. An association was considered statistically significant if $p < 0.05$. Linkage disequilibrium (LD) measurements (r^2) between rs8734 and rs3838646 were estimated by the expectation-maximization algorithm. To test whether the *CD24* gene studied polymorphisms are associated with clinical features, a univariate analysis using χ^2 or the Fisher's exact test was applied. The Montreal classification criteria were used to determine clinical variables [23]. Each clinical variable was compared with the others and with those of healthy controls. Multiple testing was corrected by false discovery rate control (p_{FDR}) or Bonferroni correction [24]. The analyses were performed using PLINK (version 1.07) to estimate odds ratios (OR) and 95% confidence intervals (95% CI) [25].

3. Results

The distributions of genotypic and allelic frequencies of the two *CD24* evaluated polymorphisms (rs8734 and rs3838646) were in HWE. No LD was observed between the SNPs (r^2 = 0.03). We observed that the "del" allele of the rs3838646 variant was significantly associated with CD (OR = 1.61, 95% CI = 1.2–2.2, p_{FDR} = 6.4E-03; Table 1) but not with UC (OR = 0.92, 95% CI = 0.6–1.4, p = 0.66; Table 2). Although we observed no significant differences in the "del/del" genotype frequencies of rs3838646 between IBD patients compared to healthy controls, we found an increase in CD patients (2.7%) compared with controls (1%), in contrast to the absence of this genotype in the UC patients.

No significant differences in the allelic frequencies of rs8734 between both CD or UC patients and controls were observed (Tables 3 and 4). We observed an increased frequency of the genotype VV of this polymorphism in both CD (8.7%; Table 3) and UC (10.9%; Table 4) patients compared with healthy controls (6.5%), but those differences were not statistically significant after Bonferroni correction (p = 0.42 in CD and p = 0.06 in UC patients).

We conducted a univariate analysis to evaluate the possible association between both *CD24* polymorphisms and the clinical manifestations of the diseases. We found that the "del" allele of rs3838646 is associated with an onset of CD between the 17 and 40 years of age (OR = 2.27, 95% CI = 1.6–3.2, p corrected = 6.6E-5), an ileocolonic location of the disease (OR = 2.42, 95% CI = 1.6–3.7, p corrected = 4.8E-4), and an inflammatory behavior of CD (OR = 2.11, 95% CI = 1.4–3.1, p corrected = 3.1E-3) (Table 1) when the clinical manifestations were compared to healthy controls. Only the comparison of the "del" allele frequency between ileocolonic

Table 2
Allele and genotype frequencies of the rs3838646 *CD24* variant in Ulcerative Colitis female patients and healthy controls

	CD24 genotype (%)			Allele frequency (%)		*p* Value (minor allele)	OR (95% CI)
	TG/TG	TG/del	del/del	TG	del		
Controls [629]	547 (87)	76 (12.1)	6 (1.0)	1170 (93)	88 (7.0)		
Ulcerative colitis [310]	270 (87.1)	40 (12.9)	0 (0)	580 (93.5)	40 (6.5)	0.66	0.92 (0.62–1.35)
Diagnosis age A **[122]**				()			
<16 (A1)	7 (100)	0 (0)	0 (0)	14 (100)	0 (0)	–	–
17–40 (A2)	66 (90.4)	7 (9.6)	0 (0)	139 (95.2)	7 (4.8)	0.32	0.67 (0.3–1.48)
>40 (A3)	35 (83.3)	7 (16.7)	0 (0)	77 (91.7)	7 (8.3)	0.64	1.21 (0.54–2.7)
Disease extension E [223]			(0)				
Ulcerative proctitis (E1)	23 (88.5)	3 (11.5)	0 (0)	49 (94.2)	3 (5.8)	0.73	0.81 (0.25–2.66)
Left-side UC (E2)	101 (87)	15 (12.9)	0 (0)	217 (93.5)	15 (6.5)	0.77	0.92 (0.52–1.62)
Extensive UC (E3)	66 (81.5)	15 (18.5)	0 (0)	147 (90.7)	15 (9.3)	0.3	1.36 (0.76–2.41)

Table 3
Allele and genotype frequencies of the rs8734 *CD24* variant in female patients with Crohn's disease and in healthy controls

	CD24 genotype (%)			Allele frequency (%)		*p* Value (minor allele)	OR (95% CI)
	AA	AV	VV	A	V		
Controls [628]	317 (50.4)	270 (43)	41 (6.5)	904 (71.9)	352 (28)		
Crohn's disease [366]	200 (54.6)	134 (36.6)	32 (8.7)	662 (89.2)	80 (10.8)	0.64	0.95 (0.78–1.17)
Diagnosis age A [270]							
<16 (A1)	13 (54.1)	9 (37.5)	2 (8.3)	35 (72.9)	13 (27.1)	0.86	0.95 (0.49–1.82)
17–40 (A2)	111 (57.2)	72 (37.1)	11 (5.7)	294 (75.7)	94 (24.2)	0.14	0.82 (0.63–1.07)
>40 (A3)	23 (44.2)	21 (40.4)	8 (15.4)	67 (64.4)	37 (35.6)	0.1	1.42 (0.93–2.16)
Disease location L [262]							
Ileal (L1)	57 (51.8)	40 (36.4)	13 (11.8)	154 (70)	66 (30)	0.55	1.1 (0.8–1.51)
Colonic (L2)	27 (51.9)	23 (44.2)	2 (3.8)	77 (74)	27 (26)	0.92	0.96 (0.43–2113)
Ileocolonic (L3)	59 (59)	31 (31)	10 (10)	149 (74.5)	51 (25.5)	0.46	0.88 (0.63–1.24)
Disease behavior B [262]							
Perforating (B3)	31 (49.2)	24 (38.1)	8 (12)	86 (68.2)	40 (31.7)	0.38	1.2 (0.8–1.78)
Structuring (B2)	29 (50)	24 (41.4)	5 (8.6)	82 (70.6)	34 (29.3)	0.77	1.1 (0.7–1.62)
Inflammatory (B1)	83 (58.9)	46 (32.6)	12 (8.5)	212 (75.1)	70 (24.8)	0.28	0.85 (0.63–1.14)

location and the others locations showed a trend of association (OR = 1.7, 95% CI = 1.0–2.7, p = 0.05, Table 1). No significant associations were observed between rs3838646 and UC clinical variables (Table 2) or between rs8734 and CD or UC clinical manifestations (Tables 3 and 4).

4. Discussion

We evaluated for the first time the influence of two previously associated autoimmunity polymorphisms (rs8734 and rs3838646) of the *CD24* gene in IBD. Our study showed evidence of association of the dinucleotide deletion in the *CD24* 3′-UTR (rs3838646) with CD but not with UC. Moreover, our results suggest that rs3838646 increased the risk to have an ileocolonic location of CD. These observations support the hypothesis that the main IBD diseases (CD and UC) share a related nature but differ in some immunologic mechanisms [21,26]. The rs3838646 was first associated as a reduced risk factor for MS and SLE [14]. After that, it was reported as a risk factor for GCA [16]. In the present study we observed that is also a risk factor to CD. Those results indicate that the rs3838646 ($P1527^{del}$) polymorphism could be a common genetic factor in autoimmunity. The opposite effect size in different diseases has been reported for other genetic factors in autoimmunity [27,28]. For example, the *PTPN22* 620W allele, a well-described risk factor for several autoimmune diseases, meanwhile is a reduced risk factor for CD [29,30]. This remarks that the same variants may influence different diseases at different levels within the complex genetic component of autoimmunity.

A recent meta-analysis of six CD genome wide association studies (GWAS) showed a signal of association (p = 4.37E-9) in the 6q21 region [31]. This suggested that the rs3838646 is involved in the genetic background of CD because the *CD24* gene is located in this region (GeneID 100133941). Moreover, two recent reports provide insights about the implication of *CD24* in the genetic background of IBD. Ahmed et al. demonstrated that *CD24* is up-regulated in regenerated epithelium in IBD, playing a role in tissue healing. The authors showed that *CD24* is an enhancer of colony formation, cell migration and cell invasion [12] By contrast, Wang et al. had previously reported that the dinucleotide deletion (rs3838646) reduced steady levels of *CD24* mRNA by more than twofold [14]. In this regard, we observed that the "del" rs3838646 allele increased the risk of an onset of CD between 17 and 40 years of age, a more extended location of the affected tissue (ileocolonic location), and an inflammatory behavior of CD (Table 1). Probably an early onset of CD involved a higher genetic component in the pathogenesis of CD than in a late onset. Indeed, a GWAS performed with early-onset IBD patients showed that there are more specific loci associated with this clinical category [32]. *CD24* seems to be an important check point in the appropriated healing mechanism on CD, and the "del" rs3838646 allele affects the expression of the gene. This suggest that this genetic variant could be directly involved in a more extended location of the tissue affected and in the development of a chronic inflammation because of a lower expression of *CD24*. Together with our findings, those results suggest that the "del" allele of the rs3838646 could alter the stability of the *CD24* mRNA driving the distorted healing mechanism in CD patients.

In summary, our study suggests that the *CD24* rs3838464 genetic variant is a risk factor for CD specifically for age of diagnosis between 17 and 40 years and the ileocolonic location. Further genetic association studies in other populations are warranted to confirm our observations, as well as more functional studies to elucidate more completely the role of the *CD24* gene in IBD.

Acknowledgments

We thank Sofia Vargas, Gema Robledo, and Sonia Garcia for their excellent technical assistance. We thank all donors, patients, and controls. We thank Banco Nacional de ADN (University of Sala-

Table 4
Allele and genotype frequencies of rs8734 *CD24* variant in Spanish ulcerative colitis (UC) patients and healthy controls

	CD24 genotype (%)			Allele frequency (%)		*p* Value (minor allele)	OR (95% CI)
	AA	AV	VV	A	V		
Controls [628]	317 (50.5)	270 (43)	41 (6.5)	904 (72)	352 (28)		
Ulcerative colitis [322]	161 (50)	126 (39.1)	35 (10.9)	448 (69.6)	196 (30.4)	0.27	1.12 (0.91–1.38)
Diagnosis age A [132]							
<16 (A1)	3 (33.3)	6 (66.7)	0 (0)	12 (66.7)	6 (33.3)	0.62	1.28 (0.48–3.45)
17–40 (A2)	35 (44.9)	35 (44.9)	8 (10.3)	105 (67.3)	51 (32.7)	0.22	1.25 (0.87–1.78)
>40 (A3)	28 (62.2)	15 (33.3)	2 (4.4)	71 (78.9)	19 (21.1)	0.16	0.69 (0.41–1.16)
Disease extension E [234]							
Ulcerative proctitis (E1)	10 (35.7)	18 (64.3)	0 (0)	38 (67.9)	18 (32.1)	0.5	1.22 (0.69–2.16)
Left-side UC (E2)	65 (53.2)	42 (34.4)	15 (12.3)	172 (70.5)	72 (29.5)	0.64	1.08 (0.8–1.45)
Extensive UC (E3)	37 (44)	40 (47.6)	7 (8.3)	114 (67.9)	54 (32.1)	0.27	1.22 (0.76–2.41)

manca, Spain) who supplied part of the control DNA samples. This work was supported by Plan Nacional de I+D, grant SAF06-00398, FIS08/1676) and Junta de Andalucía, CTS-180 group. L.-M.D.-G. was supported by the Ayudas Predoctorales de Formación en Investigación en Salud (PFIS—FI09/00544) from the Instituto de Salud Carlos III.

References

[1] Budarf ML, Labbé C, David G, Rioux JD. GWA studies: Rewriting the story of IBD. Trends Genet 2009;25:137–46.
[2] Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. Annu Rev Immunol 2010;28:573–621.
[3] Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. Finding the missing heritability of complex diseases. Nature 2009;461:747–53.
[4] Kristiansen G, Sammar M, Altevogt P. Tumour biological aspects of CD24, a mucin-like adhesion molecule. J Mol Histol 2004;35:255–62.
[5] Carsetti R, Rosado MM, Wardmann H. Peripheral development of B cells in mouse and man. Immunol Rev 2004;197:179–91.
[6] Hubbe M, Altevogt P. Heat-stable antigen/CD24 on mouse T lymphocytes: Evidence for a costimulatory function. Eur J Immunol 1994;24:731–7.
[7] Liu Y, Jones B, Aruffo A, Sullivan KM, Linsley PS, Janeway CA. Heat-stable antigen is a costimulatory molecule for CD4 T cell growth. J Exp Med 1992;175:437–45.
[8] Pruszak J, Ludwig W, Blak A, Alavian K, Isacson O. CD15, CD24, and CD29 define a surface biomarker code for neural lineage differentiation of stem cells. Stem Cells 2009;27:2928–40.
[9] Nieoullon V, Belvindrah R, Rougon G, Chazal G. Mouse CD24 is required for homeostatic cell renewal. Cell Tissue Res 2007;329:457–67.
[10] Liu Y, Zheng P. CD24: A genetic checkpoint in T cell homeostasis and autoimmune diseases. Trends Immunol 2007;28:315–20.
[11] Vermeulen L, Todaro M, de Sousa Mello F, Sprick MR, Kemper K, et al. Single-cell cloning of colon cancer stem cells reveals a multi-lineage differentiation capacity. Proc Natl Acad Sci U S A 2008;105:13427–32.
[12] Ahmed MA, Jackson D, Seth R, Robins A, Lobo DN, et al. CD24 is upregulated in inflammatory bowel disease and stimulates cell motility and colony formation. Inflamm Bowel Dis 2010;16:795–803.
[13] Zhou Q, Rammohan K, Lin S, Robinson N, Li O, et al. CD24 is a genetic modifier for risk and progression of multiple sclerosis. Proc Natl Acad Sci U S A 2003;100:15041–6.
[14] Wang L, Lin S, Rammohan KW, Liu Z, Liu JQ, et al. A dinucleotide deletion in CD24 confers protection against autoimmune diseases. PLoS Genet 2007;3:e49.
[15] Otaegui D, Sáenz A, Camaño P, Blázquez L, Goicoechea M, et al. CD24 V/V is an allele associated with the risk of developing multiple sclerosis in the Spanish population. Mult Scler 2006;12:511–4.
[16] Rueda B, Miranda-Filloy JA, Martin J, Gonzalez-Gay MA. Association of CD24 gene polymorphisms with susceptibility to biopsy-proven giant cell arteritis. J Rheumatol 2008;35:850–4.
[17] Sánchez E, Abelson AK, Sabio JM, González-Gay MA, Ortego-Centeno N, et al. Association of a CD24 gene polymorphism with susceptibility to systemic lupus erythematosus. Arthritis Rheum 2007;56:3080–6.
[18] Sánchez E, Fernández-Gutierrez B, González-Gay MA, Balsa A, García A, et al. Investigating the role of CD24 gene polymorphisms in rheumatoid arthritis. Ann Rheum Dis 2008;67:1197–8.
[19] Goris A, Maranian M, Walton A, Yeo TW, Ban M, et al. CD24 Ala/Val polymorphism and multiple sclerosis. J Neuroimmunol 2006;175:200–2.
[20] Silverberg MS, Satsangi J, Ahmad T, Arnott ID, Bernstein CN, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology. Can J Gastroenterol 2005;19(Suppl A):5–36.
[21] Díaz-Gallo LM, Palomino-Morales RJ, Gómez-García M, Cardeña C, Rodrigo L, et al. STAT4 gene influences genetic predisposition to ulcerative colitis but not Crohn's disease in the Spanish population: A replication study. Hum Immunol 2010;71:515–9.
[22] Hough MR, Rosten PM, Sexton TL, Kay R, Humphries RK. Mapping of CD24 and homologous sequences to multiple chromosomal loci. Genomics 1994;22:154–61.
[23] Gasche C, Scholmerich J, Brynskov J, D'Haens G, Hanauer SB, et al. A simple classification of Crohn's disease: Report of the Working Party for the World Congresses of gastroenterology, Vienna 1998. Inflamm Bowel Dis 2000;6:8–15.
[24] Benjamini Y, Yekutieli D. Quantitative trait loci analysis using the false discovery rate. Genetics 2005;171:783–90.
[25] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–75.
[26] Wang K, Baldassano R, Zhang H, Qu HQ, Imielinski M, et al. Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects. Hum Mol Genet 2010;15:2059–67.
[27] Gregersen PK, Olsson LM. Recent advances in the genetics of autoimmune disease. Annu Rev Immunol 2009;27:363–91.
[28] 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.
[29] Díaz-Gallo LM, Espino-Paisán L, Fransen K, Gómez-García M, van Sommeren S, et al. Differential association of two PTPN22 coding variants with Crohn's disease and Ulcerative Colitis. Inflamm Bowel Dis 2011 [Epub ahead of print].
[30] Stanford SM, Mustelin TM, Bottini N. Lymphoid tyrosine phosphatase and autoimmunity: Human genetics rediscovers tyrosine phosphatases. Semin Immunopathol 2011;32:127–36.
[31] Franke A, McGovern DP, Barrett JC, Wang K, Radford-Smith GL, et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet 2010;42:1118–25.
[32] Imielinski M, Baldassano RN, Griffiths A, Russell RK, Annese V, et al. Common variants at five new loci associated with early-onset inflammatory bowel disease. Nat Genet 2009;41:1335–40.

Association study of BAK1 gene polymorphisms in Spanish rheumatoid arthritis and systemic lupus erythematosus cohorts. Annals of Rheumatic Diseases, 2012

Letter

Association study of *BAK1* gene polymorphisms in Spanish rheumatoid arthritis and systemic lupus erythematosus cohorts

Cumulative evidence indicates that the BCL2-antagonist/killer 1 (*BAK1*) gene could be involved in several autoimmune diseases.[1 2] The proapoptotic BCL2 family members BAX and BAK are essential for regulating the number of B and T cells.[2–4] A recent study in a Colombian population suggested that the *BAK1* rs513349 and rs5745582 single nucleotide polymorphisms (SNPs) as well as the haplotype rs513349G-rs561276C-rs5745582A are significantly associated with autoimmune rheumatic diseases.[1] The aim of the present study was to evaluate the influence of these *BAK1* polymorphisms in Spanish rheumatoid arthritis (RA) and systemic lupus erythematosus (SLE) cohorts.

All the RA and SLE patients included in this case–control study met the respective American College of Rheumatology criteria, as previously described.[5 6] The study was approved by the local ethical committees from all the participating centres, and written informed consent was obtained from both patients and controls. Samples were genotyped for the *BAK1* polymorphisms via TaqMan 5′ allelic discrimination technology using predesigned SNP genotyping assays provided by Applied Biosystems (Foster City, California, USA). The genotypic frequencies of the *BAK1* SNPs were in Hardy–Weinberg equilibrium in both case and control groups ($p>0.01$, table 1). Due to the reported *BAK1* susceptibility haplotype and the related location of the *BAK1* gene with human leukocyte antigen (HLA) class II loci, it is important to determine the degree of linkage disequilibrium (LD) between the *BAK1* polymorphisms and the *HLA-DRB1* and *HLA-DQB1* loci. The LD between the three studied *BAK1* variants was moderate/weak ($r^2<0.27$), and weak with the *HLA* loci (data not shown). The fixation index was 0.0006 for whole population using the three studied SNPs of BAK1, suggesting that there is no significant population substructure in the studied population or between case and controls. No significant differences in the genotypic or allelic frequencies of *BAK1* polymorphisms were observed between patients or female patients with RA or SLE and controls or female controls (table 1). Similarly, no statistical significance was reached for the allelic combinations with a higher frequency than 5% between RA/SLE patients and controls (table 2).

The possibility that the lack of association might have been due to type II error seems highly unlikely since our cohort had enough statistical power (93% and 97% for rs5745582 in RA and SLE, respectively, and 99% for rs513349 for both diseases) to detect the previously reported effects.[1] The contradictory findings are probably a consequence of the genetic heterogeneity between populations. The allelic frequencies between our Spanish cohort and that reported for the Colombian population did not differ significantly in controls, but there was an important deviation in patients. The difference in the allelic frequencies of the cases implies that the tag of the causal variant probably varies among Latino and European populations. Indeed, previous studies have suggested that the extent of LD in both populations could differ, resulting in different sets of haplotype frequencies.[7] Data from a genome-wide association study in a European-ancestry population showed that rs513349 is not associated with RA ($p=0.17$, OR 1.08, 95% CI 0.96 to 1.23).[8] Moreover, a recent case–control study in a Spanish SLE Caucasian cohort that genotyped 6045 SNPs within the major

Table 1 Genotype and allelic frequencies for the studied variants of *BAK1* gene and RA, SLE patients and controls

SNP	AA	AG	GG	A	G	p Value	OR	95% CI
rs513349								
Controls (829)	0.249	0.538	0.212	0.519	0.481			
RA (484)	0.254	0.525	0.221	0.517	0.483	0.91	1	0.86 to 1.18
SLE (633)	0.270	0.520	0.210	0.530	0.469	0.54	0.95	0.83 to 1.11
Female controls (550)	0.251	0.538	0.211	0.520	0.480			
Female RA (309)	0.269	0.518	0.214	0.528	0.473	0.76	0.97	0.79 to 1.18
Female SLE (561)	0.260	0.522	0.218	0.521	0.479	0.95	0.99	0.84 to 1.18
SNP	**CC**	**CT**	**TT**	**C**	**T**	**p Value**	**OR**	**95% CI**
rs561276								
Controls (829)	0.960	0.040	0	0.980	0.020			
RA (484)	0.975	0.025	0	0.988	0.012	0.15	0.61	0.32 to 1.2
SLE (633)	0.957	0.041	0.002	0.978	0.022	0.68	1.11	0.67 to 1.85
Female controls (550)	0.960	0.040	0	0.980	0.020			
Female RA (309)	0.964	0.036	0	0.982	0.018	0.75	0.89	0.43 to 1.84
Female SLE (561)	0.957	0.410	0.002	0.978	0.023	0.71	1.12	0.84 to 1.99
SNP	**GG**	**GA**	**AA**	**G**	**A**	**p Value**	**OR**	**95% CI**
rs5745582								
Controls (829)	0.624	0.341	0.034	0.796	0.204			
RA (484)	0.636	0.318	0.045	0.795	0.205	0.99	1	0.86 to 1.18
SLE (633)	0.607	0.338	0.055	0.776	0.224	0.19	1.13	0.94 to 1.34
Female controls (550)	0.627	0.340	0.033	0.797	0.203			
Female RA (309)	0.641	0.320	0.039	0.801	0.199	0.85	0.98	0.76 to 1.25
Female SLE (561)	0.601	0.339	0.061	0.770	0.230	0.12	1.17	0.96 to 1.44

BAK1, BCL2-antagonist/killer 1; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus; SNP, single nucleotide polymorphism.

Table 2 Counts and frequencies of allelic combinations of the *BAK1* variants studied

Allelic combination	Controls	RA				SLE			
rs513349, rs561276, rs5745582	2n=1656 (%)	2n=968 (%)	p Value	OR	95% CI	2n=1266 (%)	p Value	OR	95% CI
ACG	859 (51.87)	500 (51.65)		Reference		671 (53.00)		Reference	
GCG	427 (25.79)	258 (26.65)	0.70	1.04	0.85 to 1.26	283 (22.35)	0.08	0.85	0.71 to 1.02
GCA	337 (20.35)	198 (20.45)	0.93	1.01	0.82 to 1.25	284 (22.43)	0.43	1.08	0.89 to 1.31

BAK1, BCL2-antagonist/killer 1; RA, rheumatoid arthritis; SLE, systemic lupus erythematosus.

histocompatibility complex region showed that rs513349 was not significantly associated with SLE (p=0.74, OR 1.03, 95% CI 0.86 to 1.24).[9] Taking into account the above together with our results, it is likely that the *BAK1* polymorphisms studied are not susceptibility loci for RA and SLE in Caucasians.

Lina-Marcela Díaz-Gallo,[1] Sonia Garcia,[1] Norberto Ortego-Centeno,[2] Juan Jiménez-Alonso,[3] Julio Sánchez-Román,[4] Enrique de Ramón,[5] María F González-Escribano,[6] Alejandro Balsa,[7] Benjamín Fernández-Gutierrez,[8] Isidoro González-Alvaro,[9] Miguel A González-Gay,[10] Javier Martin[1]

[1]Cellular Biology and Immunology Department, Instituto de Parasitología y Biomedicina 'López-Neyra', CSIC (IPBLN-CSIC), Granada, Spain
[2]Servicio de Medicina Interna, Hospital Clínico San Cecilio, Granada, Spain
[3]Servicio de Medicina Interna, Hospital Virgen de las Nieves, Granada, Spain
[4]Servicio de Medicina Interna, Hospital Virgen del Rocío, Sevilla, Spain
[5]Servicio de Medicina Interna, Hospital Carlos Haya, Málaga, Spain
[6]Servicio de Inmunología, Hospital Virgen del Rocío, Sevilla, Spain
[7]Servicio de Reumatología, Hospital Universitario La Paz, Madrid, Spain
[8]Servicio de Reumatología, Hospital Clínico San Carlos, Madrid, Spain
[9]Servicio de Reumatología, Hospital Universitario La Princesa, IIS-Princesa, Madrid, Spain
[10]Servicio de Reumatología, IFIMAV, Hospital Universitario Marqués de Valdecilla, Santander, Spain

Correspondence to Lina-Marcela Díaz-Gallo, Cellular Biology and Immunology Department, Instituto de Parasitología y Biomedicina 'López-Neyra', IPBLN-CSIC, Parque Tecnológico de Ciencias de la Salud, Avenida del Conocimiento S/N, 18100-Armilla (Granada), Spain, lina.diaz@ipb.csic.es

Contributors L-MD-G: conception and design, genotyping, analysis and interpretation of data, drafted and revised the paper, she is the guarantor. SG: monitored samples and data collection. NO-C, JJ-A, JS-R, EdR, MFG-E, AB, BF-G: identification, sampling and collection of patient samples and clinical data. IG-A, MAG-G: sampling and collection of patient samples and clinical data, critical reading of manuscript. JM: conception and design, analysis and interpretation of data, and drafted and revised the paper.

Acknowledgements We thank Gema Robledo and Sofía Vargas for their excellent technical assistance. We also thank all donors, patients and controls as well as Banco Nacional de ADN (University of Salamanca, Spain), who supplied part of the control DNA samples.

Funding This work was partially supported by RETICS Program, RD08/0075 (RIER) from Instituto de Salud Carlos III (ISCIII), within the VI PN de I+D+i 2008–2011 (FEDER). L-MD-G was supported by the 'Ayudas Predoctorales de Formación en Investigación en Salud (PFIS—FI09/00544)' from the 'Instituto de Salud Carlos III'.

Competing interests None.

Ethics approval The study was approved by the local ethical committees of all participating centres.

Provenance and peer review Not commissioned, externally peer reviewed.

Accepted 19 June 2011

Published Online First 17 August 2011

Ann Rheum Dis 2012;**71**:314–316. doi:10.1136/annrheumdis-2011-200062

Letter

REFERENCES

1. **Delgado-Vega AM**, Castiblanco J, Gómez LM, *et al*. Bcl-2 antagonist killer 1 (*BAK1*) polymorphisms influence the risk of developing autoimmune rheumatic diseases in women. *Ann Rheum Dis* 2010;**69**:462–5.
2. **Takeuchi O**, Fisher J, Suh H, *et al*. Essential role of BAX, BAK in B cell homeostasis and prevention of autoimmune disease. *Proc Natl Acad Sci USA* 2005;**102**:11272–7.
3. **Yao Y**, Marassi FM. BAX and BAK caught in the act. *Mol Cell* 2009;**36**:353–4.
4. **Rathmell JC**, Lindsten T, Zong WX, *et al*. Deficiency in Bak and Bax perturbs thymic selection and lymphoid homeostasis. *Nat Immunol* 2002;**3**:932–9.
5. **Sánchez E**, Gómez LM, Lopez-Nevot MA, *et al*. Evidence of association of macrophage migration inhibitory factor gene polymorphisms with systemic lupus erythematosus. *Genes Immun* 2006;**7**:433–6.
6. **Rodríguez-Rodríguez L**, Taib WR, Topless R, *et al*. The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in Caucasian case-control samples. *Arthritis Rheum* 2011;**63**:365–72.
7. **Sawyer SL**, Mukherjee N, Pakstis AJ, *et al*. Linkage disequilibrium patterns vary substantially among populations. *Eur J Hum Genet* 2005;**13**:677–86.
8. **Plenge RM**, Seielstad M, Padyukov L, *et al*. TRAF1-C5 as a risk locus for rheumatoid arthritis—a genomewide study. *N Engl J Med* 2007;**357**:1199–209.
9. **Fernando MMA**, Morris DL, Sanchez Rodriguez E, *et al*. Independent association of variants in the class III gene, *MSH5*, and the class II gene, *HLA-DPB1*, at the MHC in a Spanish SLE cohort. *Ann Rheum Dis* 2010;**69**(Suppl 3):133.

Evidence of new risk genetic factor to Systemic Lupus Erythematosus: the UBASH3A gene.

Plos One, 2013

OPEN ACCESS Freely available online

PLOS ONE

Evidence of New Risk Genetic Factor to Systemic Lupus Erythematosus: The *UBASH3A* Gene

Lina-Marcela Diaz-Gallo[1]*, Elena Sánchez[1], Norberto Ortego-Centeno[2], Jose Mario Sabio[3], Francisco J. García-Hernández[4], Enrique de Ramón[5], Miguel A. González-Gay[6], Torsten Witte[7], Hans-Joachim Anders[8], María F. González-Escribano[9], Javier Martin[1]

1 Cellular Biology and Immunology Department, Instituto de Parasitología y Biomedicina "López-Neyra", Consejo Superior de Investigaciones Científicas (IPBLN- Consejo Superior de Investigaciones Científicas), Granada, Spain, **2** Department of Internal Medicine, Hospital Clínico San Cecilio, Granada, Spain, **3** Department of Internal Medicine, Hospital Virgen de las Nieves, Granada, Spain, **4** Department of Internal Medicine, Hospital Virgen del Rocío, Sevilla, Spain, **5** Department of Internal Medicine, Hospital Carlos Haya, Málaga, Spain, **6** Department of Rheumatology, Instituto de Formación e Investigación Marqués de Valdecilla, Hospital Universitario Marqués de Valdecilla, Santander, Spain, **7** Department of Clinical Immunology and Rheumatology, Hannover Medical School, Hannover, Germany, **8** Medical department and policlinic IV, Klinikum der Universität, München, Munich, Germany, **9** Department of Immunology, Hospital Virgen del Rocío, Sevilla, Spain

Abstract

The ubiquitin associated and Src-homology 3 (SH3) domain containing A (*UBASH3a*) is a suppressor of T-cell receptor signaling, underscoring antigen presentation to T-cells as a critical shared mechanism of diseases pathogenesis. The aim of the present study was to determine whether the *UBASH3a* gene influence the susceptibility to systemic lupus erythematosus (SLE) in Caucasian populations. We evaluated five *UBASH3a* polymorphisms (rs2277798, rs2277800, rs9976767, rs13048049 and rs17114930), using TaqMan® allelic discrimination assays, in a discovery cohort that included 906 SLE patients and 1165 healthy controls from Spain. The SNPs that exhibit statistical significance difference were evaluated in a German replication cohort of 360 SLE patients and 379 healthy controls. The case-control analysis in the Spanish population showed a significant association between the rs9976767 and SLE (Pc = 9.9E-03 OR = 1.21 95%CI = 1.07–1.37) and a trend of association for the rs2277798 analysis (P = 0.09 OR = 0.9 95%CI = 0.79–1.02). The replication in a German cohort and the meta-analysis confirmed that the rs9976767 (Pc = 0.02; Pc = 2.4E-04, for German cohort and meta-analysis, respectively) and rs2277798 (Pc = 0.013; Pc = 4.7E-03, for German cohort and meta-analysis, respectively) *UBASH3a* variants are susceptibility factors for SLE. Finally, a conditional regression analysis suggested that the most likely genetic variation responsible for the association was the rs9976767 polymorphism. Our results suggest that *UBASH3a* gene plays a role in the susceptibility to SLE. Moreover, our study indicates that *UBASH3a* can be considered as a common genetic factor in autoimmune diseases.

Citation: Diaz-Gallo L-M, Sánchez E, Ortego-Centeno N, Sabio JM, García-Hernández FJ, et al. (2013) Evidence of New Risk Genetic Factor to Systemic Lupus Erythematosus: The *UBASH3A* Gene. PLoS ONE 8(4): e60646. doi:10.1371/journal.pone.0060646

Editor: Xiao-Ping Miao, MOE Key Laboratory of Environment and Health, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, China

Received December 18, 2012; **Accepted** March 1, 2013; **Published** April 2, 2013

Funding: This work was partially supported by RETICS Program, RD08/0075 (RIER) from Instituto de Salud Carlos III, within the VI PN de I+D+i 2008-2011 (FEDER) and grant KFO 250, TP 03, WI 1031/6-1 LMDG was supported by the "Ayudas Predoctorales de Formación en Investigación en Salud (PFIS - FI09/00544)" from the "Instituto de Salud Carlos III". The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lina.diaz@ipb.csic.es

Introduction

The T cell ubiquitin ligand proteins (TULA) family is characterized by function as suppressors of T cell receptor signalling. One of the members of the TULA family proteins is the ubiquitin associated and Src-homology 3 (SH3) domain containing A (*UBASH3a*) which is expressed only in lymphoid cells and facilitates apoptosis induced in T cells by certain stimuli, such as growth factor withdrawal [1]. *UBASH3a* gene spans 40 kb, contains 15 exons and is located on human chromosome 21q22.3 [2]. The lack of TULA proteins resulted in hyper-reactivity of T cells [1]. Evidence for both B and T lymphocyte hyper-reactivity is typically observed in autoimmune disorders [2]. These disorders are characterized by an inappropriate, ultimately excessive, inflammatory response against self, resulting in tissue destruction. Although many individuals affected by autoimmune diseases demonstrate multiorgan involvement, the primary end-organ target (e.g., autoimmune destruction of pancreatic islet cells in type 1 diabetes mellitus) typically drives the clinical presentation and disease definition. Recent studies have showed that single nucleotide polymorphisms (SNPs) of the *UBASH3a* gene are associated with some autoimmune diseases, like type 1 diabetes (T1D), celiac disease (CD), rheumatoid arthritis (RA) and vitiligo, suggesting that this gene could play an important role in the pathogenesis of autoimmune disorders [3 8].

Systemic lupus erythematosus (SLE) is a prototypic autoimmune diseases characterized by the production of autoantibodies, immune-complex deposition, and subsequent multiple organ damage. The complex aetiology of autoimmune diseases includes environmental, hormonal and genetic factors. Some of those factors remained to be defined [3,4]. Based on these insights, the aim of the present study was to evaluate the role of five *UBASH3a* polymorphism in SLE.

Table 1. Genotype and minor allele frequencies of *UBASH3a* SNPs located in Caucasian SLE patients and healthy controls from Spain, the discovery cohort.

			Genotype, N (%)			Alleles, N(%)		Allele test	
SNP	1/2	Subgroup (N)	1/1	1/2	2/2	1	2	*P*-value*	OR [CI 95%]****
rs2277798	G/A	Controls (n = 1165)	477 (40.94)	529 (45.41)	159 (13.65)	1483 (63.6)	847 (36.4)		
		SLE (n = 906)	402 (44.37)	394 (43.49)	110 (12.14)	1198 (66.1)	614 (33.9)	0.0993**	0.90 [0.79-1.02]
rs2277800	C/T	Controls (n = 1165)	1080 (92.70)	84 (7.21)	1 (0.09)	2244 (96.3)	86 (3.7)		
		SLE (n = 906)	832 (91.83)	73 (8.06)	1 (0.11)	1737 (95.9)	75 (4.1)	0.4592	1.13 [0.82–1.55]
rs9976767	A/G	Controls (n = 1165)	363 (31.16)	558 (47.90)	244 (20.94)	1284 (55.1)	1046 (44.9)		
		SLE (n = 906)	230 (25.39)	451 (49.78)	225 (24.83)	911 (503)	901 (49.7)	1.99E-03***	1.21 [1.07–1.37]
rs13048049	G/A	Controls (n = 1165)	1038 (89.10)	126 (10.82)	1 (0.09)	2202 (94.5)	128 (5.5)		
		SLE (n = 906)	808 (89.18)	96 (10.60)	2 (0.22)	1712 (94.5)	100 (5.5)	0.9719	1.01 [0.77–1.32]
rs17114930	C/G	Controls (n = 1165)	1066 (91.50)	95 (8.15)	4 (0.34)	2227 (95.6)	103 (4.4)		
		SLE (n = 906)	811 (89.51)	93 (10.26)	2 (0.22)	1715 (94.6)	97 (5.4)	0.1649	1.22 [0.92–1.63]

*All P-values have been calculated for the allelic model. ** Pc = 0.248 Benjamini & Hochberg (1995). ***Pc = 9.9E-03 Benjamini & Hochberg (1995) step-up FDR control.
****Odds ratio for the minor allele.
doi:10.1371/journal.pone.0060646.t001

Materials and Methods

Ethics Statement

Written informed consent was obtained from all participants and the respectively ethics committee approved the study according to the principles expressed in the Declaration of Helsinki.

The case-control study included 906 SLE patients and 1165 healthy controls from a white Spanish population. The replication cohort from white Germans comprehends 360 SLE patients and 379 healthy controls. All the patients met the American College of Rheumatology criteria for classification of SLE [5]. Written informed consent was obtained from all participants and the respectively ethics committee approved the study. DNA was obtained from peripheral blood using standard methods. The samples were genotyped for the *UBASH3a* rs2277798, rs2277800, rs9976767, rs13048049 and rs17114930 polymorphisms via TaqMan® 5′allelic discrimination technology using a predesigned SNPs genotyping assays provided by Applied Biosystems (assay ID: C___1724055_10, C__15885522_20, C___1724067_10, C___1724073_20 and C___25622591_10, respectively; Figure S1). At the moment of the design of the study the only confirmed case-control associated SNP with autoimmune diseases was the rs9976767 [6]. The other four SNPs were selected because they were not included in previous SLE genetic studies and they are non-synonymous changes located in different exons of the *UBASH3a* gene. Moreover, the minor allele frequency (MAF) of those SNPs was reported in Caucasian populations and they exhibited moderated LD with at least one SNP in the loci. Deviation from Hardy-Weinberg equilibrium (HWE) was tested by standard chi-square analysis. The differences in genotype distribution and allele frequency among cases and controls were calculated by contingency tables and when necessary by Fisher's exact test. Odds ratios (OR), and 95% confidence intervals (CI), were calculated according to Woolf's method. Combined data were analysed by Mantel-Haenszel tests under fixed effect model and the Breslow-Day (BD) test was used to estimate the OR heterogeneity amongst the two cohorts. An association was considered statistically significant if $P<0.05$. Benjamini & Hochberg (1995) step-up false discovery rate (FDR) control correction

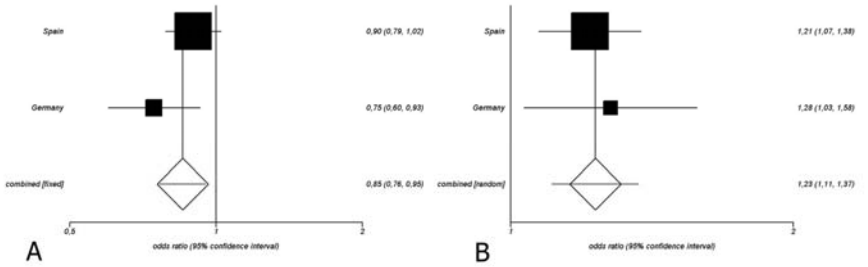


Figure 1. Graphical representation of the meta-analysis (A) Forest plot for the meta-analysis of the *UBASH3a* rs2277798 polymorphism in SLE in two Caucasian cohorts. (B) Forest plot for the meta-analysis of the *UBASH3a* rs9976767 polymorphism in SLE in two Caucasian cohorts.
doi:10.1371/journal.pone.0060646.g001

Table 2. Genotype and minor allele frequencies of *UBASH3a* SNPs located in Caucasian SLE patients and healthy controls from Germany.

			Genotype, N (%)			Alleles, N(%)		Allele test		
SNP	½	Subgroup (N)	1/1	1/2	2/2	1	2	*P*-value*	P_{FR}**	OR [CI 95%]***
rs2277798	G/A	Controls (n = 379)	184 (48.55)	132 (34.83)	63 (16.62)	448 (59.1)	310 (40.9)			
		SLE(n = 360)	149 (41.39)	163 (45.28)	48 (13.33)	475 (66)	245 (34)	0.0064	0,0128	0.75 [0.60–0.92]
rs9976767	A/G	Controls (n = 379)	186 (49.08)	136 (35.88)	57 (15.04)	458 (60.4)	300 (39.4)			
		SSc (n = 360)	180 (50.00)	106 (29.44)	74 (20.56)	392 (54.4)	328 (45.6)	0.0201	0,0201	1.28 [1.04–1.57]

*All P-values have been calculated for the allelic model. ** Benjamini & Hochberg (1995) step-up FDR control. ***Odds ratio for the minor allele.
doi:10.1371/journal.pone.0060646.t002

[7] for multiple testing was applied to the P-values in both the independent analysis and the combined meta-analysis (P_c). Linkage disequilibrium (LD) measurement (r^2) between the studied SNPs was estimated by expectation-maximization algorithm using HAPLOVIEW (version 4.2; Broad Institute of MIT and Harvard). Finally, the dependency of the association between each SNP and every studied genetic variant was determined by a conditional logistic regression analysis (considering the different cohorts as covariate). The analyses were performed using PLINK (version 1.07) [8].

Results

The distributions of genotypic and allelic frequencies of the five *UABSH3a* evaluated polymorphisms were in HWE at 5% significance level. Additionally, MAFs of the studied SNPs were similar to those reported by the HapMap project for the CEU population (http://hapmap.ncbi.nlm.nih.gov/) in both, Spanish and German cohorts. The LD structure of the five *UABSH3a* SNPs in the Spanish cohort is shown in (Figure S1). The Table 1 summarizes the results of the association analysis for the discovery cohort. The minor allele of the rs9976767 polymorphism exhibited a statistical significant association with SLE in the Spanish population (P_c = 9.9E-03, OR = 1.21, 95%CI = 1.07 1.37). In addition we observed a trend of association with the rs2277798 polymorphism (P = 0.099, P_c = 0.248, OR = 0.9, 95%CI = 0.79 1.02). The frequency of the minor alleles of the rs2277800, rs13048049 and rs17114930 *UBASH3a* polymorphisms were not statistically significantly different between SLE patients and healthy controls in the Spanish cohort.

Based on these observations, we evaluated the frequency of the rs9976767 and rs2277798 in a replication cohort from Germany (Table 2). Genotypic and allelic frequencies of both polymorphisms were in HWE. The frequency of the minor allele of both SNPs: rs9976767 and rs2277798 were statistically significant different between SLE patients and healthy controls: rs9976767 (P_c = 0.02, OR = 1.28 95%CI = 1.04 1.57) and rs227798 (P_c = 0.01, OR = 0.75, 95%CI = 0.6 0.92). Lastly, we combine both the Spanish and German cohorts through a meta-analysis in order to increase the statistical power and to determine the combine OR (Table 3 and Figure 1). This analysis showed evidence of association of the minor allele of rs9976767 with higher SLE risk (P_c = 4.7E-03, OR = 1.23 95%CI = 1.11 1.37) and the rs2277798 with lower risk to SLE (P_c = 2.4 04, OR = 0.85, 95%CI = 0.76 0.95).

Finally, we prompted out to evaluate whether one of both polymorphisms is responsible for the associations detected using a logistic regression analysis. Pair-wise conditional analysis showed that the association of the rs2277798 SNP was explained by the rs9976767effect, because only the coefficient for the test of rs9976767 remained significant (model conditioned by rs2277798P = 0.76; model conditioned by rs9976767 P = 9E-03, Table 4).

Discussion

UBASH3a is implicated in the regulation of tyrosine phosphorylation levels within T cells and is involved in facilitates the apoptosis induced in these cells. *UBASH3a* binds to the apoptosis-inducing protein AIF, which has previously been shown to function as a key factor of caspase-independent apoptosis [9]. It has also been reported that SLE T cells, compared with control T cells, undergo an increased rate of apoptosis, which contribute to SLE pathogenesis [4]. Changes in the *UBASH3a* structure or expression levels can affect the binding with AIF leading to an alteration in the apoptosis level.

Herein, we described for the first time the influence of five *UBASH3a* genetic variants in SLE susceptibility. Interestingly, the

Table 3. Meta-analysis of two *UBASH3a* genetic variants within Spanish and German SLE populations.

			Genotype, N (%)			Alleles, N(%)		Allele test		
SNP	½	Subgroup (N)	1/1	1/2	2/2	1	2	*P*-value*	P_{FDR}**	OR [CI 95%]***
rs2277798	G/A	Controls (n = 1544)	609 (39.44)	713 (46.18)	222 (14.38)	1931 (62.5)	1157 (37.5)			
		SLE (n = 1266)	565 (44.63)	543 (42.89)	158 (12.48)	1673 (66.1)	859 (33.9)	0.0047	4.7E-03	0.85 [0.76–0.95]
rs9976767	A/G	Controls (n = 1544)	499 (32.32)	744 (48.19)	301 (19.49)	1742 (56.4)	13446 (43.6)			
		SLE (n = 1266)	336 (26.54)	631 (49.84)	299 (23.62)	1303 (51.5)	1229 (48.5)	1.2E-04	2,4E-04	1.23 [1.11–1.37]

*All P-values have been calculated for the allelic model. **Benjamini & Hochberg (1995) step-up FDR control. ***Odds ratio for the minor allele.
doi:10.1371/journal.pone.0060646.t003

Table 4. Conditional logistic regression analysis for two *UBASH3a* SNPs located in SLE considering the two European populations as covariate.

Group of analysis	SNP	MAF Cases	MAF Controls	p Value: add to rs9976767	rs9976767 p value: add to SNP	r2 with rs9976767	
						Spain	Germany
SLE							
	rs2277798	0.34	0.38	0.758	0.0087	0.45	0.41

doi:10.1371/journal.pone.0060646.t004

rs9976767 polymorphism is located in the intronic region between the exons 5 and 6 while the other four studied SNPs (rs2277798, rs2277800, rs13048049 and rs17114930) are non-synonymous changes located in three different exons. The intronic regions flanking constitutive exons contain potential splicing regulatory sequences. Moreover, a study restricted to analysis of the canonical splice signals reported that 15% of point mutations disrupted splicing, a likely gross underestimate of the impact of splicing on human disease [10]. This suggests that the rs9976767 polymorphism could be affecting the expression of different *UBASH3a* isoforms consequently affecting the binding to AIF. Concerning to this we checked if there is any relation between the rs9976767 and expression of *UBASH3a* gene using expression quantitative trait loci (eQTL) databases. Interesting, there is a significant statistical correlation between the increase of *UBASH3a* expression in lymphoblastoid cell lines and the homozygotes for the minor allele of rs9976767 (rho = 0.483, P = 1.3E-05; Figure S2A) in one of the two groups of twins studied (this observation was done using Genevar 3.2.0 software) [11,12]. Furthermore the eQTL studies in asthma showed that the SNPs (rs9784215, rs3746923, rs2277797) with highest LOD score (LOD>4.5, P<1E-05) in the *UBAHS3a* locus are in moderate to high LD with rs9976767 (Figure S2B and C; this observation was done using mRNA by SNP Browser 1.0.1 http://www.sph.umich.edu/csg/liang/asthma/) [13,14]. This evidence suggested that rs9976767 could have a functional role in the regulation of the expression of *UBASH3a*. However, and according with HapMap project (http://hapmap.ncbi.nlm.nih.gov/), this SNP tags other six variants in this region (rs7278547, rs11702374, rs9976479, rs3746924, rs3761378, rs7283281; r^2>0.95) and considering the present study and the previous GWAS [15,16] we have studied approximately 15% of the genetic variation of *UBASH3a* locus. In order to cover all the genetic variation of this gene, it is necessary to genotype 181 SNPs (calculated through an aggressive tagging with 2-marker haplotypes in Haploview 4.2 software using CEPH population from HapMap project). All these together suggest that the rs9976767 is a good functional candidate risk factor to SLE, but it could be more than one variant related to SLE.

No previous reports have associated the rs9976767 *UBASH3a* polymorphism with SLE. Nevertheless, it is worth noting that the rs9976767 SNP or its six tags variants were not included in previous genome wide association studies (GWAS) in Caucasian SLE cohorts [15,16]. Although the statistical power is 96% for our meta-analysis (calculate using a p value = 0.05:OR = 1.2: MAF = 0.4), the results found in our study should be replicated in different Caucasian cohorts and other populations. Furthermore there is a need to determine whether the statistical associations are related with the involvement of *UBASH3a* in the pathogenesis of SLE and other autoimmune diseases. Regarding to this, the *UBASH3a* gene seems to be a common genetic factor in autoimmune diseases because different polymorphism of this locus has been associated with autoimmune diseases like T1D, CD, RA and vitiligo [6,17 21]. Our results showed that the minor allele of the rs9976767 *UBASH3a* polymorphism is a risk factor to SLE, as similarly observed with T1D [6]. Nevertheless, there is no evidence of association between this variant and other autoimmune diseases. This can be linked with the suggestion that common genetic factors in autoimmune diseases could match a regional level but differ in the specific genetic variant associated to each disease, like the associations observed with *IL2 IL21* and MHC loci [22]. Based on the concept of quantitative thresholds for immune-cell signalling, the effect of the rs976767 *UBAHS3a* variant could diversely affect the range of values for the stimulus-response selection of the immune cells in different autoimmune pathologies, making it more or less relevant in different diseases [3].

In conclusion, our study showed the first evidence of association of the *UBASH3a* gene with the genetic background of SLE. Together the functional role of the protein encoded by this gene, the reported data in the eQTLs databases and our results point to the *UBASH3a* gene as a new element in the pathogenic mechanism of autoimmune diseases.

Supporting Information

Figure S1 Pattern of linkage disequilibrium of the five studied SNPs and their location in the UBAHS3a gene. The values correspond to r2 calculated for the Spanish cohort. The rs2277798 polymorphism [G/A] is located in exon 1 of *UBASH3a* gene. It's a no-synonymous change in the position 18 of the protein (S[Ser]/G[Gly]). The rs2277800 polymorphism [C/T] is also located in exon 1 of *UBASH3a* gene and generate a change in the position 28 of the protein (L[Leu]/F[Phe]). In the other hand, the rs9976767 [A/G] is an intronic variant located between the exons 5 and 6 of the *UBASH3a* gene. Both variants rs13048049 [G/A] and rs17114930 [C/G] are no-synonymous changes in exons 7 and 11, respectively. The first one produce a change from arginine (R[Arg]) to glutamine (Q[Gln]) in position 286; while the rs17114930 polymorphism generates a change from aspartic acid (D[Asp]) to glutamic acid (E[Glu]) in position 428 in Caucasian population.
(TIF)

Figure S2 Results observed using different expression quantitative trait loci (eQTL) tools to evaluate if there is any relationship between the rs9976767 variant and the *UBASH3a* expression **(A) SNP-gene association plot** for the rs9976767 and the *UBASH3a* gene based on Spearman's rank correlation coefficient (rho) using the Genevar 3.2 software (http://www.sanger.ac.uk/resources/software/genevar/) [1]. The eQTL analysis was performed in lymphoblastoid cell lines from peripheral

blood sample (n = 74). The plot corresponds to one of the two twins groups studied [2]. **(B) Linkage disequilibrium (LD) plot** performed in Haploview 4.2 [3]. LD plot between rs9976767 and the rs9784215, rs3746923, rs2277797 SNPs which exhibited the highest LOD score (LOD>4.5, P<1E-05) in the *UBAHS3a* locus showed in **(C) Snapshot of observed eQTLs related with *UBASH3a* gene** from the mRNA by SNP Browser 1.0.1 software (http://www.sph.umich.edu/csg/liang/asthma/) based on eQTL studies in asthma [4,5]. The LOD scores and P values for those SNPs are: rs9784215, LOD = 4.909 P = 2E-06; rs3746923, LOD = 4.905 P = 2E-06; rs2277797, LOD = 4.68 P = 3.4E-06. They are signalled as red dots in the LOD plot. 1. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26: 2474-2476. 2. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 7: e1002003. 3. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21: 263-265. 4. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nat Genet 39: 1202-1207. 5. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 448: 470-473. (TIF)

Acknowledgments

We thank to GemaRobledo, Sofia Vargas and Sonia Garcia for their excellent technical assistance. we thank to all donors, patients and controls. We thank BancoNacional de ADN (University of Salamanca, Spain) who supplied part of the control DNA samples.

Author Contributions

Colection of the samples and clinical information: NOC JMS FJGH EDG MAGG TW HJA MFGE. Review of the manuscript: ES NOC JMS FJGH EDG MAGG TW HJA MFGE. Conceived and designed the experiments: LMDG ES JM. Performed the experiments: LMDG ES. Analyzed the data: LMDG. Contributed reagents/materials/analysis tools: LMDG ES NOC JMS FJGH EDG MAGG TW HJA MFGE JM. Wrote the paper: LMDG JM.

References

1. Tsygankov AY (2009) TULA-family proteins: an odd couple. Cell Mol Life Sci 66: 2949-2952.
2. Zenewicz LA, Abraham C, Flavell RA, Cho JH (2010) Unraveling the genetics of autoimmunity. Cell 140: 791-797.
3. Cho JH, Gregersen PK (2011) Genomics and the multifactorial nature of human autoimmune disease.N Engl J Med. 2011/10/28 ed. 1612-1623.
4. Guerra SG, Vyse TJ, Cunninghame Graham DS (2012) The genetics of lupus: a functional perspective. Arthritis Res Ther 14: 211.
5. Hochberg MC (1997) Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 40: 1725.
6. Grant SF, Qu HQ, Bradfield JP, Marchand L, Kim CE, et al. (2009) Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. Diabetes 58: 290-295.
7. Benjamini Y HY (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57: 289-300.
8. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559-575.
9. Collingwood TS, Smirnova EV, Bogush M, Carpino N, Annan RS, et al. (2007) T-cell ubiquitin ligand affects cell death through a functional interaction with apoptosis-inducing factor, a key factor of caspase-independent apoptosis. J Biol Chem 282: 30920-30928.
10. Yeo GW, Van Nostrand EL, Liang TY (2007) Discovery and analysis of evolutionarily conserved intronic splicing regulatory elements. PLoS Genet 3: e85.
11. Nica AC, Parts L, Glass D, Nisbet J, Barrett A, et al. (2011) The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. PLoS Genet 7: e1002003.
12. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26: 2474-2476.
13. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, et al. (2007) A genome-wide association study of global gene expression. Nat Genet 39: 1202-1207.
14. Moffatt MF, Kabesch M, Liang L, Dixon AL, Strachan D, et al. (2007) Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma. Nature 448: 470-473.
15. Harley JB, Alarcon-Riquelme ME, Criswell LA, Jacob CO, Kimberly RP, et al. (2008) Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40: 204-210.
16. Kozyrev SV, Abelson AK, Wojcik J, Zaghlool A, Linga Reddy MV, et al. (2008) Functional variants in the B-cell gene BANK1 are associated with systemic lupus erythematosus. Nat Genet 40: 211-216.
17. Concannon P, Onengut-Gumuscu S, Todd JA, Smyth DJ, Pociot F, et al. (2008) A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3. Diabetes 57: 2858-2861.
18. Jin Y, Birlea SA, Fain PR, Gowan K, Riccardi SL, et al. (2010) Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo. N Engl J Med 362: 1686-1697.
19. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, et al. (2008) Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med 359: 2767-2777.
20. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508-514.
21. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, et al. (2011) Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS Genet 7: e1002004.
22. Diaz-Gallo LM, Martin J (2012) Common genes in autoimmune diseases: a link between immune-mediated diseases. Expert Rev Clin Immunol 8: 107-109.

The supplement information of this manuscript Figure S1 and Figure S2 correspond to the Figures 11 and 12 of the discussion of this thesis.

DISCUSSION

New evidence of shared genetic components between autoimmune diseases

This Thesis provides evidence for the shared genetic component of AIDs, using the candidate gene association approach (Table 2) [209-215]. We described new associations in SSc, SLE and IBD with polymorphisms previous described in other AIDs. Besides the conclusions obtain from our results, we will discuss important features of the genetic association studies in AIDs that we could observed and compared with different evidence.

PTPN22, SSc and IBD

For the ***PTPN22***, our studies gather together previous results of the relation between the classical functional variant rs2476601, on SSc and IBD [209, 211], demonstrating that this variation has a very low influence in the genetic susceptibility to SSc in contrast to the effect observed for CD in our study, T1D, RA or SLE in previous reports. The effect of the rs2476601 polymorphism is more evident for ACA positive SSc patients and ileal location of the affection in CD. In other studies of this variant with autoantibodies presence in AIDs have show similar results. The rs2476601 *PTPN22* polymorphism has been shown to be more strongly associated with ACCP positive RA, with an OR=1.91 compared with the OR=1.20 for ACCP negative patients [216]. Something similar was observed in anti-dsDNA positive (OR=1.41) and dsDNA negative SLE patients (OR=1.14) [130]. A recent study in discordant twin pairs that evaluated the relationship between risk alleles and occurrence of autoantibodies showed that there is a high concordance of ACCP positivity in MZ twins (78.6%) but significantly less in DZ (25%), suggesting an important genetic determinant in the production of these autoantibodies [217]. Although, the authors concluded that the rs2476601 *PTPN22* allele is not an associated factor with the occurrence of ACCP, none of the healthy co-MZ twins were T-risk allele carriers. This suggested that could be an additive effect of the occurrence of autoantibodies and the minor allele of the rs2476601 in the risk to develop an AID, but the production of the autoantibodies could be influence by other genetic variants.

Table 2. Summary of the evaluated genetic variants and the results observed for the minor allele of each variant in the studies [209-215].

Locus	Chromosome location	Polymorphism	Phenotype	Subphenotype	Sample Size cases:controls	P value	OR	95% CI	Other associated AIDs
PTPN22	1p11	rs2476601	SSc		4063:4406	0.03*	1.15	1.03-1.28	RA, SLE, T1D and Grave's diseases
				ACA+	1376:4126	0.022*	1.22	1.05-1.42	
			CD		9254:8766	7.4E-06*	0.81	0.75-0.89	
				Ileal location	589:3082	9E-03*	0.64	0.49-0.84	
			UC		5695:8766	0.88	0.98	0.85-1.15	
		rs33996649	SSc		3422:3638	0.36	0.89	0.72-1.12	RA, SLE
			CD		1903:3107	0.22	1.16	0.91-1.47	
			UC		1677:3107	0.013*	0.69	0.51-0.93	
STAT4	2q32	rs7574865	CD		820:2012	0.71	1.07	0.74-1.54	RA, SLE, SSc, primary biliary cirrhosis
			UC		754:2012	0.012*	1.2	1.04-1.39	
IL2-IL21	4q27	rs6822844	SSc		4407:5792	6.6E-04*	0.86	0.79-0.93	RA, SLE, IBD, T1D, Grave's diseases, celiac disease
				lcSSc	2977:5792	6E-04*	0.84	0.76-0.92	
CD24	6q21	rs8734	CD		366:628	0.64	0.95	0.78-1.17	RA, SLE, Multiple sclerosis, giant cell arteritis
			UC		322:628	0.27	1.12	0.91-1.38	
		rs3838646	CD		371:629	6.4E-03	1.61	1.71-2.21	
				diagnosis age from 17 to 40 years old	268:629	6.62E-05	2.27	1.59-3.24	
			UC		310:629	0.66	0.92	0.62-1.35	
BAK1	6p33	rs513349, rs561276, rs5745582	RA		484:829	NS			
			SLE		633:829	NS			
UBASH3a	21q43	rs997676	SLE		1266:1544	2.4E-04*	1.23	1.11-1.37	RA, T1D and celiac disease

* The p values, OR and 95% CI correspond to the results of the combine analysis with more than one Caucasian population. NS: non statistically significant.

Other interesting characteristic of the rs2476601 *PTPN22* polymorphism association with AIDs, is the opposite direction of OR of the minor allele in CD (Figure 6) [209]. A genetic comparative study of previous associated variants for T1D, CD and UC reported opposite directions of OR in associates' *loci* [87]. The study confirmed the opposite association for rs2476601 variant between T1D (OR=1.99) and CD (OR=0.72); and revealed new evidence of similar phenomenon for the variants: rs3024505 in the *IL10 locus* (T1D OR=0.76, CD OR=1.24), rs917997 in the *IL18RAP* region, rs4788084 in the *NUPR1/IL27* and several MHC SNPs. A similar example is the association found between the rs2076530 *BTNL2* variant, T1D (OR=1.38), RA (OR=1.35) and SLE (OR=0.62) in European descendent population [218]. Sirota *et al.*[219] implemented a novel analysis with multiple SNPs in an allele-specific fashion comparing multiple sclerosis, ankylosing spondylitis, autoimmune thyroid disease, RA, CD and T1D [219]. The purpose of the authors was to refine the genetic variation profile of a certain disease to determined differences and similarities between related diseases. They proposed a notion of genetic variation score (GVS) which captures both the strength of association (p value) of a given SNP and whether an allele is protective (OR<1) or susceptible (OR>1). Based on this there are negative or positive disease-disease relationship, the first one defines by two phenotypes strongly associated to the same SNP, where the minor allele is oppositely associated to each trait. The positive disease-disease relationship is the reverse case, two phenotypes strongly associated to the same SNP which minor allele exerts in both the same effect. Applying this to our results, we could say that SSc and CD have a negative disease-disease relationship that would be reflected in a negative GVS value. This negative disease-disease relationship could be related with the idea that the susceptibility *loci* for AIDs (both inside and outside of MHC) are potentially under a balancing selection that are dependent on heterogeneity in environmental factors. For example, as we mentioned in the discussion of our manuscript of *PTPN22* and IBD [209]; the minor allele of the rs2476601 SNP seems to play a protective role against tuberculosis [220-222], illustrating the pleiotropic role of this SNP. In summary, it has suggested that the balancing selection does not act on the phenotype (in this case AID) per se, because that requires tens of thousands of years; but impacts immune responses towards different infectious agents, predisposing to different diseases, that act during recent times [87].

Finally, the association of the rs2476601 *PTPN22* with AIDs has been observed mainly in European descendent populations and in some South Latin American population studies [223, 224], but not in Asian or African descendent populations because it is not polymorphic in these populations [97, 104, 225, 226]. Five novel SNPs were found in Japanese and Korean populations when the *PTPN22* gene was sequenced [225]; and one of them (rs2488457) has been suggestive associated with T1D and RA in these populations [225, 226]. This represents an example of the genetic heterogeneity between populations that adds an important factor to the complexity of the aetiology of AIDs. As we mentioned in the Introduction the MHC associations are also restricted to genetic heterogeneity between human populations; other examples are the genes *PADI4*, *SLC22A4*, *PDCD1* [227]. This highlights the importance of the replication in different ancestral groups. A good example is the recent published fine mapping of the Xq28 *locus* for SLE in multiple ancestral groups [228]. The strongest associated SNPs for each population were specific. Nevertheless, a risk haplotype shared in the four populations point out the *locus* that contains the genes *TMEM187- IRAK1-MECP2*. Using a conditional analysis and a meta-analysis the authors observed that the rs1059102 could best explain the association signals detected in the region.

We demonstrated for the first time, that the minor allele of the functional polymorphisms **rs33996649** of the ***PTPN22*** gene is protective against UC (Table 2) [209]. This result suggests that there is a positive disease-disease relationship between RA, SLE and UC [115, 116, 209]. Our results showed that the *PTPN22 locus* is related with IBD but in a differential fashion between CD and UC. Coming back to the mentioned lessons taken from the MHC region, our observations showed something similar in a smaller scale because the precise allele and its effects seems to be specific for each disease. The location and amino acid change of both genetic variants, rs2476601 and rs33996649, have direct implications in the function of LYP. Although the mechanisms of action of both polymorphisms in AIDs are not clear; some studies have shown that the amino acid (W) encoded by the minor allele of rs247660, is a gain-of-function form of LYP, and carriers of LYP-W620 show reduced TCR signalling [229, 230]. On the other hand, the evidence suggest that the amino acid (Q) encoded by the minor allele of rs33996649 is a rare loss-of-function mutant of LYP [115] (Figure 6). Together with our results, this proposes another paradox, because both a reduction and

an activation of the TCR signalling would protect against the development of CD and UC, respectively.

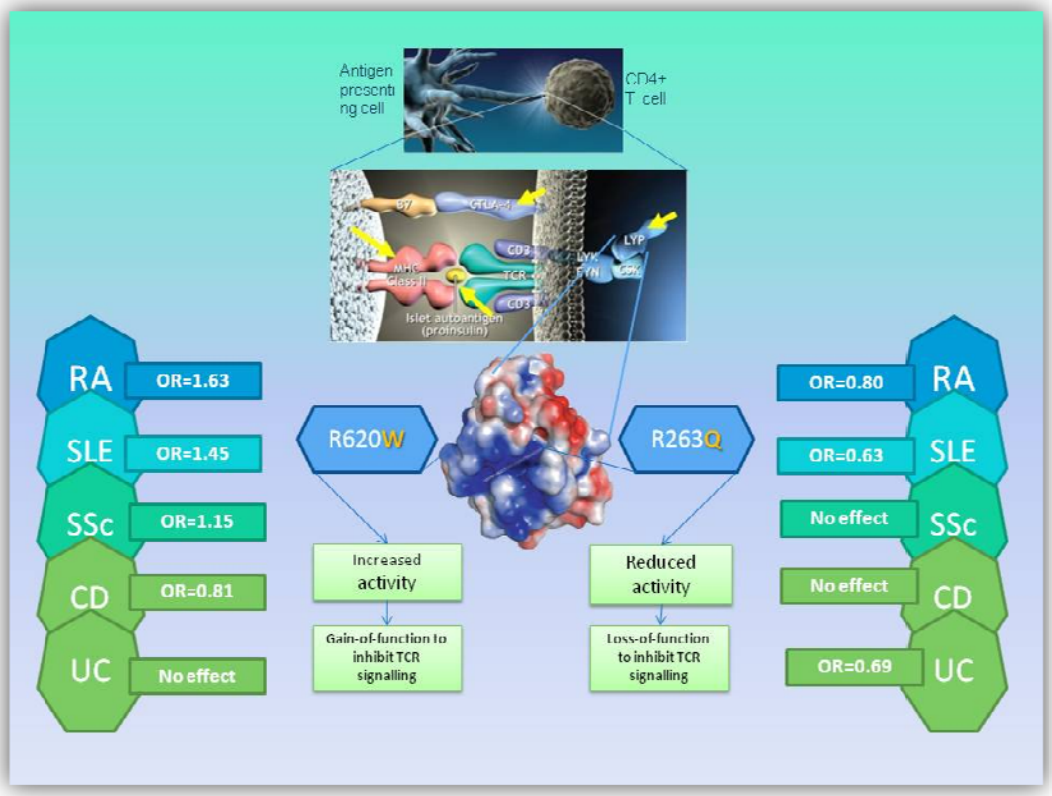


Figure 6. Representation of LYP (encoded by *PTPN22* gene) and the functional mutations associated with the development of AIDs. In the immunological synapses are involved the antigen presenting cell and the T cell. Among the important molecules of the interaction are the MHC and TCR; and one of the proteins that interfere in the regulation signalling of the TCR is LYP. Two functional polymorphisms rs2476601 (R620W) and rs33996649 (R263Q) in the *PTPN22* gene that encodes LYP, have been associated with the genetic susceptibility to AIDs. The abbreviations for the AIDs studied in this Thesis and their respective reported OR in Caucasian populations are indicated [106, 115, 116]. As part of the contribution of this Thesis in the shared genetic component of AIDs are the study of both *PTPN22* variants in IBD (CD and UC) and SSc [209, 211].

STAT4 & IBD

We observed through a meta-analysis in a Spanish population that the minor allele of the **rs7574865** polymorphisms in the ***STAT4*** gene is a risk factor to develop UC [213]. Previous to our study, a CGAS for this variant in an IBD Spanish case-control cohort

showed that the minor allele exert a risk effect for both CD and UC, in the same direction that the previous studies had reported in RA, SLE, SSc and Sjögren's syndrome [121, 122, 126-128]. The aim of our study was to evaluate the *STAT4* variant in our IBD cohort and increase the statistical power pooling both studies. The result was replicated for UC but not for CD (Table 2). We observed heterogeneity in the meta-analysis of CD, due to the inverse OR in both groups (See Figure 1 in [213]). The frequency of the minor allele of the rs7574865 was slightly higher in healthy controls compared to CD patients in our cohort compared with Martinez *et al* study [128] finding. Posterior to the publication of our study, another CGAS in IBD European descendent population that included the rs7574865 *STAT4* observed a lack of a association in both CD and UC [231]. Nevertheless, they observed a trend of association with the minor allele of rs7574865 SNP and CD (p=0.047, OR=0.86 95%CI=0.74-0.99). On the other hand, a study that analyzed eight *STAT4* polymorphisms in IBD Korean population, suggested an association in the dominant model for the rs925843 *STAT4* variant in UC but not an influence of the rs7574865; moreover the variants are not in LD [232]. Another study in Tunisian population that evaluated the role of rs7574865 in IBD showed that the minor allele of the variant had a marginal association with UC but not with CD [233]. Because there is not consistent results among these studies, we performed a meta-analysis of the five studies and we observed an association of the minor allele of the rs7574865 with UC (p value=0.019, OR=1.13, 95%CI=1.02-1.25); while there was not effect of this allele on CD (p value=0.36, OR=0.95, 95%CI=0.86-1.05). The heterogeneity measured by the Breslow-Day (BD) test, was not significant for UC (p value= 0.2, BD=5.91) while it was for CD (p value=0.016, BD=12.17) (Figure 7). Together, these suggest that there is an effect of the rs7574865 *STAT4* variant on UC but not in CD. If we take in account our observations in *PTPN22* and SSc, we can suggest that the effect of the rs75748656 *STAT4* variant in UC represents a minor genetic susceptibility factor such it is the rs2476601 *PTPN22* for SSc. Moreover, it is possible that the association for this *locus* in IBD fall on different SNPs of the *STAT4 locus*. Indeed the functional evidence has suggested that *STAT4* is implicated in IBD [117, 234-236]. For example, *STAT4* is required for the fully functional Th1 cells development, STAT4-deficient mice are protected from the effects of T-cell-mediated autoimmune diseases in several models of autoimmune diseases [117], among them colitis induced by transplantation [237]. A study with IBD patients and healthy controls showed that there is a significant increase

in *STAT4* mRNA level in the colonic mucosa and peripheral bone marrow cells (PBMCs) of the patients compared with controls [236]. A difference that the authors of the study did not mention on their manuscript and it is in line with our results of association, it is that the over expression seems to be higher for UC than CD patients, according to the graphics and the p-values. Moreover they reported a reduce DNA methylation in the -172 *STAT4* promoter region in the homozygous for the risk allele of the rs7574865. Also, the authors observed an overexpression of *STAT4*'s mRNA of the rs7574865 risk allele carriers. Taken into account that *STAT4* enhance the Th1 response that is related with IBD in animal models, that the *STAT4* overexpression could be higher in UC patients than CD or healthy controls, that there is a correlation with the *STAT4* expression and the presence of the rs7574865 risk allele and in turn this allele has been associated with an increase risk to UC, we can suggest that the rs7574865 *STAT4* polymorphism plays a role that is mark in UC but not in CD. Finally, it is interesting highlight that we observed in the study of *PTPN22* and IBD a differential association between both UC and CD and according with the previous discussion there is a similar phenomenon with the *STAT4* rs7574865 SNP in these diseases. The associations of the *PTPN22* and *STAT4* studied genetic variants have similar effects in UC as in RA, SLE or SSc while it seems to be an opposite effect in CD compared with these AIDs.

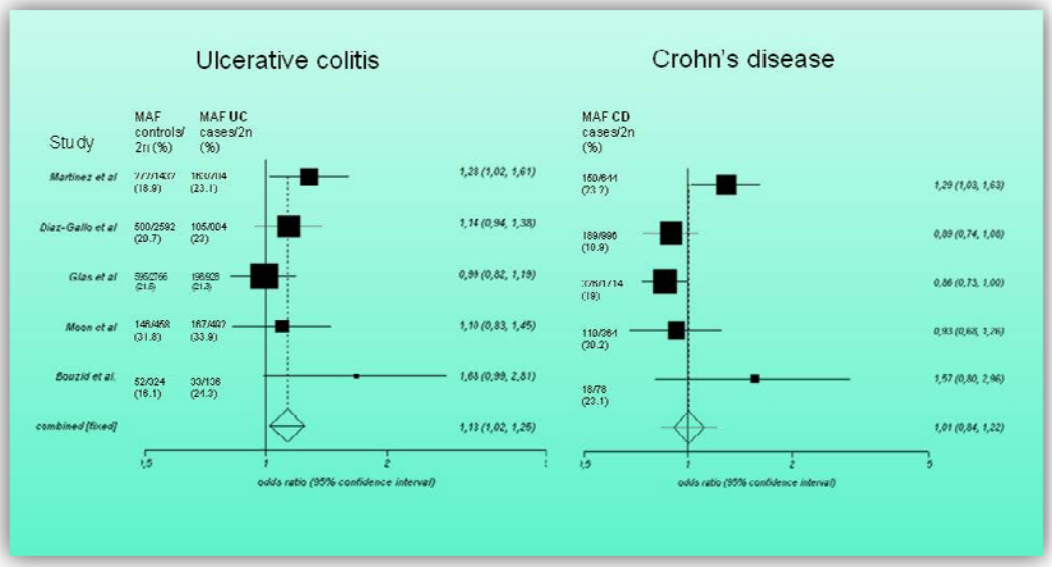


Figure 7. Forest plot for meta-analysis of the T allele frequency of the rs7574865 polymorphism of *STAT4* in IBD. We pooled together the cohorts included in our publication and the studies that evaluated the influence of the same polymorphism in

European descendent, Korean and Tunisian populations [128, 213, 231-233]. The left forest plot represents the meta-analysis of the five studies in Ulcerative colitis (UC); there was not heterogeneity between the studies measured by Breslow-day (BD) test=5.91, p value=0.205. The right forest plot represents the meta-analysis of the five studies in Crohn's disease (CD) that exhibited heterogeneity BD=12.17, p value=0.016.

IL2-IL21 & SSc

Most of the genetic associations described for SSc have also been reported to play a role in the susceptibility to SLE [94]. For that reason we evaluated in SSc the highly significant associated SNPs of ***IL2-IL21*** *loci* in SLE, that have been found using a fine mapping approach (Figure 8) [146]. Our study reported for the first time that the minor allele of the rs6822844 is associated with a decrease risk to SSc (Table 2) [215]. The location of the rs6822844 SNP together with the known immunologic role of the *IL2* and *IL21* genes, suggest that this SNP could be implicated in the pathogenic mechanisms of the associated AIDs. There is no evidence that directly relates the rs6822844 variant with a dysregulation of any of these cytokines; nevertheless, *in silico* approach that have grouped the known genetic AIDs associated factors into immunologic pathways agree in highlight the T and B cells differentiation pathways, involving the *IL2* and *IL21* genes [5, 238]. *IL2* is known to elicit T cell proliferation, survival and differentiation of effectors Th1 and Th2 cells [239]. Moreover, *IL2* is implied in maintain peripheral T cell tolerance [239, 240]. On the other hand *IL21* can induce the differentiation and activation of NK cells, promote NK T cell proliferation, cytokine release an effector function, and enhance the differentiation of Th17 cells [241]. B10 cells are a rare population of B cells that are able to express *IL10* and to negatively regulate inflammation and AID in both mice and humans [242]. Interestingly, a recent reported showed that the negative regulation of AIDs severity by B10 cells requires *IL21* [243]. Data from murine models showed that a haplotype of *IL2* SNPs predisposes to organ-specific autoimmune disease by reducing IL2 production from antigen-specific T cells [148]. Other type of murine models where there is an IL21 and IL17 deregulated production can lead to either lupus-like disease or RA-like symptoms depending on the genetic background [244].

The minor allele of the rs6822844 *IL2-IL21* variant exhibited a protective effect to SSc and lcSSc in our study (Table 2) [215]. The same effect direction has been observed in the previous associated AIDs with the *IL2-IL21* polymorphisms or variants in high LD with it [134-146]. Although, the variant rs6822844 is also associated with SLE, the SNPs rs6835457 and rs907715 could explain the association of the *IL2-IL21* region in this disease according with the logistic regression analysis performed by the authors of the fine mapping in SLE [146]. Our analysis reveals that rs6822844 could explain the association of the four studied SNPs in SSc [215]. It is important to highlight that LD between the SNPs rs6835457, rs907715 and rs6822844 is low (r^2=0.33) in European descendent populations (Figure 8). This suggested that into the same *locus* could be specific variants associated to each disease. We have mentioned such phenomenon in the introduction related to the MHC associations. Then the fact that different SNPs form the same *locus* associated with different phenotypes, mark the relevance of the *locus* in the mechanisms of the pathogenesis for that group of diseases. But this leaves the question about the specific role of the SNPs in the associated phenotype. The recent evidence suggests that the answer is not simple and probably is not either a single one. Beyond the classic discussion between the common-disease common-variant model [245] and common-diseases rare-variant model [246], there are new arguments based on evidence that support a causal role of the associated SNPs at the same time that other arguments support a marker-of-the-causal role of the associated SNPs. It is possible that the common variants, as the rs6822844 in *IL2-IL21 locus* related here, principally found in GWAS may be the result of synthetic associations. This means that the observed association arise from the sum of multiple low-frequency ungenotype markers, as it is the case for the *NOD2* association signal from GWAS in CD that could be explain by three rare functional variants form the same *locus* [247]. Although the synthetic associations could explain some of the common variants associated to AIDs, the comparisons between GWAS and modest linkage studies have shown few overlapping of signals and this fact limited the field of action of the synthetic associations [248]. Most of the geneticists agree based on the evidence collected until now, that the associated SNPs are tags that reside in the proximity of the relevant functional genomics element [2, 5, 47, 248]. On the other hand, the non-coding associated SNP could be direct implicated in the mechanisms that affect the phenotype. For example, one of the alleles of a SNP highly associated with human skin, hair and eye colour was demonstrated to alter the DNA looping in that genomic region, in turn affecting the

expression of proxy gene involved in pigmentation [249]. This associated SNP resides in the intron of a different gene, 21 kilobases upstream of the gene implicated in pigmentation. The authors of the study demonstrated that this hit SNP region is an enhancer-promoter looping involved in the expression regulation of the gene directly implicated with the colour phenotype. These suggested that associated SNPs could be marking the three-dimensional structure of DNA that has been recognized as playing an important role in the regulation of the gene expression [250]. The rs6822844 genetic variant could be implicated in the AIDs pathogenic mechanisms affecting the three-dimensional structure of the DNA in that *locus*, but to prove such hypothesis further and specific studies are required.

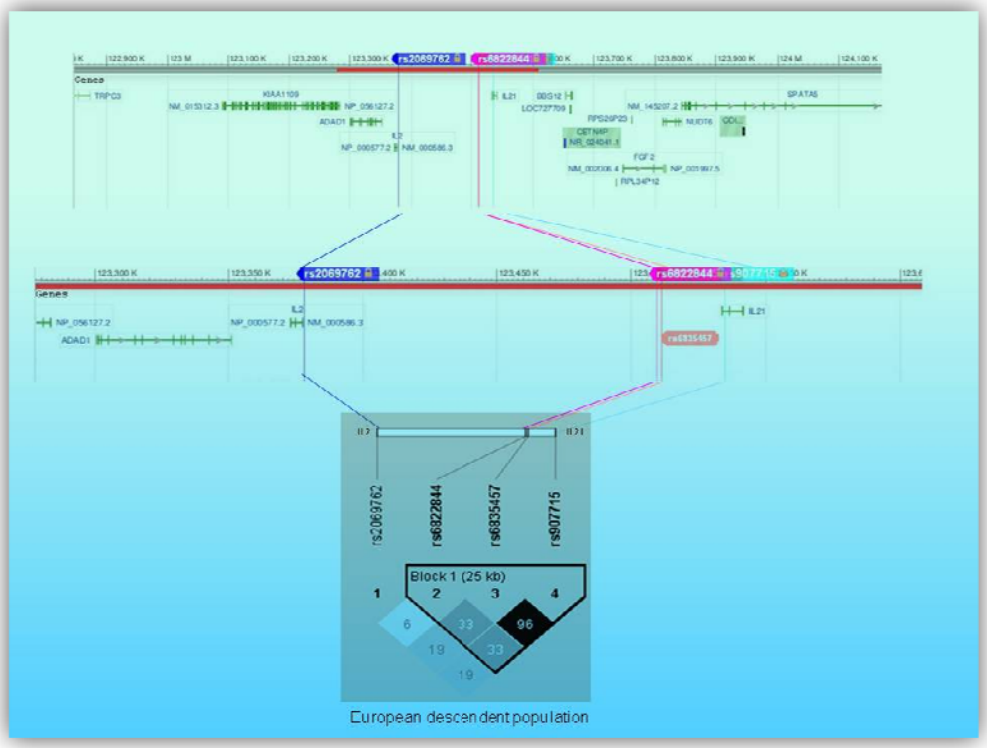


Figure 8. Location of the *IL2-IL21 locus* SNPs studied in SSc and their respective LD plot. The upper part of the diagram showed the position of the *IL2* and *IL21* genes and their closer genes among them: *KIAA1109*, *ADAD1*, *FGF2* and *NUDT6*. The middle part of the figure shows a zoom of the location of the *IL2*, *IL21* genes and the four studied SNPs. The rs2069762 variant is located in the flanking 5'-untranslated (UTR) region of *IL2* gene. The rs6822844 and rs6835457 SNPs are located in the flanking 3'-UTR region of *IL21* gene. The rs907715 polymorphism is located in the intron 3of the *IL21* gene. In the lower part is represented the LD plot for these four variants in the

European descendent populations that were evaluated in our study [215]. The scheme colour and the values show represent the r^2 values.

CD24 & IBD

As we mention in the Introduction *CD24* is related in the homeostasis of T cells, and is known that an imbalance of this equilibrium is related with develop of AIDs [152]. For that reason *CD24* was a good candidate for genetic association studies in autoimmune pathologies. Two *CD24* functional polymorphisms were described and associated with multiple sclerosis and SLE (Figure 9) [151, 161]. Our group evaluated both polymorphisms rs8734 and rs3838646 in RA, SLE and giant cell arteritis finding significant associations [158-160]. These observations together with the immunologic role of *CD24* make us think that we could be studding a good genetic risk factors common for AIDs and we decided to evaluate the association of both variants in a modest Spanish IBD cohort [212]. We observed for the first time a differential result related with UC and CD. There was a significant association between the rs3838646 and CD but not with UC. Moreover the association indicated a risk effect of the minor allele of the rs3838646 to CD, in opposition to the decreased risk observed in multiple sclerosis and SLE. To this point of our discussion that observation is not surprising because it is in line with the associations of the studied variants of *STAT4* and *PTPN22* in IBD. The relevance of *CD24* in CD is discussed in our manuscript, now we want to comment a particularity about *CD24* gene sequence and the mentioned variants. We checked through an alignment and using human build 37 genome database, which fragment of the human genome the primers used would amplify. The primers presented an identity of 99% with the region of the chromosomes 6 and Y that contain the gene and pseudogene of *CD24,* respectively. For that reason we included only females in our study, ensuring that the region amplified correspond to the chromosome six. At that time both SNPs in the NCBI were referenced in the chromosome 6. But now, there is a missing annotation of the reference assembly on NCBI's build 37.3 because of a gap in the region of the *CD24* that was initially located in 6q21 [153]; and until now, there is not information available for the sequence of *CD24* in this region from the 1000 Genomes Project [97]. That explains in part why the information related to these polymorphisms in AIDs came from CGAS but not from GWAS. Since the *CD24* gene is part of a segmental duplication, special care is required for the identification and

genotyping of SNPs. The SNPs from segmental duplications are well described and can be detected through the Hardy-Weinberg equilibrium test (HW) and in abnormalities in the genotyping patterns [251]. We observed that our IBD studied exhibit normal genotyping patterns and the studied cohort were in HW equilibrium, this together with the functional reports suggest that could be an important role of the *CD24* gene in the pathogenesis of CD and other AIDs. Nevertheless, these results should be taken carefully. Direct sequencing of this gene for the identification of better genetic markers is required to refine the *CD24* genetic role in AIDs.

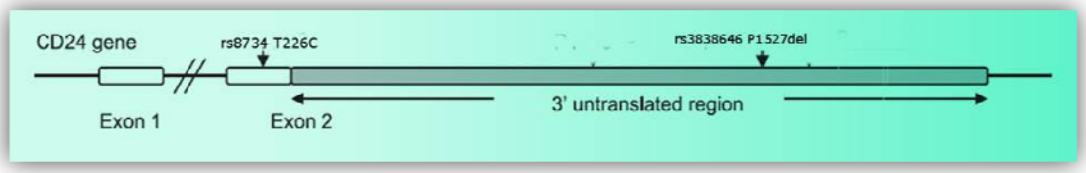


Figure 9. Representation of the *CD24* gene and the location of rs8734 and rs3838646 genetic variants. The cDNA of the *CD24* gene has particular features. Around 10% of its ~2 kb of mRNA, is the protein coding portion; the leading 90% constitute a long 3´UTR [153, 154]. Modified from Wang *et al.* [151]

BAK1, RA & SLE

As we mentioned before a dysregulation in apoptotic process can affect the homeostasis immunologic cells leading the escape of self-antigens activated cells that can produce AIDs. In turn genes like *BAK1* that regulate apoptosis are good candidates to explore associations between polymorphisms in those *loci* and AIDs. We investigate the relationship between three polymorphisms (rs513349, rs561276 and rs5745582) in the *BAK1* gene, RA and SLE in Spanish population (Figure 10) [210]. These SNPs were previously associated with RA, SLE and Sjögren's syndrome in a Colombian population [168]. We observed that these genetic variants were not associated with RA either with SLE in the Spanish population studied. The first possibility that could explain the opposite results is the difference among the genetic structure of Latin American and European descendent populations. We mentioned before the best know example of this, the *PTPN22* variant that have been associated with different AIDs in

European descendent populations but it is not polymorphic in Asian and African descendent populations. Although, there is around 10% of variation of the minor allele frequencies of three *BAK1* polymorphisms between Latino and European descendent populations, the rs513349, rs561276 and rs5745582 SNPs are polymorphic in both populations [97]. Then the difference could be attributed to variation in the LD structure between both studied descendent groups so the genetic markers from the *BAK1* gene are different for Latino and European groups. As we detailed for the *STAT4* gene other kind of evidence, like murine models and measures of the expression levels of the genes in AIDs patients and healthy donors, indicate that these genes have a role in the pathogenic mechanisms but the polymorphisms studied are not the same between different populations. These remark one of the intriguing features of genetics in AIDs, the difference among different racial origins.

There is no evidence from GWAS that polymorphisms from the *BAK1* gene are associated with AIDs [96, 97]. It is prossible that the rs513349, rs561276 and rs5745582 SNPs from this *locus* have a small effect in AIDs and to be detected is required a bigger sample size than the used in our study, as we have showed for the rs2476601 *PTPN22* variant in SSc. The differences in the results observed between Latino and European association studies of *BAK1* polymorphisms in AIDs could be explained also by difference in the predominant subphenotypes from each cohort. The *BAK1* variants could influence a specific manifestation of AIDs that could be better represented in the Latino cohort than in our Spanish population studied. Interestingly, genetic variants from the *BAK1* gene have been consistently associated in GWAS with haematological trait variation in different populations and among the associated SNPs is the rs5745582, one of the associated variants in AIDs in the Latin American population [252-255]. The association of the *BAK1* polymorphisms is specifically with the count of platelets, and there is evidence that showed the relevance of *BAK* family genes in the control of platelet survival and life span [256]. Moreover it is known that hematologic disorders are important manifestation in AIDs; the platelets levels are related with the disease activity in RA and SLE [257, 258]. Together these suggested that the effect of the *BAK1* genetic variants could be related with a very specific hematologic subphenotype of RA and SLE, and difference in the frequency of patients that suffer such clinical manifestations between the studied cohorts could lead the difference in the

results. Unfortunately we do not have enough clinical information from both cohorts to determine that the opposite results are due to such explanation.

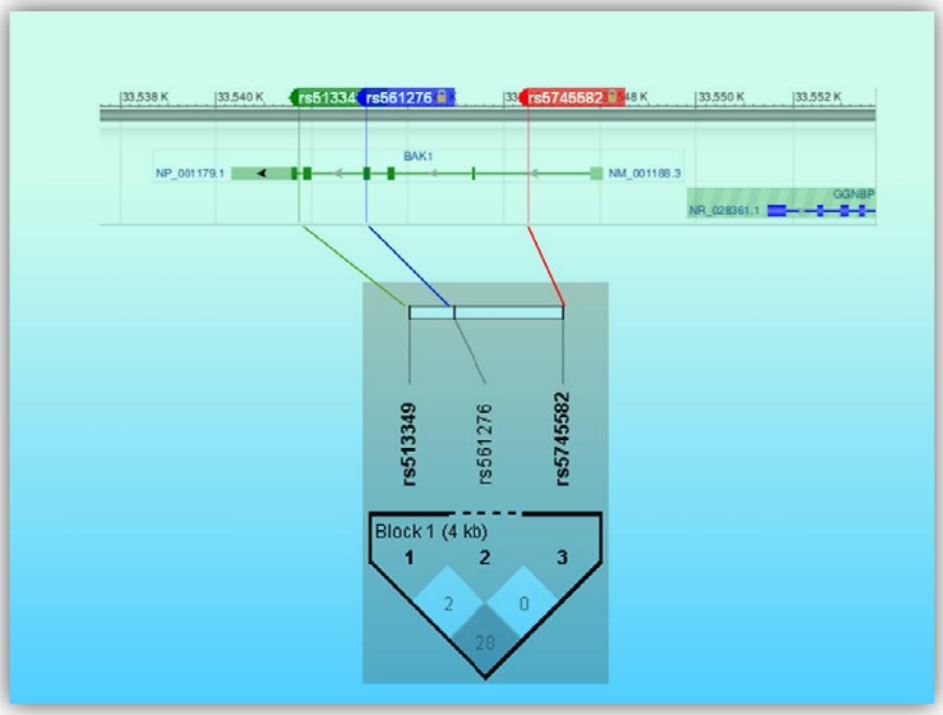


Figure 10. Pattern of LD of the three studied SNPs and their location in the *BAK1* gene. The tree variants show in the figure were associated as genetic susceptibility factors in a Colombian population [168], but the results of our study suggest that these polymorphisms do not have the same effect in RA and SLE patients from Spanish origin [210].

UBASH3a & SLE

UBASH3a is a good example of how the genetic studies in AIDs have helped to find unexplored pathogenic pathways. The protein ubiquitination, analogous to the protein phosphorylation, is regulating immune signal mechanism based on the addition and removal of ubiquitin from proteins[259]. We studied five polymorphisms (rs2277798, rs2277800, rs9976767, rs13048049 and rs17114930) of the *UBASH3a* gene in SLE in a Spanish population. When we started the study and selected the SNPs, the intronic rs9976767 SNP was a strong association signal observed in GWAS in type 1 diabetes [172] so we decided to evaluate its influence in SLE. We included in our study the other four SNPs because they are non-synonymous variants and they had reported

frequency information in the HapMap project (Figure 11) [76]. Our results reported for the first time the influence of *UBASH3a* polymorphisms in SLE European descendent populations [214]. We observed that the minor alleles of the rs2277798 and rs9976767 genetic variants were associated with protection and risk to SLE, respectively. The rs2277798 (A136G) generates a change from serine to glycine at position 18 in exon 1; this change could affect the function of the protein that is involved in apoptotic process. Nevertheless, a regression logistic analysis indicated that the rs9976767, located in the intronic region between the exons 5 and 6, is the SNP that better explain the associations among the studied genetic variants. As we discussed in our manuscript the rs9976767 variant could alter splicing mechanisms of the *UBAHS3a* and the homozygotes for the minor allele are correlated with a higher expression levels of this gene in lymphoblastoid cell lines (Figure 12).

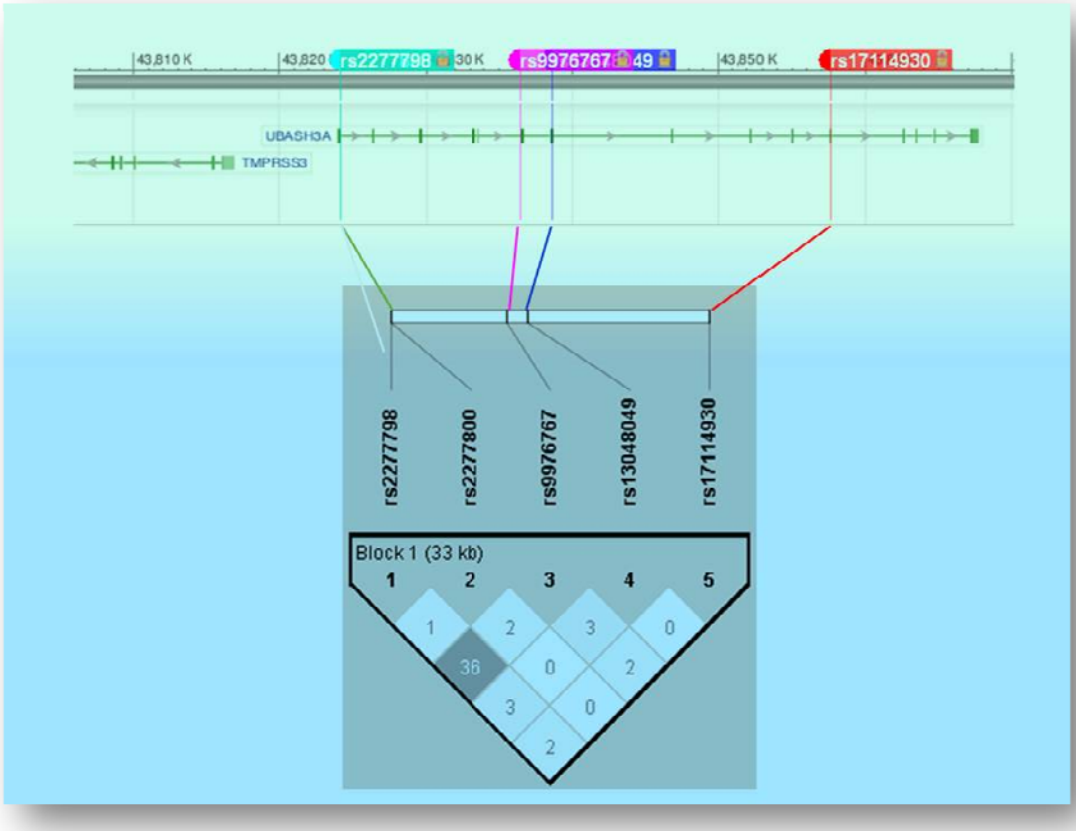


Figure 11. Pattern of LD of the five studied SNPs and their location in the *UBAHS3a* gene. The values correspond to r2 calculated for the Spanish cohort. The rs2277798 polymorphism [G/A] is located in exon 1 of *UBASH3a* gene. It is a non-synonymous change in the position 18 of the protein (S[Ser]/G[Gly]). The rs2277800 polymorphism [C/T] is also located in exon 1 of *UBASH3a* gene and generate a change

in the position 28 of the protein (L[Leu]/F[Phe]). In the other hand, the rs9976767 [A/G] is an intronic variant located between the exons 5 and 6 of the *UBASH3a* gene. Both variants rs13048049 [G/A] and rs17114930 [C/G] are non-synonymous changes in exons 7 and 11, respectively. The first one produce a change from arginine (R[Arg]) to glutamine (Q[Gln]) in position 286; while the rs17114930 polymorphism generates a change from aspartic acid (D[Asp]) to glutamic acid (E[Glu]) in position 428 in Caucasian population.

Opposite to the hypothesis that the non-synonymous *UBASH3a* variants could be strongly associated with SLE, we could highlight that the intronic rs9976767 SNP better explains the association signal form the studied SNPs. This is in line with one of the key results from the study of the genetic of human traits; approximately 40% of the trait-associated SNPs are located in intergenic regions, and another 40% fall in non-coding introns. The other 20% of the associated genetic variants occurred in, or are in tight LD with, protein-coding regions of genes [260]. Indeed the most surprising finding of the sequencing of the human genome is that the majority of functional sequence does not encode proteins. Protein-coding sequences, which comprise only ~1.5% of the genome, are thus dwarfed by functional conserved non-coding elements (CNEs) [9-11]. The association of *UBASH3a* genetic variants with SLE, type 1 diabetes and other AIDs points out that many common genetic association signals in AIDs have been revealed. But because we are in the beginning of the understanding of the function of the human genome the genetic associations with AIDs are like red flags in specific points of a huge field that show us where we should start to look for the treasures.

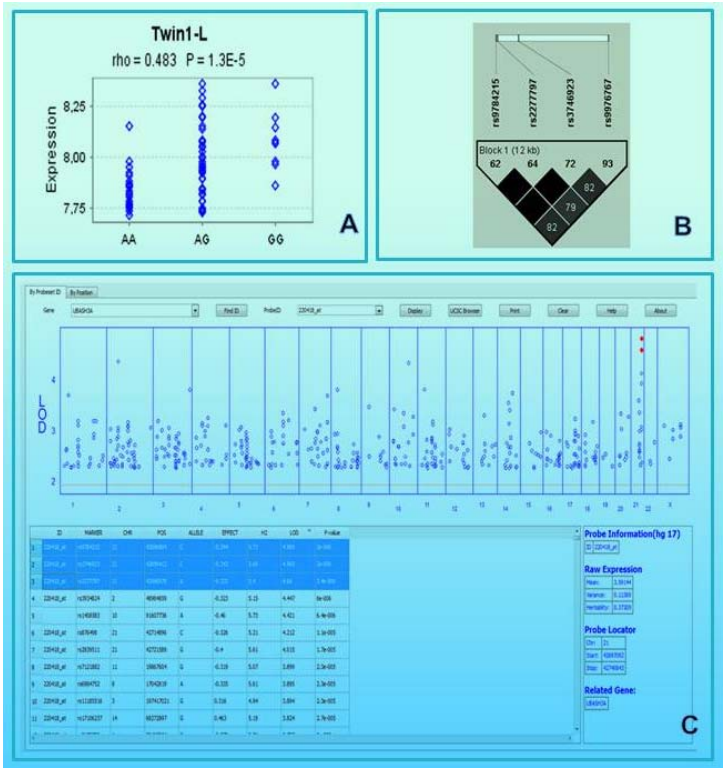


Figure 12. Expression quantitative trait *loci* (eQTL) and *UBASH3a*. Results observed using different eQTL tools to evaluate if there is any relationship between the rs9976767 variant and the *UBASH3a* expression **(A)** SNP-gene association plot for the rs9976767 and the *UBASH3a* gene based on Spearman's rank correlation coefficient (rho) using the Genevar 3.2 software [261]. The eQTL analysis was performed in lymphoblastoid cell lines from peripheral blood sample (n= 74). The plot corresponds to one of the two twins groups studied [262]. **(B)** Linkage disequilibrium (LD) plot performed in Haploview 4.2 [263]. LD plot between rs9976767 and the rs9784215, rs3746923, rs2277797 SNPs which exhibited the highest LOD score (LOD>4.5, P<1E-05) in the *UBAHS3a locus* showed in **(C)** Snapshot of observed eQTLs related with *UBASH3a* gene from the mRNA by SNP Browser 1.0.1 software based on eQTL studies in asthma [264, 265]. The LOD scores and P values for those SNPs are: rs9784215, LOD=4.909 P=2E-06; rs3746923, LOD=4.905 P=2E-06; rs2277797, LOD=4.68 P=3.4E-06. They are signalled as red dots in the LOD plot.

From the SNPs to the Immunologic pathways

The most important contribution of the genetic association studies for the AIDs etiopathogenesis is to highlight possible involved immunologic pathways, based on the genetic association signals [5]. In turn we can cluster the results of this Thesis into i) AIDs associated SNPs of genes involved in T and B cell differentiation signalling: the rs7574865 *STAT4*, the rs6822844 *IL2-IL21* and the rs3838646 *CD24* polymorphisms. ii)

AIDs associated genetic variants of genes involved in immune cell signalling: the rs2476601 and rs33996649 *PTPN22* variants and iii) AIDs associated polymorphisms of genes involved in immune-cell homeostasis: the rs9976767 *UBASH3a* SNP. However, how could this evidence be functional linked with the appearance of an AID or other; for example, the diseases concerning this Thesis RA, SLE, SSc, UC or CD? This is a question which answer could be driven by different hypothesis but still need to be solved. One of these hypotheses is the quantitative thresholds for immune-cell signalling [2, 4, 266]. The autoimmunity is a normal process in the body that is kept in check by a variety of mechanisms, many of which appear to be altered by genetic *loci* such *STAT4*, *IL2-IL21*, *CD24*, *PTPN22* and *UBAHS3a*. The self-reactivity of the immune system is controlled by cell-signal events that are normally regulated within a range of potency that may vary among persons and cell types, due to the genetic and environmental variability and the interaction between them (epigenetics). This has been group in the concept of quantitative thresholds for immune regulation, which means that the immune regulation and reactivity is a trait that exists on a continuum and is set by thresholds for cell activation and reaction [2, 266]. The threshold liability can take any shape, determinated by multiple variable factors, among them the cumulative content of disease susceptibility that an individual inherit (Figure 13) [4].

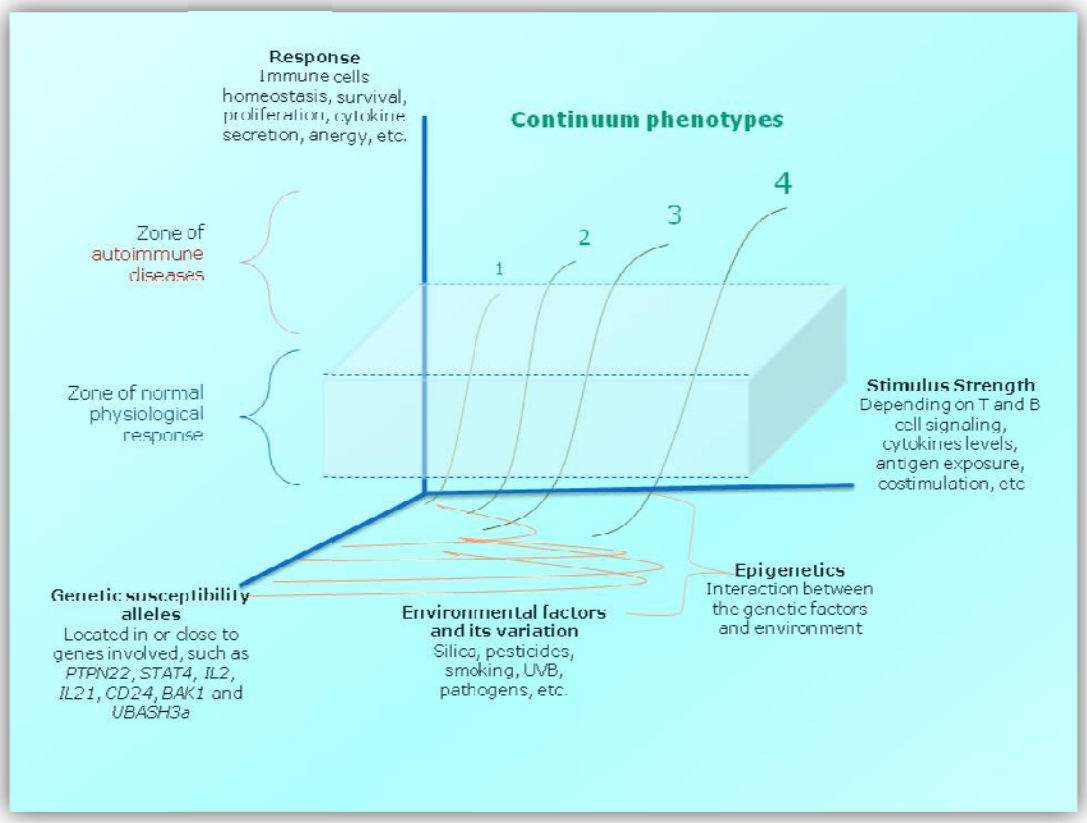


Figure 13. Complex quantitative thresholds model behind the aetiology of AIDs. The immune regulation and reactivity (represented by the axis called **Response**) is a trait that exists on a continuum (**Continuum phenotypes**) and is set by thresholds for cell activation and reaction (represented by the axis **Stimulus Strength**) [2, 266]. The self-reactivity of the immune system is controlled by cell-signal events that are normally regulated within a range of potency that may vary among persons and cell types (**Continuum phenotypes**), due to the genetic (represented by the axis **Genetic Susceptibility alleles**) and environmental variability (red curve lines called **Environmental factors and its variation**) and the interaction between them (**Epigenetics**). The threshold liability (represented as the three-dimensional zone marked as **Zone of normal physiological response** and **Zone of autoimmune diseases**) can take any shape (**Continuum phenotypes**), determinated by multiple variable factors, among them the cumulative content of disease susceptibility that an individual inherit [4].

Other multidimensional factor that adds more complexity to the quantitative thresholds model is the epigenetics. This phenomenon could link the environmental and genetic factors in AIDs. Moreover, part of the explanation of common genetic background in different AIDs, could be given by epigenetics. Due to this mechanism different phenotypes can be originated from identical DNA genomic sequence [267]. Epigenetic

modifications include DNA methylation, histone modifications and nucleosome positioning [32]. There are l*oci* harbouring genetic variants that influence methylation state [268, 269], representing a link between the associated variants and their influence of the regulation of immune related genes expression. Studies of epigenetics and AIDs have shown interesting results of this relation [32, 270, 271]. For example, studies on discordant MZ twins for SLE showed widespread changes in the DNA methylation status of a significant number of genes [270]. T cells from patients with SLE or RA, as well as synovial fibroblasts from individuals with RA, have a lower content of 5‑methylcytosine than their healthy equivalents. That means an overexpression of methylation-regulated genes, implying T cell autoreactivity leading also anti-dsDNA antibody production (reviewed in [32]). Studies in AIDs that integrate the information from candidate gene association studies, GWAs and environmental epidemiologic studies with epigenome-wide association studies could give light on the connexion between the associated SNPs and their role in immune mechanisms.

What is common and different between these autoimmune diseases?

This question could be partially answered along with the background and results of this Thesis. Using a global view, the immunological mechanisms underlying AIDs overlap between them and other inflammatory diseases [5]. The overlapping is beyond the genetic component because environmental factors such infections; crystalline silica and smoking are susceptibility factors for more than one AID [15]. These tree vectors: immunologic mechanisms underlying by genetic component and activated by environmental factors have the phenotypic reflexion in similar clinical manifestations and similar response to the same treatments [44]. Using a microscopic view, the details of the same immunologic mechanisms seem to be driven in different directions in each disease compare to the other. Refinements of the genetic associations identify specific SNPs and genes with opposite risk profile that can cluster the AIDs in subgroups [219]. Moreover, when the genetic polymorphisms are evaluated in subgroups of patients with better define phenotype, specifically with the presence or absence of autoantibodies; the risk to develop that subphenotype increase [82]. The same environmental factor can also have opposite risk profile in different diseases (for example, UV protects against multiple sclerosis but affects the severity of SLE). Then, the phenotypic reflexions of

the microscopic view are presence of different autoantibodies, affectation of different tissues and organs and patients that do not respond to the treatments.

Finally, it is important to remember the main goals of the study of the genetic component of AIDs. They can be group in two [12]: The primary goal is to transform the treatment of common diseases through an understanding of the underlying molecular pathways; which can lead to therapies with broad utility. The secondary goal consists to provide patients with personalized risk prediction. Although the partial risk prediction seems to be feasible and medically useful in some cases in the future, there are expected to be basic limits on precise prediction due to the complex architecture of common traits; common variants with tiny effect, rare variants that cannot be fully enumerated, complex epistatic interactions, environmental factors, as well as the complex to be discover from the regulations systems of the human genome.

FURTHER DIRECTIONS

To think about the next steps to unravel the complete inherited component of AIDs requires stop and evaluate what we have done until now. Data from genetic association studies in AIDs have led and are providing a catalogue of polymorphisms related with these diseases. We have contributed to the expansion of this catalogue through this Thesis, with new data about common genetic component among AIDs. From these premises different questions could be made, such: Is the catalogue of the genetic variants associated with AIDs complete? And what is the role of the associated polymorphisms in the etiopathogenesis of these phenotypes? The questions are simple and barely the same from the beginning of the study of this field. However, the answers have become increasingly complex and remain a work in progress.

The answer for the first question is obvious, around 80% of the AIDs genetic background remains to be discovered [8]. The fulfilment of it requires to move to the next level of complexity that are strictly related with our understanding of the function of the human genome [7, 12, 84, 272]. Also, it is necessary to improve the genetic association studies with different strategies that are starting to be implemented, among them:

1. The replication and meta-analysis of polymorphisms that have been suggested as genetic factors for AIDs mainly in European descendent groups, in the same and different populations. This will help to refine the associations for specific loci, identify difference between populations, and give insight about the causative variant(s). This should be combine with gene-based approach and pathway-based analysis [273].
2. The evaluation of known associated polymorphisms with some AIDs in those diseases no previously studied, as we implemented in this thesis is helping to described specific characteristics of the polymorphisms associated with each disease, to find new associations, to understand better the similarities and differences of the AIDs genetic background. The principle of this strategy is the base of pan-meta-analysis, which is the combination of GWAS of different AIDs [86-89].

3. The analysis of specific subphenotypes of the diseases, grouped by measurable variables such the presence of autoantibodies or the response to the treatment is helping to find stronger associations. In the future, these will be combine with other factors (for example, environmental) to generate prediction models and better treatments of the diseases [82, 274].
4. The interaction between different associated variants and also with environmental factors is one of the areas where the stronger effects of susceptibility to AIDs could be, as like as the relationship between the polymorphisms associated and the functional role of them or the loci where they reside [275].
5. The sequencing of the AIDs associated loci in large sample size groups will help to identify causal variant(s), to reveal bigger effects of rare variants, small effects of common variants and the combination of them probably could be another important part of the missing heritability. Indeed, recent studies argue that the missing heritability of common AIDs could be a result of many common variant loci with weak effect [276, 277].
6. The integration of the information from candidate gene association studies, GWAs and environmental epidemiologic studies with epigenome-wide association studies are also an important way to understand the interaction between the different variables of the complex aetiology of AIDs [32].

The answer to the question about the functional role of the AIDs associated variants is even more incomplete than the number of associated variants itself. As mentioned before, most of the associated variants to AIDs reside in intronic or intergenic regions. The interpretation of the molecular mechanisms of disease-associated loci can be a great challenge; given the diversity of noncoding functions, the incomplete annotation of regulatory elements and the potential existence of unknown mechanisms of regulatory control in the human genome. Nevertheless, some mechanisms through which noncoding variants influence human diseases has been proposed [272]. For example, splicing mechanisms can be altered by polymorphisms located in intronic regions. These alterations include aberrant inclusion of introns, skip of exons or an imbalance of the expression of different isoforms of a gene. Studies in AIDs of alternative splicing have shown the first insights [278, 279]. In fact, preliminary results from one of our

ongoing projects, indicate that there is a differential expression of isoforms of two genes in the chromosome X among SSc, RA patients and healthy donors; moreover it seems to be a relation between the expression of the isoforms and a previous AIDs associated polymorphism [67]. The study of noncoding variants mechanisms on human diseases partially relies on the systematic annotation of regulatory regions. In answer to this necessity, a joint of international efforts is working in the ENCODE consortium that is developing a comprehensive annotation of the noncoding genome [280]. This area should be part of our coming work.

CONCLUSIONS

1. Our data suggest that the minor allele of the *PTPN22* functional variant R620W (rs2476601) confers a moderate risk to SSc and the ACA positive subphenotype in European descendent populations. On the other hand, the rare and functional R263Q SNP (rs33996649), previously associated with RA and SLE, does not have a significant effect in SSc in our study.

2. According with our observations, there is a differential effect of these *PTPN22* variants in the two main subtypes of IBD:

   2.1. The minor allele of the R620W polymorphism exerts a protective effect against CD, while does not have effect in UC in European descendent populations.
   2.2. On the other hand, the minor allele of the R263Q SNP seems to confer risk to UC while is not significant associated with CD.

3. A similar differential association is observed with the rs7574865 polymorphism of the *STAT4* gene in both subphenotypes of IBD in our study. The minor allele seems to be a risk factor to UC, while is not significantly associated with CD.

4. Another observation from our investigation suggests, once again, a differential effect between CD and UC: the minor allele of the rs3838646 *CD24* variant is a risk factor for CD, but it does not show a significant effect on UC. Although functional evidence supports a role of *CD24* in UC, the association studies on this variant should be taken with caution due to the segmental duplication of this gene in the human genome.

5. Our results indicate for the first time that the minor allele of the rs6822844 polymorphism, located in the *IL2-IL21 locus*, is a protective factor for SSc and the lcSSc subphenotype in European descendent populations.

6. The three studied polymorphisms of the *BAK1* gene (rs513349, rs561276 and rs5745582), formerly related with Sjögren syndrome, RA and SLE in the

Colombian population, are not significantly associated with SLE or RA in the Spanish population, according to our study. This fact highlights the heterogeneity across different descendant populations in the AIDs genetic basis.

7. We report for the first time that the minor allele of the rs9976767 intronic variant of the *UBASH3a* gene may represent a risk factor for SLE in populations of European ancestry, pointing the ubiquitination as an important mechanism contributing to the development of AIDs.

8. Taken together, our results support a shared genetic background among different AIDs. Therefore, the study of genetic polymorphisms known to be susceptibility factors of a particular AID, it may represent a good approach to unravel part of the missing heritability of related immune conditions.

ACKNOWLEDGEMENTS

"Los cielos cuentan la gloria de Yahveh,
el firmamento proclama la obra de sus manos.
un día cuenta al otro la noticia,
una noche a la otra se la hace saber.
Sin palabras,
sin lenguaje,
sin una voz perceptible,
por toda la tierra
resuena si eco,
¡sus palabras llegan hasta los confines del mundo!.." Salmo 19:1-4

"The heavens declare the glory of Yahveh,
the skies proclaim the work of his hands.
Day after day they pour forth speech;
night after night they reveal knowledge.
They have no speech, they use no words;
No sound is heard from them.
Yet their voice goes out into all the earth,
Their words to the ends of the word…"
Psalm19:1-4

My first acknowledgement is to this Thesis itself, because it has led me to the most important encounter of all my life: I have met the love that Jesus Christ has for all of us. Moreover, I have got friends that are like family and good friendships in many of my work colleagues. I have matured (not enough) and learned about the amazing value of the simple things of life, through difficult times and invaluable relationships. I have been really happy…Not least; I have met the man of my life, Gustaf Norlén.

I don't have words to thank and declare how important my family back in Colombia has been - all through my life but especially during the last five years. To me they are joy, hope, love, strength and creativity in the shape of human beings. My "Pa and Ma" Marcos and Lucy, my grandma Cecilia, my aunties Tatica, Clarita, Lilo, Yoli, Zulay and Apa; my uncles "Tio J", Tutico and Daniel. The little ones, that nowadays are taller than me (although that's not so difficult); Sophi, Nani, David and Daniel. I also want to say thanks for the different kinds of support that I have received from my new family, that is "new" in time although it feels like I have belonged to them forever: Lars, Marianne, Görel, Börje, Hedvig, Matilda, Magnus, Axel, Emilia and Albin.

I thank my supervisor Javier Martín for the opportunity that he gave me to be part of his lab, for his guidance and support in many aspects. I thank him especially for the opportunity to go to Sweden and for his particular command for me to fulfil there (that I strictly obeyed). I thank all our national and international collaborators, without whom our work wouldn't have been possible: especially to Rosario Gómez-García, Miguel Angel López Nevot, Carlos Cardeña, Norberto Ortego-Centeno, Bobby Koeleman, Marta Alarcón- Riquelme, Angélica Maria Delgado-Vega, Rebecca Roberts and Annika Nordin. I thank Juan-Manuel Anaya for introducing me to the immune-genetics of the AIDs word. I particularly thank Adriana Rojas, Ruben Mantilla and Ricardo Pineda for being my guarantors to get the economic resources for the first period of my PhD studies. I am grateful to Leonid Padyukov and his team for a really nutritive, from both the scientific and human point of view, long stay at the Karolinska Institute.

I want to show particularly appreciation to all my wonderful "lab 206" colleagues. Mercedes, thanks "mi MergaB" for always cheering me up when I've been down, for offering me a different and positive perspective in the difficult moments. Mercedes, Ana and David, thanks for the proofreading of some of the sections of this Thesis. Sofi and David, thanks for your help during that awful visits to the hospital. Lara Susana, thanks for your unique sense of humour that helps to dress up your sweet heart in that "evil mask" (that is also fun). Mamen, thanks for making every day at work more fun, for sharing your knowledge of immunology and for the delicious meals at your home. Maria, thanks for the discussions about science and life, for the care and availability to always help, for being a role model when it comes to organisation and for the innumerable rides in your car. Eze, thanks for always being available to teach and help me with bioinformatics, bureaucratic process and scientific doubts. Elena, thanks for the dances, the bike rides and the laughs. Sonia, thanks for sharing your adventures with Lucia. Gema, thanks for the constant cultural interchange and the daily update of the news. Raquel, thanks for the interesting conversations. Blanca, thanks for teaching me about the lab when I was new, the samples, the analysis and also for the nice time in Alcala de la Real. Dr. Ilein, thanks for being the google about everything, the deadlines, the offers, the concerts, the SLE genes, etc. Roge, thanks for showing me the easy way to do forestplots and your very practical way to do the things. It has been a really good adventure to work with all of you and I'm looking forward to the possibility of

collaborating with you in the future. Also, thanks to all other colleagues from the IPBLN, among them Dianita, Maria Morell, Marta P, Darien, Almu, Maria P, Virginia, Kiko, David, Michael, Enrique, Dorot, Paco.

My dear friends, those who are like brothers and sisters. I will just mention you with hope of having thousands of moments together to show thankfulness. Osmany, Lucía, Luciana, Wes and Celia, Miriam and Nico, Juan and Noe, Javi, Marlena, Nubia, Adri, Agus, Ana, Rubita, Esther and Keren, Andy and Lisi, Mariene, Melissa, Dani, April and Timo, Maryla, Damaris and J, Lamia and Scott, José Daniel and Sara, Isa and Josúe, Shanda, the Sambuc family, Andrea, Chichi, Eva, Ingrid, Emilie, the Carrmet family, Antonio and Pilar, Pipe, Marta Roa and her family, Aninha and Rogerio, Cleo, Marta López Solé, and many others that have brought light to my life.

[Yes, Lara and Eze were right this section of my Thesis is long.]

BIBLIOGRAPHY

1. Cooper, G.S. and B.C. Stroehla, *The epidemiology of autoimmune diseases.* Autoimmun Rev, 2003. **2**(3): p. 119-25.
2. Cho, J.H. and P.K. Gregersen, *Genomics and the multifactorial nature of human autoimmune disease.* N Engl J Med, 2011. **365**(17): p. 1612-23.
3. Marrack, P., J. Kappler, and B.L. Kotzin, *Autoimmune disease: why and where it occurs.* Nat Med, 2001. **7**(8): p. 899-905.
4. Wandstrat, A. and E. Wakeland, *The genetics of complex autoimmune diseases: non-MHC susceptibility genes.* Nat Immunol, 2001. **2**(9): p. 802-9.
5. Zhernakova, A., C.C. van Diemen, and C. Wijmenga, *Detecting shared pathogenesis from the shared genetics of immune-related diseases.* Nat Rev Genet, 2009. **10**(1): p. 43-55.
6. Diaz-Gallo, L.M. and J. Martin, *Common genes in autoimmune diseases: a link between immune-mediated diseases.* Expert Rev Clin Immunol, 2012. **8**(2): p. 107-9.
7. Feero, W.G., A.E. Guttmacher, and F.S. Collins, *Genomic medicine--an updated primer.* N Engl J Med, 2010. **362**(21): p. 2001-11.
8. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.
9. *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.
10. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
11. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
12. Lander, E.S., *Initial impact of the sequencing of the human genome.* Nature, 2011. **470**(7333): p. 187-97.
13. Cotsapas, C., et al., *Pervasive sharing of genetic effects in autoimmune disease.* PLoS Genet, 2011. **7**(8): p. e1002254.
14. Jacobson, D.L., et al., *Epidemiology and estimated population burden of selected autoimmune diseases in the United States.* Clin Immunol Immunopathol, 1997. **84**(3): p. 223-43.
15. Miller, F.W., et al., *Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop.* J Autoimmun, 2012. **39**(4): p. 259-71.
16. Silverstein, A.M., *Autoimmunity versus horror autotoxicus: the struggle for recognition.* Nat Immunol, 2001. **2**(4): p. 279-81.
17. Witebsky, E., N.R. Rose, and S. Shulman, *The autoantibody nature of the thyroiditis antibody and the role of thyroglobulin in the reaction.* Lancet, 1958. **1**(7025): p. 808-10.
18. Davies, A.J., *Immunological tolerance and the autoimmune response.* Autoimmun Rev, 2008. **7**(7): p. 538-43.
19. Rioux, J.D. and A.K. Abbas, *Paths to understanding the genetic basis of autoimmune disease.* Nature, 2005. **435**(7042): p. 584-9.
20. Walker, L.S. and A.K. Abbas, *The enemy within: keeping self-reactive T cells at bay in the periphery.* Nat Rev Immunol, 2002. **2**(1): p. 11-9.
21. Gregersen, P.K. and T.W. Behrens, *Genetics of autoimmune diseases--disorders of immune homeostasis.* Nat Rev Genet, 2006. **7**(12): p. 917-28.
22. Rose, N.R. and C. Bona, *Defining criteria for autoimmune diseases (Witebsky's postulates revisited).* Immunol Today, 1993. **14**(9): p. 426-30.

23. Evoli, A., *Acquired myasthenia gravis in childhood.* Curr Opin Neurol, 2010. **23**(5): p. 536-40.
24. Ambrosi, A. and M. Wahren-Herlenius, *Congenital heart block: evidence for a pathogenic role of maternal autoantibodies.* Arthritis Res Ther, 2012. **14**(2): p. 208.
25. Seok, J., et al., *Genomic responses in mouse models poorly mimic human inflammatory diseases.* Proc Natl Acad Sci U S A, 2013. **110**(9): p. 3507-12.
26. Alarcon-Segovia, D., et al., *Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort.* Arthritis Rheum, 2005. **52**(4): p. 1138-47.
27. Vyse, T.J. and J.A. Todd, *Genetic analysis of autoimmune disease.* Cell, 1996. **85**(3): p. 311-8.
28. Cardenas-Roldan, J., A. Rojas-Villarraga, and J.M. Anaya, *How do autoimmune diseases cluster in families? A systematic review and meta-analysis.* BMC Med, 2013. **11**(1): p. 73.
29. Humbert, P. and J.L. Dupond, *[Multiple autoimmune syndromes].* Ann Med Interne (Paris), 1988. **139**(3): p. 159-68.
30. Anaya, J.M., et al., *The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity.* Expert Rev Clin Immunol, 2007. **3**(4): p. 623-35.
31. Somers, E.C., et al., *Autoimmune diseases co-occurring within individuals and within families: a systematic review.* Epidemiology, 2006. **17**(2): p. 202-17.
32. Ballestar, E., *Epigenetic alterations in autoimmune rheumatic diseases.* Nat Rev Rheumatol, 2011. **7**(5): p. 263-71.
33. Invernizzi, P., et al., *Female predominance and X chromosome defects in autoimmune diseases.* J Autoimmun, 2009. **33**(1): p. 12-6.
34. Quintero, O.L., et al., *Autoimmune disease and gender: plausible mechanisms for the female predominance of autoimmunity.* J Autoimmun, 2012. **38**(2-3): p. J109-19.
35. Andreoli, L., et al., *Pregnancy in autoimmune rheumatic diseases: the importance of counselling for old and new challenges.* Autoimmun Rev, 2010. **10**(1): p. 51-4.
36. Markle, J.G., et al., *Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity.* Science, 2013. **339**(6123): p. 1084-8.
37. Gabrielli, A., E.V. Avvedimento, and T. Krieg, *Scleroderma.* N Engl J Med, 2009. **360**(19): p. 1989-2003.
38. Hughes, M. and A. Herrick, *Systemic sclerosis.* Br J Hosp Med (Lond), 2012. **73**(9): p. 509-10, 511-6.
39. Tsokos, G.C., *Systemic lupus erythematosus.* N Engl J Med, 2011. **365**(22): p. 2110-21.
40. Klareskog, L., A.I. Catrina, and S. Paget, *Rheumatoid arthritis.* Lancet, 2009. **373**(9664): p. 659-72.
41. McInnes, I.B. and G. Schett, *The pathogenesis of rheumatoid arthritis.* N Engl J Med, 2011. **365**(23): p. 2205-19.
42. Ananthakrishnan, A.N. and R.J. Xavier, *How Does Genotype Influence Disease Phenotype in Inflammatory Bowel Disease?* Inflamm Bowel Dis, 2013.
43. Abraham, C. and J.H. Cho, *Inflammatory bowel disease.* N Engl J Med, 2009. **361**(21): p. 2066-78.
44. Anaya, J.M., *Common mechanisms of autoimmune diseases (the autoimmune tautology).* Autoimmun Rev, 2012. **11**(11): p. 781-4.
45. Anaya, J.M., L. Gomez, and J. Castiblanco, *Is there a common genetic basis for autoimmune diseases?* Clin Dev Immunol, 2006. **13**(2-4): p. 185-95.
46. Gregersen, P.K. and L.M. Olsson, *Recent advances in the genetics of autoimmune disease.* Annu Rev Immunol, 2009. **27**: p. 363-91.
47. Lessard, C.J., et al., *The genomics of autoimmune disease in the era of genome-wide association studies and beyond.* Autoimmun Rev, 2012. **11**(4): p. 267-75.

48. Satsangi, J., et al., *Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease.* Lancet, 1996. **347**(9010): p. 1212-7.
49. Shoenfeld, Y., *The kaleidoscope of autoimmunity.* Autoimmunity, 1993. **15**(3): p. 245-52.
50. Silman, A.J. and J.E. Pearson, *Epidemiology and genetics of rheumatoid arthritis.* Arthritis Res, 2002. **4 Suppl 3**: p. S265-72.
51. Germolec, D., et al., *Animal models used to examine the role of the environment in the development of autoimmune disease: findings from an NIEHS Expert Panel Workshop.* J Autoimmun, 2012. **39**(4): p. 285-93.
52. Cooper, G.S., et al., *Evidence of autoimmune-related effects of trichloroethylene exposure from studies in mice and humans.* Environ Health Perspect, 2009. **117**(5): p. 696-702.
53. Klareskog, L., et al., *A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination.* Arthritis Rheum, 2006. **54**(1): p. 38-46.
54. Baka, Z., E. Buzas, and G. Nagy, *Rheumatoid arthritis and smoking: putting the pieces together.* Arthritis Res Ther, 2009. **11**(4): p. 238.
55. Caliskan, R., et al., *The relationship between arthritis and human parvovirus B19 infection.* Rheumatol Int, 2005. **26**(1): p. 7-11.
56. Anzilotti, C., et al., *Antibodies to viral citrullinated peptide in rheumatoid arthritis.* J Rheumatol, 2006. **33**(4): p. 647-51.
57. Deo, S.S., et al., *Detection of Viral Citrullinated Peptide Antibodies Directed Against EBV or VCP: In Early Rheumatoid Arthritis Patients of Indian Origin.* J Lab Physicians, 2010. **2**(2): p. 93-9.
58. James, J.A., et al., *Systemic lupus erythematosus in adults is associated with previous Epstein-Barr virus exposure.* Arthritis Rheum, 2001. **44**(5): p. 1122-6.
59. Yu, S.F., et al., *Detecting Epstein-Barr virus DNA from peripheral blood mononuclear cells in adult patients with systemic lupus erythematosus in Taiwan.* Med Microbiol Immunol, 2005. **194**(3): p. 115-20.
60. Ram, M., et al., *The putative protective role of hepatitis B virus (HBV) infection from autoimmune disorders.* Autoimmun Rev, 2008. **7**(8): p. 621-5.
61. Faria, C., R. Zakout, and M. Araujo, *Helicobacter pylori and autoimmune diseases.* Biomed Pharmacother, 2013. **67**(4): p. 347-9.
62. Fourneau, J.M., et al., *The elusive case for a role of mimicry in autoimmune diseases.* Mol Immunol, 2004. **40**(14-15): p. 1095-102.
63. Proal, A.D., P.J. Albert, and T.G. Marshall, *The human microbiome and autoimmunity.* Curr Opin Rheumatol, 2013. **25**(2): p. 234-40.
64. Ascherio, A. and K.L. Munger, *99th Dahlem conference on infection, inflammation and chronic inflammatory disorders: Epstein-Barr virus and multiple sclerosis: epidemiological evidence.* Clin Exp Immunol, 2010. **160**(1): p. 120-4.
65. Carrel, L. and H.F. Willard, *X-inactivation profile reveals extensive variability in X-linked gene expression in females.* Nature, 2005. **434**(7031): p. 400-4.
66. Carmona, F.D., et al., *New insight on the Xq28 association with systemic sclerosis.* Ann Rheum Dis, 2013.
67. Diaz-Gallo, L.M., Shchetynsky K., Nordin A., Catrina A.I., Martin J., Padyukov L. *Expression of two genes from the candidate locus of chromosome X in Rheumatoid Arthritis and Systemic Sclerosis.* in *Functional Genomics Workshop*. 2013. London, UK: CMM-Karolinska Institutet, Instituto de Parasitologia y Biomedicina Lopez-Neyra-CSIC.
68. Chabchoub, G., et al., *Analysis of skewed X-chromosome inactivation in females with rheumatoid arthritis and autoimmune thyroid diseases.* Arthritis Res Ther, 2009. **11**(4): p. R106.

69. Uz, E., et al., *Skewed X-chromosome inactivation in scleroderma.* Clin Rev Allergy Immunol, 2008. **34**(3): p. 352-5.
70. Broen, J.C., et al., *Skewed X chromosomal inactivation impacts T regulatory cell function in systemic sclerosis.* Ann Rheum Dis, 2010. **69**(12): p. 2213-6.
71. Evans, P.C., et al., *Long-term fetal microchimerism in peripheral blood mononuclear cell subsets in healthy women and women with scleroderma.* Blood, 1999. **93**(6): p. 2033-7.
72. Chan, W.F., et al., *Microchimerism in the rheumatoid nodules of patients with rheumatoid arthritis.* Arthritis Rheum, 2012. **64**(2): p. 380-8.
73. Kremer Hovinga, I.C., et al., *Pregnancy, chimerism and lupus nephritis: a multi-centre study.* Lupus, 2008. **17**(6): p. 541-7.
74. Sahin, A., et al., *Peripheral blood mononuclear cell microchimerism in Turkish female patients with systemic sclerosis.* Mod Rheumatol, 2013.
75. Guttmacher, A.E. and F.S. Collins, *Genomic medicine--a primer.* N Engl J Med, 2002. **347**(19): p. 1512-20.
76. *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.
77. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.
78. Blank, M. and M.E. Gershwin, *Autoimmunity: from the mosaic to the kaleidoscope.* J Autoimmun, 2008. **30**(1-2): p. 1-4.
79. Fernando, M.M., et al., *Transancestral mapping of the MHC region in systemic lupus erythematosus identifies new independent and interacting loci at MSH5, HLA-DPB1 and HLA-G.* Ann Rheum Dis, 2012. **71**(5): p. 777-84.
80. Rioux, J.D., et al., *Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18680-5.
81. Fernando, M.M., et al., *Defining the role of the MHC in autoimmunity: a review and pooled analysis.* PLoS Genet, 2008. **4**(4): p. e1000024.
82. Gorlova, O., et al., *Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy.* PLoS Genet, 2011. **7**(7): p. e1002178.
83. Raychaudhuri, S., et al., *Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis.* Nat Genet, 2012. **44**(3): p. 291-6.
84. Baker, M., *Genomics: The search for association.* Nature, 2010. **467**(7319): p. 1135-8.
85. Hindorff LA, M.J.E.B.I., Morales J (European Bioinformatics Institute), Junkins HA, Hall PN, Klemm AK, and Manolio TA. *A Catalog of Published Genome-Wide Association Studies.* 09/05/2013]; Available from: www.genome.gov/gwastudies.
86. Martin, J.E., et al., *Systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci.* Hum Mol Genet, 2013. **[Epub ahead of print]**.
87. Wang, K., et al., *Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects.* Hum Mol Genet, 2010. **19**(10): p. 2059-67.
88. Festen, E.A., 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**(1): p. e1001283.
89. Zhernakova, A., 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**(2): p. e1002004.
90. Viatte, S., D. Plant, and S. Raychaudhuri, *Genetics and epigenetics of rheumatoid arthritis.* Nat Rev Rheumatol, 2013. **9**(3): p. 141-53.
91. Deng, Y. and B.P. Tsao, *Genetic susceptibility to systemic lupus erythematosus in the genomic era.* Nat Rev Rheumatol, 2010. **6**(12): p. 683-92.

92. Guerra, S.G., T.J. Vyse, and D.S. Cunninghame Graham, *The genetics of lupus: a functional perspective.* Arthritis Res Ther, 2012. **14**(3): p. 211.
93. Khor, B., A. Gardet, and R.J. Xavier, *Genetics and pathogenesis of inflammatory bowel disease.* Nature, 2011. **474**(7351): p. 307-17.
94. Martin, J.E., L. Bossini-Castillo, and J. Martin, *Unraveling the genetic component of systemic sclerosis.* Hum Genet, 2012.
95. Lettre, G. and J.D. Rioux, *Autoimmune diseases: insights from genome-wide association studies.* Hum Mol Genet, 2008. **17**(R2): p. R116-21.
96. Todd, J.A., Burren O., Schofield E., Achuthan P., Giel W. *ImmunoBase*. [cited 2013 09/05/2013]; Available from: https://www.immunobase.org.
97. *The 1000 Genomes Browser*. 2013 10/05/2013]; Available from: http://browser.1000genomes.org/index.html.
98. Cloutier, J.F. and A. Veillette, *Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase.* J Exp Med, 1999. **189**(1): p. 111-21.
99. Hasegawa, K., et al., *PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells.* Science, 2004. **303**(5658): p. 685-9.
100. Wu, J., et al., *Identification of substrates of human protein-tyrosine phosphatase PTPN22.* J Biol Chem, 2006. **281**(16): p. 11002-10.
101. Menard, L., et al., *The PTPN22 allele encoding an R620W variant interferes with the removal of developing autoreactive B cells in humans.* J Clin Invest, 2011. **121**(9): p. 3635-44.
102. Zhang, J., 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**(9): p. 902-7.
103. Bottini, N., et al., *A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes.* Nat Genet, 2004. **36**(4): p. 337-8.
104. Begovich, A.B., et al., *A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis.* Am J Hum Genet, 2004. **75**(2): p. 330-7.
105. Lee, Y.H., et al., *The PTPN22 C1858T functional polymorphism and autoimmune diseases--a meta-analysis.* Rheumatology (Oxford), 2007. **46**(1): p. 49-56.
106. Orozco, G., et al., *Association of a functional single-nucleotide polymorphism of PTPN22, encoding lymphoid protein phosphatase, with rheumatoid arthritis and systemic lupus erythematosus.* Arthritis Rheum, 2005. **52**(1): p. 219-24.
107. Zhernakova, A., et al., *Differential association of the PTPN22 coding variant with autoimmune diseases in a Dutch population.* Genes Immun, 2005. **6**(6): p. 459-61.
108. Barrett, J.C., et al., *Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease.* Nat Genet, 2008. **40**(8): p. 955-62.
109. Franke, A., et al., *Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci.* Nat Genet. **42**(12): p. 1118-25.
110. Gregersen, P.K., et al., *REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis.* Nat Genet, 2009. **41**(7): p. 820-3.
111. Harley, J.B., et al., *Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci.* Nat Genet, 2008. **40**(2): p. 204-10.
112. Plenge, R.M., et al., *TRAF1-C5 as a risk locus for rheumatoid arthritis--a genomewide study.* N Engl J Med, 2007. **357**(12): p. 1199-209.
113. Radstake, T.R., et al., *Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus.* Nat Genet, 2010. **42**(5): p. 426-9.
114. Carlton, V.E., et al., *PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis.* Am J Hum Genet, 2005. **77**(4): p. 567-81.

115. Orru, V., et al., *A loss-of-function variant of PTPN22 is associated with reduced risk of systemic lupus erythematosus.* Hum Mol Genet, 2009. **18**(3): p. 569-79.
116. Rodriguez-Rodriguez, L., et al., *The PTPN22 R263Q polymorphism is a risk factor for rheumatoid arthritis in Caucasian case-control samples.* Arthritis Rheum, 2011. **63**(2): p. 365-72.
117. Kaplan, M.H., *STAT4: a critical regulator of inflammation in vivo.* Immunol Res, 2005. **31**(3): p. 231-42.
118. Amos, C.I., et al., *High-density SNP analysis of 642 Caucasian families with rheumatoid arthritis identifies two new linkage regions on 11p12 and 2q33.* Genes Immun, 2006. **7**(4): p. 277-86.
119. Osorio, Y.F.J., et al., *Dense genome-wide linkage analysis of rheumatoid arthritis, including covariates.* Arthritis Rheum, 2004. **50**(9): p. 2757-65.
120. Cantor, R.M., et al., *Systemic lupus erythematosus genome scan: support for linkage at 1q23, 2q33, 16q12-13, and 17q21-23 and novel evidence at 3p24, 10q23-24, 13q32, and 18q22-23.* Arthritis Rheum, 2004. **50**(10): p. 3203-10.
121. Remmers, E.F., et al., *STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus.* N Engl J Med, 2007. **357**(10): p. 977-86.
122. Abelson, A.K., et al., *STAT4 associates with systemic lupus erythematosus through two independent effects that correlate with gene expression and act additively with IRF5 to increase risk.* Ann Rheum Dis, 2009. **68**(11): p. 1746-53.
123. Lee, H.S., et al., *Association of STAT4 with rheumatoid arthritis in the Korean population.* Mol Med, 2007. **13**(9-10): p. 455-60.
124. Palomino-Morales, R.J., et al., *STAT4 but not TRAF1/C5 variants influence the risk of developing rheumatoid arthritis and systemic lupus erythematosus in Colombians.* Genes Immun, 2008. **9**(4): p. 379-82.
125. Korman, B.D., et al., *Variant form of STAT4 is associated with primary Sjogren's syndrome.* Genes Immun, 2008. **9**(3): p. 267-70.
126. Palomino-Morales, R.J., et al., *Influence of STAT4 polymorphism in primary Sjogren's syndrome.* J Rheumatol, 2010. **37**(5): p. 1016-9.
127. Rueda, B., et al., *The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype.* Hum Mol Genet, 2009. **18**(11): p. 2071-7.
128. Martinez, A., et al., *Association of the STAT4 gene with increased susceptibility for some immune-mediated diseases.* Arthritis Rheum, 2008. **58**(9): p. 2598-602.
129. Allanore, Y., et al., *Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis.* PLoS Genet, 2011. **7**(7): p. e1002091.
130. Chung, S.A., et al., *Differential genetic associations for systemic lupus erythematosus based on anti-dsDNA autoantibody production.* PLoS Genet, 2011. **7**(3): p. e1001323.
131. Han, J.W., et al., *Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus.* Nat Genet, 2009. **41**(11): p. 1234-7.
132. Hom, G., et al., *Association of systemic lupus erythematosus with C8orf13-BLK and ITGAM-ITGAX.* N Engl J Med, 2008. **358**(9): p. 900-9.
133. Kochi, Y., et al., *A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility.* Nat Genet, 2010. **42**(6): p. 515-9.
134. *Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls.* Nature, 2007. **447**(7145): p. 661-78.
135. Todd, J.A., et al., *Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes.* Nat Genet, 2007. **39**(7): p. 857-64.
136. van Heel, D.A., et al., *A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21.* Nat Genet, 2007. **39**(7): p. 827-9.
137. Sawalha, A.H., et al., *Genetic association of interleukin-21 polymorphisms with systemic lupus erythematosus.* Ann Rheum Dis, 2008. **67**(4): p. 458-61.

138. Zhernakova, A., et al., *Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases.* Am J Hum Genet, 2007. **81**(6): p. 1284-8.
139. Barton, A., et al., *Identification of AF4/FMR2 family, member 3 (AFF3) as a novel rheumatoid arthritis susceptibility locus and confirmation of two further pan-autoimmune susceptibility genes.* Hum Mol Genet, 2009. **18**(13): p. 2518-22.
140. Hollis-Moffatt, J.E., et al., *Only one independent genetic association with rheumatoid arthritis within the KIAA1109-TENR-IL2-IL21 locus in Caucasian sample sets: confirmation of association of rs6822844 with rheumatoid arthritis at a genome-wide level of significance.* Arthritis Res Ther, 2010. **12**(3): p. R116.
141. Festen, E.A., et al., *Genetic variants in the region harbouring IL2/IL21 associated with ulcerative colitis.* Gut, 2009. **58**(6): p. 799-804.
142. Glas, J., et al., *Novel genetic risk markers for ulcerative colitis in the IL2/IL21 region are in epistasis with IL23R and suggest a common genetic background for ulcerative colitis and celiac disease.* Am J Gastroenterol, 2009. **104**(7): p. 1737-44.
143. Hollis-Moffatt, J.E., et al., *Consolidation of evidence for association of the KIAA1109-TENR-IL2-IL21 rs6822844 variant with Crohn's disease.* Am J Gastroenterol, 2010. **105**(5): p. 1204-5; author reply 1206-7.
144. Marquez, A., et al., *Novel association of the interleukin 2-interleukin 21 region with inflammatory bowel disease.* Am J Gastroenterol, 2009. **104**(8): p. 1968-75.
145. Warren, R.B., et al., *A systematic investigation of confirmed autoimmune loci in early-onset psoriasis reveals an association with IL2/IL21.* Br J Dermatol, 2011. **164**(3): p. 660-4.
146. Hughes, T., et al., *Fine-mapping and transethnic genotyping establish IL2/IL21 genetic association with lupus and localize this genetic effect to IL21.* Arthritis Rheum, 2011. **63**(6): p. 1689-97.
147. Crispin, J.C. and G.C. Tsokos, *Transcriptional regulation of IL-2 in health and autoimmunity.* Autoimmun Rev, 2009. **8**(3): p. 190-5.
148. Yamanouchi, J., et al., *Interleukin-2 gene variation impairs regulatory T cell function and causes autoimmunity.* Nat Genet, 2007. **39**(3): p. 329-37.
149. Wei, L., et al., *IL-21 is produced by Th17 cells and drives IL-17 production in a STAT3-dependent manner.* J Biol Chem, 2007. **282**(48): p. 34605-10.
150. Ettinger, R., S. Kuchen, and P.E. Lipsky, *Interleukin 21 as a target of intervention in autoimmune disease.* Ann Rheum Dis, 2008. **67 Suppl 3**: p. iii83-6.
151. Wang, L., et al., *A dinucleotide deletion in CD24 confers protection against autoimmune diseases.* PLoS Genet, 2007. **3**(4): p. e49.
152. Liu, Y. and P. Zheng, *CD24: a genetic checkpoint in T cell homeostasis and autoimmune diseases.* Trends Immunol, 2007. **28**(7): p. 315-20.
153. *CD24, GeneID 100133941*. 2013 10/05/2013]; Available from: http://www-ncbi-nlm-nih-gov.proxy.kib.ki.se/gene/?term=100133941.
154. Hough, M.R., et al., *Mapping of CD24 and homologous sequences to multiple chromosomal loci.* Genomics, 1994. **22**(1): p. 154-61.
155. Inoue, N., et al., *The association between a functional polymorphism in the CD24 gene and the development of autoimmune thyroid diseases.* Tissue Antigens, 2013. **81**(3): p. 161-3.
156. Kollaee, A., et al., *Investigation of CD24 and its expression in Iranian relapsing-remitting multiple sclerosis.* Int J Neurosci, 2011. **121**(12): p. 684-90.
157. Otaegui, D., et al., *CD24 V/V is an allele associated with the risk of developing multiple sclerosis in the Spanish population.* Mult Scler, 2006. **12**(4): p. 511-4.
158. Rueda, B., et al., *Association of CD24 gene polymorphisms with susceptibility to biopsy-proven giant cell arteritis.* J Rheumatol, 2008. **35**(5): p. 850-4.

159. Sanchez, E., et al., *Association of a CD24 gene polymorphism with susceptibility to systemic lupus erythematosus.* Arthritis Rheum, 2007. **56**(9): p. 3080-6.
160. Sanchez, E., et al., *Investigating the role of CD24 gene polymorphisms in rheumatoid arthritis.* Ann Rheum Dis, 2008. **67**(8): p. 1197-8.
161. Zhou, Q., et al., *CD24 is a genetic modifier for risk and progression of multiple sclerosis.* Proc Natl Acad Sci U S A, 2003. **100**(25): p. 15041-6.
162. Strasser, A., *The role of BH3-only proteins in the immune system.* Nat Rev Immunol, 2005. **5**(3): p. 189-200.
163. Takeuchi, O., et al., *Essential role of BAX,BAK in B cell homeostasis and prevention of autoimmune disease.* Proc Natl Acad Sci U S A, 2005. **102**(32): p. 11272-7.
164. Anaya, J.M., R.D. Mantilla, and P.A. Correa, *Immunogenetics of primary Sjogren's syndrome in Colombians.* Semin Arthritis Rheum, 2005. **34**(5): p. 735-43.
165. Banisor, I. and B. Kalman, *Bcl-2 and its homologues in the brain of patients with multiple sclerosis.* Mult Scler, 2004. **10**(2): p. 176-81.
166. Chernavsky, A.C., et al., *Evidences of the involvement of Bak, a member of the Bcl-2 family of proteins, in active coeliac disease.* Autoimmunity, 2002. **35**(1): p. 29-37.
167. Hiromatsu, Y., et al., *Immunohistochemical analysis of bcl-2, Bax and Bak expression in thyroid glands from patients with Graves' disease.* Endocr J, 2004. **51**(4): p. 399-405.
168. Delgado-Vega, A.M., et al., *Bcl-2 antagonist killer 1 (BAK1) polymorphisms influence the risk of developing autoimmune rheumatic diseases in women.* Ann Rheum Dis, 2010. **69**(2): p. 462-5.
169. Collingwood, T.S., et al., *T-cell ubiquitin ligand affects cell death through a functional interaction with apoptosis-inducing factor, a key factor of caspase-independent apoptosis.* J Biol Chem, 2007. **282**(42): p. 30920-8.
170. Tsygankov, A.Y., *TULA-family proteins: an odd couple.* Cell Mol Life Sci, 2009. **66**(17): p. 2949-52.
171. Zenewicz, L.A., et al., *Unraveling the genetics of autoimmunity.* Cell, 2010. **140**(6): p. 791-7.
172. Grant, S.F., et al., *Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes.* Diabetes, 2009. **58**(1): p. 290-5.
173. Jin, Y., et al., *Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo.* N Engl J Med, 2010. **362**(18): p. 1686-97.
174. Smyth, D.J., et al., *Shared and distinct genetic variants in type 1 diabetes and celiac disease.* N Engl J Med, 2008. **359**(26): p. 2767-77.
175. OMIM®. *Mendelian Inheritance in Man. A Catalog of Human Genes and Genetic Disorders.* 17/04/2013]; Available from: http://omim.org/.
176. Klareskog, L., et al., *Immunity to citrullinated proteins in rheumatoid arthritis.* Annu Rev Immunol, 2008. **26**: p. 651-75.
177. Nishimura, K., et al., *Meta-analysis: diagnostic accuracy of anti-cyclic citrullinated peptide antibody and rheumatoid factor for rheumatoid arthritis.* Ann Intern Med, 2007. **146**(11): p. 797-808.
178. Nielen, M.M., et al., *Specific autoantibodies precede the symptoms of rheumatoid arthritis: a study of serial measurements in blood donors.* Arthritis Rheum, 2004. **50**(2): p. 380-6.
179. Gregersen, P.K., J. Silver, and R.J. Winchester, *The shared epitope hypothesis. An approach to understanding the molecular genetics of susceptibility to rheumatoid arthritis.* Arthritis Rheum, 1987. **30**(11): p. 1205-13.
180. Mahdi, H., et al., *Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis.* Nat Genet, 2009. **41**(12): p. 1319-24.
181. MacGregor, A.J., et al., *Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins.* Arthritis Rheum, 2000. **43**(1): p. 30-7.

182. Arnett, F.C., et al., *The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis.* Arthritis Rheum, 1988. **31**(3): p. 315-24.
183. Hochberg, M.C., *Updating the American College of Rheumatology revised criteria for the classification of systemic lupus erythematosus.* Arthritis Rheum, 1997. **40**(9): p. 1725.
184. *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**(5): p. 581-90.
185. LeRoy, E.C., et al., *Scleroderma (systemic sclerosis): classification, subsets and pathogenesis.* J Rheumatol, 1988. **15**(2): p. 202-5.
186. Satsangi, J., et al., *The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications.* Gut, 2006. **55**(6): p. 749-53.
187. Jolly, M., *How does quality of life of patients with systemic lupus erythematosus compare with that of other common chronic illnesses?* J Rheumatol, 2005. **32**(9): p. 1706-8.
188. Pons-Estel, G.J., et al., *Understanding the epidemiology and progression of systemic lupus erythematosus.* Semin Arthritis Rheum, 2010. **39**(4): p. 257-68.
189. Rahman, A. and D.A. Isenberg, *Systemic lupus erythematosus.* N Engl J Med, 2008. **358**(9): p. 929-39.
190. Deapen, D., et al., *A revised estimate of twin concordance in systemic lupus erythematosus.* Arthritis Rheum, 1992. **35**(3): p. 311-8.
191. Moser, K.L., et al., *Recent insights into the genetic basis of systemic lupus erythematosus.* Genes Immun, 2009. **10**(5): p. 373-9.
192. Castano-Rodriguez, N., et al., *Meta-analysis of HLA-DRB1 and HLA-DQB1 polymorphisms in Latin American patients with systemic lupus erythematosus.* Autoimmun Rev, 2008. **7**(4): p. 322-30.
193. Walker, J.G., et al., *The development of systemic sclerosis classification criteria.* Clin Rheumatol, 2007. **26**(9): p. 1401-9.
194. Nihtyanova, S.I. and C.P. Denton, *Autoantibodies as predictive tools in systemic sclerosis.* Nat Rev Rheumatol, 2010. **6**(2): p. 112-6.
195. Reveille, J.D., *Ethnicity and race and systemic sclerosis: how it affects susceptibility, severity, antibody genetics, and clinical manifestations.* Curr Rheumatol Rep, 2003. **5**(2): p. 160-7.
196. Mayes, M.D., et al., *Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population.* Arthritis Rheum, 2003. **48**(8): p. 2246-55.
197. Feghali-Bostwick, C., T.A. Medsger, Jr., and T.M. Wright, *Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies.* Arthritis Rheum, 2003. **48**(7): p. 1956-63.
198. Arora-Singh, R.K., et al., *Autoimmune diseases and autoantibodies in the first degree relatives of patients with systemic sclerosis.* J Autoimmun, 2010. **35**(1): p. 52-7.
199. Hudson, M., et al., *Polyautoimmunity and familial autoimmunity in systemic sclerosis.* J Autoimmun, 2008. **31**(2): p. 156-9.
200. Koumakis, E., et al., *Familial autoimmunity in systemic sclerosis -- results of a French-based case-control family study.* J Rheumatol, 2012. **39**(3): p. 532-8.
201. Zhou, X., 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**(12): p. 3807-14.
202. Behr, M.A., M. Divangahi, and J.D. Lalande, *What's in a name? The (mis)labelling of Crohn's as an autoimmune disease.* Lancet, 2010. **376**(9736): p. 202-3.
203. Kuna, A.T., *Serological markers of inflammatory bowel disease.* Biochem Med (Zagreb), 2013. **23**(1): p. 28-42.

204. Loftus, E.V., Jr., *Clinical epidemiology of inflammatory bowel disease: Incidence, prevalence, and environmental influences.* Gastroenterology, 2004. **126**(6): p. 1504-17.
205. Silverberg, M.S., et al., *Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: Report of a Working Party of the 2005 Montreal World Congress of Gastroenterology.* Can J Gastroenterol, 2005. **19 Suppl A**: p. 5-36.
206. Halfvarson, J., et al., *Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics.* Gastroenterology, 2003. **124**(7): p. 1767-73.
207. Orholm, M., et al., *Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study.* Scand J Gastroenterol, 2000. **35**(10): p. 1075-81.
208. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.
209. Diaz-Gallo, L.M., et al., *Differential association of two PTPN22 coding variants with Crohn's disease and ulcerative colitis.* Inflamm Bowel Dis, 2011. **17**(11): p. 2287-94.
210. Diaz-Gallo, L.M., et al., *Association study of BAK1 gene polymorphisms in Spanish rheumatoid arthritis and systemic lupus erythematosus cohorts.* Ann Rheum Dis, 2012. **71**(2): p. 314-6.
211. Diaz-Gallo, L.M., et al., *Analysis of the influence of PTPN22 gene polymorphisms in systemic sclerosis.* Ann Rheum Dis, 2011. **70**(3): p. 454-62.
212. Diaz-Gallo, L.M., et al., *Analysis of the influence of two CD24 genetic variants in Crohn's disease and ulcerative colitis.* Hum Immunol, 2011. **72**(10): p. 969-72.
213. Diaz-Gallo, L.M., et al., *STAT4 gene influences genetic predisposition to ulcerative colitis but not Crohn's disease in the Spanish population: a replication study.* Hum Immunol, 2010. **71**(5): p. 515-9.
214. Diaz-Gallo, L.M., et al., *Evidence of new risk genetic factor to systemic lupus erythematosus: the UBASH3A gene.* PLoS One, 2013. **8**(4): p. e60646.
215. Diaz-Gallo, L.M., et al., *Implication of IL-2/IL-21 region in systemic sclerosis genetic susceptibility.* Ann Rheum Dis, 2012.
216. Viatte, S., et al., *Genetic markers of rheumatoid arthritis susceptibility in anti-citrullinated peptide antibody negative patients.* Ann Rheum Dis, 2012. **71**(12): p. 1984-90.
217. Svendsen, A.J., et al., *The impact of genes on the occurrence of autoantibodies in rheumatoid arthritis. A study on disease discordant twin pairs.* J Autoimmun, 2013. **41**: p. 120-5.
218. Orozco, G., et al., *Analysis of a functional BTNL2 polymorphism in type 1 diabetes, rheumatoid arthritis, and systemic lupus erythematosus.* Hum Immunol, 2005. **66**(12): p. 1235-41.
219. Sirota, M., et al., *Autoimmune disease classification by inverse association with SNP alleles.* PLoS Genet, 2009. **5**(12): p. e1000792.
220. Boechat, A.L., et al., *Association between the PTPN22 1858C/T gene polymorphism and tuberculosis resistance.* Infect Genet Evol, 2013. **16C**: p. 310-313.
221. Gomez, L.M., J.M. Anaya, and J. Martin, *Genetic influence of PTPN22 R620W polymorphism in tuberculosis.* Hum Immunol, 2005. **66**(12): p. 1242-7.
222. Lamsyah, H., et al., *Association of PTPN22 gene functional variants with development of pulmonary tuberculosis in Moroccan population.* Tissue Antigens, 2009. **74**(3): p. 228-32.
223. Ramirez, M., et al., *The PTPN22 C1858T variant as a risk factor for rheumatoid arthritis and systemic lupus erythematosus but not for systemic sclerosis in the Colombian population.* Clin Exp Rheumatol, 2012. **30**(4): p. 520-4.
224. Gomez, L.M., et al., *PTPN22 C1858T polymorphism in Colombian patients with autoimmune diseases.* Genes Immun, 2005. **6**(7): p. 628-31.

225. Kawasaki, E., et al., *Systematic search for single nucleotide polymorphisms in a lymphoid tyrosine phosphatase gene (PTPN22): association between a promoter polymorphism and type 1 diabetes in Asian populations.* Am J Med Genet A, 2006. **140**(6): p. 586-93.
226. Huang, J.J., et al., *A PTPN22 promoter polymorphism -1123G>C is associated with RA pathogenesis in Chinese.* Rheumatol Int, 2012. **32**(3): p. 767-71.
227. Mori, M., et al., *Ethnic differences in allele frequency of autoimmune-disease-associated SNPs.* J Hum Genet, 2005. **50**(5): p. 264-6.
228. Kaufman, K.M., et al., *Fine mapping of Xq28: both MECP2 and IRAK1 contribute to risk for systemic lupus erythematosus in multiple ancestral groups.* Ann Rheum Dis, 2013. **72**(3): p. 437-44.
229. Vang, T., et al., *Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant.* Nat Genet, 2005. **37**(12): p. 1317-9.
230. Rieck, M., et al., *Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes.* J Immunol, 2007. **179**(7): p. 4704-10.
231. Glas, J., et al., *Evidence for STAT4 as a common autoimmune gene: rs7574865 is associated with colonic Crohn's disease and early disease onset.* PLoS One, 2010. **5**(4): p. e10373.
232. Moon, C.M., et al., *Association of signal transducer and activator of transcription 4 genetic variants with extra-intestinal manifestations in inflammatory bowel disease.* Life Sci, 2010. **86**(17-18): p. 661-7.
233. Bouzid, D., et al., *The CREM gene is involved in genetic predisposition to inflammatory bowel disease in the Tunisian population.* Hum Immunol, 2011. **72**(12): p. 1204-9.
234. Mudter, J., et al., *Activation pattern of signal transducers and activators of transcription (STAT) factors in inflammatory bowel diseases.* Am J Gastroenterol, 2005. **100**(1): p. 64-72.
235. Pang, Y.H., et al., *Increased expression and activation of IL-12-induced Stat4 signaling in the mucosa of ulcerative colitis patients.* Cell Immunol, 2007. **248**(2): p. 115-20.
236. Kim, S.W., et al., *Abnormal genetic and epigenetic changes in signal transducer and activator of transcription 4 in the pathogenesis of inflammatory bowel diseases.* Dig Dis Sci, 2012. **57**(10): p. 2600-7.
237. Simpson, S.J., et al., *T cell-mediated pathology in two models of experimental colitis depends predominantly on the interleukin 12/Signal transducer and activator of transcription (Stat)-4 pathway, but is not conditional on interferon gamma expression by T cells.* J Exp Med, 1998. **187**(8): p. 1225-34.
238. Tuller, T., et al., *Common and specific signatures of gene expression and protein-protein interactions in autoimmune diseases.* Genes Immun, 2013. **14**(2): p. 67-82.
239. Malek, T.R. and A.L. Bayer, *Tolerance, not immunity, crucially depends on IL-2.* Nat Rev Immunol, 2004. **4**(9): p. 665-74.
240. Fontenot, J.D., et al., *A function for interleukin 2 in Foxp3-expressing regulatory T cells.* Nat Immunol, 2005. **6**(11): p. 1142-51.
241. Spolski, R. and W.J. Leonard, *Interleukin-21: basic biology and implications for cancer and autoimmunity.* Annu Rev Immunol, 2008. **26**: p. 57-79.
242. Iwata, Y., et al., *Characterization of a rare IL-10-competent B-cell subset in humans that parallels mouse regulatory B10 cells.* Blood, 2011. **117**(2): p. 530-41.
243. Yoshizaki, A., et al., *Regulatory B cells control T-cell autoimmunity through IL-21-dependent cognate interactions.* Nature, 2012. **491**(7423): p. 264-8.
244. Biswas, P.S., et al., *A murine autoimmune model of rheumatoid arthritis and systemic lupus erythematosus associated with deregulated production of IL-17 and IL-21.* Methods Mol Biol, 2012. **900**: p. 233-51.
245. Becker, K.G., *The common variants/multiple disease hypothesis of common complex genetic disorders.* Med Hypotheses, 2004. **62**(2): p. 309-17.

246. Pritchard, J.K., *Are rare variants responsible for susceptibility to complex diseases?* Am J Hum Genet, 2001. **69**(1): p. 124-37.
247. Dickson, S.P., et al., *Rare variants create synthetic genome-wide associations.* PLoS Biol, 2010. **8**(1): p. e1000294.
248. Orozco, G., J.C. Barrett, and E. Zeggini, *Synthetic associations in the context of genome-wide association scan signals.* Hum Mol Genet, 2010. **19**(R2): p. R137-44.
249. Visser, M., M. Kayser, and R.J. Palstra, *HERC2 rs12913832 modulates human pigmentation by attenuating chromatin-loop formation between a long-range enhancer and the OCA2 promoter.* Genome Res, 2012. **22**(3): p. 446-55.
250. Parker, S.C., et al., *Local DNA topography correlates with functional noncoding regions of the human genome.* Science, 2009. **324**(5925): p. 389-92.
251. Fredman, D., et al., *Complex SNP-related sequence variation in segmental genome duplications.* Nat Genet, 2004. **36**(8): p. 861-6.
252. Kamatani, Y., et al., *Genome-wide association study of hematological and biochemical traits in a Japanese population.* Nat Genet, 2010. **42**(3): p. 210-5.
253. Li, J., et al., *GWAS of blood cell traits identifies novel associated loci and epistatic interactions in Caucasian and African-American children.* Hum Mol Genet, 2013. **22**(7): p. 1457-64.
254. Lo, K.S., et al., *Genetic association analysis highlights new loci that modulate hematological trait variation in Caucasians and African Americans.* Hum Genet, 2011. **129**(3): p. 307-17.
255. Soranzo, N., et al., *A genome-wide meta-analysis identifies 22 loci associated with eight hematological parameters in the HaemGen consortium.* Nat Genet, 2009. **41**(11): p. 1182-90.
256. Mason, K.D., et al., *Programmed anuclear cell death delimits platelet life span.* Cell, 2007. **128**(6): p. 1173-86.
257. Yazici, S., et al., *The platelet indices in patients with rheumatoid arthritis: mean platelet volume reflects disease activity.* Platelets, 2010. **21**(2): p. 122-5.
258. Ziakas, P.D., et al., *Thrombocytopaenia in lupus as a marker of adverse outcome--seeking Ariadne's thread.* Rheumatology (Oxford), 2006. **45**(10): p. 1261-5.
259. Bhoj, V.G. and Z.J. Chen, *Ubiquitylation in innate and adaptive immunity.* Nature, 2009. **458**(7237): p. 430-7.
260. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.
261. Yang, T.P., et al., *Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies.* Bioinformatics, 2010. **26**(19): p. 2474-6.
262. Nica, A.C., et al., *The architecture of gene regulatory variation across multiple human tissues: the MuTHER study.* PLoS Genet, 2011. **7**(2): p. e1002003.
263. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
264. Dixon, A.L., et al., *A genome-wide association study of global gene expression.* Nat Genet, 2007. **39**(10): p. 1202-7.
265. Moffatt, M.F., et al., *Genetic variants regulating ORMDL3 expression contribute to the risk of childhood asthma.* Nature, 2007. **448**(7152): p. 470-3.
266. Liston, A., et al., *Genetic lesions in T-cell tolerance and thresholds for autoimmunity.* Immunol Rev, 2005. **204**: p. 87-101.
267. Fraga, M.F., et al., *Epigenetic differences arise during the lifetime of monozygotic twins.* Proc Natl Acad Sci U S A, 2005. **102**(30): p. 10604-9.
268. Kerkel, K., et al., *Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation.* Nat Genet, 2008. **40**(7): p. 904-8.

269. Bell, J.T., et al., *DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines.* Genome Biol, 2011. **12**(1): p. R10.
270. Javierre, B.M., et al., *Changes in the pattern of DNA methylation associate with twin discordance in systemic lupus erythematosus.* Genome Res, 2010. **20**(2): p. 170-9.
271. Liu, Y., et al., *Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis.* Nat Biotechnol, 2013. **31**(2): p. 142-7.
272. Ward, L.D. and M. Kellis, *Interpreting noncoding genetic variation in complex traits and human disease.* Nat Biotechnol, 2012. **30**(11): p. 1095-106.
273. Fu, J., E.A. Festen, and C. Wijmenga, *Multi-ethnic studies in complex traits.* Hum Mol Genet, 2011. **20**(R2): p. R206-13.
274. Taylor, K.E., et al., *Risk alleles for systemic lupus erythematosus in a large case-control collection and associations with clinical subphenotypes.* PLoS Genet, 2011. **7**(2): p. e1001311.
275. Delgado-Vega, A.M., et al., *Fine mapping and conditional analysis identify a new mutation in the autoimmunity susceptibility gene BLK that leads to reduced half-life of the BLK protein.* Ann Rheum Dis, 2012. **71**(7): p. 1219-26.
276. Hunt, K.A., et al., *Negligible impact of rare autoimmune-locus coding-region variants on missing heritability.* Nature, 2013.
277. Stahl, E.A., et al., *Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis.* Nat Genet, 2012. **44**(5): p. 483-9.
278. Diaz-Gallo, L.M. and J. Martin, *PTPN22 splice forms: a new role in rheumatoid arthritis.* Genome Med, 2012. **4**(2): p. 13.
279. Moulton, V.R., et al., *Splicing factor SF2/ASF rescues IL-2 production in T cells from systemic lupus erythematosus patients by activating IL-2 transcription.* Proc Natl Acad Sci U S A, 2013. **110**(5): p. 1845-50.
280. Dunham, I., et al., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.