Title: Genetic predictors of systemic sclerosis-associated interstitial lung disease: a review of recent literature

Running title: Genetics of SSc-ILD

Carmel J.W. Stock, Elisabetta A. Renzoni

AFFILIATIONS:

Interstitial Lung Disease Unit, Royal Brompton Hospital and Imperial College London

Corresponding author:

Dr Carmel Stock, Royal Brompton and Harefield NHS Foundation Trust, Sydney Street,
London, SW3 6NP, United Kingdom
Tel: +44 207 3518456

e-mail: c.stock@imperial.ac.uk

ACKNOWLEDGEMENTS

This work was funded by the Arthritis and Research UK and the Scleroderma and Raynaud's UK (formed by merger of Raynaud's and Scleroderma Association and the Scleroderma Society). It was supported by the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

ABSTRACT

The interplay between genetic and environmental factors is likely involved in the pathogenesis of systemic sclerosis (SSc). Interstitial lung disease associated in the context of SSc (SSc-ILD) is associated with significant morbidity, and is the leading cause of death in SSc. The spectrum of SSc-ILD severity is wide, ranging from patients with only limited and inherently stable pulmonary involvement, to those with extensive and progressive pulmonary fibrosis. In order to provide accurate prognostic information for patients, and to initiate appropriate monitoring and treatment regimens, the ability to identify patients at risk of developing severe ILD early in the disease course is crucial. Identification of genetic variants involved in disease pathogenesis can not only potentially provide diagnostic/prognostic markers, but can also highlight dysregulated molecular pathways for therapeutic targeting. A number of genetic associations have been established for susceptibility to SSc, but far fewer studies have investigated genetic susceptibility to SSc-ILD specifically. In this review we present a summary of the studies assessing genetic associations with SSc-ILD.

KEYWORDS: Systemic sclerosis, SSc-ILD, pulmonary fibrosis, genetics, polymorphisms

INTRODUCTION

Systemic sclerosis (SSc) is a connective tissue disease characterised by immune activation, fibrosis of the skin and internal organs, and widespread vasculopathy. The pattern of internal organ involvement and the natural history of the disease are highly variable. The reported frequency of interstitial lung disease in SSc (SSc-ILD) varies from 25% to 90%, depending on the detection method and disease definition.[1,2] SSc-ILD is more common in patients with the diffuse form of skin involvement, and with anti topo-isomerase autoantibodies (ATA),[3] although at least half of patients with SSc-ILD do not have ATA antibodies.[4] The prominent pathological ILD pattern is non-specific interstitial pneumonia (NSIP).[5] The progression of SSc-ILD is highly variable, with stable and limited disease observed in the majority of patients, and severe progressive disease in a substantial minority.[6]

Evidence for a genetic predisposition to SSc includes the observation that disease prevalence in relatives of patients with SSc is significantly higher than in the general population, with a reported relative risk of disease of 13 in first degree relatives, and of 15 in siblings.[7] Prevalence also varies according to ethnicity. In a large US population study, the prevalence of SSc was higher in individuals of African descent compared to European descent, with an adjusted prevalence ratio of 1.15.[8] Choctaw native Americans have the highest reported prevalence in any population (66/100 000).[9] Compared to patients of African, Japanese, and Choctaw descent, the frequency of ILD is lower in SSc patients of European descent, who also seem to have slower decline in lung function and better survival rates.[10]

Specific non-overlapping antinuclear antibodies (ANAs), including anti-centromere antibodies (ACA) and ATA, also known as Scl-70, are associated with different subsets of SSc. ATA autoantibodies are strongly associated with the development of SSc-ILD, while ACA are protective for ILD.[11] Twin studies have shown a high concordance for ANA specificity, with 90% concordance in monozygotic twins compared to 40% concordance in dizygotic twins, demonstrating a strong genetic influence on ANA status.[12]

Genetic associations with SSc as a whole have been recently extensively reviewed elsewhere.[13,14] Similarly to autoimmune diseases, a predominant genetic effect is observed within the human leukocyte antigen (HLA) region. However, HLA region associations are mainly confined to subgroups of patients possessing specific autoantibodies. Non-HLA genes consistently associated with SSc comprise genes involved in innate immunity as well as B and T cell activation, including the highly repeatable associations with interferon regulatory factor 5 (*IRF5*), signal transducer and activator of transcription 4 (*STAT4*), and cell receptor CD3ζ (*CD247*). [15,13,14]

GENETIC ASSOCIATION STUDIES WITH SSC-ILD

Since the discovery in the 80s that ATA autoantibodies are strongly associated with SSc-ILD, there has been limited progress in enabling prediction of which SSc patients will develop significant ILD. A staging system, based on the extent of fibrosis on HRCT, integrated with pulmonary function as needed, provides accurate prognostic information on the clinical course of SSc-ILD.[6] However, this tool can only be utilised once interstitial lung disease has developed. Identification of biological or genetic markers to enable, at the time of SSc diagnosis, the discrimination of patients at higher risk of developing ILD, and prediction of disease progression, would result in improved clinical management of these patients.

Major histocompatibility complex

A number of HLA alleles have been associated with SSc-ILD, summarised in Table 1. However, many of these studies include only small numbers of patients with SSc-ILD. Selected studies, including some of the larger ones, are discussed below.

Fanning *et al.* reported that the strongest risk factor for SSc-ILD in a UK population (47 SSc-ILD/83 non-ILD) was a combination of ATA positivity, dcSSc, and HLA DRB1*11 (RR=21.9, p=0.0002). In the absence of these three risk factors, DRB1*301 was a risk marker for SSc-ILD, with the highest relative risk seen in ATA negative patients (RR=7.5, p=0.0001).[16] The HLA-DRB1*11 association with SSc-ILD has also been demonstrated in a number of different populations including Spanish,[17] and Black South African.[18] In both an initial and a separate Japanese replication cohort (1st cohort - 41 SSc-ILD/147 controls, 2nd cohort - 40 SSc-ILD/83 controls), the DRB5*0105 allele was significantly more common in SSc-ILD patients compared to healthy controls (OR=8.07, p<0.001 and OR=17.39, p=0.009 respectively).[19] A number of studies of HLA alleles in Han Chinese patients have recently been published. The DQB1*0501 allele was significantly more frequent in SSc-ILD (OR=5.03, p=6x10^{-7}) compared to healthy controls in the study by Zhou et al. (134 SSc-ILD/239 controls). However, DQB1*0501 was also found to be associated with SSc as a whole, and there was no frequency difference between the patients with and without ILD (p=0.9), indicating that this association may not be subtype specific. In a study of the DPB1 locus by Wang et al., (199 SSc-ILD/ 78 SSc no-ILD/480 controls), DPB1*0301 was associated specifically with SSc-ILD (OR=3.86, p<10^{-7}), with no difference in allele frequency between patients without ILD and healthy controls (p=0.79), and a significant difference when the two patient groups were directly compared (OR=3.56, p=0.0069). DPB1*1301 was also more common in the patient group with ILD than the controls (OR=2.25, p<3.3x10^{-4}), but not in patients without ILD (p=0.17).[20] In a study of the DRB1 locus (295 SSc-ILD/ 138 SSc no-ILD/ 458 controls), three alleles were all significantly more common in SSc-ILD compared to controls, but only DRB1*0301 was not also significantly more common in the patients without lung involvement compared to controls (OR=2.47, p=0.0026).[21]

Genome-wide association studies (GWAS)

Although a number of genome-wide association studies (GWAS)[22,23,24,25] and Immunochip studies[26,27] have targeted SSc as a whole, to date none have been specifically designed to assess genetic determinants of SSc-ILD, possibly due to the limitations on achievable cohort sizes. However, post-hoc analyses of data from one of the GWAS studies was performed to investigate the impact of SSc-associated single nucleotide polymorphisms (SNPs) on survival and severity of ILD,[23,28] discussed below in the section on *IRF5*.

CANDIDATE GENE STUDIES

The details of the candidate gene studies discussed in this review are summarised in Table 2.

IRF5

The transcription factor interferon regulatory factor 5 (IRF5) induces expression of interferon A and B genes and pro-inflammatory cytokines, and is critical for antiviral immunity.[29] In a French population (280 SSc-ILD/760 controls), *IRF5* rs2004640 was significantly associated with SSc-ILD, even after adjusting for disease duration, cutaneous involvement, and ANA on multivariate analysis (OR=1.38, p=0.016).[30] A similar association was observed in a Han Chinese population (227 SSc-ILD/502 controls, OR=1.38, p=0.028).[31] A three SNP haplotype containing rs2004640, as well as rs3757385 and rs10954213, is a marker for a five base-pair insertion/deletion polymorphism in intron 1 of *IRF5*. Analysis of the individual SNPs of this haplotype showed that rs3757385 (OR=1.42, $p=5.5x10^{-3}$) and rs2004640 (OR=1.54, $p=9.2x10^{-5}$) were significantly associated with SSc-ILD (292 SSc-ILD/989 controls), although only rs2004640 remained significant following conditional regression analysis. Haplotype analysis of the three SNPs showed the haplotype comprising the protective allele of each SNP was significantly less common in SSc-ILD compared to controls (OR=0.64, p=3.7x10^{-4}), and compared to non-ILD SSc patients (n=397, p=0.018).[32] However, analysis of data from the 2010 GWAS study[23] to investigate the impact of SSc-associated SNPs on survival and severity of ILD, using % predicted FVC as a surrogate marker of ILD severity (1 443 SSc in survival analysis, 914 SSc in FVC% linear regression analysis), did not find rs2004640, or the three SNP haplotype, to be associated with survival or ILD severity. However, the minor allele of *IRF5* rs4728142 was associated with improved survival (HR=0.75, p=0.002), independent of age of onset, gender, cutaneous involvement, and ANA.[33] The minor allele was also associated with less severe ILD after taking disease duration into account (mean difference=2.64, p=0.019). In addition, the number of rs4728142 minor alleles was associated with lower expression of *IRF5* in monocytes from both patients and controls.[33] Meta-analysis of data from five European populations (total of 883 SSc-ILD/4 012 controls), tested the above mentioned *IRF5* SNPs rs2004640 and rs4728142, plus an additional SNP, rs10488631, and found all three to be significantly associated with SSc-ILD compared to controls. However, all three SNPs were also significantly associated with each of the other subtypes tested (lcSSc, dcSSc, ATA, ACA, no ILD), and there was no difference in allele frequencies when the patients with and without each phenotype, including with and without ILD (883 SSc-ILD/1 797 SSc no-ILD), were compared directly, suggesting that these *IRF5* polymorphisms may be associated with SSc as a whole rather than with any specific subtype.[34]

STAT4

Signal transducer and activator of transcription 4 (STAT4) is a transcription factor associated with expression of type 1 interferons, IL-12, and IL-23. *STAT4* rs7574865 is associated with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA).[35] This polymorphism has also been associated with SSc-ILD (316 SSc-ILD/964 controls, OR=1.42, p=0.006), with an additive effect of the *IRF5* SNP rs2004640, where carriage of at least three risk alleles of these two SNPs is strongly associated with SSc-ILD (OR=1.79, p=0.002), with dcSSc and ATA autoantibody being independent risk factors.[36] In a study of three *STAT4* SNPs in a Han Chinese population (237 SSc-ILD/534 controls), rs7574865 and rs10168266 were both significantly associated with SSc-ILD compared to controls (OR=1.86, $p=1.2x10^{-4}$ and OR=1.73, $p=7.7x10^{-4}$ respectively). The third SNP tested, rs3821236, was also associated with SSc-ILD, but significance was lost following Bonferroni correction (p=0.015, OR =1.54).[37] However, in a study of six populations of European ancestry (total of 450 SSc-ILD/3 113 controls), rs7574865 was not associated with SSc-ILD in any of the populations individually or in a meta-analysis.[38]

CD226

CD226 encodes DNAX accessory molecule 1, involved in cell-mediated cytotoxicity of T and NK cells. The non-synonymous *CD226* SNP, rs763361, has been associated with a number of autoimmune diseases including type 1 diabetes mellitus, multiple sclerosis, and RA.[39] A meta-analysis of three European populations (total of 662 SSc-ILD/1 642 controls) found this SNP to be associated with SSc-ILD (OR=1.27, $p=2.98x10^{-4}$). A trend towards a significant association with SSc-ILD was also seen when the populations were analysed separately.[40] A haplotype of three SNPs in *CD226*, rs763361, rs34794968, and rs727088, has been significantly associated with SLE and correlated with expression levels in T cells.[41] Meta-analysis testing of this haplotype in seven European populations (729 SSc-ILD/3 966 controls) found none of the individual SNPs to be associated with SSc-ILD, but did find that one of the haplotypes containing the previously associated allele of rs763361, was overrepresented in the SSc-ILD subgroup compared to controls (OR=1.27, p=0.032). A trend towards a significant difference in frequency of this haplotype between SSc patients with and without ILD was also seen (p=0.069).[42]

NLRP1

NLR family, pyrin domain containing 1 (*NLRP1*) is the activating platform required for formation of the NALP1 inflammasome, involved in activation of inflammatory processes. In a three-population meta-analysis study investigating five *NRLP1* SNPs (674 SSc-ILD/1 587 controls), rs8182352 was significantly associated with SSc-ILD compared to controls (OR=1.19, p=0.0065), and compared to the non-ILD subgroup (n=1 255, OR not stated, p=0.046). An additive effect of *NRLP1* rs8182352 with the *IRF5* rs2004640 and *STAT4* rs7574865 risk alleles was identified, resulting in a 1.33-fold increase in OR for SSc-ILD with each additional risk allele.[43]

IRAK1

Like many autoimmune diseases, SSc is characterised by female predominance, approximately 4.6:1.[44] Interleukin-1 receptor-associated kinase 1 (*IRAK1*), a protein kinase involved in signalling through the Toll-like receptors/IL-1R is located on the X chromosome. Two non-synonymous SNPs, rs1059702 (Phe196Ser) and rs1059703 (Leu532Ser) are in complete linkage disequilibrium, and the variant forms result in increased NFκ-B activity in inflammatory responses.[45] The *IRAK1* variant rs1059702, was investigated in a large study of SSc in three European populations. In the Italian cohort (167 SSc-ILD/ 509 controls) both the T allele and TT genotype were significantly associated with SSc-ILD (OR=2.19, p=0.007 and

OR=2.19, p=0.039 respectively). Only the allelic association reached statistical significance (OR=1.11, p=0.047) in the German cohort (167 SSc-ILD/1 083 controls), although the TT genotype frequency was also non-significantly increased in the SSc-ILD group. In the French cohort (334 SSc-ILD/625 controls), the frequency of both the rs1059702 T allele and the TT genotype of were increased in SSc-ILD compared to controls, but neither reached statistical significance (p=0.14 for allele, p-value for genotype not stated). When the three cohorts were analysed together in a meta-analysis, both the T allele and the TT genotype were significantly associated with SSc-ILD (OR=1.37, 1.99×10^{-4} and OR=2.09, 9.05×10^{-4} respectively).[46] The findings of this study have been replicated in a subsequent study of women from four European cohorts (461 SSc-ILD/2 043 controls, only meta-analysis of the cohorts reported), which also found rs1059702 to be significantly associated with SSc-ILD when compared to both controls (OR=1.30, $p=8.46 \times 10^{-3}$) and patients without ILD (OR=1.26, p=0.025).[47]

CTGF

Connective tissue growth factor (CTGF) induces myofibroblast differentiation and increased extracellular matrix (ECM) production. Serum levels of CTGF correlate with the extent of pulmonary fibrosis SSc-ILD.[48] In the study by Fonseca and Lindahl *et al*, the GG genotype of *CTGF* rs6918698 was significantly associated with SSc-ILD compared to controls (207 SSc-ILD/500 controls), even after adjusting for gender and ANA (OR=2.0, p<0.05). The disease associated G allele results in significantly higher transcriptional activity, with allele specific differential binding of the transcription factors Sp1 and Sp3 to this locus.[49] This association was confirmed in a Japanese cohort (188 SSc-ILD/269 controls, OR=2.0, p<0.001).[50] However, in a study of seven populations of European ancestry, no significant association was detected in any of the populations whether tested separately, or together in a meta-analysis (total of 1 180 SSc/1 784 controls), although no further information, including patient numbers, is provided with regards to the subtype analyses.[51] The most recently published study of this polymorphism was performed in a small Thai cohort (34 SSc-ILD/99 controls) with no association identified with SSc-ILD compared to controls.[52]

CD247

The *CD247* gene encodes the T-cell surface glycoprotein zeta chain (CD3ζ), a signalling component of the T cell receptor (TCR)/CD3 complex. In a French population, *CD247* rs2056626 was found to be associated with SSc-ILD compared to controls (346 SSc-ILD/990 controls, OR=0.65, p=6.8x10^{-3}), and not as strongly associated in patients with no lung disease compared to controls (n=554, p=0.01).[53] This finding was however not replicated in a study in a Han Chinese population (198 SSc-ILD/523 controls, p=0.83).[54]

UNREPLICATED STUDIES WITH SMALL COHORT SIZES

There are a number of additional studies identifying genetic associations with SSc-ILD, but in cohorts which are too small to allow meaningful conclusions, and which have not been repeated in additional cohorts. These studies have been included in Table 2 for completeness, but the small number of patients and lack of replication must been borne in mind while interpreting these associations.

DISCUSSION

For many of the associations presented in this review there have either been conflicting results published from replication studies, or, following the initial association, there have been no further studies published in independent cohorts. However, in recent years there has been a move towards published association studies including both discovery and internal replication cohorts with meta-analysis performed on the combined cohorts, allowing greater confidence in the results compared to those from small, single cohort studies. SSc-ILD is a complex disease with a number of genetic factors expected to be involved in susceptibility, each with only relatively modest effects. As SSc-ILD is relatively rare, most of the published studies are hampered by insufficient power to detect associations when SSc phenotypic subgroups are analysed separately. This must be taken into account when interpreting negative association results. The majority of published studies have been performed in populations of European descent. However, the prevalence of ILD is lower in SSc patients of European descent than in patients of African or Japanese descent. More studies in these non-European populations may aid discovery of SSc-ILD associated genes. A large collaborative project entitled ‘Genome Research in African American Scleroderma Patients’, led by the National Human Genome Institute, is currently ongoing, with the aim of discovering common and low-frequency variants associated with SSc susceptibility in African Americans.[55]

When studying clinical subgroups, the careful definition of phenotypes is crucial to allow appropriate comparisons between patients with and without a phenotype, as well as between different studies. In the field of SSc-ILD genetics this has so far been hampered by the lack of a standardised definition of SSc-ILD, with studies using variable definitions for the presence of ILD, including the presence of ground glass or reticular shadowing on HRCT, evidence of fibrosis on chest radiograph, or impaired lung function.

The disease course of SSc-ILD is highly variable. Identification of specific genetic predictors of severe/progressive SSc-ILD is crucial, both from a pathogenesis and a clinical management perspective. Use of longitudinal clinical data to further define the SSc-ILD phenotype in terms of severity or rate of progression would enable investigation of genetic variants in relation to likelihood of ILD progression and severity. The recent staging system proposed by Goh et al.,[6] which subgroups SSc-ILD as limited or extensive based on rapid estimation of CT extent, supplemented, if necessary, with FVC levels, has been shown to provide accurate prospective prognostic separation. This system could be used to provide prognostic information, even when only limited clinical data is available. The ability of the Goh staging system to predict mortality is further increased when combined with short term pulmonary function trends.[56] Use of this surrogate of disease mortality means that long term follow-up data may not be required to investigate association of genetic variants with SSc-ILD outcome.

Finally, in most studies published so far, it is difficult to disentangle the association with autoantibodies linked with SSc-ILD, such as ATA, and associations with SSc-ILD per se. Although ATA autoantibodies have a high degree of specificity for the development of ILD in SSc, they are not a sensitive marker, as more than half of SSc-ILD patients are ATA autoantibody negative.[4] Therefore, subgroup analysis of SSc-ILD cohorts according to ANA status is required to allow separation of genetic variants associated with ATA or other antibodies and those associated specifically with development of lung fibrosis.

In SSc as a whole, the genetic risk appears to be mainly linked to immune pathway genes. Whether this is the same for the genetic risks for severe or progressive SSc-ILD remains to be determined. The genetic basis for SSc-ILD would seem to be different from that of the idiopathic interstitial pneumonias, as no association is observed with the *MUC5B* variant strongly associated with IPF.[57] The fact that immunosuppressants are observed to stabilise disease in the majority of patients with progressive lung fibrosis in the context of SSc suggests that immune mediated pathways are key in driving the fibrotic process, but how this translates into genetic predisposition will require further study.

Considering the expected small effect size from each individual genetic loci, and the need to analyse SSc-ILD subgroups according to clinical and serological phenotypes, the requirement for sufficiently large sample sizes with well characterised phenotypes is clear. National and international collaborations will be indispensable to study genetic associations specific to SSc-ILD, in order to enable collection of sufficiently large patient cohorts. It is also important that replication of association studies is followed by functional work to determine the biological significance of disease-associated genetic variants.

CONCLUSIONS

From the published literature presented in this review, genetic variation seems to be involved in susceptibility to SSc-ILD. However, to date, no specific genetic variant has been unequivocally associated with SSc-ILD and/or likelihood of ILD progression. By studying sufficiently large cohorts of SSc with and without ILD, carefully staged, with reliable longitudinal data, we should place ourselves in a better position to identify genes associated with the development and rate of progression of SSc-ILD. Knowledge of the genetic susceptibility to SSc-ILD should represent a stepping stone towards a better understanding of the pathobiology of severe/progressive SSc-ILD, and should enable the identification of prognostic and therapeutic targets in this debilitating and potentially fatal disease.

ACKNOWLEDGEMENTS

This work was funded by Arthritis Research UK, and Scleroderma and Raynaud's UK (formed by merger of the Raynaud's and Scleroderma Association and the Scleroderma Society). It was supported by the NIHR Respiratory Disease Biomedical Research Unit at the Royal Brompton and Harefield NHS Foundation Trust and Imperial College London.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

REFERENCES

1 Schurawitzki H, Stiglbauer R, Graninger W *et al.* Interstitial lung disease in progressive systemic sclerosis: high-resolution CT versus radiography. *Radiology* 1990; **176**: 755-759.

2 Steen VD, Conte C, Owens GR, Medsger TA, Jr.. Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum.* 1994; **37**: 1283-1289.

3 Morelli S, Barbieri C, Sgreccia A *et al.* Relationship between cutaneous and pulmonary involvement in systemic sclerosis. *J.Rheumatol.* 1997; **24**: 81-85.

4 Gilchrist FC, Bunn C, Foley PJ *et al.* Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun.* 2001; **2**: 76-81.

5 Bouros D, Wells AU, Nicholson AG *et al.* Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am.J.Respir.Crit Care Med.* 2002; **165**: 1581-1586.

6 Goh NS, Desai SR, Veeraraghavan S *et al.* Interstitial lung disease in systemic sclerosis: a simple staging system. *Am.J.Respir.Crit Care Med.* 2008; **177**: 1248-1254.

7 Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. *Arthritis Rheum.* 2001; **44**: 1359-1362.

8 Mayes MD, Lacey JV, Jr., Beebe-Dimmer J *et al.* Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. *Arthritis Rheum.* 2003; **48**: 2246-2255.

9 Arnett FC, Howard RF, Tan F *et al.* Increased prevalence of systemic sclerosis in a Native American tribe in Oklahoma. Association with an Amerindian HLA haplotype. *Arthritis Rheum.* 1996; **39**: 1362-1370.

10 Kuwana M, Kaburaki J, Arnett FC, Howard RF, Medsger TA, Jr., Wright TM. Influence of ethnic background on clinical and serologic features in patients with systemic sclerosis and anti-DNA topoisomerase I antibody. *Arthritis Rheum.* 1999; **42**: 465-474.

11 Hesselstrand R, Scheja A, Shen GQ, Wiik A, Akesson A. The association of antinuclear antibodies with organ involvement and survival in systemic sclerosis. *Rheumatology.(Oxford)* 2003; **42**: 534-540.

12 Feghali-Bostwick C, Medsger TA, Jr., Wright TM. Analysis of systemic sclerosis in twins reveals low concordance for disease and high concordance for the presence of antinuclear antibodies. *Arthritis Rheum.* 2003; **48**: 1956-1963.

13 Murdaca G, Contatore M, Gulli R, Mandich P, Puppo F. Genetic factors and systemic sclerosis. *Autoimmun.Rev.* 2016

14 Chairta P, Nicolaou P, Christodoulou K. Genomic and genetic studies of systemic sclerosis: A systematic review. *Hum.Immunol.* 2017; **78**: 153-165.

15 Ramos PS, Silver RM, Feghali-Bostwick CA. Genetics of systemic sclerosis: recent advances. *Curr.Opin.Rheumatol.* 2015; **27**: 521-529.

16 Fanning GC, Welsh KI, Bunn C, Du BR, Black CM. HLA associations in three mutually exclusive autoantibody subgroups in UK systemic sclerosis patients. *Br.J.Rheumatol.* 1998; **37**: 201-207.

17 Simeon CP, Fonollosa V, Tolosa C *et al.* Association of HLA class II genes with systemic sclerosis in Spanish patients. *J.Rheumatol.* 2009; **36**: 2733-2736.

18 Tikly M, Rands A, McHugh N, Wordsworth P, Welsh K. Human leukocyte antigen class II associations with systemic sclerosis in South Africans. *Tissue Antigens* 2004; **63**: 487-490.

19 Odani T, Yasuda S, Ota Y *et al.* Up-regulated expression of HLA-DRB5 transcripts and high frequency of the HLA-DRB5*01:05 allele in scleroderma patients with interstitial lung disease. *Rheumatology.(Oxford)* 2012; **51**: 1765-1774.

20 Wang J, Guo X, Yi L *et al.* Association of HLA-DPB1 with scleroderma and its clinical features in Chinese population. *PLoS.One.* 2014; **9**: e87363-

21 He D, Wang J, Yi L *et al.* Association of the HLA-DRB1 with scleroderma in Chinese population. *PLoS.One.* 2014; **9**: e106939-

22 Zhou X, Tan FK, Wang N *et al.* Genome-wide association study for regions of systemic sclerosis susceptibility in a Choctaw Indian population with high disease prevalence. *Arthritis Rheum.* 2003; **48**: 2585-2592.

23 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**: 426-429.

24 Allanore Y, Saad M, Dieude P *et al.* Genome-wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. *PLoS.Genet.* 2011; **7**: e1002091-

25 Zhou X, Lee JE, Arnett FC *et al.* HLA-DPB1 and DPB2 are genetic loci for systemic sclerosis: a genome-wide association study in Koreans with replication in North Americans. *Arthritis Rheum.* 2009; **60**: 3807-3814.

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

27 Zochling J, Newell F, Charlesworth JC *et al.* An Immunochip based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. *Arthritis Res.Ther.* 2014; **16**: 438

28 Gorlova O, Martin JE, Rueda B *et al.* Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. *PLoS.Genet.* 2011; **7**: e1002178

29 Yanai H, Chen HM, Inuzuka T *et al.* Role of IFN regulatory factor 5 transcription factor in antiviral immunity and tumor suppression. *Proc.Natl.Acad.Sci.U.S.A* 2007; **104**: 3402-3407.

30 Dieude 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**: 225-233.

31 Wang J, Yi L, Guo X *et al.* Association of the IRF5 SNP rs2004640 with systemic sclerosis in Han Chinese. *Int.J.Immunopathol.Pharmacol.* 2014; **27**: 635-638.

32 Dieude P, Dawidowicz K, Guedj M *et al.* Phenotype-haplotype correlation of IRF5 in systemic sclerosis: role of 2 haplotypes in disease severity. *J.Rheumatol.* 2010; **37**: 987-992.

33 Sharif R, Mayes MD, Tan FK *et al.* IRF5 polymorphism predicts prognosis in patients with systemic sclerosis. *Ann.Rheum.Dis.* 2012; **71**: 1197-1202.

34 Carmona FD, Martin JE, Beretta L *et al.* The systemic lupus erythematosus IRF5 risk haplotype is associated with systemic sclerosis. *PLoS.One.* 2013; **8**: e54419

35 Remmers EF, Plenge RM, Lee AT *et al.* STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. *N.Engl.J.Med.* 2007; **357**: 977-986.

36 Dieude 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**: 2472-2479.

37 Yi L, Wang JC, Guo XJ *et al.* STAT4 is a genetic risk factor for systemic sclerosis in a Chinese population. *Int.J.Immunopathol.Pharmacol.* 2013; **26**: 473-478.

38 Rueda B, Broen J, Simeon C *et al.* The STAT4 gene influences the genetic predisposition to systemic sclerosis phenotype. *Hum.Mol.Genet.* 2009; **18**: 2071-2077.

39 Hafler JP, Maier LM, Cooper JD *et al.* CD226 Gly307Ser association with multiple autoimmune diseases. *Genes Immun.* 2009; **10**: 5-10.

40 Dieude P, Guedj M, Truchetet ME *et al.* Association of the CD226 Ser(307) variant with systemic sclerosis: evidence of a contribution of costimulation pathways in systemic sclerosis pathogenesis. *Arthritis Rheum.* 2011; **63**: 1097-1105.

41 Lofgren SE, Delgado-Vega AM, Gallant CJ *et al.* A 3'-untranslated region variant is associated with impaired expression of CD226 in T and natural killer T cells and is associated with susceptibility to systemic lupus erythematosus. *Arthritis Rheum.* 2010; **62**: 3404-3414.

42 Bossini-Castillo L, Simeon CP, Beretta L *et al.* A multicenter study confirms CD226 gene association with systemic sclerosis-related pulmonary fibrosis. *Arthritis Res.Ther.* 2012; **14**: R85

43 Dieude P, Guedj M, Wipff J *et al.* NLRP1 influences the systemic sclerosis phenotype: a new clue for the contribution of innate immunity in systemic sclerosis-related fibrosing alveolitis pathogenesis. *Ann.Rheum.Dis.* 2011; **70**: 668-674.

44 Arora-Singh RK, Assassi S, del Junco DJ *et al.* Autoimmune diseases and autoantibodies in the first degree relatives of patients with systemic sclerosis. *J.Autoimmun.* 2010; **35**: 52-57.

45 Liu G, Tsuruta Y, Gao Z, Park YJ, Abraham E. Variant IL-1 receptor-associated kinase-1 mediates increased NF-kappa B activity. *J.Immunol.* 2007; **179**: 4125-4134.

46 Dieude P, Bouaziz M, Guedj M *et al.* Evidence of the contribution of the X chromosome to systemic sclerosis susceptibility: association with the functional IRAK1 196Phe/532Ser haplotype. *Arthritis Rheum.* 2011; **63**: 3979-3987.

47 Carmona FD, Cenit MC, Diaz-Gallo LM *et al.* New insight on the Xq28 association with systemic sclerosis. *Ann.Rheum.Dis.* 2013

48 Sato S, Nagaoka T, Hasegawa M *et al.* Serum levels of connective tissue growth factor are elevated in patients with systemic sclerosis: association with extent of skin sclerosis and severity of pulmonary fibrosis. *J.Rheumatol.* 2000; **27**: 149-154.

49 Fonseca C, Lindahl GE, Ponticos M *et al.* A polymorphism in the CTGF promoter region associated with systemic sclerosis. *N.Engl.J.Med.* 2007; **357**: 1210-1220.

50 Kawaguchi Y, Ota Y, Kawamoto M *et al.* Association study of a polymorphism of the CTGF gene and susceptibility to systemic sclerosis in the Japanese population. *Ann.Rheum.Dis.* 2009; **68**: 1921-1924.

51 Rueda B, Simeon C, Hesselstrand R *et al.* A large multicentre analysis of CTGF -945 promoter polymorphism does not confirm association with systemic sclerosis susceptibility or phenotype. *Ann.Rheum.Dis.* 2009; **68**: 1618-1620.

52 Louthrenoo W, Kasitanon N, Wichainun R *et al.* Lack of CTGF*-945C/G Dimorphism in Thai Patients with Systemic Sclerosis. *Open.Rheumatol.J.* 2011; **5**: 59-63.

53 Dieude P, Boileau C, Guedj M *et al.* Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. *Ann.Rheum.Dis.* 2011; **70**: 1695-1696.

54 Wang J, Yi L, Guo X *et al.* Lack of Association of the CD247 SNP rs2056626 with Systemic Sclerosis in Han Chinese. *Open.Rheumatol.J.* 2014; **8**: 43-45.

55 Genome Research in African American Scleroderma Patients http://www.srfcure.org/research/featured-research

56 Goh NS, Hoyles RK, Denton CP *et al.* Short term pulmonary function trends are predictive of mortality in interstitial lung disease associated with systemic sclerosis. *Arthritis Rheumatol.* 2017 [Epub ahead of print]

57 Stock CJ, Sato H, Fonseca C *et al.* Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. *Thorax* 2013; **68**: 436-441.

58 Gladman DD, Kung TN, Siannis F, Pellett F, Farewell VT, Lee P. HLA markers for susceptibility and expression in scleroderma. *J.Rheumatol.* 2005; **32**: 1481-1487.

59 Zhou XD, Yi L, Guo XJ *et al.* Association of HLA-DQB1*0501 with scleroderma and its clinical features in Chinese population. *Int.J.Immunopathol.Pharmacol.* 2013; **26**: 747-751.

60 Briggs DC, Vaughan RW, Welsh KI, Myers A, duBois RM, Black CM. Immunogenetic prediction of pulmonary fibrosis in systemic sclerosis. *Lancet* 1991; **338**: 661-662.

61 Zhao W, Yue X, Liu K *et al.* The status of pulmonary fibrosis in systemic sclerosis is associated with IRF5, STAT4, IRAK1, and CTGF polymorphisms. *Rheumatol.Int.* 2017

62 Kowal-Bielecka O, Chwiesko-Minarowska S, Bernatowicz PL *et al.* The arachidonate 5-lipoxygenase activating protein gene polymorphism is associated with the risk of

scleroderma-related interstitial lung disease: a multicentre European Scleroderma Trials and Research group (EUSTAR) study. *Rheumatology.(Oxford)* 2017; **56**: 844-852.

63 Hoshino K, Satoh T, Kawaguchi Y, Kuwana M. Association of hepatocyte growth factor promoter polymorphism with severity of interstitial lung disease in Japanese patients with systemic sclerosis. *Arthritis Rheum.* 2011; **63**: 2465-2472.

64 Kawaguchi Y, Tochimoto A, Ichikawa N *et al.* Association of IL1A gene polymorphisms with susceptibility to and severity of systemic sclerosis in the Japanese population. *Arthritis Rheum.* 2003; **48**: 186-192.

65 Beretta L, Bertolotti F, Cappiello F *et al.* Interleukin-1 gene complex polymorphisms in systemic sclerosis patients with severe restrictive lung physiology. *Hum.Immunol.* 2007; **68**: 603-609.

66 Rech TF, Moraes SB, Bredemeier M *et al.* Matrix metalloproteinase gene polymorphisms and susceptibility to systemic sclerosis. *Genet.Mol.Res.* 2016; **15**:

67 Manetti M, Ibba-Manneschi L, Fatini C *et al.* Association of a functional polymorphism in the matrix metalloproteinase-12 promoter region with systemic sclerosis in an Italian population. *J.Rheumatol.* 2010; **37**: 1852-1857.

68 Sumita Y, Sugiura T, Kawaguchi Y *et al.* Genetic polymorphisms in the surfactant proteins in systemic sclerosis in Japanese: T/T genotype at 1580 C/T (Thr131Ile) in the SP-B gene reduces the risk of interstitial lung disease. *Rheumatology.(Oxford)* 2008; **47**: 289-291.

69 Ates O, Musellim B, Ongen G, Topal-Sarikaya A. NRAMP1 (SLC11A1): a plausible candidate gene for systemic sclerosis (SSc) with interstitial lung involvement. *J.Clin.Immunol.* 2008; **28**: 73-77.

Table 1. HLA associations with SSc-ILD

Corrected p values given where available. ORs are shown as OR (95% confidence interval), where available.

Table 2. Non-HLA associations with SSc-ILD

Corrected p values given where available. ORs are shown as OR (95% confidence interval), where available. [†]= meta-analysis or previously published studies. [§]= total number of SSc patients, when SSc-ILD number not given. [¶]= meta-analysis of the different populations included.