OPEN ACCESS

TRANSLATIONAL SCIENCE

Comprehensive analysis of the major histocompatibility complex in systemic sclerosis identifies differential HLA associations by clinical and serological subtypes

Marialbert Acosta-Herrera [1], Martin Kerick,[1] Elena Lopéz-Isac,[1] Shervin Assassi [2], Lorenzo Beretta [3], Carmen Pilar Simeón-Aznar,[4] Norberto Ortego-Centeno,[5] International SSc Group, Susanna M Proudman,[6] Australian Scleroderma Interest Group (ASIG), Nicolas Hunzelmann,[7] Gianluca Moroncini,[8] Jeska K de Vries-Bouwstra,[9] Gisela Orozco,[10,11] Anne Barton [10,11], Ariane L Herrick,[12] Chikashi Terao,[13] Yannick Allanore,[14] Matthew A Brown [15], Timothy RDJ Radstake,[16] Carmen Fonseca,[17] Christopher P Denton,[17] Maureen D Mayes,[2] Javier Martin [1]

Check for updates

ABSTRACT

Objective The greatest genetic effect reported for systemic sclerosis (SSc) lies in the major histocompatibility complex (MHC) locus. Leveraging the largest SSc genome-wide association study, we aimed to fine-map this region to identify novel human leucocyte antigen (HLA) genetic variants associated with SSc susceptibility and its main clinical and serological subtypes.

Methods 9095 patients with SSc and 17 584 controls genome-wide genotyped were used to impute and test single-nucleotide polymorphisms (SNPs) across the MHC, classical HLA alleles and their composite amino acid residues. Additionally, patients were stratified according to their clinical and serological status, namely, limited cutaneous systemic sclerosis (lcSSc), diffuse cutaneous systemic sclerosis (dcSSc), anticentromere (ACA), antitopoisomerase (ATA) and anti-RNApolIII autoantibodies (ARA).

Results Sequential conditional analyses showed nine SNPs, nine classical alleles and seven amino acids that modelled the observed associations with SSc. This confirmed previously reported associations with *HLA-DRB1*11:04* and *HLA-DPB1*13:01*, and revealed a novel association of *HLA-B*08:01*. Stratified analyses showed specific associations of *HLA-DQA1*02:01* with lcSSc, and an exclusive association of *HLA-DQA1*05:01* with dcSSc. Similarly, private associations were detected in *HLA-DRB1*08:01* and confirmed the previously reported association of *HLA-DRB1*07:01* with ACA-positive patients, as opposed to the *HLA-DPA1*02:01* and *HLA-DQB1*03:01* alleles associated with ATA presentation.

Conclusions This study confirms the contribution of HLA class II and reveals a novel association of HLA class I with SSc, suggesting novel pathways of disease pathogenesis. Furthermore, we describe specific HLA associations with SSc clinical and serological subtypes that could serve as biomarkers of disease severity and progression.

Key messages

What is already known about this subject?
► The major histocompatibility complex is the genomic region shown to have the greatest genetic effect in several autoimmune diseases such as systemic sclerosis (SSc).

What does this study add?
► Taking advantage of the largest genetic study in SSc, we conducted an extensive fine-mapping of the region by assessing single nucleotide polymorphisms, human leucocyte antigen (HLA) classical alleles and polymorphic amino acid residues associated with SSc.
► We have confirmed the strong contribution of HLA class II in SSc susceptibility and showed for the first time the independent association of HLA class I, suggesting novel pathways of disease pathogenesis.
► We have identified specific associations in the different clinical forms of the disease, as well as private associations regarding autoantibodies presentation.

How might this impact on clinical practice or future developments?
► These findings may improve our knowledge of disease onset and progression, as well as assist in the identification of biomarkers that allow early and specific interventions.

INTRODUCTION

Genetic variation within the major histocompatibility complex (MHC) has been associated with many human conditions, particularly autoimmune and infectious diseases or those with a central immunological component.[1 2] Systemic sclerosis (SSc) or scleroderma is a rare systemic immune-mediated

inflammatory disease (IMID), with a broad spectrum of clinical forms, affecting primarily connective tissues.[3] It is characterised by an immunological disturbance leading to the production of autoantibodies, vascular damage, and widespread fibrosis of the skin and internal organs.[3–5] Regarding the clinical characteristics of the disease, patients with SSc are classified depending on the extent of the dermal fibrosis as either limited cutaneous systemic sclerosis (lcSSc) and diffuse cutaneous systemic sclerosis (dcSSc).[3 5 6] Further classifications are performed according to the immunological dysregulation and the mutually exclusive production of autoantibodies in anticentromere (ACA), antitopoisomerase (ATA) and anti-RNA polymerase III (ARA) antibodies.[3 6]

SSc is a complex disease, in which the contributions of environmental and genetic factors are crucial for disease onset and progression.[7–9] Several genome-wide association studies (GWASs) and Immunochip analyses have shed light into this genetic component.[10–12] Interestingly, a recent GWAS in SSc confirmed that the greatest genetic contribution to the disease described thus far comes from the human leucocyte antigen (HLA) region.[13] Genetic variations in the HLA system may determine their binding affinity for specific antigens and their presentation to antigen-presenting cells, leading to the activation of autoreactive T-helper and B cells and the production of autoantibodies.[14] These genetic variations are detected in seropositive IMIDs, and several of them have been described to be shared among them.[15 16] HLA fine-mapping studies have been carried out successfully in several IMIDs, including rheumatoid arthritis (RA),[17] systemic lupus erythematosus (SLE)[18] and myositis,[19] among others, and have been proven useful in identifying the strongest genetic risk factors in autoimmune diseases.[2] In SSc previous assessments identified polymorphic amino acid positions and single-nucleotide polymorphisms (SNPs) that modelled the observed associations in populations of European descent,[11 12] and a recent study in African and European confirmed an African ancestry-predominant allele and a transancestry association with individuals of European ancestry.[20 21]

Taking the aforementioned into consideration and leveraging the enhanced power provided by the most recent GWAS in SSc, we conducted a broad analysis of the MHC region to evaluate SNPs, classical HLA alleles and their polymorphic amino acid positions, with SSc and its clinical and serological subphenotypes. We also functionally explored the associated variants, finding evidence of colocalisation with expression quantitative trait loci (eQTLs).

MATERIALS AND METHODS

Study population

This study included genome-wide genotyped data from 9846 patients with SSc and 18 333 healthy individuals from the same source population.[13] The patients fulfilled the 2013 American College of Rheumatology/the European League Against Rheumatism classification criteria or the criteria proposed by LeRoy and Medsger for early SSc.[22 23] In addition, patients were stratified by the main clinical classifications (lcSSc or dcSSc) and main autoantibody status (ACA, ATA or ARA). Details of the cohorts, genotyping methods and quality control (QC) for genotyped data are described elsewhere.[13]

SNP and HLA imputation

After genotyping QC, SNPs, classical HLA alleles and amino acid variants, were all imputed for each case–control dataset separately in the extended MHC region in chromosome 6.[24] The

SNP2HLA[25] software was used for imputation using a reference panel consisting of 5225 European individuals in the Type 1 Diabetes Genetic Consortium,[26] containing data of 8961 variants across the MHC region, and two and four digit-resolution allelic identities of the HLA class I (*HLA-A*, *HLA-B* and *HLA-C*) and II genes (*HLA-DPA1*, *HLA-DPB1*, *HLA-DQA1*, *HLA-DQB1* and *HLA-DRB1*) as well as their amino acid make-up. Imputed data were also filtered for 95% success call rate for alleles and amino acids, deviation from Hardy-Weinberg equilibrium (HWE) considering a *p*-value of <0.001 for SNPs in controls and 95% total call rate for individuals. The total numbers of imputed variants per case–control set are specified in online supplemental table 1.

Statistical analysis

Association analyses were performed with PLINK[27] using logistic regressions in each of the 14 independent datasets, including sex and the five first principal components (PCs) as covariates. Briefly, PC analysis was performed using ~100 000 quality-filtered independent SNPs outside the MHC region using PLINK and GCTA64. Outliers were identified and removed as described elsewhere.[13] We tested SNPs, classical HLA alleles and all possible combinations of amino acid residues per position. Inverse variance fixed effects meta-analysis was conducted with PLINK to evaluate the consistency of effects across studies. The genome-wide significance was established at a *p*-value $\leq 5 \times 10^{-8}$.

Considering the main clinical SSc subtypes and serological classifications, stratification of cases was performed following the same procedure as for the global analysis and comparisons were made with the control group and intracases, namely, dcSSc with lcSSc, and ATA with ACA (patients without available data or positive for both autoantibodies were excluded from the analysis). Only classical alleles whose results outperformed those from the global analysis and that were significantly associated in both comparisons were declared as private.

To identify independent signals within the region, sequential conditional association analyses were performed with the software GCTA-COJO[19 28 29] using the summary statistics from the meta-analysis (global and stratified by clinical and serological subtypes) and separately for each variant type (SNPs, alleles, and amino acids). The Manhattan plot was obtained with an in-house R script. The Protein Data Bank entries 3pdo, 1a1m, 3lqz and 2bvp were used to produce the 3D models of the HLA molecules with the UCSF Chimaera software.[30]

Functional assessment of associated variants

In order to assign a biological meaning of our association results at the SNP level, we performed a colocalisation analysis using COLOC[31] and the Genotype-Tissue Expression (GTEx) project release V.8 (dbGaP Accession phs000424.v8.p2). Colocalisation analysis evaluates if two independent studies at the same locus consistently share a causal variant; if so, the probability of a causal association increases.

RESULTS

A total of 9095 patients with SSc and 17 584 healthy individuals fulfilled the QCs and 8339 variants were meta-analysed, including SNPs, classical alleles and amino acid positions within the MHC region (online supplemental table 1), identifying 1273 reaching the genome-wide level of significance (figure 1).
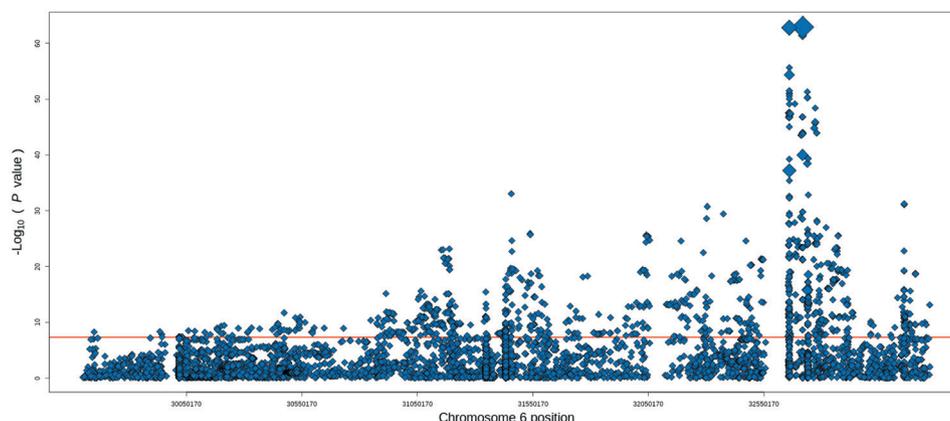
Figure 1 Association signals for systemic sclerosis in the human leucocyte antigen region. The $-\log_{10}$ of the meta-analysis p values are plotted against their chromosomal position. The red line represents the genome-wide level of significance (p value=5×10^{-08}). The size of the diamond indicates the degree of linkage disequilibrium with the strongest association from the meta-analysis (rs1048372).

SNP and HLA associations

Within this region, the global meta-analysis yielded 1082 significantly associated SNPs, from which nine were independent and modelled the observed SNPs associations in the region (table 1) after the sequential conditional analysis. The most associated signal corresponded to a protective synonymous coding SNP in the *HLA-DQA1* gene (rs1048372, OR=0.70, 95% CI 0.67 to 0.73, p value=1.29×10^{-63}). In addition, another synonymous coding SNP in the same gene was independently associated with SSc (rs1142338, OR=1.86, 95% CI 1.67 to 2.07, p value=3.16×10^{-12}) and a truncating SNP mapping in the

HLA-DPB1 gene (rs1126511, OR=1.21, 95% CI 1.16 to 1.27, p value=2.37×10^{-25}). All the remaining SNPs were non-coding intronic/intergenic, potentially involved in the regulation of gene expression. Given the complex linkage disequilibrium (LD) structure in the region, we assessed the relationship among the associated variants, and rs2844532 and rs17500468 were not in LD with any classical alleles and amino acid residues (figure 2).

In the global meta-analysis and regarding the classical alleles, a total of 21 four-digit classical alleles were significantly associated with SSc at the genome-wide level with strong signals within the HLA class II. The strongest association observed

Table 1 Independent association results from the global analysis comparing scleroderma and controls after the sequential conditional analysis

Gene	Variation	BP	N	OR (95% CI)	P value	Conditioned P value
HLA-DQA1	rs1048372T*	32 642 659	14	0.70 (0.67 to 0.73)	1.29E-63	–
HLA-DRB5/HLA-DQA1	rs482044G	32 608 287	14	0.73 (0.70 to 0.76)	7.09E-50	1.10E-35
HLA-DPB1	rs1126511T*	33 080 689	14	1.21 (1.16 to 1.27)	2.01E-16	2.37E-25
COL11A2	rs9469378C	33 191 887	14	1.44 (1.31 to 1.58)	8.02E-14	2.16E-14
HLA-DQA1	rs1142338T*	32 641 545	14	1.86 (1.67 to 2.07)	3.12E-29	3.16E-12
HLA-B/MICA	rs2844532G	2 685 662	14	0.77 (0.73 to 0.81)	2.11E-23	1.54E-11
HLA-DQA2	rs17500468G	32 743 401	14	1.36 (1.28 to 1.43)	1.10E-27	3.64E-10
MICA/MICB	rs3094228G	31 462 150	14	1.29 (1.23 to 1.36)	1.30E-26	5.42E-09
BTNL2/HLA-DRA	rs9268515C	32 411 518	14	1.08 (1.03 to 1.14)	3.01E-03	1.76E-09
HLA-DRB1	DRB1*11:04	32 584 287	13	2.11 (1.92 to 2.31)	2.52E-56	–
HLA-DQB1	DQB1*02:02	32 663 284	14	0.56 (0.51 to 0.60)	5.79E-51	3.84E-45
HLA-DPB1	DPB1*13:01	33 081 591	14	2.05 (1.82 to 2.31)	6.10E-32	9.77E-30
HLA-DQA1	DQA1*04:01	32 640 529	14	1.86 (1.67 to 2.07)	3.12E-29	2.97E-28
HLA-DRB1	DRB1*13:01	32 584 287	14	0.68 (0.62 to 0.75)	2.15E-16	3.00E-14
HLA-B	B*08:01	31 355 516	14	1.22 (1.15 to 1.30)	1.29E-10	1.79E-12
HLA-DQB1	DQB1*05:01	32 663 284	14	1.20 (1.14 to 1.27)	3.25E-10	1.33E-12
HLA-DPB1	DPB1*03:01	33 081 591	14	1.19 (1.12 to 1.27)	3.15E-08	4.81E-08
HLA-DPB1	DPB1*06:01	33 081 591	13	1.47 (1.27 to 1.70)	2.19E-07	2.15E-08
HLA-DRB1	AA Ile67	32 584 192	14	0.70 (0.67 to 0.73)	1.70E-63	–
HLA-DPB1	AA Ile76	33 080 885	14	1.74 (1.56 to 1.93)	1.73E-23	3.70E-29
HLA-DRB1	AA Tyr60	32 584 213	14	0.65 (0.62 to 0.69)	3.50E-52	7.03E-22
HLA-DQA1	AA Thr69	32 641 502	14	1.85 (1.66 to 2.05)	1.65E-30	1.90E-19
HLA-DRB1	AA Ala58	32 584 219	14	1.46 (1.37 to 1.55)	2.88E-33	2.56E-14
HLA-DPB1	AA Leu11	33 080 690	14	1.21 (1.16 to 1.27)	2.01E-16	1.65E-11

BP position based on build hg38.

Sequential conditional association analyses were performed separately for each variant type.

*Coding single-nucleotide polymorphisms: rs1126511 (SNP_DPB1_33156444), rs1142338 (SNP_DQA1_32717300) and rs1048372 (SNP_DQA1_32718414).

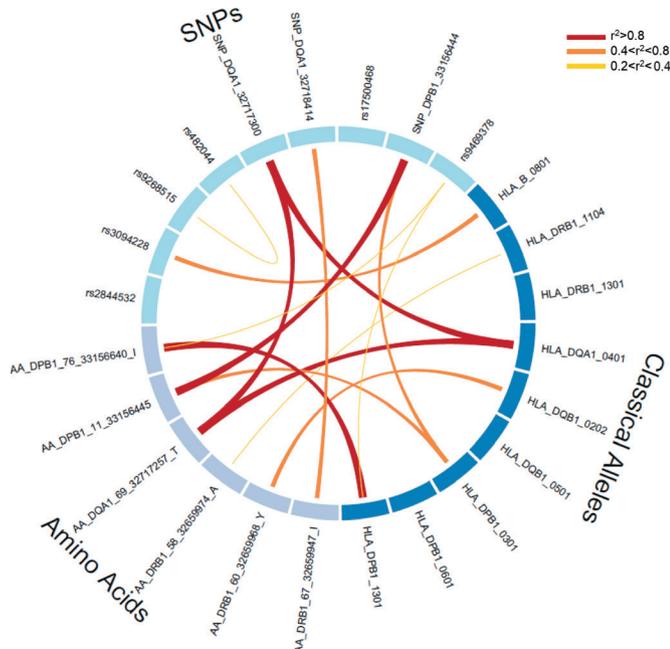BP, base pair; N, number of cohorts where the variant was meta-analysed.

Figure 2 LD among the independent variants. Circos plot depicting the LD relationship among the SNPs, four-digit classical HLA alleles and HLA amino acid residues independently associated from the sequential conditional analysis. HLA, human leucocyte antigen; LD, linkage disequilibrium; SNP, single-nucleotide polymorphism.

was with *HLA-DRB1*11:04* (OR=2.11, 95% CI 1.92 to 2.31, p value=2.52×10^{-56}). After sequential conditional analysis controlling for the effect of the most associated HLA alleles, seven additional class II genes were independently associated, including three *HLA-DPB1*, two *HLA-DQB1*, one *HLA-DQA1* and one *HLA-DRB1* alleles (table 1). Interestingly, here we describe for the first time the independent association with SSc of an HLA class I conferring risk, which belongs to the classical allele *HLA-B*08:01* (OR=1.22, 95% CI 1.15 to 1.30, p value=1.79×10^{-12}) (table 1). The *HLA-DRB1*13:01*, *HLA-DQB1*05:01* and the *HLA-DPB1*06:01* were independent ($r^2 < 0.2$) from the other variants irrespective of their nature (amino acid residues and SNPs) (figure 2).

Moreover, we performed amino acid analysis with a total of 170 polymorphic amino acid residues significantly associated in the global SSc meta-analysis. The most significant amino acid residue associated with SSc was the Ile67 of the HLA-DRβ1 (OR=0.70, 95% CI 0.67 to 0.73, p value=1.70×10^{-63}). A summary of the independent associations after the stepwise conditional model is shown in table 1 and online supplemental figure S1A–C. All the associated amino acid residues were in moderate to high LD ($0.4 < r^2 < 0.8$) with the reported classical alleles and SNPs (online supplemental table 2).

Functional annotation of associated SNPs

To functionally characterise the associations from the meta-analysis and their proxies ($r^2 \geq 0.8$) at the SNP level, they were tested against the eQTLs from the 49 tissues contained in GTEx by a colocalisation approach.[31] We identified 70 SNPs affecting the expression of 82 eGenes with a posterior probability of 80% in 40 tissues (online supplemental table 3). Then, we further assessed their overlap with the independent variants or any proxies and identified five of them affecting the expression of 11 eGenes in relevant cells and tissues involved in the disease,

including lymphocytes, fibroblasts, colon and oesophagus, among others (table 2).

Clinically restricted subphenotype analysis

The numbers of patients in each subgroup are summarised in online supplemental table 4. Given that previous studies reported genetic differential susceptibility to SSc depending on its subtype,[12 13 32] we performed stratified analyses comparing with the control group and the results are summarised in the online supplemental tables 5–9. For lcSSc, a total of six classical alleles were identified as independently associated (table 3 and online supplemental table 5). Among them, *HLA-DQA1*02:01* was only associated with lcSSc (OR=0.54, p value=5.23×10^{-51}), and this was further confirmed when compared with the patients with dcSSc (OR=0.71, p value=2.08×10^{-8}; online supplemental table 10). Regarding dcSSc, four classical alleles were independently associated with this subphenotype when compared with the healthy individuals (table 3 and online supplemental table 6). *HLA-DQA1*05:01* was exclusively associated with dcSSc (OR=1.49, *p* value=1.59×10^{-11}). This was confirmed when comparing these patients with patients with lcSSc (OR=1.30, p value=1.76×10^{-11}) (online supplemental table 10).

Serologically restricted analysis

Taking into account the serologically restricted subphenotypes, we conducted different analyses to compare ACA, ATA and ARA positive cases and controls in this locus. In the case of ACA, *HLA-DRB1*08:01* (OR=3.18, p value=4.00×10^{-64}) and *HLA-DRB1*07:01* (OR=0.36, p value=1.84×10^{-45}) were the classical alleles independently and exclusively associated with the presence of this autoantibody (online supplemental table 7), and this was verified when comparing with the ATA-positive patients (OR=2.17, p value=1.42×10^{-10} and OR=0.42, p value=3.85×10^{-27}, respectively) (table 3 and online supplemental table 10). Concerning the analysis in the ATA subgroup, two classical alleles, namely, *HLA-DPA1*02:01* and *HLA-DQB1*03:01* were significantly and exclusively associated with this phenotype (OR=1.87, p value=2.93×10^{-19} and OR=1.86, p value=7.00×10^{-19}, respectively) (online supplemental table 8), which was confirmed in the intracases comparison (OR=2.41, p value=1.09×10^{-40} and OR=1.67, p value=1.73×10^{-22}, respectively) (table 3). Regarding the ARA-positive analysis, only *HLA-DRB1*11:04* was significantly associated with the presence of this autoantibody (online supplemental table 9).

Given the known correlation between the subphenotypes and the autoantibodies,[3] it is worth noting the overlap of *HLA-DRB1*08:01* as associated with lcSSc (OR=2.18, p value=8.07×10^{-29}) and with ACA-positive patients (OR=3.18, p value=4.00×10^{-64}); however, this association was no longer significant for lcSSc when compared with dcSSc (OR=1.49, p value=3.43×10^{-5}) (online supplemental tables 5 and 7).

DISCUSSION

Leveraging the largest genetic study conducted in SSc, we performed a comprehensive analysis of the MHC locus by fine-mapping approaches involving SNPs and imputed four-digit classical HLA alleles and their amino acid residues. Our results showed strong evidence for the substantial contribution of the HLA class II region in the pathophysiology of SSc, with strong associations with *HLA-DRB1*11:04*, *HLA-DQB1*02:02* and *HLA-DPB1*13:01* alleles. Furthermore, we revealed for the first time the genome-wide significant association of a class I HLA

Table 2 Colocalisation analysis for the independently associated SNPs

SNP	Association P value	eGENE	Tissue	GTEx P value
rs482044	1.10E-35	*TNXA*	Oesophagus (mucosa)	8.46E-05
		STK19P	Kidney cortex	7.10E-07
		HLA-DRB1	Adipose (visceral)	1.70E-18
			Brain	3.50E-16
			Cardiac ventricle	1.20E-22
			Liver	1.50E-06
		HLA-DRB6	Brain	3.92E-21
		BRD2	Nerve (tibial)	1.05E-04
rs1126511 (SNP_DPB1_33156444)	2.37E-25	*HLA-DPA2*	Brain	4.46E-07
		LEMD2	Colon transverse	1.04E-04
rs9469378	1.16E-14	*RING1*	Adipose (subcutaneous)	2.70E-05
			Cultured fibroblasts	2.20E-06
			Oesophagus (muscularis)	1.30E-05
			Skin	5.50E-06
		ITPR3	Oesophagus (mucosa)	7.87E-06
rs1142338 (SNP_DQA1_32717300)*	3.16E-12	*C2*	Transformed lymphocytes	6.03E-08
rs3094228	5.42E-09	*DDR1*	Oesophagus (muscularis)	3.85E-05

eGENE is the gene modulated by SNP.

*Colocalisation was found for the proxy of rs1142338 (SNP_DQA1_32717300), the rs4713586 (r^2=0.95).

GTEx, Genotype-Tissue Expression; SNP, single-nucleotide polymorphism.

gene, the *HLA-B*08:01*. In addition, we identified associated amino acid residues in several HLA type II genes, mapping in the peptide binding pocket and non-coding variants involved in gene expression regulation. Importantly, the stratified analysis showed *HLA-DQA1*02:01* as associated with the lcSSc subtype, in contrast to *HLA-DQA1*05:01* that was private for dcSSc. Likewise, the serological stratification also showed exclusive associations; for instance, *HLA-DRB1*08:01* and *HLA-DRB1*07:01* alleles were significantly associated in ACA-positive patients unlike in the ATA-positive patients, where associations with the *HLA-DPA1*02:01* and *HLA-DQB1*03:01* alleles were detected. In ARA-positive patients only *HLA-DRB1*11:04* was significantly associated with this autoantibody presentation.

In the global analysis, three of the most associated alleles in the model, that is, *HLA-DRB1*11:04*, *HLA-DQB1*02:02* and *HLA-DPB1*13:01*, have been previously reported as associated with SSc in different populations,[21 33–36] confirming them as robustly associated with the disease. In addition, the risk allele rs17500468*G, which is an intronic variant mapping in the *HLA-DQA2* gene, was previously reported in an Immunochip study[11] and is in tight LD (D'=1.0) with rs2857130 identified in a further Immunochip,[12] validating its association in European population.[11] The current study revealed the genome-wide significant association of *HLA-B*08:01*. Other autoimmune

diseases such as RA, SLE, myositis, Sjögren's syndrome (SjS) and primary sclerosing cholangitis[17 37 38] with strong HLA class II associations, have also shown HLA class I associations, as described here for SSc. A haplotypic block containing this allele was previously nominally associated in Mexican patients with SSc.[35] These *HLA-B*08* associations have been attributed to the long ancestral 8.1 haplotype, supporting a common genetic background in autoimmunity.[16 39–41] This allele is in high LD (r^2=0.998) with the amino acid residue Asp9 located in the peptide binding groove of *HLA-B*, with a potential functional impact on antigen presentation.[17]

The association with SSc of independent signals in HLA classes I and II may suggest novel mechanisms for disease pathogenesis, including the involvement of not only CD4+ but also CD8+ T cells.[2 42] Interestingly, genes associated with CD8+ T-cell biology have been reported to be deregulated in skin biopsies of active SSc lesions, and these cells have been described to produce proinflammatory cytokines, contributing to the overproduction of collagen by fibroblasts and excessive fibrosis.[43] A recent report by Maehara *et al* assessed T-cell infiltrates in the skin of early dcSSc and showed that CD4+ cytotoxic T cells and CD8+ T cells are responsible for these infiltrates and induce apoptotic death of endothelial cells, contributing to the vasculopathy and fibrotic environment observed in SSc.[44 45]

Table 3 Summary of the independent association results from the stratified analysis

Gene	Alleles	OR*	P value*	Conditioned P value	Phenotype	OR†	Intracase P value†
HLA-DQA1	DQA1*02:01	0.54	5.23E-51	NA‡	lcSSc	0.71	2.08E-08
HLA-DQA1	DQA1*05:01	1.49	1.16E-30	1.59E-11	dcSSc	1.30	1.76E-11
HLA-DRB1	DRB1*08:01	3.18	9.73E-57	4.00E-64	ACA	2.17	1.42E-10
HLA-DRB1	DRB1*07:01	0.36	1.17E-63	1.84E-45	ACA	0.42	3.85E-27
HLA-DPA1	DPA1*02:01	1.87	7.91E-43	2.93E-19	ATA	2.41	1.09E-40
HLA-DQB1	DQB1*03:01	1.86	7.11E-47	7.00E-19	ATA	1.67	1.73E-22

*Association effect and p value compared with the control group.

†Association effect and p value in the intracase comparisons (dcSSc with lcSSc and ATA with ACA).

‡Not available as it was the most significant allele in the sequential conditional model.

ACA, anticentromere; ATA, antitopoisomerase; dcSSc, diffuse cutaneous systemic sclerosis; lcSSc, limited cutaneous systemic sclerosis.

Functional maturation defects have been detected in regulatory CD8+ lymphocytes from an ex vivo model of SSc[46] and differential regulatory programmes of IFN-associated genes in CD4+ and CD8+ T cells have been shown to lead to elevated serum interferon levels in patients with SSc.[47] Taken altogether, further studies on the contribution of CD8+ T cells in SSc may bear great therapeutic value, due to either its connection with the development of fibrosis or the assessment of the subpopulation of regulatory CD8+ cells in these patients.

Given that most of the independent SNPs in the global analysis mapped in non-coding regions of the genome, we performed a colocalisation study to assess if the associated variants were modulating gene expression. Our data showed immunity-related genes such as *HLA-DRB1*, *HLA-DRB6*, *HLA-DPA2* and the complement gene *C2* as eGENES regulated by the associated variants (table 2). Specifically, the risk allele of rs4713586 is correlated with an increased expression of the *C2* gene in transformed lymphocytes. This gene was previously associated with SLE[48] and psoriasis.[49] Interestingly, genetic variations on the complement genes have been recently described to contribute to the sex-biased susceptibility in highly related diseases like SLE and SjS,[50] and may be further explored in SSc. This could be seen as a limitation of our study because the reference panel used here does not allow the imputation of these structural variations.

One potential application of genetic studies is the identification of variants associated with clinical and serological subtypes to assist in patient stratification, and potentially to anticipate their progression and to propose specific therapeutic interventions. The determination of classical HLA alleles is routine in immunology laboratories for autoimmune diseases such as coeliac disease, ankylosing spondylitis and type 1 diabetes, and could be extended to others like SSc. To this aim, our stratified analysis showed that *HLA-DQA1*02:01* was significantly associated with lcSSc, while *HLA-DQA1*05:01* was exclusively associated with dcSSc. Regarding the serological stratifications, *HLA-DRB1*08:01* and *HLA-DRB1*07:01* were associated with ACA positive patients, further confirming associations reported in previous candidate gene GWAS, and Immunochip studies.[11 12 51] In the ATA-positive SSc subgroup, *HLA-DPA1*02:01* and *HLA-DQB1*03:01* showed exclusive and significant associations, and the latter was also reported in an Immunochip study.[11]

It is worth noting that the private associations were stronger when stratifying by the clinical and serological group of patients, despite the considerable loss of statistical power (online supplemental table 10). These results highlight the importance of analysing homogeneous groups of patients, reducing the loss of power due to phenotypical heterogeneity.[52] As expected, these alleles were significantly different even when comparing the group of patients among them and not with the control group, reinforcing the idea that they are present in specific clinical and serological subtypes of patients. Overall, the risk alleles identified thus far bear modest effects and a better understanding of the genetic structure of the disease will include interactions between several risk factors. Further studies warrant the simultaneous qualitative and quantitative assessments of allele-specific expression of the genes in order to detect context-specific regulatory effects.[53 54] Genotyping equivalent SNPs to the associated HLA alleles may also be of clinical utility, as SNP genotyping is straightforward and cost-efficient and has been proven to be very valuable to infer classical alleles for this and other rheumatic diseases.[21 55 56]

In summary, our extensive study of the HLA genes has confirmed and revealed novel associations with SSc susceptibility, highlighting for the first time the involvement of HLA class I genes in the pathogenesis of the disease. In addition, our data points to specific allelic associations that may serve as molecular biomarkers of clinical disease and serological subphenotypes. This evidence may eventually lead to early interventions that are crucial to avoid the devastating effects of the disease, and to develop specific and effective therapeutic options for patients with SSc.

Author affiliations
[1]Department of Cell Biology and Immunology, Institute of Parasitology and Biomedicine López-Neyra, CSIC, Granada, Andalucía, Spain
[2]Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center at Houston, Houston, Texas, USA
[3]Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy
[4]Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain
[5]Department of Internal Medicine, Clinic University Hospital, Granada, Spain
[6]Department of Rheumatology, Royal Adelaide Hospital, Adelaide, Victoria, Australia
[7]Department of Dermatology, University of Cologne, Cologne, Germany
[8]Department of Clinical and Molecular Science, Università Politecnica delle Marche and Ospedali Riuniti, Ancona, Italy
[9]Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands
[10]Centre for Genetics and Genomics Versus Arthritis, Division of Musculoskeletal and Dermatological Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, The University of Manchester, Manchester, UK
[11]NIHR Manchester Biomedical Research Centre, Manchester University NHS Foundation Trust, Manchester, UK
[12]Division of Musculoskeletal and Dermatological Sciences, The University of Manchester, Salford Royal NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK
[13]Laboratory for Statistical and Translational Genetics, RIKEN Center for Integrative Medical Sciences, Yokohama, Kanagawa, Japan
[14]Department of Rheumatology A, Hospital Cochin, Paris, Île-de-France, France
[15]NIHR Biomedical Research Centre, Guy's and Saint Thomas' NHS Foundation Trust and King's College, London, UK
[16]Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands
[17]Centre for Rheumatology, Royal Free and University College Medical School, London, UK

Collaborators International SSc Group: P Carreira, Department of Rheumatology, 12 de Octubre University Hospital, Madrid, Spain; I Castellvi, Department of Rheumatology, Santa Creu i Sant Pau University Hospital, Barcelona, Spain; R Ríos, Department of Internal Medicine, San Cecilio Clinic University Hospital, Granada, Spain; J L Callejas, Department of Internal Medicine, San Cecilio Clinic University Hospital, Granada, Spain; R García Portales, Department of Rheumatology, Virgen de la Victoria Hospital, Málaga, Spain; A Fernández-Nebro, Department of Rheumatology, Carlos Haya Hospital, Málaga, Spain; F J García-Hernández, Department of Internal Medicine, Virgen del Rocío Hospital, Sevilla, Spain; M A Aguirre, Department of Rheumatology, Reina Sofía/IMIBIC Hospital, Córdoba, Spain; B Fernández-Gutiérrez, Department of Rheumatology, San Carlos Clinic Hospital, Madrid, Spain; L Rodríguez-Rodríguez, Department of Rheumatology, San Carlos Clinic Hospital, Madrid, Spain; P García de la Peña, Department of Rheumatology, Madrid Norte Sanchinarro Hospital, Madrid, Spain; E Vicente, Department of Rheumatology, La Princesa Hospital, Madrid, Spain; J L Andreu, Department of Rheumatology, Puerta de Hierro Hospital-Majadahonda, Madrid, Spain; M Fernández de Castro, Department of Rheumatology, Puerta de Hierro Hospital-Majadahonda, Madrid, Spain; F J López-Longo, Department of Rheumatology, Gregorio Marañón University Hospital, Madrid, Spain; V Fonollosa, Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain; A Guillén, Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain; G Espinosa, Department of Internal Medicine, Clinic Hospital, Barcelona, Spain; C Tolosa, Department of Internal Medicine, Parc Tauli Hospital, Sabadell, Spain; A Pros, Department of Rheumatology, Hospital Del Mar, Barcelona, Spain; E Beltrán, Department of Rheumatology, Hospital Del Mar, Barcelona, Spain; M Rodríguez Carballeira, Department of Internal Medicine, Hospital Universitari Mútua Terrasa, Barcelona, Spain; F J Narváez, Department of Rheumatology, Bellvitge University Hospital, Barcelona, Spain; M Rubio Rivas, Department of Internal Medicine, Bellvitge University Hospital, Barcelona, Spain; V Ortiz-Santamaría, Department of Rheumatology, Granollers Hospital, Granollers,

Spain; A B Madroñero, Department of Internal Medicine, Hospital General San Jorge, Huesca, Spain; M A González-Gay, Epidemiology, Genetics and Atherosclerosis Research Group on Systemic Inflammatory Diseases, DIVAL, University of Cantabria, Santander, Spain; B Díaz, Department of Internal Medicine, Hospital Central de Asturias, Oviedo, Spain; L Trapiella, Department of Internal Medicine, Hospital Central de Asturias, Oviedo, Spain; M V Egurbide, Department of Internal Medicine, Hospital Universitario Cruces, Barakaldo, Spain; P Fanlo-Mateo, Department of Internal Medicine, Hospital Virgen del Camino, Pamplona, Spain; L Saez-Comet, Department of Internal Medicine, Hospital Universitario Miguel Servet, Zaragoza, Spain; F Díaz, Department of Rheumatology, Hospital Universitario de Canarias, Tenerife, Spain; J A Roman-Ivorra, Department of Rheumatology, Hospital Universitari i Politecnic La Fe, Valencia, Spain; J J Alegre Sancho, Department of Rheumatology, Hospital Universitari Doctor Peset, Valencia, Spain; M Freire, Department of Internal Medicine, Thrombosis and Vasculitis Unit, Complexo Hospitalario Universitario de Vigo, Vigo, Spain; F J Blanco Garcia, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña, Spain; N Oreiro, Department of Rheumatology, INIBIC-Hospital Universitario A Coruña, La Coruña, Spain; T Witte, Department of Clinical Immunology, Hannover Medical School, Hannover, Germany; A Kreuter, Department of Dermatology, Josefs-Hospital, Ruhr University Bochum, Bochum, Germany; G Riemekasten, Clinic of Rheumatology, University of Lübeck, Lübeck, Germany; P Airo, Service of Rheumatology and Clinic Immunology Spedali Civili, Brescia, Italy; C Magro, Department of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands; A E Voskuyl, Department of Rheumatology, VU University Medical Center, Amsterdam, The Netherlands; M C Vonk, Department of Rheumatology, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands; R Hesselstrand, Department of Rheumatology, Lund University, Lund, Sweden; A Nordin, Division of Rheumatology, Department of Medicine, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden; C Lunardi, Department of Medicine, Università degli Studi di Verona, Verona, Italy; A Gabrielli, Istituto di Clinica Medica Generale, Ematologia ed Immunologia Clinica, Università Politecnica delle Marche, Ancona, Italy; A Hoffmann-Vold, Department of Rheumatology, Oslo University Hospital, Oslo, Norway; J H W Distler, Department of Internal Medicine 3, Institute for Clinical Immunology, University of Erlangen-Nuremberg, Erlangen, Germany; L Padyukov, Division of Rheumatology, Department of Medicine, Karolinska University Hospital, Karolinska Institute, Stockholm, Sweden; B P C Koeleman, University Medical Center Utrecht, Utrecht, The Netherlands. Australian Scleroderma Interest Group: W Stevens, St. Vincent's Hospital, Melbourne, Victoria, Australia; M Nikpour, The University of Melbourne at St. Vincent's Hospital, Melbourne, Victoria, Australia; J Zochling, Menzies Research Institute Tasmania, University of Tasmania, Hobart, TAS, Australia; J Sahhar, Department Rheumatology, Monash Medical Centre, Melbourne, VIC, Australia; J Roddy, Rheumatology, Royal Perth Hospital, Perth, WA, Australia; P Nash, Research Unit, Sunshine Coast Rheumatology, Maroochydore, QLD, Australia; K Tymms, Canberra Rheumatology, Canberra, ACT, Australia; M Rischmueller, Department Rheumatology, The Queen Elizabeth Hospital, Woodville, SA, Australia; S Lester, Department Rheumatology, The Queen Elizabeth Hospital, Woodville, SA, Australia.

ORCID iDs
Marialbert Acosta-Herrera http://orcid.org/0000-0002-9868-6535
Shervin Assassi http://orcid.org/0000-0002-8059-9978
Lorenzo Beretta http://orcid.org/0000-0002-6529-6258
Anne Barton http://orcid.org/0000-0003-3316-2527
Matthew A Brown http://orcid.org/0000-0003-0538-8211
Javier Martin http://orcid.org/0000-0002-2202-0622

REFERENCES

1 Trowsdale J, Knight JC. Major histocompatibility complex genomics and human disease. *Annu Rev Genomics Hum Genet* 2013;14:301–23.
2 Matzaraki V, Kumar V, Wijmenga C, et al. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol* 2017;18:76.
3 Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. *N Engl J Med* 2009;360:1989–2003.
4 Allanore Y, Simms R, Distler O, et al. Systemic sclerosis. *Nat Rev Dis Primers* 2015;1:15002.
5 Denton CP, Khanna D. Systemic sclerosis. *Lancet* 2017;390:1685–99.
6 Katsumoto TR, Whitfield ML, Connolly MK. The pathogenesis of systemic sclerosis. *Annu Rev Pathol* 2011;6:509–37.
7 Bossini-Castillo L, López-Isac E, Martín J. Immunogenetics of systemic sclerosis: defining heritability, functional variants and shared-autoimmunity pathways. *J Autoimmun* 2015;64:53–65.
8 Angiolilli C, Marut W, van der Kroef M, et al. New insights into the genetics and epigenetics of systemic sclerosis. *Nat Rev Rheumatol* 2018;14:657–73.
9 Acosta-Herrera M, López-Isac E, Martín J. Towards a better classification and novel therapies based on the genetics of systemic sclerosis. *Curr Rheumatol Rep* 2019;21:44.
10 Radstake TRDJ, 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.
11 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.
12 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.
13 López-Isac E, Acosta-Herrera M, Kerick M, et al. Gwas for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. *Nat Commun* 2019;10:4955.
14 Dendrou CA, Petersen J, Rossjohn J, et al. Hla variation and disease. *Nat Rev Immunol* 2018;18:325–39.
15 Barturen G, Beretta L, Cervera R, et al. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. *Nat Rev Rheumatol* 2018;14:75–93.
16 Acosta-Herrera M, Kerick M, González-Serna D, et al. Genome-wide meta-analysis reveals shared new *loci* in systemic seropositive rheumatic diseases. *Ann Rheum Dis* 2019;78:311–9.
17 Raychaudhuri S, Sandor C, Stahl EA, et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291–6.
18 Morris DL, Taylor KE, Fernando MMA, et al. Unraveling multiple MHC gene associations with systemic lupus erythematosus: model choice indicates a role for HLA alleles and non-HLA genes in Europeans. *Am J Hum Genet* 2012;91:778–93.
19 Rothwell S, Chinoy H, Lamb JA, et al. Focused HLA analysis in Caucasians with myositis identifies significant associations with autoantibody subgroups. *Ann Rheum Dis* 2019;78:996–1002.
20 Sharif R, Fritzler MJ, Mayes MD, et al. Anti-fibrillarin antibody in African American patients with systemic sclerosis: immunogenetics, clinical features, and survival analysis. *J Rheumatol* 2011;38:1622–30.
21 Gourh P, Safran SA, Alexander T, et al. HLA and autoantibodies define scleroderma subtypes and risk in African and European Americans and suggest a role for molecular mimicry. *Proc Natl Acad Sci U S A* 2020;117:552–62.
22 van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American College of rheumatology/European League against rheumatism collaborative initiative. *Ann Rheum Dis* 2013;72:1747–55.
23 LeRoy EC, Medsger TA. Criteria for the classification of early systemic sclerosis. *J Rheumatol* 2001;28:1573–6.

24 de Bakker PIW, McVean G, Sabeti PC, *et al*. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat Genet* 2006;38:1166–72.

25 Jia X, Han B, Onengut-Gumuscu S, *et al*. Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS One* 2013;8:e64683.

26 Brown WM, Pierce J, Hilner JE, *et al*. Overview of the MHC fine mapping data. *Diabetes Obes Metab* 2009;11 Suppl 1:2–7.

27 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.

28 Yang J, Lee SH, Goddard ME, *et al*. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.

29 Yang J, Ferreira T, Morris AP, *et al*. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* 2012;44:369–75.

30 Pettersen EF, Goddard TD, Huang CC, *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* 2004;25:1605–12.

31 Giambartolomei C, Vukcevic D, Schadt EE, *et al*. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014;10:e1004383.

32 Gorlova O, Martin J-E, 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.

33 Arnett FC, Gourh P, Shete S, *et al*. Major histocompatibility complex (MHC) class II alleles, haplotypes and epitopes which confer susceptibility or protection in systemic sclerosis: analyses in 1300 Caucasian, African-American and Hispanic cases and 1000 controls. *Ann Rheum Dis* 2010;69:822–7.

34 Beretta L, Rueda B, Marchini M, *et al*. Analysis of class II human leucocyte antigens in Italian and Spanish systemic sclerosis. *Rheumatology* 2012;51:52–9.

35 Rodriguez-Reyna TS, Mercado-Velázquez P, Yu N, *et al*. Hla class I and II blocks are associated to susceptibility, clinical subtypes and autoantibodies in Mexican systemic sclerosis (SSC) patients. *PLoS One* 2015;10:e0126727.

36 González-Serna D, López-Isac E, Yilmaz N, *et al*. Analysis of the genetic component of systemic sclerosis in Iranian and Turkish populations through a genome-wide association study. *Rheumatology* 2019;58:289–98.

37 Liu JZ, Hov JR, Folseraas T, *et al*. Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis. *Nat Genet* 2013;45:670–5.

38 Gambino CM, Aiello A, Accardi G, *et al*. Autoimmune diseases and 8.1 ancestral haplotype: an update. *HLA* 2018;92:137–43.

39 Zhernakova A, Withoff S, Wijmenga C. Clinical implications of shared genetics and pathogenesis in autoimmune diseases. *Nat Rev Endocrinol* 2013;9:646–59.

40 Li YR, Li J, Zhao SD, *et al*. Meta-Analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nat Med* 2015;21:1018–27.

41 Ellinghaus D, Jostins L, Spain SL, *et al*. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat Genet* 2016;48:510–8.

42 Deng Q, Luo Y, Chang C, *et al*. The emerging epigenetic role of CD8+T cells in autoimmune diseases: a systematic review. *Front Immunol* 2019;10:856.

43 Klein M, Schmalzing M, Almanzar G, *et al*. Contribution of CD8+ T cells to inflammatory cytokine production in systemic sclerosis (SSC). *Autoimmunity* 2016;49:532–46.

44 Maehara T, Kaneko N, Perugino CA, *et al*. Cytotoxic CD4+ T lymphocytes may induce endothelial cell apoptosis in systemic sclerosis. *J Clin Invest* 2020;130:2451–64.

45 Fuschiotti P. T cells in SSC skin lesions: knowing your enemy. *Nat Rev Rheumatol* 2020;16:253–4.

46 Negrini S, Fenoglio D, Parodi A, *et al*. Phenotypic alterations involved in CD8+ Treg impairment in systemic sclerosis. *Front Immunol* 2017;8:18.

47 Ding W, Pu W, Wang L, *et al*. Genome-Wide DNA Methylation Analysis in Systemic Sclerosis Reveals Hypomethylation of IFN-Associated Genes in CD4$^+$ and CD8$^+$ T Cells. *J Invest Dermatol* 2018;138:1069–77.

48 Armstrong DL, Zidovetzki R, Alarcón-Riquelme ME, *et al*. Gwas identifies novel SLE susceptibility genes and explains the association of the HLA region. *Genes Immun* 2014;15:347–54.

49 Lee K-Y, Leung K-S, Tang NLS, *et al*. Discovering genetic factors for psoriasis through exhaustively searching for significant second order SNP-SNP interactions. *Sci Rep* 2018;8:15186.

50 Kamitaki N, Sekar A, Handsaker RE, *et al*. Complement genes contribute sex-biased vulnerability in diverse disorders. *Nature* 2020;582:577–81.

51 Simeón CP, Fonollosa V, Tolosa C, *et al*. Association of HLA class II genes with systemic sclerosis in Spanish patients. *J Rheumatol* 2009;36:2733–6.

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

53 Gutierrez-Arcelus M, Baglaenko Y, Arora J, *et al*. Allele-Specific expression changes dynamically during T cell activation in HLA and other autoimmune loci. *Nat Genet* 2020;52:247–53.

54 Chun S, Bang S-Y, Ha E, *et al*. Allele-Specific quantification of HLA-DRB1 transcripts reveals imbalanced allelic expression that modifies the amino acid effects in HLA-DRβ1. *Arthritis Rheumatol* 2021;73:381–91.

55 Hinks A, Bowes J, Cobb J, *et al*. Fine-Mapping the MHC locus in juvenile idiopathic arthritis (JIA) reveals genetic heterogeneity corresponding to distinct adult inflammatory arthritic diseases. *Ann Rheum Dis* 2017;76:765–72.

56 Terao C, Brynedal B, Chen Z, *et al*. Distinct HLA associations with rheumatoid arthritis subsets defined by serological subphenotype. *Am J Hum Genet* 2019;105:616–24.