RHEUMATOLOGY

Rheumatology 2019;58:289–298
doi:10.1093/rheumatology/key281
Advance Access publication 21 September 2018

Original article

Analysis of the genetic component of systemic sclerosis in Iranian and Turkish populations through a genome-wide association study

David González-Serna[1], Elena López-Isac[1], Neslihan Yilmaz[2], Farhad Gharibdoost[3], Ahmadreza Jamshidi[3], Hoda Kavosi[3], Shiva Poursani[3], Faraneh Farsad[4], Haner Direskeneli[5], Guhrer Saruhan-Direskeneli[6], Sofia Vargas[1], Amr H. Sawalha[7,8], Matthew A. Brown[9], Sule Yavuz[2], Mahdi Mahmoudi[3] and Javier Martin[1]

Abstract

Objectives. SSc is an autoimmune disease characterized by alteration of the immune response, vasculopathy and fibrosis. Most genetic studies on SSc have been performed in European-ancestry populations. The aim of this study was to analyse the genetic component of SSc in Middle Eastern patients from Iran and Turkey through a genome-wide association study.

Methods. This study analysed data from a total of 834 patients diagnosed with SSc and 1455 healthy controls from Iran and Turkey. DNA was genotyped using high-throughput genotyping platforms. The data generated were imputed using the Michigan Imputation Server, and the Haplotype Reference Consortium as a reference panel. A meta-analysis combining both case–control sets was conducted by the inverse variance method.

Results. The highest peak of association belonged to the HLA region in both the Iranian and Turkish populations. Strong and independent associations between the classical alleles *HLA-DRB1*11: 04* [$P = 2.10 \times 10^{-24}$, odds ratio (OR) = 3.14] and *DPB1*13: 01* ($P = 5.37 \times 10^{-14}$, OR = 5.75) and SSc were observed in the Iranian population. *HLA-DRB1*11: 04* ($P = 4.90 \times 10^{-11}$, OR = 2.93) was the only independent signal associated in the Turkish cohort. An omnibus test yielded HLA-DRB1 58 and HLA-DPB1 76 as relevant amino acid positions for this disease. Concerning the meta-analysis, we also identified two associations close to the genome-wide significance level outside the HLA region, corresponding to *IRF5-TNPO3* rs17424921-C ($P = 1.34 \times 10^{-7}$, OR = 1.68) and *NFKB1* rs4648133-C ($P = 3.11 \times 10^{-7}$, OR = 1.47).

Conclusion. We identified significant associations in the HLA region and suggestive associations in *IRF5-TNPO3* and *NFKB1* loci in Iranian and Turkish patients affected by SSc through a genome-wide association study and an extensive HLA analysis.

Key words: SSc, GWAS, Iranian and Turkish populations, risk loci

Rheumatology key messages

- The human leucocyte antigen region is strongly associated with SSc susceptibility in Iranian and Turkish populations.
- *IRF5-TNPO3* and *NFKB1* loci showed suggestive association with SSc.
- *GOT1-NKX2.3* emerged as a new potential susceptibility locus in SSc.

[1]Cell Biology and Immunology Department, Institute of Parasitology and Biomedicine López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain, [2]Department of Rheumatology, Istanbul Bilim University, Istanbul, Turkey, [3]Rheumatology Research Center, Tehran University of Medical Sciences, [4]Department of Rheumatology, Loghman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran, [5]Division of Rheumatology, Marmara University, [6]Department of Physiology, Istanbul Medical Faculty, Istanbul University, Istanbul, Turkey, [7]Division of Rheumatology, Department of Internal Medicine, [8]Department of Computational Medicine and Bioinformatics, The Center for Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor MI, USA and [9]Institute of Health and Biomedical Innovation, Queensland University

Submitted 25 April 2018; accepted 17 July 2018

Correspondence to: Javier Martin, Institute of Parasitology and Biomedicine López-Neyra, Consejo Superior de Investigaciones Científicas, Granada, Spain.
E-mail: javiermartin@ipb.csic.es

© The Author(s) 2018. Published by Oxford University Press on behalf of the British Society for Rheumatology. All rights reserved. For permissions, please email: journals.permissions@oup.com

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

CLINICAL SCIENCE

Introduction

SSc is an uncommon autoimmune disease with high mortality and morbidity, characterized by altered immune response, vascular damage and abnormal fibrosis of skin and internal organs [1]. SSc aetiology is complex but, similar to most autoimmune conditions, it is thought that a combination of environmental and a multiplicity of genetic factors leads to its development [2]. The disease is clinically classified into two main groups: the limited cutaneous form (lcSSc) and the diffuse cutaneous form (dcSSc). SSc is also characterized by production of specific autoantibodies, ACA and anti-topoisomerase antibody (ATA) (anti-Scl-70) being the two most common autoantibodies [1].

Large genetic studies, including genome-wide association studies (GWASs) and Immunochip analyses, have confirmed the HLA class II region as the most significant genetic region associated with SSc [3–6] and its different clinical and serological features [5, 6]. In addition, evidence has indicated the significant genetic component of the non-HLA region in SSc, with a total of 14 non-HLA loci showing association at the genome-wide level of significance (P values $< 5.0 \times 1\ 0^{-8}$) to date [3–12]. However, most of the previous GWASs and HLA analyses for SSc have been performed in European-descent populations. Therefore, the study of the genetic component in different ethnic groups could help to identify unknown, or pinpoint known, disease-associated susceptibility alleles, taking advantage of differences in linkage disequilibrium (LD) structure or allele frequencies between populations.

Thus, we decided to perform the first GWAS for SSc in Iranian and Turkish populations. Even though the Turkish genetic structure is unique, genetic overlap with other Middle East populations has been clearly established [13]. As the HLA region is the most significant genetic association reported for SSc, we also decided to perform an extensive analysis of the HLA region.

Methods

Study population

A series of 834 patients diagnosed with SSc (547 from Iran and 287 from Turkey) and 1455 unaffected and unrelated controls (830 from Iran and 625 from Turkey) were included in this study. All case samples fulfilled the American College of Rheumatology classification criteria for SSc [14]. Written informed consent was obtained from all the participants. The study was approved by local ethical committees from the different participant centres, and was carried out in accordance with the guidelines of our institution (Spanish Research Council) ethical committee, and the Declaration of Helsinki 1975, as revised in 1983.

Clinical information regarding subtypes of SSc defined by LeRoy *et al* [15], and genetic data relating to the presence or absence of ACA and ATA were collected. The clinical information was selected based on previous genetic studies determining specific associations with SSc subtypes or autoantibodies status (Supplementary Table S1, available at *Rheumatology* online). Some SSc patients showed other complex forms of the disease that could not be classified into dcSSc or lcSSc.

Genotyping

Genomic DNA was extracted from saliva samples or whole blood by standard methods. The GWAS genotyping of the SSc cases and controls was performed as follows: the Iranian and Turkish SSc cases, together with 136 Turkish controls were genotyped using the Illumina Infinium HumanCore-12v1 BeadChip. The remaining Turkish controls were genotyped using the Illumina HumanOmni1-Quad v1 BeadChip. The control group from Iran was genotyped using the Illumina Infinium CoreExome-24 BeadChip.

Quality control

We applied the same quality control (QC) criteria for both the Iranian and the Turkish GWAS data. Single-nucleotide polymorphisms (SNPs) and subjects with call rates lower than 98% and 95%, respectively, were removed using PLINK V.1.9 [16]. SNPs with minor allele frequencies <0.01 and those that were not in Hardy–Weinberg equilibrium (P values < 0.001) were also excluded. In addition, one subject per duplicate pair and per pair of first-degree relatives was also removed via the Genome function in PLINK V.1.9 with a Pi-HAT threshold of 0.4. Principal component (PC) analyses were performed to identify and exclude outliers based on their ethnicity by using PLINK V.1.9, and the GCTA64 and R-base under GNU Public license V.2. We calculated the 10 first PCs using $\sim$100 000 quality-filtered independent SNPs. Those subjects showing more than six standard deviations from the cluster centroids were considered as outliers. After QC, 764 SSc patients (505 from Iran and 259 from Turkey) and 1343 controls (770 from Iran and 573 from Turkey) were included in the analysis. A total of 242 501 and 186 435 genotyped SNPs remained after QC filtering of the Iranian and Turkish datasets, respectively (Supplementary Table S1, available at *Rheumatology* online).

Imputation of GWAS data

Imputation was performed using the Michigan Imputation Server [17]. The software SHAPEIT [18] was used in order to estimate the haplotypes, and the Haplotype Reference Consortium r1.1 [19] was used as the reference panel for both Turkish and Iranian genotyped data. Imputation was carried out in individual chunks of 50 000 Mb, covering whole-genome regions with a probability threshold for merging genotypes of 0.9, to maximize the quality of imputed variants. Imputed data were also subjected to the above-mentioned QC filters in PLINK V.1.9. A total of 6 313 908 and 5 885 622 SNPs were finally analysed in the Iranian and Turkish GWASs, respectively (Supplementary Table S1, available at *Rheumatology* online).

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

HLA imputation

Considering the previously reported strong HLA association with SSc, a more extensive analysis of the HLA region was conducted in the Turkish and Iranian cohorts. We extracted the extended MHC region (29 000 000–34 000 000 bp in chromosome 6) from the non-imputed data and imputed a total of 424 classical HLA alleles at two- and four-digits, 7261 SNPs and 1276 polymorphic amino acids. We used the SNP2HLA method with the Beagle software package [20] and a reference panel collected by the Type 1 Diabetes Genetics Consortium composed of 5225 individuals of European origin [21]. Alleles and amino acids with call rates <95%, and SNPs in deviation from Hardy–Weinberg equilibrium (P values < 0.001), were removed using PLINK V.1.9.

Statistical analyses

The statistical analyses were performed with PLINK and R. First, the Iranian and the Turkish cohorts were independently analysed by logistic regression on the best-guess genotypes (>0.9 probability), assuming an additive model and including the first 10 PCs as covariates as a correcting method for population stratification. Odds ratios (ORs) and 95% CIs were calculated according to Woolf's method. The genomic-inflation factor (λ) was estimated at 1.02 in both the Iranian and the Turkish cohorts (Q-Q plot, Supplementary Figure S1, available at *Rheumatology* online). We also performed stratified analysis considering the different clinical and serological features (dcSSc, lcSSc, ACA^+SSc and ATA^+SSc). Subsequently, the Iranian and the Turkish GWAS datasets were combined by inverse variance-weighted fixed-effects meta-analysis to integrate the two different association studies. Cochran's Q and I^2 tests were used to measure the heterogeneity of the ORs across studies. SNPs showing association P-values $< 5.0 \times 10^{-8}$ were regarded as significant. The presence of independent effects was examined using a stepwise logistic regression by conditioning on the lead SNP and the first 10 PCs. The statistical power of the meta-analysis, Iranian and Turkish GWASs for determining significance at P values of 5×10^{-8} (minor allele frequency = 30% and OR = 1.7) was 99%, 80% and 24%, respectively.

To determine the influence of the polymorphic amino acid positions on disease susceptibility, an omnibus association test [22] was carried out in the Iranian and the Turkish cohorts as described [23]. A null generalized linear model was established, including the 10 first PCs as covariates. Next, an alternative model, including the same variables and all the possible alleles in the analysed amino acid positions was built. Finally, to compare both models a likelihood ratio test was conducted.

Results

Following QC and imputation, a total of 5 698 748 SNPs in 764 SSc patients (505 from Iran and 259 from Turkey) and 1343 controls (770 from Iran and 573 from Turkey) were analysed.

As shown in Fig. 1, a high association peak corresponding to the HLA class II region reached the GWAS significance threshold (5×10^{-8}) in both the Iranian and Turkish cohorts. Specifically, for the Iranian cohort, the top associated signal belonged to the SNP rs9268923 ($P = 6.55 \times 10^{-18}$, OR = 2.29) (Fig. 1A). Out of the HLA region, some suggestive level associations were found, highlighting the *IRF5-TNPO3* (interferon regulatory factor 5-transportin 3) and the *NFKB1* (Nuclear Factor Kappa B Subunit 1) loci (Table 1), which represent previously reported susceptibility loci for SSc. The top associated signal for the *IRF5-TNPO3* locus corresponds to the SNP rs17424921 ($P = 5.28 \times 10^{-5}$, OR = 1.64), located in an intergenic region and 5' upstream from *TNPO3* gene. This SNP is in strong LD (r^2=0.92) in our dataset with rs10488631 ($P = 6.15 \times 10^{-5}$, OR = 1.62), a previously reported SNP associated with SSc [3, 7], located in the intergenic region between *IRF5* and *TNPO3*. Regarding the *NFKB1* locus, the top associated SNP was the rs4648133 ($P = 1.52 \times 10^{-5}$, OR = 1.48), located in an intronic position. Interestingly, this SNP is in moderate LD (r^2=0.73) with rs1598859 ($P = 6.23 \times 10^{-3}$, OR = 1.27), a previously reported SNP associated with SSc, located in an intronic position of the *NFKB1* gene [9].

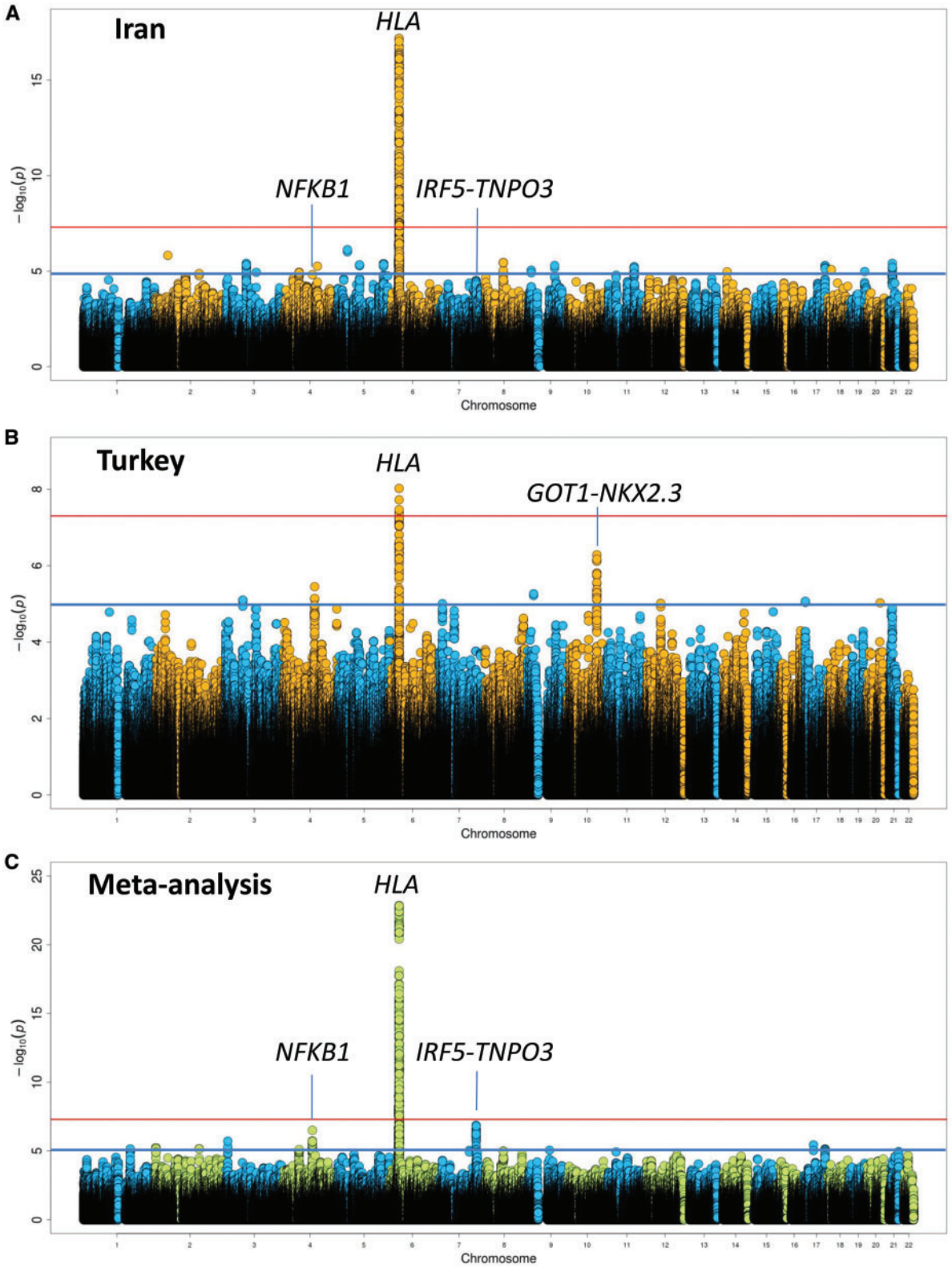
In the Turkish cohort (Fig. 1B), the top associated SNP corresponding to the HLA class II region was rs34297496 ($P = 9.46 \times 10^{-9}$, OR = 2.16). Out of the HLA region, one signal corresponding to the SNP rs7095491 reached the suggestive level of significance ($P = 5.17 \times 10^{-7}$, OR = 1.88). This SNP is located in an intergenic region between *GOT1* (glutamic-oxaloacetic transaminase 1) and *NKX2.3* (NK2 Homeobox 3). This locus was specifically associated with SSc in the Turkish cohort, but did not reach a suggestive level of significance in either the Iranian cohort ($P = 3.14 \times 10^{-2}$, OR = 0.89) or the meta-analysis ($P = 3.21 \times 10^{-1}$, OR = 1.07).

Meta-analysis

We decided to perform a meta-analysis in order to identify susceptibility loci to SSc that are associated with moderate effect sizes in the Iranian and the Turkish cohorts, independently. The HLA class II region was the highest association peak observed in the meta-analysis (Fig. 1C), and the top associated signal belonged to the SNP rs28746976 ($P = 1.41 \times 10^{-23}$, OR = 2.09). Out of the HLA region, two signals corresponding to the *IRF5-TNPO3* and *NFKB1* loci almost reached the genome-wide significance threshold (Table 1). The top associated signals corresponded to the same SNPs that were observed in the Iranian cohort for both loci: *IRF5-TNPO3* rs17424921 ($P = 1.34 \times 10^{-7}$, OR = 1.68) and *NFKB1* rs4648133 ($P = 3.11 \times 10^{-7}$, OR = 1.47) (Supplementary Figs S2 and S3, available at *Rheumatology* online). Non-HLA loci showing genome-wide significant associations ($P < 5 \times 10^{-8}$) were reported for their associations in our datasets (Supplementary Table S2, available at *Rheumatology* online).

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

FIG. 1 Manhattan plot of the genome-wide association study results from the Iranian cohort (**A**), Turkish cohort (**B**) and meta-analysis (**C**)



The values on the *y*-axes denote the $-\log_{10}$ transformed *P*-values. Genomic positions for each SNP for 22 autosomes are plotted on the *x*-axis. The red line denotes the *a priori* threshold for genome-wide significance ($P = 5 \times 10^{-8}$). The suggestive level of significance ($P = 1 \times 10^{-5}$) is highlighted in blue. SNP: single-nucleotide polymorphism.

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

Table 1 Non-HLA loci associated with SSc at the suggestive significance level ($P < 1 \times 10^{-5}$) in the meta-analysis

Loci	Chr	BP[a]	Most significant SNP	Minor allele	Iran		Turkey		Meta-analysis			
					P-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	Q	I^2
IRF5-TNPO3	7	128 708 122	rs17424921	C	5.28×10^{-5}	1.64 (1.29, 2.08)	6.57×10^{-4}	1.77 (1.27, 2.46)	1.34×10^{-7}	1.68 (1.39, 2.03)	0.69	0
	7	128 594 183	rs10488631	C	6.15×10^{-5}	1.62 (1.28, 2.05)	3.05×10^{-3}	1.66 (1.18, 2.33)	6.35×10^{-7}	1.63 (1.34, 1.97)	0.90	0
NFKB1	4	103 536 413	rs4648133	C	1.52×10^{-5}	1.48 (1.24, 1.77)	6.29×10^{-3}	1.45 (1.11, 1.89)	3.11×10^{-7}	1.47 (1.27, 1.70)	0.90	0
GRM7-LOC101927394	3	7 907 077	rs9821717	T	2.52×10^{-5}	0.68 (0.57, 0.82)	2.23×10^{-2}	0.74 (0.57, 0.96)	1.91×10^{-6}	0.70 (0.60, 0.80)	0.59	0
UNC45B	17	33 501 644	rs12452554	C	3.77×10^{-4}	1.74 (1.28, 2.35)	2.54×10^{-3}	2.00 (1.28, 3.13)	3.57×10^{-6}	1.81 (1.41, 2.33)	0.61	0
PXDN-MYT1L	2	1 785 659	rs2059413	C	2.29×10^{-4}	0.71 (0.60, 0.85)	8.01×10^{-3}	0.70 (0.54, 0.91)	5.70×10^{-6}	0.71 (0.61, 0.82)	0.92	0
LOC101929282-RBM43	2	151 773 679	rs917238	G	1.73×10^{-5}	1.54 (1.27, 1.88)	9.15×10^{-2}	1.28 (0.96, 1.72)	6.86×10^{-6}	1.45 (1.23, 1.71)	0.30	6.49
ATF6-OLFML2B	1	161 937 892	rs3002626	A	7.41×10^{-4}	0.73 (0.60, 0.87)	2.38×10^{-3}	0.65 (0.50, 0.86)	6.98×10^{-6}	0.70 (0.60, 0.82)	0.52	0
HOPX-SPINK2	4	57 558 089	rs9968446	A	1.21×10^{-5}	1.44 (1.22, 1.70)	1.46×10^{-1}	1.21 (0.93, 1.56)	7.92×10^{-6}	1.37 (1.19, 1.57)	0.25	22.08
OSTF1-PCSK5	9	78 042 927	rs473299	A	1.42×10^{-4}	1.44 (1.19, 1.74)	2.02×10^{-2}	1.38 (1.05, 1.80)	8.69×10^{-6}	1.42 (1.21, 1.65)	0.78	0

[a]BP corresponding to the NCBI build 37. BP: base pair; Chr: chromosome; SNP: single nucleotide polymorphism; OR: odds ratio; Q and I2: heterogeneity value.

HLA analysis

In order to extensively analyse the association of the HLA region and SSc in both the Iranian and the Turkish populations, a comprehensive HLA imputation was performed. As stated above, a strong association signal was observed within the HLA class II region in both the Iranian and the Turkish GWAS datasets (Table 2). With respect to the Iranian cohort, the top associated signal belonged to the classical allele *HLA-DRB1*11: 04* ($P = 2.10 \times 10^{-24}$, OR = 3.14). After controlling for *HLA-DRB1*11: 04*, an independent secondary effect was found in the HLA-DPB1 region. Specifically, the top associated signal was the *HLA-DPB1*13: 01* ($P = 5.37 \times 10^{-14}$, OR = 5.75), and no additional independent associations were observed after conditioning on both signals, *DRB1*11: 04* and *DPB1*13: 01* (Fig. 2). Significant insights were obtained from the ATA^+SSc vs. controls stratified analysis, because a stronger association between *HLA-DRB1*11: 04* and the disease was evident ($P = 2.49 \times 10^{-34}$, OR = 4.92). In addition, after controlling for this classical allele, *HLA-DPB1*13: 01* also showed a stronger association ($P = 6.44 \times 10^{-21}$, OR = 10.60) (Supplementary Table S3, available at *Rheumatology* online). We also performed a stratified analysis comparing ATA^+SSc vs. ATA^-SSc, showing a strong association of *HLA-DRB1*11: 04* with ATA^+SSc ($P = 2.69 \times 10^{-14}$, OR = 5.06). After conditioning on *HLA-DRB1*11: 04*, *HLA-DPB1*13: 01* also showed a statistically significant association with ATA^+SSc ($P = 5.02 \times 1\ 0^{-8}$, OR = 10.21). Finally, we performed a stratified analysis comparing ATA^-SSc vs. controls, in which we did not find any significant association, either for *HLA-DRB1*11: 04* [$P = 7.78 \times 10^{-1}$, OR = 1.05 (0.73–1.53)] or for *HLA-DPB1*13: 01* [$P = 6.48 \times 10^{-1}$, OR = 1.22 (0.52–2.86)]. In the Turkish cohort, *DRB1*11: 04* ($P = 4.90 \times 10^{-11}$, OR = 2.93) showed the most significant association, and no independent secondary effects were found after controlling for this HLA classical allele (Fig. 3). All statistically significant HLA associations (classical alleles, amino acid variants and SNPs) found in the Iranian and the Turkish HLA analyses are summarized in Supplementary Table S4, available at *Rheumatology* online.

Subsequently, specific amino acid positions that could be responsible for the association observed for these classical alleles were examined by means of an omnibus test. The most relevant amino acid positions for disease risk in the Iranian cohort were positions 58 ($P_{LRT} = 6.24 \times 10^{-24}$) and 67 ($P_{LRT} = 5.76 \times 10^{-22}$) of the HLA-DRβ1 molecule (Table 3), which were in strong LD (r^2=0.83). After conditioning on the strongest association (position 58), position 76 ($P_{LRT} = 1.57 \times 10^{-13}$) of the HLA-DPβ1 protein remained independently associated. None of the other signals remained significant after conditioning on positions 58 and 76 of the HLA-DRβ1 and HLA-DPβ1 molecules, respectively. The most associated amino acid residues for the previously mentioned positions were Glu ($P = 5.39 \times 10^{-22}$, OR = 2.57) and Phe ($P = 1.08 \times 10^{-20}$, OR = 2.38) in positions 58 and 67 of the HLA-DRβ1 molecule, respectively, and Ile ($P = 1.52 \times 10^{-12}$, OR = 4.60) in

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

TABLE 2 Classical four digit HLA alleles showing the strongest association with SSc

	Iran				Turkey			
	Unconditioned		Conditioned on *HLA-DRB1*11: 04*		Unconditioned		Conditioned on *HLA-DRB1*11: 04*	
HLA allele	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)	*P*-value	OR (95% CI)
*DRB1*11: 04*	2.10×10^{-24}	3.14 (2.52, 3.91)	NA	NA	4.90×10^{-11}	2.93 (2.12, 4.03)	NA	NA
*DQB1*03: 01*	2.04×10^{-17}	2.22 (1.84, 2.66)	3.08×10^{-2}	1.31 (1.03, 1.68)	9.16×10^{-7}	1.93 (1.49, 2.51)	4.51×10^{-1}	1.15 (0.80, 1.64)
*DPB1*13: 01*	7.23×10^{-13}	5.05 (3.24, 7.86)	5.37×10^{-14}	5.75 (3.65, 9.07)	1.79×10^{-3}	2.93 (1.49, 5.75)	4.48×10^{-4}	3.43 (1.72, 6.84)
*DQA1*05: 01*	1.08×10^{-10}	1.77 (1.49, 2.11)	6.54×10^{-1}	1.05 (0.84, 1.31)	1.08×10^{-5}	1.77 (1.37, 2.28)	3.46×10^{-1}	1.16 (0.85, 1.59)

position 76 of the HLA-DPβ1 (Supplementary Table S4, available at *Rheumatology* online). In the case of the Turkish cohort, position 58 of the HLA-DRβ1 molecule was the only signal reaching the significance threshold ($P_{LRT} = 4.96 \times 10^{-8}$) (Table 3), Ala being the most associated amino acid residue in that position ($P = 9.42 \times 10^{-8}$, OR = 2.13). All statistically significant amino acid positions found in the Iranian and the Turkish cohorts are summarized in Supplementary Table S5, available at *Rheumatology* online.

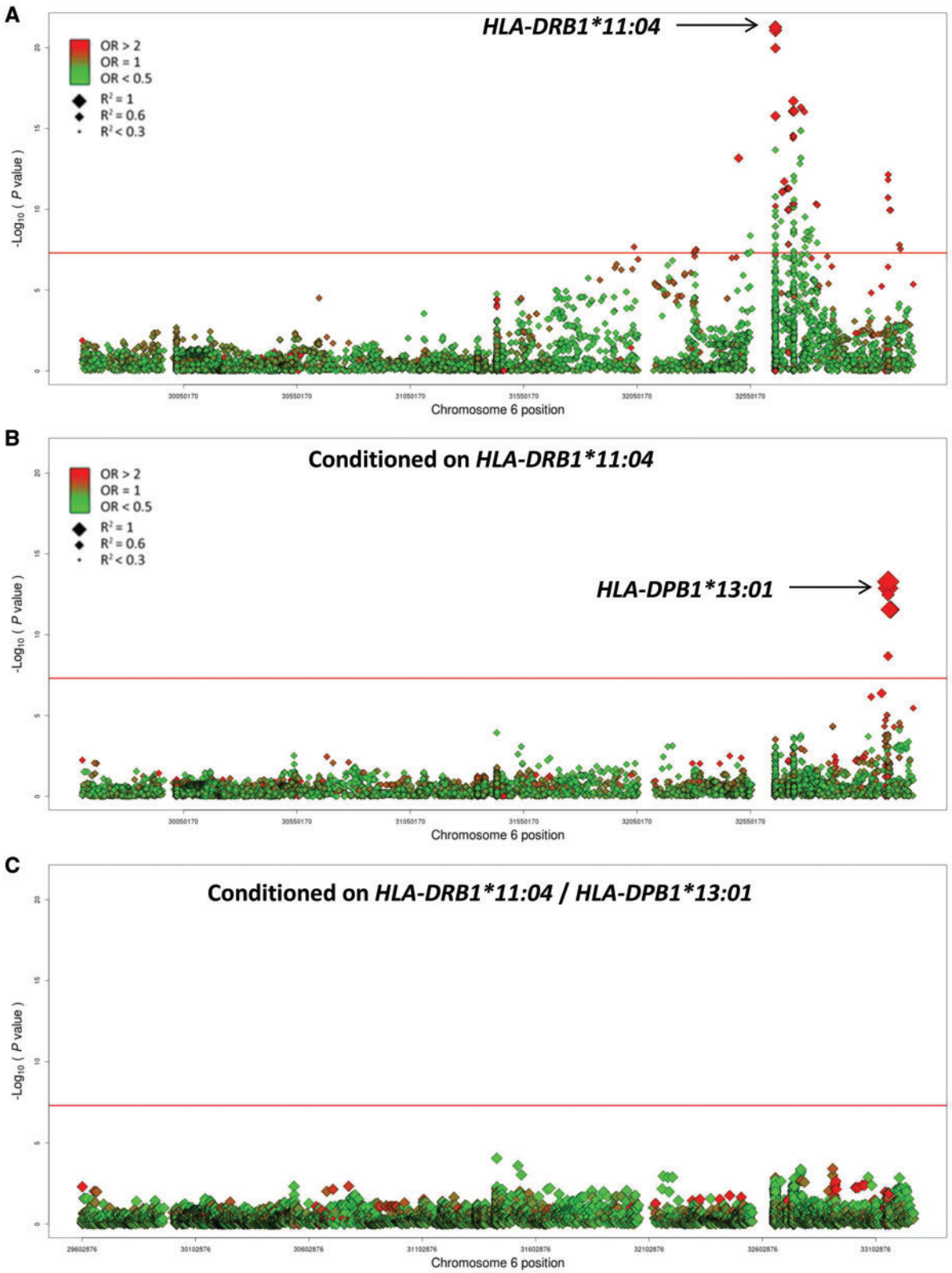
Discussion

This study represents the first GWAS of SSc performed in Iranian and Turkish populations. Given the relevance of the HLA region in SSc predisposition, and the lack of information about the role of HLA genes in SSc in Iranian or Turkish populations, we performed an extensive analysis of the HLA region, identifying the HLA class II region as the most strongly associated loci to SSc in both Middle Eastern populations. Interestingly, we observed that the *HLA-DRB1*11: 04* classical allele showed the most significant associations in both populations. Our data reinforce the role of the *HLA-DRB1*11: 04* allele in susceptibility to SSc in different ethnic populations, including Caucasians [24–26], African-Americans [24] and Mexicans [27]. Furthermore, our results suggest for the first time that this effect could be driven by the amino acid Glu-58. Nevertheless, amino acid Phe-67, which is in high LD with Glu-58 in our study, has been previously set as a relevant amino acid in susceptibility to SSc [5]. After conditioning on *HLA-DRB1*11: 04*, the classical allele *HLA-DPB1*13: 01*, which has been reported as a susceptibility allele in Caucasians [24] and Koreans [28], remains significant in the Iranian cohort. Our results also suggest that the association between *HLA-DPB1*13: 01* and SSc could be driven by amino acid Ile-76, which is in line with previous results [5]. Furthermore, amino acid 76 is part of the binding pocket of the HLA-DPβ1 molecule [29]. However, associations of *HLA-DRB1*11: 04* and *HLA-DPB1*13: 01* with SSc were explained by its strong correlation with ATA in previously mentioned studies. This specific correlation with ATA$^+$SSc was verified in our stratified analysis. In this sense, the presence of *HLA-DRB1*11: 04* and *HLA-DPB1*13: 01* classical alleles being specifically associated with ATA in the non-stratified analysis could be owing to a clinical predominance of ATA$^+$SSc patients in the Iranian population. On the other hand, the lack of association of *HLA-DPB1*13: 01* with ATA+SSc in the Turkish cohort could be due to the lower sample size as compared with the Iranian cohort. Owing to the low allele frequency of *HLA-DPB1*13: 01* in our Turkish samples (4%), small changes in the sample size can result in huge changes in the statistical power; e.g. for an OR= 3.2 and an allele frequency of 4%, the statistical power in the Turkish cohort was 25% and in the Iranian cohort was 85%.

Outside of the HLA region, two suggestive associations at *IRF5-TNPO3* and *NFKB1* loci were found. *IRF5-TNPO3* was one of the first associated loci identified in SSc, and

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

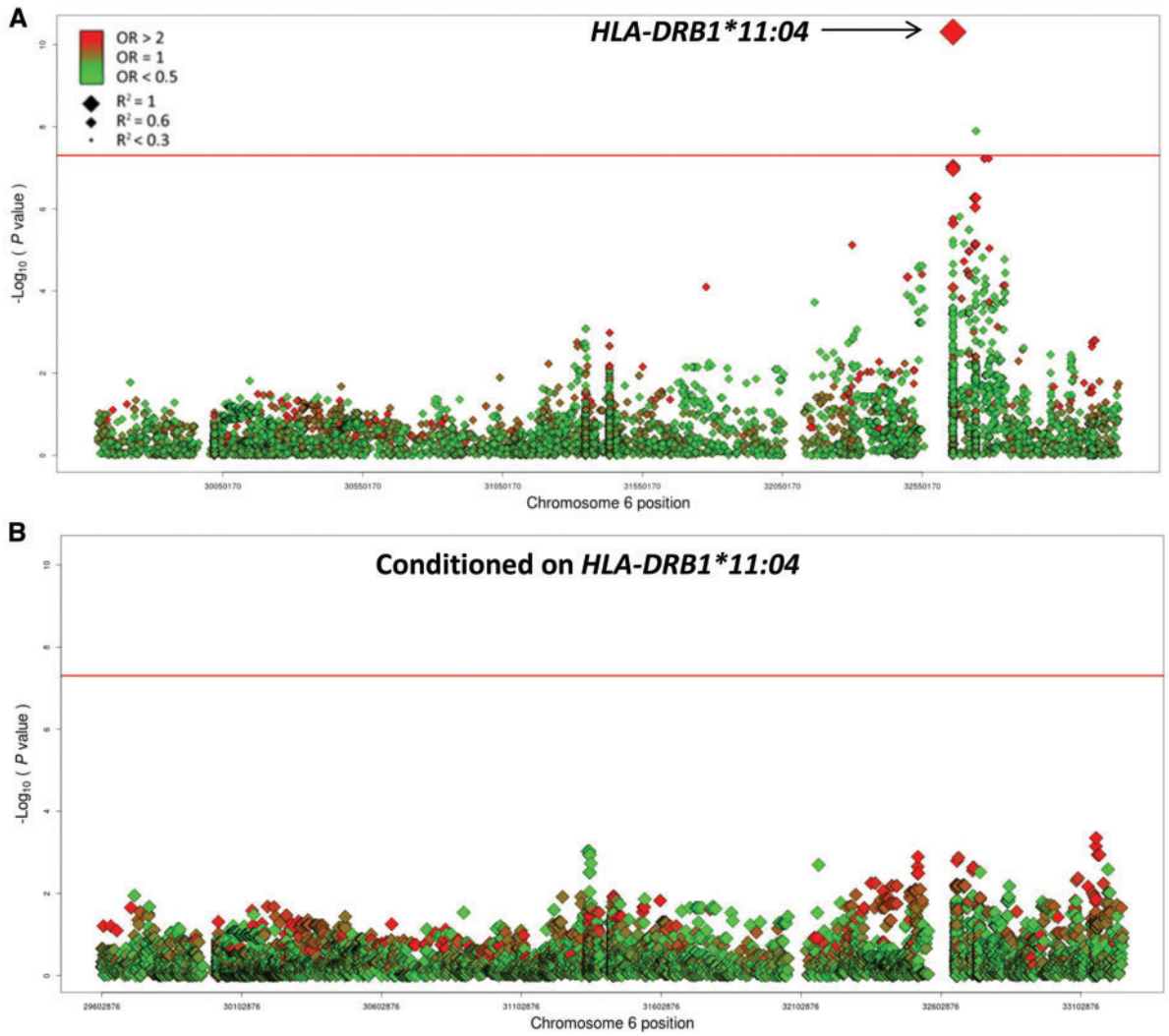
FIG. 2 Association results for the HLA región in Iranian patients with SSc



(**A**) Unconditioned test of the HLA region. (**B**) Results after conditioning on the *HLA-DRB1*11: 04* classical allele. (**C**) Results after conditioning on the *HLA-DRB1*11: 04* and *DPB1*13: 01* alleles. The red/green colour gradient represents the effect direction of each analysed variant (red for risk and green for protection). The size of the diamonds indicates the degree of linkage disequilibrium with the classical allele *HLA-DRB1*11: 04* and *DPB1*13: 01* in A and B, respectively. The red line represents the genome-wide level of significance ($P = 5 \times 10^{-8}$).

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

FIG. 3 Association results for the HLA region in Turkish patients with SSc



(**A**) Unconditioned test of the HLA region. (**B**) Results after conditioning on the *HLA-DRB1*11: 04* classical allele. The red/green colour gradient represents the effect direction of each analysed variant (red for risk and green for protection). The red line represents the genome-wide level of significance ($P = 5 \times 10^{-8}$).

this has been replicated in various studies and ethnicities [3–7, 30, 31], indicating it is a firm susceptibility factor for SSc and other autoimmune diseases such as SLE and RA [32, 33]. On the other hand, the *NFKB1* locus was identified as a susceptibility gene in SSc in a GWAS meta-analysis and in a subsequent candidate gene approach analysis in a Caucasian population [9, 34]. NF-κB has been broadly described as controlling the inflammatory process, and its role in autoimmunity is widely accepted [35]. Furthermore, the interaction of *NFKB1* with other well-defined susceptibility genes in SSc, such as *TNFAIP3* (Tumor Necrosis Factor Inducible Protein A20) [36], which encodes a protein that inhibits NF-κB activation, suggests that it could be a good candidate gene to be involved in SSc. Regarding the Turkish cohort, a suggestive level associated signal corresponding to the *GOT1-NKX2.3* locus emerged. This locus is a well-established signal associated with other immune-mediated diseases, such as ulcerative colitis and Crohn's disease [37, 38], suggesting a potential role of the *GOT1-NKX2.3* locus in autoimmunity. Furthermore, the strongest association reported for the *GOT1-NKX2.3* locus in both studies (rs4409764) is in strong LD (r^2=0.93) with the top associated SNP observed in our Turkish cohort (rs7095491).

Despite the successful identification of SSc susceptibility genes in Iranian and Turkish populations, our study had some limitations. In this regard, larger cohorts could help to elucidate whether the two suggestive associations observed in the meta-analysis (*IRF5-TNPO3* and *NFKB1*) reach the genome-wide significance threshold. Nevertheless, the results in four of the six most relevant SSc hits outside the HLA obtained by Carmona *et al*. [39]

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

TABLE 3 Amino acid positions showing the strongest association with SSc after omnibus test; LRT: likelihood ratio test

HLA molecule	Amino acid position	Centre codon position	Tested alleles	Iran		Turkey	
				P_{LRT} (unconditioned)	P_{LRT} conditioned on DRβ1 position 58	P_{LRT} (unconditioned)	P_{LRT} conditioned on DRβ1 position 58
DRβ1	58	32659974	2	6.24×10^{-24}	NA	4.96×10^{-8}	NA
DRβ1	67	32659947	3	5.76×10^{-22}	4.18×10^{-2}	6.19×10^{-6}	8.40×10^{-1}
DQβ1	45	32740702	2	1.14×10^{-17}	4.67×10^{-1}	3.68×10^{-7}	4.15×10^{-1}
DQβ1	55	32740672	3	1.25×10^{-16}	1.59×10^{-2}	6.95×10^{-3}	2.72×10^{-1}
DQβ1	140	32737868	2	2.47×10^{-16}	3.68×10^{-2}	3.97×10^{-3}	4.05×10^{-1}
DQβ1	182	32737742	2	2.47×10^{-16}	3.68×10^{-2}	3.97×10^{-3}	4.05×10^{-1}
DPβ1	76	33156640	3	8.06×10^{-14}	1.57×10^{-13}	8.40×10^{-3}	3.93×10^{-3}

in a Turkish population were very similar to our GWAS results (Supplementary Table S6, available at *Rheumatology* online), highlighting the reproducibility of our study.

In summary, our results confirm previous reports of association of the HLA region with SSc susceptibility, as well as two associations almost reaching the significance threshold outside the HLA region, in Iranian and Turkish populations. This study sheds light on the unexplored genetic background of SSc in these populations, which contributes to a better understanding of genetic structure and pathogenesis of the disease.

Acknowledgements

This work is part of the Doctoral Thesis Bases Genéticas de la Esclerosis Sistémica: Integrando Genómica y Transcriptómica. We thank Sonia García and Gema Robledo for their excellent technical assistance, and all the patients and healthy controls for kindly participating in this research.

Funding: This work was supported by the Spanish Ministry of Economy and Competitiveness [SAF2015-66761-P to J.M. and SAF2015-66761-P to D.G.S.] and The Cooperative Research Thematic Network (RETICS) program, from the Instituto de Salud Carlos III (ISCIII, Health Ministry, Madrid, Spain) [RD16/0012/0004].

Disclosure statement: The authors have declared no conflicts of interests.

Supplementary data

Supplementary data are available at *Rheumatology* online.

References

1 Gabrielli A, Avvedimento E, Krieg T. Scleroderma. N Engl J Med 2009;360:1989–2003.

2 Allanore Y, Simms R, Distler O *et al*. Systemic sclerosis. Nat Rev Dis Primers 2015;1:15002.

3 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.

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

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

6 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.

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

8 Bossini-Castillo L, Martin JE, Broen J *et al*. A GWAS follow-up study reveals the association of the *IL12RB2* gene with systemic sclerosis in Caucasian populations. Hum Mol Genet 2012;21:926–33.

9 Martin JE, Broen JC, Carmona FD *et al*. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Hum Mol Genet 2012;21:2825–35.

10 López-Isac E, Bossini-Castillo L, Guerra S *et al*. Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus. Arthritis Rheum 2014;66:3521–3.

11 López-Isac E, Campillo-Davo D, Bossini-Castillo L *et al*. Influence of *TYK2* in systemic sclerosis susceptibility: a new *locus* in the IL-12 pathway. Ann Rheum Dis 2016;75:1521–6.

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

13 Hodoğlugil U, Mahley RW. Turkish population structure and genetic ancestry reveal relatedness among Eurasian populations. Ann Hum Genet 2012;76:128–41.

14 Masi AT, Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria Committee. Preliminary criteria for the

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025

classification of systemic sclerosis (scleroderma). Arthritis Rheum 1980;23:581–90.

15 LeRoy EC, Black C, Fleischmajer R *et al*. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J Rheumatol 1988;15:202–5.

16 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.

17 Das S, Forer L, Schönherr S *et al*. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–7.

18 Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods 2011;9:179–81.

19 McCarthy S, Das S, Kretzschmar W *et al*. A reference panel of 64, 976 haplotypes for genotype imputation. Nat Genet 2016;48:1279–83.

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

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

22 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.

23 Carmona FD, Mackie SL, Martín JE *et al*. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. Am J Hum Genet 2015;96:565–80.

24 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.

25 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.

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

27 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.

28 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;60:3807–14.

29 Dai S, Murphy GA, Crawford F *et al*. Crystal structure of HLA-DP2 and implications for chronic beryllium disease. Proc Natl Acad Sci U S A 2010;107:7425–30.

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;60:225–33.

31 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.

32 Hom G, Graham R, Modrek B *et al*. Association of systemic lupus erythematosus with C8orf13–BLK and ITGAM–ITGAX. N Engl J Med 2008;358:900–9.

33 Okada Y, Wu D, Trynka G *et al*. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 2014;506:376–81.

34 Arismendi M, Giraud M, Ruzehaji N *et al*. Identification of *NF-κB* and *PLCL2* as new susceptibility genes and highlights on a potential role of *IRF8* through interferon signature modulation in systemic sclerosis. Arthritis Res Ther 2015;17:71.

35 Sun S-C, Chang J-H, Jin J. Regulation of nuclear factor-κB in autoimmunity. Trends Immunol 2013;34:282–9.

36 Bhattacharyya S, Wang W, Graham LV, Varga J. A20 suppresses canonical Smad-dependent fibroblast activation: novel function for an endogenous inflammatory modulator. Arthritis Res Ther 2016;18:216.

37 Liu JZ, van Sommeren S, Huang H *et al*. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979–86.

38 de Lange KM, Moutsianas L, Lee JC *et al*. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256–61.

39 Carmona FD, Onat AM, Fernández-Aranguren T *et al*. Analysis of Systemic Sclerosis-associated Genes in a Turkish Population. J Rheumatol 2016;43:1376–9.

Downloaded from https://academic.oup.com/rheumatology/article/58/2/289/5105787 by Universidad de Granada - Biblioteca user on 23 January 2025