Martínez-López Javier (Orcid ID: 0000-0003-0184-9543)
Rangel-Peláez Carlos (Orcid ID: 0000-0001-7697-1696)
Distler Oliver (Orcid ID: 0000-0002-2413-1959)
Proudman Susanna M (Orcid ID: 0000-0002-3046-9884)
Herrick Ariane L (Orcid ID: 0000-0003-4941-7926)
Beretta Lorenzo (Orcid ID: 0000-0002-6529-6258)
Assassi Shervin (Orcid ID: 0000-0002-8059-9978)
Denton Christopher P (Orcid ID: 0000-0003-3975-8938)
Mayes Maureen D. (Orcid ID: 0000-0001-5070-2535)
Acosta-Herrera Marialbert (Orcid ID: 0000-0002-9868-6535)

Running title: *C4* copy numbers in Scleroderma subtypes

Title: Distinct effects of complement *C4A* and *C4B* copy number in Systemic Sclerosis serological and clinical subtypes

Javier Martínez-López[1,2], MSc; Carlos Rangel-Peláez[1], MSc; Inmaculada Rodriguez-Martin[1], MSc; Alfredo Guillen-Del-Castillo[3], MD, PhD; Carmen P Simeón-Aznar[3], MD, PhD; José L Callejas[2], MD, PhD; Oliver Distler[4], MD, Prof.; International SSc Group; PRECISESADs Clinical Consortium; Susanna M Proudman[5], MD, Prof.; Mandana Nikpour[6,7], MD, Prof.; Australian Scleroderma Interest Group (ASIG); Nicolas Hunzelmann[8], MD, Prof.; Gianluca Moroncini[9,10], MD, PhD; Jeska K de Vries-Bouwstra[11], MD, PhD; Ariane L Herrick[12], MD, Prof.; Yannick Allanore[13], MD, Prof.; Marta E. Alarcon-Riquelme[14], MD, Prof.; Lorenzo Beretta[15], MD, PhD; Shervin Assassi[16], MD, Prof.; Christopher P Denton[17], MD, Prof.; Maureen D Mayes[16], MD, Prof.; Javier Martin[1,*], MD, Prof.; Marialbert Acosta-Herrera[1,*], PhD; Martin Kerick[1,*,c], PhD

1. Institute of Parasitology and Biomedicine López-Neyra, CSIC, 18016 Granada, Spain
2. Systemic Autoimmune Disease Unit, Hospital Clínico San Cecilio, Instituto de Investigación Biosanitaria Ibs. GRANADA, Granada, Spain
3. Unit of Systemic Autoimmune Diseases, Department of Internal Medicine, Hospital Universitari Vall d'Hebron, Barcelona, Spain
4. Department of Rheumatology, University Hospital Zurich, University of Zurich, Zurich, Switzerland
5. Royal Adelaide Hospital and the University of Adelaide, Adelaide, South Australia, Australia
6. Department of Medicine, University of Melbourne, Fitzroy, Victoria, Australia
7. The University of Sydney School of Public Health, Sydney, New South Wales, Australia
8. Department of Dermatology, University of Cologne, Cologne, Germany
9. Department of Clinical and Molecular Science, Marche Polytechnic University, Ancona, Italy
10. Department of Internal Medicine, Marche University Hospital, Ancona, Italy
11. The Department of Rheumatology, Leiden University Medical Center, Leiden, the Netherlands.
12. The University of Manchester, Northern Care Alliance NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, United Kingdom.
13. Department of Rheumatology, Université Paris Cité UFR de Médecine, Paris, France.
14. Centre for Genomics and Oncological Research, Pfizer, University of Granada/Andalusian Regional Government, Granada, Spain
15. Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy
16. Department of Internal Medicine, Division of Rheumatology, The University of Texas Health

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/art.43433

Science Center at Houston, Houston, TX, USA
17. Centre for Rheumatology, Royal Free Hospital, UCL, London, UK
✱ These authors share senior authorship
C Corresponding author

Correspondence to:
Martin Kerick, PhD
Institute of Parasitology and Biomedicine López-Neyra, (IPBLN)
Consejo Superior de Investigaciones Científicas (CSIC)
Parque Tecnológico de Ciencias de la Salud, Avenida del Conocimiento, nº 17,
18016, Armilla, Granada, Spain.
e-mail: mkerick@ipb.csic.es

Funding

This work was supported by Instituto de Salud Carlos III (ISCIII) (RD21/0002/0039, RD24/0007/0016, PI22/00092) and by the Spanish Ministry of Science and Innovation (PID2022-13929208-I00). It was also funded by the Innovative Medicines Initiative Joint Undertaking (grant 115565; PRECISESADS project), resources of which are composed of financial contribution from the European Union's Seventh Framework Program (FP7/2007–2013) and European Federation of Pharmaceutical Industries and Associations companies' in-kind contribution. JML was supported by a predoctoral contract in health related research (PFIS) (FI23/00231) from ISCIII. CRP was funded by Grant PREP2022-000747 funded by MICIU/AEI/10.13039/501100011033 and by ESF+. IRM was supported by the programme FPU (FPU21/02746) funded by the Spanish Ministry of University. This research was supported by the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Centre (NIHR203308).

Conflict of interest

The authors declare no conflicts of interest.

Abstract

Objective

Complement component 4 (C4), encoded by *C4A* and *C4B* within the major histocompatibility complex (MHC) on chromosome 6, regulates the immune response and clears immune complexes. Variable copy number (CN) of C4 genes and retroviral HERV-K element influence its function. Given the relationship of C4 CN with SSc risk, we assessed associations with SSc clinical and serological subtypes.

Methods

We compared imputed *C4* CNs across SSc subgroups (4,049 ACA^+; 2,200 ATA^+; 577 ARA^+; 1,078 triple negative patients (TN); 6,295 limited cutaneous (lcSSc); 2,946 diffuse cutaneous (dcSSc)) and 17,991 controls. We evaluated associations with SSc subtypes, identifying C4-independent HLA alleles.

Results

Lower *C4* CN and higher HERV-K CN were associated with increased risk in all SSc subgroups. ATA^+ patients showed the strongest association, particularly with *C4A* (OR=1.88) and differences in *C4A CN* association were more pronounced between autoantibody subgroups (ATA^+ vs ACA^+, $p = 4x10^{-11}$) than between clinical subgroups (dcSSc vs lcSSc, $p = 1x10^{-4}$). In ACA^+ patients, only low *C4B* CN showed a significant association to SSc risk ($p=1.23x10^{-5}$). We also observed sex-biased associations: dcSSc, ATA^+ and ARA^+ males showed stronger effects for *C4A* and ACA^+ and lcSSc females for *C4B*. Finally, our results suggest that the HLA alleles associated with SSc subgroups are independent of C4 CN.

Conclusion

This study highlights distinct genetic contributions of *C4A* and *C4B* in SSc subtypes susceptibility. Our findings suggest that lower *C4* CNs, particularly *C4A*, increase the risk of the severe dcSSc subtype, potentially through a mechanism involving immune complex clearance.

Introduction

Systemic Sclerosis (SSc) is an heterogeneous immune-mediated inflammatory disease (IMID) that affects the connective tissue, occurring more frequently in females than males[1]. It is mainly characterized by the appearance of fibrosis in the skin and internal organs[2]. In fact, the extent of fibrosis categorizes patients into limited cutaneous SSc (lcSSc) and diffuse cutaneous SSc (dcSSc), with dcSSc patients having a poorer prognosis[2]. Interestingly, SSc clinical subtypes correlate with differences in serological profiles, where lcSSc patients most frequently carry anti centromere (ACA^+) autoantibodies and dcSSc patients likely to carry anti topoisomerase I (ATA^+) or anti RNA polymerase (ARA^+)[3]. Detection of these autoantibodies emerged as a promising diagnostic tool and a novel classification criterion, as up to 95% of SSc patients produce autoantibodies with detectable levels in 50% of patients before the clinical onset of the disease[4].

SSc etiology is complex and genetics play a relevant role in its development[5]. Thus, genome-wide association studies (GWAS) have identified multiple genetic risk loci in SSc and distinct genomic associations in its main clinical and serological subtypes, including specific HLA alleles[6,7]. Despite recent advances, GWAS studies explain only a fraction of SSc heritability, with another portion explained by structural variations, such as copy number (CN) polymorphisms, which are not captured by GWAS[8].

Complement component 4 (*C4*) gene encodes a protein that plays a key role in the classical and lectin complement pathways[9]**.** Partial or complete deficiency in C4 has been related to autoimmunity, contributing to its pathogenesis through multiple mechanisms such as defective immune complex and apoptotic cell debris clearance[10]. The genetics of *C4* gene is complex, as it is encoded by *C4A* and *C4B*, two polymorphic genes located within the major histocompatibility complex (MHC) class III region on chromosome 6. Each individual carries multiple copies of *C4*, which may be *C4A* or *C4B*, and each copy might be in both long and short forms depending on the presence of the 6.4kb human endogenous retrovirus K (HERV-K) element in the ninth intron[11]. Fewer copies and a greater proportion of the long forms with HERV-K leads to reduced gene expression and protein levels, resembling *C4* deficiency[12].

Recent studies revealed a genetic association between *C4* CNs and the main serological and clinical subtypes of IMIDs such as systemic lupus erythematosus (SLE), Sjogrens' syndrome (SjS) and idiopathic inflammatory myopathies (IIM)[13,14]. Our group

showed that lower CNs of *C4* are associated with an increased risk of developing SSc[13]. However, the relationship between C4 CNs and SSc clinical subtypes or autoantibodies remains unexplored. Therefore, our study aims to elucidate the role of *C4* CN variation across SSc clinical stratifications and serotypes and assess the independence of HLA alleles from *C4* CNs, given its location within the MHC region.

Material and methods

Patients and study cohort

All patients included in the study met the classification criteria of the 2013 American College of Rheumatology (ACR), the European League Against Rheumatism, or the criteria proposed by LeRoy and Medsger for early SSc[15,16]. This study received approval from the CSIC's Ethics Committee, and written informed consent was collected from all patients in accordance with the tenets of the Declaration of Helsinki.

A European ancestry cohort of 10,657 SSc patients, stratified by ACA^+, ATA^+, ARA^+, triple negative patients (TN), lcSSc and dcSSc; and 17,991 unaffected individuals, was recruited through an international collaborative effort involving clinicians and centers across 11 countries. These multiple cohorts were analyzed in this study, adding 7.5% new samples to our prior GWAS in SSc[6]. TN patients were defined as those negative for ACA, ARA and ATA autoantibodies. A principal component (PCA) analysis of ~100,000 independent SNPs that passed quality controls was performed using GCTA software, as described in Lopez-Isac E et al[6]. Ancestry outliers were defined as samples with >4 standard deviations from the cluster centroid of each country cohort and were subsequently removed from further analyses. Additionally, duplicated samples or the inclusion of relatives were assessed with identity-by-descent (IBD) estimation using PLINK v1.07. One of the relatives (Pi_Hat>0.45) or duplicates (Pi_Hat>0.99) were also filtered out. Then, after merging the resulting datasets from all country cohorts, the PCA was repeated as previously and the first two principal components are plotted in **Suppl. Fig. 1**.

Imputation

We followed a two step imputation strategy. First, we imputed the SNPs from the C4 reference dataset in each cohort to avoid a cohort specific C4 imputation bias due to the different genotyping arrays used. Second, 7,172 reference SNPs imputed in all cohorts at r2>0.3 were used to impute the C4 alleles. In detail:

*Single nucleotide polymorphisms (*SNPs) on chromosome 6 were extracted from the cohorts and the allelic information from non-genotyped variants was imputed using the TOPMed reference panel with default settings (https://imputation.biodatacatalyst.nhlbi.nih.gov/). To ensure imputation quality, pre-

imputation quality control (QC) filters were applied to the merged dataset: SNPs with call rates < 0.98, minor allele frequencies (MAFs) <0.01, and those that deviated from Hardy–Weinberg equilibrium (HWE; P < 0.001 in both case and control subjects) were filtered out; samples with call rates <0.95 were also removed. After imputation, an additional set of QCs were applied: r^2>0.3, MAF>0.01 and HWE (p<0.001).

C4 haplotypes. TOPMed imputed SNPs across all individuals and overlapping the *C4* CN reference panel were selected. The resulting 7,172 SNPs were used as input for the Imputec4[17] software, which wraps Beagle, that was employed to impute the *C4* haplotypes using the reference panel downloaded from dbGaP with the study accession identifier phs001992.v1.p1. As a measure of overall C4 haplotype imputation quality, we averaged all C4 CN imputation r^2 values, weighted by haplotype allele frequency, yielding an overall score of 0.90.

C4 copy numbers. A *C4* haplotype consists of a defined number of *C4* isoforms (*C4A* and/or *C4B*) and HERV-K. The dosage for each element was obtained by multiplying the allele dosage of the structural haplotype by the count of C4 isoform and HERV-K copies it carried, in accordance with the approach described by Kerick et al. 2022[18]. The reference panel allows the imputation of 29 distinct haplotypes, each carrying up to 3 *C4* copies (0–2 *C4A* and 0–2 *C4B*, 1–3 total *C4*). With two haplotypes inherited per individual, the possible combinations range from 0–4 *C4A*, 0–4 *C4B*, and 2–6 total *C4* copies.

C4 copy number association analysis

A logistic regression was performed to assess the association of *C4*, *C4A*, *C4B* and *HERV-K* dosages with SSc clinical and serological subtypes using glm function in R. We used CNs dosages in all analyses to include some measure of imputation uncertainty. Logistic regression models included sex and five principal components (PCs) as covariates to account for population structure and potential confounders. We applied a nominal p-value cutoff of p<0.05 for C4 CN analyses to be consistent with our previous study[18]. Odds ratios (ORs) were calculated as the exponential of the beta values from the regressions (e^{β}) with signs inverted so that they represent the effect of decreasing CN. In detail we calculated three distinct models with C4, C4A, C4B and HERV-K representing copy number dosages:

I: *disease* ~ PC1-5 + sex + *C4* + *HERV-K*

II: *disease* ~ PC1-5 + sex + *C4A* + *C4B* + *HERV-K*

III: *disease* ~ PC1-5 + sex + *C4A* + *C4B* + *HERV-K* + sex:*C4A* + sex:*C4B* + sex:*HERV-K*

Disease was coded as a binary variable: SSc = 1, controls (Ctrl) = 0. This coding was adapted for each comparison. For example: for ACA^+ vs Controls: ACA^+ = 1, Controls = 0; others excluded. Model II was run separately for each sex. Model III tested whether sex differences were significant.

We assessed the relationship among subtypes by performing Spearman correlation analyses on the ORs from each subtype comparison. Correlations with p-values below 0.05 were considered statistically significant.

HLA alleles imputation and independence analysis

Classical HLA allele dosages were imputed using the Four-digit Multi-ethnic HLA v1 (2021) reference panel on the Michigan imputation server (https://imputationserver.sph.umich.edu/) with default settings and performing the same pre- and post-imputation quality filters as described above[19,20]. We performed conditional association analysis for imputed classical HLA alleles to determine their *C4*-independent association with SSc. This model includes the dosages of *C4A*, *C4B* and HERV-K next to five PCs and sex as covariates. All alleles that were associated with the disease at a GWAS level threshold of $p<5x10^{-8}$ after conditioning analysis were considered independent for that specific subtype.

Results

We conducted a logistic regression analysis for *C4*, HERV-K, *C4A* and *C4B* CNs in serological (ACA^+, ATA^+, ARA^+) and clinical (lcSSc, dcSSc) subsets of SSc patients. After QCs, 28,648 individuals were included: 10,657 SSc patients (classified into 4,049 ACA^+, 2,200 ATA^+, 577 ARA^+, 6,295 lcSSc and 2,946 dcSSc) and 17,991 unaffected individuals with their corresponding demographic characteristics summarised in **Suppl. Table 1**. Individuals had 0-4 copies of C4A and C4B and 0-6 copies of HERV-K with similar distributions in all subgroups (**Suppl. Table 2**).

Association of C4 copy numbers in SSc serological stratifications

Our results showed that low *C4* CN was significantly associated with increased risk in all serological groups compared to unaffected individuals, with distinct effect sizes between them (**Fig. 1**). Specifically, lower *C4* CNs significantly increases the risk in ATA^+ patients compared to ACA^+ patients (p=0.017), suggesting differences in *C4* CNs impact between subtypes **(Fig. 1)**. As expected, lower HERV-K CN, whose presence constitutes the long form of the *C4* gene and reduces the synthesis of *C4*, presented a protective effect for SSc serotypes, contrary to *C4* **(Fig. 1)**. Again, this effect was more pronounced in ATA^+ patients than in ACA^+, and remained significant when comparing both stratifications (p=0.05) (**Fig. 1**).

Given the biological differences between *C4A* and *C4B* in their affinity for distinct ligands, we assessed their specific contributions to SSc serological stratifications. Our analysis revealed that lower *C4A* CNs provided a significant risk effect exclusively in ATA^+ patients (OR=1.88, $p=7.12x10^{-26}$), as no significant association was observed in ACA^+ and ARA^+ patients (**Fig. 1**). The comparison ATA^+ versus ACA^+ and ARA^+ subgroups was highly significant ($p=9.06x10^{-18}$, $p=2.88x10^{-9}$), emphasizing *C4A* impact in ATA^+ patients (**Fig. 1**). Conversely, *C4B* showed a significant association in ACA^+ patients (OR=1.20, $p=1.61x10^{-6}$) and ARA^+ patients (OR=1.26, p=0.012) (**Fig. 1**), but the comparison between ATA^+ and ACA^+ groups did not reach statistical significance (**Fig. 1**). On the contrary, *C4B* showed a significant association in ARA^+ patients as opposed to ATA^+ (**Fig. 1**). Our results report thus a biased effect of *C4A* towards ATA^+ patients and of *C4B* towards ACA^+ and ARA^+ patients.

To support our previous analysis, we compared auto-antibody positive versus TN, albeit with reduced statistical power (**Suppl. table 3**). Again, we found low *C4A* CN to be

strongly associated with the ATA$^+$ subgroup (OR=1.32, p=6.96x10^{-3}). When comparing ACA+ patients with TN patients, we observed a C4A association similar to that seen in the ATA+ versus ACA+ comparison (Suppl. Table 3). Overall, the pattern showed that the strongest association with C4A CN is in ATA+ patients, followed by ARA+, then TN, while ACA+ patients are similar to healthy controls.

Associations of C4 copy numbers in SSc clinical subtypes

We found that lower *C4* CNs were significantly associated with increased risk for both dcSSc and lcSSc, with a stronger effect observed in dcSSc. Indeed, this difference in effect between the two clinical subtypes was statistically significant (p=0.03). Again, HERV-K CN showed effects opposite to *C4*, with significant associations in both dcSSc and lcSSc **(Fig. 2)**, although less pronounced than in the serological subtype analysis. Overall, these results exhibit an analogous trend for both SSc clinical and serological stratifications.

Analysis of *C4* genes across clinical subtypes revealed that dcSSc patients exhibited a stronger association with *C4A* (OR=1.63, p=2.22x10^{-22}) compared to *C4B* CNs (OR=1.21, p=1.48x10^{-5}) **(Fig. 2)**, whereas lcSSc patients showed similar associations with both *C4A* and *C4B* CN **(Fig. 2)**. Notably, low *C4A* CNs significantly increased the risk of developing dcSSc compared to lcSSc patients (OR=1.27, p=3.14x10^{-5}), while no such difference was observed for *C4B* (p=0.55). Hence, dcSSc mirrored ATA$^+$ in both *C4A* and *C4B* patterns, while lcSSc patients only align with ACA$^+$ with respect to *C4B* but not *C4A*.

Sex biased effects of C4 copy numbers in SSc stratifications

In a previous study we reported that low *C4A* CN had a stronger association to disease risk in male SSc patients while low *C4B* CN was found to be of risk in women only[18] and here we extend these results investigating SSc clinical and serological subgroups.

Regarding serological subtypes, we found low *C4A* to have the strongest risk effect in ATA$^+$ and ARA$^+$ males, while in ATA$^+$ females the effect was significantly smaller than in ATA$^+$ males (p=1.06x10^{-4}). In ARA$^+$ females and both ACA$^+$ females and males we did not find any significant effect of *C4A*. In contrast, low *C4B* CN was associated with SSc risk in ACA$^+$ and ARA$^+$ females (**Suppl. Table 3**), with no significant association found in ACA$^+$ and ARA$^+$ males.

CN analysis of clinical subtypes corroborated the association between *C4A* and males, with dcSSc males showing the largest effect (OR = 2.08, p =1.32x10^{-11}) (**Suppl. Table 3**).

Interestingly, low *C4A* CN conferred risk also for lcSSc, again with a more pronounced effect in males (OR = 1.48, p = 8.79×10^{-5}) (**Suppl. Table 3**). Regarding *C4B*, it was associated with risk in females for both lcSSc and dcSSc, showing similar effect sizes (**Fig. 3**). No significant association was found for *C4B* in males for either subtype **(Fig. 3)**. Finally, we investigated if the stratified ORs of low *C4A*/*C4B* /HERV-K CN correlate between serological and clinical subtypes. We found a positive correlation for the odds ratios in dcSSc and ATA$^+$ (R^2: 0.99, p=1.26×10^{-4}), dcSSc and ARA$^+$ (R^2: 0.89, p=0.02), and to a lesser extent in lcSSc and ACA$^+$ (R^2: 0.83, p=0.042) (**Fig. 3**).

Independence of C4 copy numbers from HLA alleles

Recent studies in SLE, Sjs and SSc have shown that C4 genetics contribute to disease etiology independent of HLA classical alleles[18,21]. We analysed 326 HLA alleles and conducted conditional analyses for each clinical and serological subtype, revealing a total of 44 *HLA* alleles independent from *C4* CNs in at least one SSc subtype (**Suppl. Table 4**). Of these, 32 were *HLA* class II alleles (9 *HLA-DQB1*, 8 *HLA-DRB1*, 7 *HLA-DQA1*, 6 *HLA-DPB1*, 2 *HLA-DPA1*), and 12 belong to *HLA* class I alleles (7 to *HLA-B*, 3 to *HLA-C* and 2 to *HLA-A*) (**Suppl. Table 4**).

These alleles predominantly showed association to either ATA$^+$ and dcSSc or ACA$^+$ and lcSSc, with consistent directional effects across subtypes (**Suppl. Table 4**). Our conditional analyses additionally confirmed the results of our previous study[7] to be independent of *C4* CNs (highlighted in bold in **Suppl. Table 4**).

Discussion

Multiple pathogenic pathways involving complement deficiency have been linked to autoimmunity, including impaired clearance of autoantigens and autoantibodies, along with disrupted cytokine and interferon regulation[22,23]. Interestingly, functional differences between *C4* genes further relate with disease risk, where *C4A* has a higher affinity for immune complexes than *C4B*, which in turn shows higher affinity to carbohydrate antigens[12]. Our study reports that lower CNs of *C4* are associated with increased risk of different clinical and serological subtypes of SSc. Specifically, *C4A* CNs show stronger association with ATA$^+$ serology (typically linked to dcSSc), while *C4B* CN shows stronger association with ACA$^+$ serology (associated with lcSSc). These results extend our prior findings, which identified *C4A* and *C4B* CN as contributors to overall SSc risk[18].

It is well described that serological subtypes of SSc have distinct molecular profiles, although the exact pathogenic mechanisms and their role in the different forms of the disease are not fully understood[24]. Indeed, ATA$^+$ patients showed increased differentiation of B-cells and autoantibody production, which accelerate disease progression[25]. In this line, we here report differential associations between *C4A* CNs and SSc autoantibodies, where reduced CN of *C4A* increased disease risk in ATA$^+$ patients. It has been suggested that *C4A* deficiency leads to the accumulation of immune complexes[22], exacerbating the pathogenic process of the disease by promoting vascular damage[26]. Indeed, comparing ATA$^+$ patients with TN patients revealed that lower *C4A* CN is of risk in ATA$^+$, which is in line with this thesis. Of interest, ATA$^+$ and ARA$^+$ male patients, both associated with *C4A*, represent the SSc subgroups with worse progressions as they are prone to develop dcSSc, the most severe clinical form of the disease.

Furthermore, we highlight the association of low *C4B* CN with ACA$^+$ SSc patients but not with ATA$^+$ patients, however this was not detected when comparing lcSSc with dcSSc, where low *C4B* CN was of risk in both groups. A study on rheumatoid arthritis suggested *C4B* to play a role in the clearance of citrullinated antigens contributing to the maintenance of immune tolerance[27], which advocates an immune complex mediated pathogenesis, similar to *C4A*. Indeed, *C4A* does not appear to play a role with ACA$^+$ patients, as no significant association was observed compared to controls, and ACA$^+$ was significantly different from ATA$^+$ and TN subgroups. These findings support the idea that SSc serological subtypes may involve distinct disease processes. Accordingly, previous studies have identified distinct gene

expression profiles in SSc, with lcSSc patients, who frequently present ACA autoantibodies, being more likely to exhibit a proinflammatory signature[28].

As in our previous study we found a sex biased association to the subgroups of SSc, reporting a trend relating *C4B* CN with lcSSc and ACA$^+$ female patients, reinforcing the rationale that the *C4*-SSc relationship is distinct for both sexes[18]. Indeed, lower *C4A* CNs significantly increases risk in ATA$^+$, lcSSc and dcSSc male patients, with ATA$^+$ patients exhibiting the highest effect size reported in the study. Sex differences in *C4* have been previously described in SLE, SjS and SSc with males showing greater genetic risk effects and higher C4 concentration in serum[18,21]. Interestingly, this tendency is also observed in male-predominant diseases, such as schizophrenia[21]. The exact mechanism is not completely understood, but there is molecular evidence that supports these sex differences. For instance, IFN gamma, a sex biased pathway relevant in SSc, enhances the stability of the mRNA encoding C4[29,30]. Therefore, our study further supports that C4 may contribute to the differences between sexes in SSc[31].

C4 is located within the MHC, the region with the strongest association to SSc[6]. However, the interplay between *C4* CNs and HLA remains enigmatic. In a previous study we reported that C4 genetics together with classical alleles of *HLA-DRB1* and *HLA-DPB1* can explain most associations with SSc in the MHC region and here we report 44 classical HLA alleles with *C4* independent associations for at least one SSc subgroup analyzed. Additionally, we establish that all HLA alleles previously associated with SSc subtypes are independent of *C4* CNs[18], reinforcing the potential implication of these alleles as independent contributors to the disease in the different stratifications. Overall, these results point towards a combined contribution of both the *HLA* and *C4* to the development of SSc subgroups, although confirmation in multi-ancestry cohorts is still required due to linkage-disequilibrium differences across populations.

Our study has certain limitations. First, the exact pathogenic process of *C4*, *C4A* and *C4B* CNs in SSc subtypes remains uncertain, in spite of the significant contribution to SSc risk. Second, given the low prevalence of males in the disease, the statistical power of sex-stratified analyses is limited, which emphasizes the significant results presented. Third, we compared ACA$^+$ or ATA$^+$ against TN patients, which, although underpowered, allowed us to identify specific significant associations with ATA$^+$. Nevertheless, our analysis might still be slightly confounded by an enrichment of an unknown antibody which is not ACA, ATA or ARA.

Finally, although the *C4* CN imputation quality was high, the reference panel does not capture haplotypes with more than four *C4A* or *C4B* copies. Although these haplotypes are extremely rare with reported frequencies ranging from 0.005 to <0.0002[13,32], their absence could slightly affect the effect size estimates.

Despite these constraints, it is worth mentioning that gaining a deeper understanding of the complement system in SSc could provide valuable clinical insights. For instance, hypocomplementemia is a well-established biomarker in related diseases, such as SjS[33]. Additionally, complement-targeted therapies, such as eculizumab, which targets C5, have demonstrated efficacy in IIM and in different cases of SSc-associated renal crisis[34–36]. Our findings support the potential of complement-modulating therapies for this SSc complication, as we observed a significant association between *C4* CN and ARA$^+$ patients, who are frequently associated with SSc-related renal crisis[3]. Given that low C4 CN is associated with increased SSc risk, particularly in dcSSc and ATA$^+$ patients, it could be hypothesized that a therapeutic strategy aiming to reduce C4 consumption driven by autoantibodies, or restoring C4 availability, might be beneficial for the disease. However, considering the incomplete understanding of the molecular involvement of C4 in the disease, functional assays and further longitudinal studies of this pathway are required to explore its potential in the clinical management of the disease.

In conclusion, this study highlights the distinct contribution of *C4* CN to the susceptibility of SSc subgroups. We emphasize the distinct autoantibody profiles, which are stable across the clinical trajectory of SSc patients[37], showed greater differences than lcSSc and dcSSc. *C4A* CN were significantly associated with ATA$^+$ and male ARA$^+$ patients, while *C4B* CN were significantly associated with ACA$^+$, suggesting different roles for each C4 isotype across SSc subgroups, which reinforces the idea of distinct endophenotypes in the disease represented by SSc serotypes. Ultimately, our findings reveal that lower *C4A* CNs increases the risk of dcSSc, the more severe form of SSc, suggesting that it may prevent disease progression by promoting immune complex clearance in ATA$^+$ patients.

Data availability

Genotype data are unavailable due to privacy and consent concerns. All other data in the article, supplementary materials, are available upon reasonable request to the authors.

Authors' contributions

Conceptualization: JM, MAH, MK. Formal analysis: JML, CRP, IRM, MK. Funding acquisition: MEAR, JM, MAH. Resources & Data Curation: AGDC, CPSA, JLC, OD, SMP, MN, NH, GM, JKVB, ALH, YA, MEAR, LB, SA, CPD, MDM. Interpretation of the data: JML, JM, MAH, MK. Writing – original draft: JML, JM, MAH, MK. Writing – review & editing: all authors reviewed and edited the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Sofia Vargas and Gema Robledo for their excellent technical assistance. We also want to thank all the patients and control donors for their essential collaboration. We also thank all of the participants who provided the data for their essential collaboration and all members of the PRECISESADS consortium. This research was supported by the National Institute for Health and Care Research (NIHR) Manchester Biomedical Research Centre (NIHR203308). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. This research is part of the doctoral degree awarded to JML within the biomedicine program from the University of Granada.

References

[1] Denton CP, Khana D. Systemic sclerosis. The Lancet 2017;390:1685–99.

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

[3] Mahler M, Hudson M, Bentow C, et al. Autoantibodies to stratify systemic sclerosis patients into clinically actionable subsets. Autoimmun Rev 2020;19:102583.

[4] Cavazzana I, Vojinovic T, Airo' P, et al. Systemic Sclerosis-Specific Antibodies: Novel and Classical Biomarkers. Clin Rev Allergy Immunol 2023;64:412–30.

[5] 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.

[6] 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.

[7] Acosta-Herrera M, Kerick M, Lopéz-Isac E, et al. Comprehensive analysis of the major histocompatibility complex in systemic sclerosis identifies differential HLA associations by clinical and serological subtypes. Ann Rheum Dis 2021;80:1040–7.

[8] Harris L, McDonagh EM, Zhang X, et al. Genome-wide association testing beyond SNPs. Nat Rev Genet 2025;26:156–70.

[9] Mastellos DC, Hajishengallis G, Lambris JD. A guide to complement biology, pathology and therapeutic opportunity. Nat Rev Immunol 2024;24:118–41.

[10] Coss SL, Zhou D, Chua GT, et al. The complement system and human autoimmune diseases. J Autoimmun 2023;137:102979.

[11] Saxena K, Kitzmiller KJ, Wu YL, et al. Great genotypic and phenotypic diversities associated with copy-number variations of complement C4 and RP-C4-CYP21-TNX (RCCX) modules: a comparison of Asian-Indian and European American populations. Mol Immunol 2009;46:1289–303.

[12] Yang Y, Chung EK, Zhou B, et al. Diversity in intrinsic strengths of the human complement system: serum C4 protein concentrations correlate with C4 gene size and polygenic variations, hemolytic activities, and body mass index. J Immunol 2003;171:2734–45.

[13] Lundtoft C, Pucholt P, Martin M, et al. Complement C4 Copy Number Variation is Linked to SSA/Ro and SSB/La Autoantibodies in Systemic Inflammatory Autoimmune Diseases. Arthritis Rheumatol 2022;74:1440–50.

[14] Zhou D, King EH, Rothwell S, et al. Low copy numbers of complement and deficiency are risk factors for myositis, its subgroups and autoantibodies. Ann Rheum Dis 2023;82:235–45.

[15] 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. Arthritis Rheum 2013;65:2737–47.

[16] LeRoy EC, Medsger TA Jr. Criteria for the classification of early systemic sclerosis. J Rheumatol 2001;28:1573–6.

[17] Sekar A, Bialas AR, de Rivera H, et al. Schizophrenia risk from complex variation of complement component 4. Nature 2016;530:177–83.

[18] Kerick M, Acosta-Herrera M, Simeón-Aznar CP, et al. Complement component C4 structural variation and quantitative traits contribute to sex-biased vulnerability in systemic sclerosis. NPJ Genom Med 2022;7:57.

[19] Das S, Forer L, Schönherr S, Sidore C, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–7.

[20] Luo Y, Kanai M, Choi W, et al. A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. Nat Genet 2021;53:1504–16.

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

[22] Sturfelt G, Truedsson L. Complement in the immunopathogenesis of rheumatic disease. Nat Rev Rheumatol 2012;8:458–68.

[23] Dijkstra DJ, Joeloemsingh JV, Bajema IM, et al. Complement activation and regulation in rheumatic disease. Seminars in Immunology 2019;45:101339.

[24] Rouvière B, Le Dantec C, Bettacchioli E, et al. Stratification according to autoantibody status in systemic sclerosis reveals distinct molecular signatures. Ann Rheum Dis 2024; doi: 10.1136/ard-2024-225925.

[25] Jia G, Ramalingam TR, Heiden JV, et al. An interleukin 6 responsive plasma cell signature is associated with disease progression in systemic sclerosis interstitial lung disease. iScience 2023;26:108133.

[26] Raschi E, Privitera D, Bodio C, et al. Scleroderma-specific autoantibodies embedded in immune complexes mediate endothelial damage: an early event in the pathogenesis of systemic sclerosis. Arthritis Res Ther 2020;22:265.

[27] Rigby WFC, Wu YL, Zan M, et al. Increased frequency of complement C4B deficiency in rheumatoid arthritis. Arthritis Rheum 2012;64:1338–44.

[28] Franks JM, Martyanov V, Cai G, et al. A Machine Learning Classifier for Assigning Individual Patients With Systemic Sclerosis to Intrinsic Molecular Subsets. Arthritis Rheumatol 2019;71:1701–10.

[29] Fernandez-Ruiz R, Niewold TB. Type I Interferons in Autoimmunity. J Invest Dermatol 2022;142:793–803.

[30] Forsyth KS, Jiwrajka N, Lovell CD, et al. The conneXion between sex and immune

responses. Nat Rev Immunol 2024;24:487–502.

[31] Hughes M, Pauling JD, Armstrong-James L, et al. Gender-related differences in systemic sclerosis. Autoimmun Rev 2020;19:102494.

[32] Yang Y, Chung EK, Wu YL, et al. Gene Copy-Number Variation and Associated Polymorphisms of Complement Component C4 in Human Systemic Lupus Erythematosus (SLE): Low Copy Number Is a Risk Factor for and High Copy Number Is a Protective Factor against SLE Susceptibility in European Americans. The American Journal of Human Genetics 2007;80:1037–54.

[33] Mavragani CP, Moutsopoulos HM. Sjögren's syndrome. Annu Rev Pathol 2014;9:273–85.

[34] Cole A, Ong VH, Denton CP. Renal Disease and Systemic Sclerosis: an Update on Scleroderma Renal Crisis. Clin Rev Allergy Immunol 2023;64:378–91.

[35] Gouin A, Ribes D, Colombat M, et al. Role of C5 inhibition in Idiopathic Inflammatory Myopathies and Scleroderma Renal Crisis-Induced Thrombotic Microangiopathies. Kidney Int Rep 2021;6:1015–21.

[36] Toker Dincer Z, Dincer MT, Sen G, et al. Targeting complement dysregulation: eculizumab in scleroderma renal crisis management—a case-based review. Rheumatology International 2024;44:3135–40.

[37] Burcu A, Michele D, Daniel EF, et al. Characterising the autoantibody repertoire in systemic sclerosis following myeloablative haematopoietic stem cell transplantation. Annals of the Rheumatic Diseases 2023;82:670–80.

Figures legend

Figure 1. Forest plots illustrating the association between *C4* gene copy numbers and systemic sclerosis (SSc) serological subtypes. Each panel represents a specific comparison indicated in the right part and displays the odds ratios (ORs) and corresponding 95% confidence intervals (CIs) for the effects of total *C4*, the HERV-K insertion (long form of the gene), and the two C4 genes (*C4A* and *C4B*). The vertical dashed line indicates the null value (OR = 1), representing no effect. Colors differentiate the *C4* genes, and asterisks denote levels of statistical significance (*, p-value<0.05; **, p-value<0.01; ***, p-value<0.001). ATA, anti topoisomerase patients; ACA, anticentromere patients; ARA, anti RNA polymerase patients.

Figure 2. Forest plots illustrating the association between C4 gene copy numbers and systemic sclerosis (SSc) clinical subtypes. Each panel corresponds to one of the clinical comparisons as labeled to the right. Points show odds ratios (ORs) and horizontal bars depict 95% confidence intervals (CIs) for total *C4* CN, the HERV-K long form, and the two C4 isotypes (*C4A* and *C4B*). The vertical dashed line marks the null effect (OR=1). Colors distinguish each *C4* isotypes, whereas asterisks indicate significance levels (*, p-value<0.05; **, p-value<0.01; ***, p-value<0.001). lcSSc, limited cutaneous SSc patients; dcSSc, diffuse cutaneous SSc patients.

Figure 3. Scatter plot comparing odds ratios (ORs) between Systemic Sclerosis (SSc) serological subtypes and their corresponding clinical subtypes in sex-stratified analyses. Each point represents a sex-specific comparison, with color indicating whether the data correspond to female (purple) or male (green) patients. The vertical and horizontal solid lines represent the null effect (OR=1), while the diagonal dashed line represents equal effects in serological and clinical ORs. Each point is mapped to highlight the relative effect of a genetic factor in serological versus clinical subtypes, enabling visual comparison of its impact across subtype categories when stratified by sex. ATA, anti topoisomerase patients; ACA, anticentromere patients; ARA, anti RNA polymerase patients; lcSSc, limited cutaneous SSc patients; dcSSc, diffuse cutaneous SSc patients.

Tables

Table 1. *C4*, *C4A*, *C4B*, and HERV-K copy number (CN) association analysis of SSc serological and clinical subgroups versus unaffected individuals, without sex stratification.

SSc subgroup	C4 Variable	p-value	OR [CI95]	SE
ACA	*C4*	4.53×10^{-04}	1.13 [1.05 - 1.21]	0.03
	HERV-K	3.88×10^{-08}	0.86 [0.82 - 0.91]	0.03
	C4A	8.81×10^{-01}	1.01 [0.89 - 1.15]	0.07
	C4B	1.61×10^{-06}	1.20 [1.11 - 1.29]	0.04
ATA	*C4*	2.58×10^{-08}	1.30 [1.18 - 1.42]	0.05
	HERV-K	2.55×10^{-37}	0.64 [0.60 - 0.68]	0.03
	C4A	7.12×10^{-26}	1.88 [1.67 - 2.11]	0.06
	C4B	1.94×10^{-01}	1.07 [0.96 - 1.19]	0.05
ARA	*C4*	1.67×10^{-02}	1.22 [1.04 - 1.44]	0.08
	HERV-K	2.79×10^{-01}	0.93 [0.82 - 1.06]	0.06
	C4A	1.30×10^{-01}	1.17 [0.95 - 1.44]	0.11
	C4B	1.16×10^{-02}	1.26 [1.05 - 1.50]	0.09
lcSSc	*C4*	1.22×10^{-10}	1.21 [1.14 - 1.28]	0.03
	HERV-K	1.65×10^{-07}	0.89 [0.85 - 0.93]	0.02
	C4A	6.93×10^{-09}	1.25 [1.16 - 1.34]	0.04
	C4B	1.50×10^{-08}	1.20 [1.12 - 1.27]	0.03
dcSSc	*C4*	6.41×10^{-13}	1.34 [1.23 - 1.45]	0.04
	HERV-K	1.20×10^{-14}	0.79 [0.74 - 0.84]	0.03
	C4A	2.22×10^{-22}	1.63 [1.48 - 1.80]	0.05
	C4B	1.48×10^{-05}	1.21 [1.11 - 1.32]	0.04

OR, Odds ratio; CI95, 95% confidence interval; SE, standard error; SSc, Systemic Sclerosis; ACA, anti centromere positive; ATA, anti topoisomerase I; lcSSc, limited cutaneous SSc; dcSSc, diffuse cutaneous SSc.

Table 2. Association results of HLA alleles with SSc subtypes before and after conditioning on C4 copy numbers (CN), restricted to alleles previously reported in Acosta-Herrera M, et al; 2021; Ann Rheum Dis.

		***C4* CN conditioned analysis**			**Acosta-Herrera M, et al.; 2021; Ann Rheum Dis**		
HLA Allele	**MAF**	**p-value**	**OR**	**SSc subtypes**	**p-value**	**OR**	**SSc subtypes**
[illegible]08:01	0.12	$1{,}78 \times 10^{-12}$	1.36	**lcSSc**, dcSSc	1.29×10^{-10}	1.22	SSc
DPA1*02:01	0.15	$3{,}55 \times 10^{-65}$	2.08	**ATA**, dcSSc	7.91×10^{-43}	1.87	ATA
[illegible]31*13:01	0.02	$3{,}72 \times 10^{-173}$	7.77	**ATA**, lcSSc, dcSSc	6.10×10^{-32}	2.05	SSc
DQA1*05:01	0.26	$3{,}24 \times 10^{-30}$	1.57	**ATA**, dcSSc	1.16×10^{-30}	1.49	dcSSc
DQB1*03:01	0.2	$1{,}15 \times 10^{-31}$	1.60	**ATA**, dcSSc	7.11×10^{-47}	1.86	ATA
DRB1*08:01	0.02	$2{,}57 \times 10^{-55}$	3.00	**ACA**, lcSSc	9.73×10^{-57}	3.18	ACA
[illegible]31*11:04	0.03	$3{,}16 \times 10^{-105}$	4.14	**ATA**, lcSSc, dcSSc	2.52×10^{-56}	2.11	SSc

p-values and odds ratios (OR) correspond to the most significant subtype, highlighted in bold. HLA, human leukocyte antigen; MAF, minor allele frequency; SSc, Systemic Sclerosis; ACA, anti centromere positive; ATA, anti topoisomerase I; lcSSc, limited cutaneous SSc; dcSSc, diffuse cutaneous SSc.

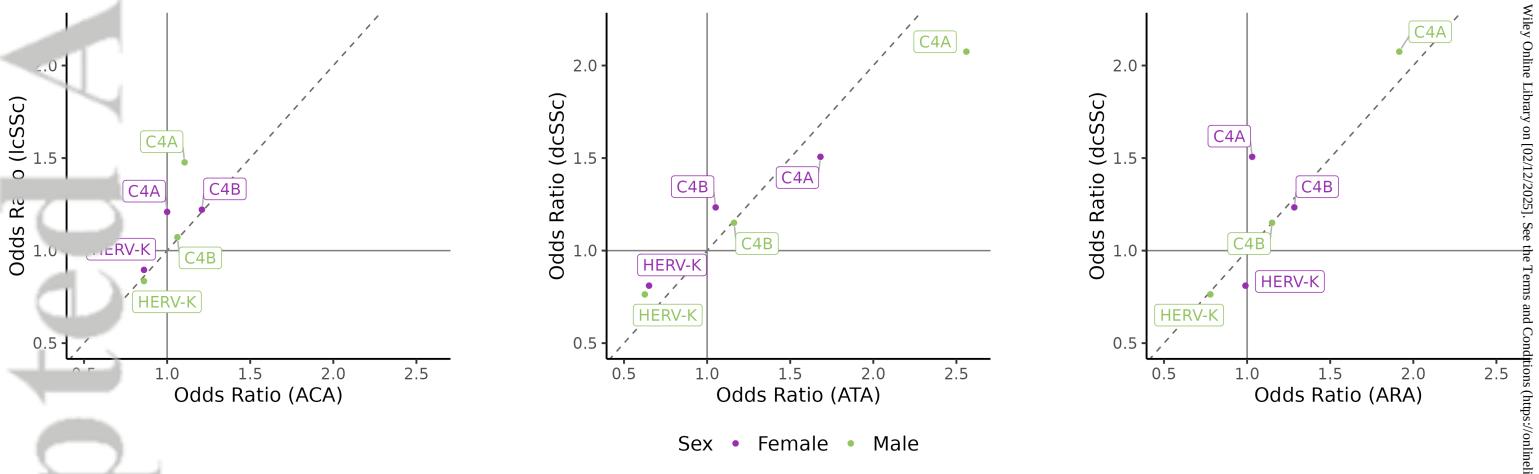
Supporting Information

Supplementary Table 1. Study cohort stratified by countries and disease subtypes.
Supplementary Table 2. Mean copy number of *C4* and its isotypes in systemic sclerosis subgroups.
Supplementary Table 3. *C4*, HERV-K, *C4A* and *C4B* associations among all comparisons.
Supplementary Table 4. HLA alleles independent from *C4* in one systemic sclerosis subgroup at least and their associations with systemic sclerosis subgroups after conditioning for *C4* copy numbers.

Supplementary Table S1

ART_43433_f1_F1.tiff

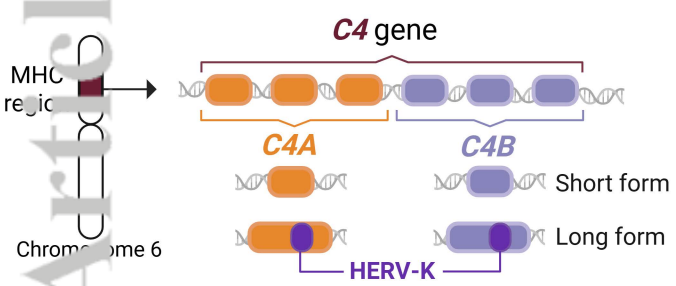
ART_43433_f2_F2.tiff



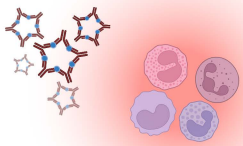
ART_43433_f3_F3.tiff

Distinct effects of *C4* copy numbers in SSc subtypes

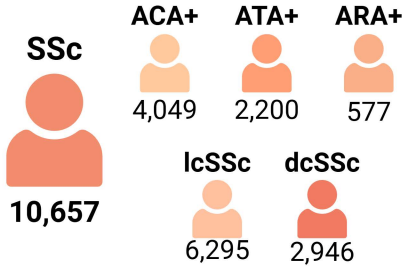
Association analysis of *C4* copy number (CN) variation with SSc subtypes



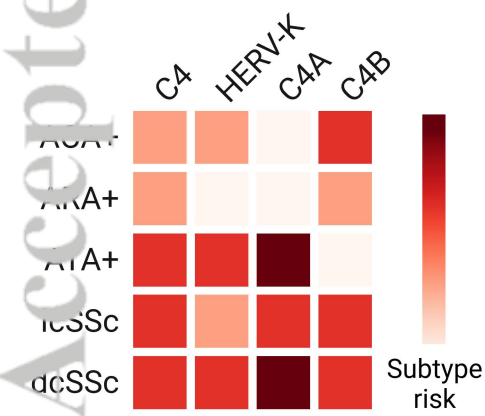
C4 is involved in immune complex clearance and immune regulation



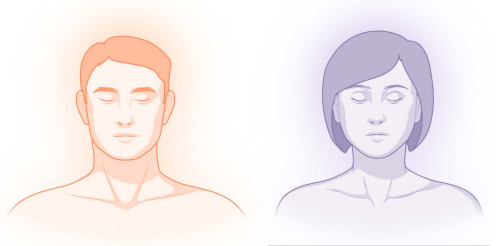
Study cohort



Distinct *C4A* and *C4B* CNs impact SSc subtypes



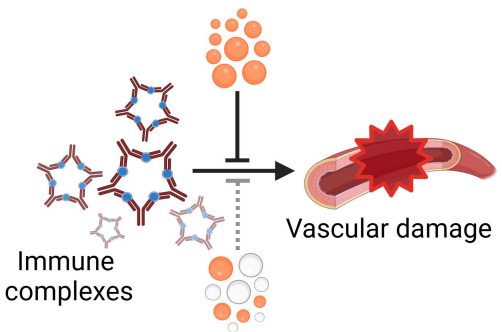
Stronger effect of *C4A* in dcSSc, ATA+, and ARA+ males; and *C4B* in ACA+ and lcSSc females



ATA+
ARA+
dcSSc

ACA+
lcSSc

Lower *C4A* CN increase risk of severe SSc subtypes potentially due to impaired immune complex clearance



Low C4A synthesis

Martinez-Lopez J, Rangel-Pelaez C, Rodriguez-Martin I, et al. Distinct effects of complement *C4A* and *C4B* copy number in systemic sclerosis serological and clinical subtypes. Arthritis Rheumatol 2025.

Arthritis & Rheumatology
ACR American College of Rheumatology