Tesis Doctoral

Deciphering the genetic basis of systemic sclerosis

Memoria presentada por el Diplomado en Enfermería y Graduado en Bioquímica Gonzalo Villanueva Martín para optar al grado de Doctor Internacional por la Universidad de Granada.

Programa de Doctorado en Biomedicina.

Directores:

Javier Martín Ibáñez, Profesor de Investigación del CSIC.

Lara María Bossini Castillo, Profesora Contratada Doctora Indefinida de la Universidad de Granada.

CSIC
Consejo Superior de Investigaciones Científicas

UNIVERSIDAD DE GRANADA

Instituto de Parasitología y Biomedicina López-Neyra, CSIC.

Granada, Noviembre de 2023

Editor: Universidad de Granada. Tesis Doctorales
Autor: Gonzalo Villanueva Martín
ISBN: 978-84-1195-184-5
URI: https://hdl.handle.net/10481/89461

El doctorando / *The doctoral candidate*, Gonzalo Villanueva Martín y los directores de la tesis / *and the thesis supervisors*: Javier Martín Ibáñez y Lara María Bossini Castillo,

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la supervisión de los directores y, hasta donde nuestro conocimiento alcanza, se han respetado los derechos de otros autores a ser citados cuando se han utilizado sus resultados o publicaciones.

Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisors and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.

Asimismo, aseguramos que las tres publicaciones presentadas no se han utilizado en la defensa de ninguna otra tesis en España u otro país y que no será utilizada con tal propósito, quedando así reflejado en el apartado de agradecimientos de la publicación.

Lugar y fecha / Place and date:

Granada, 31 de octubre de 2023.

Directores de la tesis/ *Thesis supervisors* Doctorando / Doctoral candidate

Firma / Signed

Firma / Signed

Prof. Javier Martín Ibáñez Gonzalo Villanueva Martín

Dr. Lara María Bossini Castillo

Agradecimientos

No llores porque ya se terminó, sonríe porque sucedió.

Gabriel García Márquez.

Como en la mayoría de las cosas que empiezo, he tenido que hacer una lista para esta sección. Porque cuando se termina una tesis hay que hacer una lista de todos los que te han ayudado, te han apoyado y te han alentado en este camino largo y difícil. Si me dejo a alguien, lo siento, pero cuentas con mi agradecimiento.

La vida en la ciencia es muy complicada, las oportunidades vienen y van. A veces es importante dejar pasar unas sin saber si encontrarás otra más adelante. Por eso quiero agradecer a mi director de tesis, Javier Martín, por darme esa oportunidad en el momento oportuno. Mi agradecimiento sincero por eso, por orientarme y enseñarme en este mundo tan complicado. Pero sobre todo, por acogerme en su laboratorio y ponerme a trabajar a destajo. También quiero agradecer a mi directora de tesis, Lara Bossini Castillo. A veces Gru, a veces el mismo diablo. Pero siempre una ayuda, una palabra amable y una guía infalible en lo que haga falta. Si algún día necesitas un riñón o un hígado, yo te consigo uno. No el mío, pero uno. A los dos gracias.

A todos los pacientes que han dado un pedacito de sí para los estudios. Hay un altruismo infinito en esos gestos, que tanto dolor llevan, pero que tanto bien hacen. A los médicos que nos ayudan a hacer nuestro trabajo, normalmente a costa de tener ellos un trabajo extra. Me gustaría mencionar especialmente a Norberto Ortego, Carmen Pilar Simeón y Alfredo Guillén, con quienes he tenido el privilegio de trabajar más estrechamente.

He tenido la suerte de trabajar en un centro de investigación maravilloso lleno de gente maravillosa, el Instituto de Parasitología y Biomedicina López-Neyra. A la gente que trabaja ahí, tienen todo mi cariño y agradecimiento por todos estos años. Desde el conserje Javi, pasando por Nicolás, hasta Carmen de recursos humanos.

Por supuesto en el López he tenido la suerte de formar parte del maravilloso equipo de los laboratorios 206-214. Agradezco profundamente a Ana Marquez, Marialbert y Lourdes por sus consejos y ayuda cada vez que lo he necesitado. Sin duda trabajar cerca de vosotras ha sido enriquecedor y una suerte. A Martín, por su invaluable ayuda y pasión por la pantalla negra. También a Gema y a Sofi, las dos personas que resuelven cualquier problema y que siempre tienen un rato para escucharme.

A los doctorandos que han tenido la suerte y la desgracia de trabajar codo con codo conmigo durante años. Aguantarme es tarea complicada, pero aguantaros a vosotros… eso ha sido un privilegio que repetiría toda la vida. Bueno, quizás no tanto. Al Club de los Cinco: David, Desi, Elkyn y Elio. David pasó dos años a mi espalda, chocándonos y compartiendo momentos. Al final has sido el más listo de todos. Desi, que me acogió justo cuando llegué y que es una verdadera amiga. Ya mismo voy para allá. Elkyn, gracias por enseñarme lo diferente que puede ser el mundo, por ser un amigo y un apoyo. Tu calidad humana iguala a tu calidad científica. Elio, siempre serás un poco más viejo que yo y lo dejo aquí escrito. Eso sí, eres una persona excepcional con la que he compartido este viaje al completo. Gracias. Te veré al otro lado del charco, titán. Pasamos a la siguiente generación. Y no hay nadie como Javi, que es un pesado, cuyo optimismo me saca totalmente de mí. Pero con quién he disfrutado, me he reído hasta reventar y que cada vez que no está se le echa en falta. Gracias por ser así de insufrible y no cambies nunca. Gonzalo B., una de las personas más brillantes que he conocido, y lo

mejor, es que su corazón es tan grande como su inteligencia. Gracias por todo, y eso incluye los dibujos de las células. Llegarás tan lejos como te lo propongas. A Inma, una fuente de energía y fuerza inagotable, con una determinación que ya quisiera yo para mí. Te has ganado todo mi cariño y estoy convencido de que vas a ser una científica a la altura de lo que tú te propongas. Gracias por guardarme la espalda. A los Carlos os deseo la mejor de las suertes, la tarea que tenéis por delante es dura, pero a vuestra altura.

El siguiente punto en mi lista es la "gente de Alemania", tanto los miembros del SeqCore como el resto de gente. An das gesamte SeqCore-Team richte ich diese Dankesworte auf Deutsch, für den Fall, dass Sven im Urlaub ist. An Bernd Timmermann, der mir zweimal die Türen des SeqCore geöffnet und mich ohne Vorbehalte aufgenommen hat. An Norbert, weil er mir nicht nur die Geheimnisse der Einzelzellsequenzierung beigebracht hat, sondern auch ein Freund und eine Stütze geworden ist. An Sven und Stefan, eine Inspiration für alle, die Bioinformatik lieben. An Sonia, die mir mit meinem Deutsch geholfen und sich um mich gekümmert hat, und außerdem eine ausgezeichnete Tortilla gemacht hat. An Myriam, die mit mir geweint hat, als es nötig war. An Ilona und Daniela, die immer für mich da waren. Vielen Dank an euch alle, ich werde die Herzlichkeit, mit der ihr mich aufgenommen habt und wie ihr mich zu einem Teil von euch gemacht habt, nicht vergessen. Und an Helene, für ihre unschätzbare Hilfe und Ratschläge, sowohl dort als auch danach. And finally, I would like to thank all the people from the Guest House: Rosi, Anita, Sai, Ste. My friends, thank you for supporting me, for the dinners talking about science and everything else.

En Barcelona empezó este camino, de una forma o de otra. Por ello, permitidme que me detenga para agradecer a toda la gente del grupo de Judith Armstrong, donde se encendió mi chispa de la bioinformática con Jan y Guerau. También a Silvia, Ainhoa y Clara, personas maravillosas con las que

compartí momentos increíbles. A la gente de Bellvitge, sois muchos y sé que no os hago justicia con unas pocas líneas. Alba, Adri, María, Gema, Ana Vidal, Laura y la gente de Serveis. Gracias por todo, por seguir ahí aunque esté tan lejos. Cristian, gracias por enseñarme tantísimo de tantísimas cosas. Cualquiera que aprenda de mí a hacer cultivos lleva tu marca, amigo mío. Y me dejo a Pere para el final. Hay familia que se elige y tu eres mi hermano. Te quiero Pere, eres un referente en mi vida y eso no va a cambiar nunca. Gracias a todos.

A mis amigos de la carrera, los Erlenmeyers. Estamos esturreados por el mundo, pero la amistad perdura y aguanta. Os quiero a todos. Elena, Irene y Ana, que eligieron mí mismo camino, espero que os compense tanto como me ha compensado a mí. A los que os fuisteis a la empresa, vosotros pagáis, que tenéis las pelas. A mis amigos del máster en genética. Elia, la mejor repostera que ha dado la ciencia y una amiga que sin duda no me merezco, pero doy gracias a que tengo. Inma, soy capaz de oír tu risa con solo cerrar los ojos. Camille, la persona más dura y decidida que ha dado esta tierra. A Víctor, la persona que siempre quiero a mi lado, especialmente si hay que planificar una broma. Pero para todo lo demás también.

Y como esto ya me está quedando muy largo voy cerrando. Gracias a mi familia, que siempre ha estado a mi lado. Álvaro, mi hermano que no tiene ni idea de lo que hago, pero que le da igual porque está orgulloso de mí. A mis padres, que me han dado todas las oportunidades que les he pedido y que siempre con infinita paciencia me han escuchado y guiado. Gracias, soy lo que soy gracias a vuestro amor. Vosotros tenéis el mío para siempre. Os quiero.

A Anna, mi equipo, mi compañera y mi amor. Gracias por ayudarme en esto y todo lo demás, gracias por llevar la carga conmigo, por soportarme

y ayudarme. Estoy aquí porque tú me has ayudado y levantado cada vez que lo he necesitado. Gracias de corazón. Te quiero.

Gracias a los que estuvieron, pero no están. Gracias abuela, gracias Laura.

Gracias a ti, lector. Espero que esta tesis sea lo que esperas. Y si no, es culpa tuya por esperar algo.

A mis padres,

A Anna,

A aquellos que caminaron con nosotros
y ahora aguardan a que lleguemos

Index

Abbreviations

AD: Autoimmune disorder

ACA: Anticentromere antibody

ARA: Anti-RNA polymerase III antibodies

ATA: Anti-topoisomerase I antibodies

ATAC-seq: Assay for Transposase-Accessible Chromatin using sequencing

AUC: Area under the curve

BBkNN: Batch Balanced k-Nearest Neighbour

BMI: Body-Mass Index

CAD: Coronary artery disease

CD: Crohn's disease

cDNA: complementary deoxyribonucleic acid

CEU: Utah residents (CEPH) with Northern and Western European ancestry

CI: Confidence interval

cMo: Classical Monocyte

CMV: Cytomegalovirus

CRP: C-reactive protein

CTRL: Control

CT-SLEB: Cutaneous T-cell lymphoma associated antigen SLEB

DC: Dendritic Cells

dcSSc: Diffuse cutaneous systemic sclerosis

DE: Differentially expressed

DEG: Differentially expressed genes

DNA: Deoxyribonucleic acid

EBV: Epstein-Barr virus

EC: Endothelial cells

ECM: Extracellular matrix

eQTL: Expression quantitative trait *loci*

FDR: False discovery rate

G-PROB: Genetic Probability tool

GBP: Guanylate binding protein

GBR: British in England and Scotland

GI: Gastrointestinal involvement

GIANT: Genetic Investigation of Anthropometric Traits consortium

GLM: Generalized linear model

GO: Gene Ontology

GRS: Genomic risk score

GWAS: Genome Wide Association Study

Hi-C: High-throughput chromosome conformation capture

HLA: Human leukocyte antigen

HVG: Highly variable genes

ILD: Interstitial lung disease

IMD: Immune mediated disease

IMID: Immune mediated inflammatory disease

iMo: Intermediate monocyte

IFN: Interferon

IV: Instrumental Variable

IVW: Inverse Variance Weighted

KEGG: Kyoto Encyclopedia of Genes and Genomes

lcSSc: Limited cutaneous Systemic Sclerosis

LD: Linkage Disequilibrium

LDSC: Linkage Disequilibrium regression score

MHC: Major Histocompatibility Complex

Mo: Monocyte

MR: Mendelian Randomization

mRNA: Messenger Ribonucleic acid sequencing

MS: Multiple sclerosis

MVMR: Multivariable Mendelian Randomization

ncMo: Non-classical monocyte

NK: Natural Killer

OR: Odds Ratio

PAH: Pulmonary arterial hypertension

PB: Peripheral Blood

PBMC: Peripheral Blood Mononuclear cell

PCA: Principal Component Analysis

PGE_2: Prostaglandin E2

pHESS: Heritability Estimation from Summary Statistics

PRS: Polygenic risk score

QC: Quality Control

R2: squared correlation

RA: Rheumatoid Arthritis

RNA-seq: Ribonucleic acid sequencing

ROC: Receiver operating characteristic

scRNA-seq: Single Cell RNA sequencing

SJS: Sjögren's Syndrome

SLE: Systemic Lupus Erythematosus

SNP: Single nucleotide polymorphism

SSc: Systemic Sclerosis

T1D: Type 1 diabetes

T2D: Type 2 diabetes

TCR: T cell Receptor

Thf: T helper follicular cell

Th1: T helper 1 cell

Th2: T helper 2 cell

Th17: T helper 17 cell

Th22: T helper 22 cell

Treg: T regulatory cell

UC: Ulcerative colitis

UKBB: United Kingdom Biobank

UMAP: Uniform Manifold Approximation and Projection

UMI: Unique molecular identifier

WAT: White adipocyte tissue

WHR: Waist to hip ratio

WHRadjBMI: Waist to hip ratio adjusted for body-mass index

Genes

AHNAK: AHNAK nucleoprotein

ANXA6: Annexin A6

ARHGAP31: Rho GTPase-activating protein 31

C1QA: Complement C1q subcomponent A

C1QB: Complement C1q subcomponent B

C1QC: Complement C1q subcomponent C

CCR3: C-C Motif Chemokine Receptor 3

CD4: T-Cell Surface Glycoprotein CD4

CD14: Monocyte Differentiation Antigen CD14

CD36: Leukocyte Differentiation Antigen CD36

CD52: Human Epididymis-Specific Protein 5

CD207: C-Type Lectin Domain Family 4 Member K

CD247: T-Cell Surface Glycoprotein CD3 Zeta Chain

CEACAM3: Carcinoembryonic Antigen-related Cell Adhesion Molecule 3

CELA2A: Chymotrypsin-like elastase family member 2A

CLEC12A: C-type lectin domain family 12, member A

COTL1: Coactosin-like protein

CSF1R: Colony Stimulating Factor 1 Receptor

CSF3R: Colony Stimulating Factor 3 Receptor

CSK: C-Src kinase

CXCL4: C-X-C motif chemokine 4

CXCR5: C-X-C chemokine receptor type 5

DDX6: DEAD-box helicase 6

DNASE1L3: Deoxyribonuclease 1-like 3

FCER1A: Fc fragment of IgE receptor Ia

FCGR3A: Fc fragment of IgG receptor IIIa

FCN-1: Ficolin-1

GBP1: Guanylate-binding protein 1

GBP2: Guanylate-binding protein 2

GBP3: Guanylate-binding protein 3

GBP4: Guanylate-binding protein 4

GBP5: Guanylate-binding protein 5

GRB2: Growth factor receptor-bound protein 2

GSDMA: Gasdermin A

GSDMB: Gasdermin B

HLA-A: Human Leukocyte Antigen-A

HLA-B08:01: Human Leukocyte Antigen-B08:01

HLA-DQA1: Human Leukocyte Antigen-DQA1

HLA-DQB103:01: Human Leukocyte Antigen-DQB1*03:01

HLA-DRA: Human Leukocyte Antigen-DRA

HLA-DRB1: Human Leukocyte Antigen-DRB1

HLA-DRB5: Human Leukocyte Antigen-DRB5

IFI6: Interferon alpha-inducible protein 6

IFITM2: Interferon-induced transmembrane protein 2

IFITM3: Interferon-induced transmembrane protein 3

IFNGR2: Interferon gamma receptor 2

IL-12: Interleukin-12

IL-17: Interleukin-17

IL-6: Interleukin-6

IL12A: Interleukin-12 subunit alpha

IL12RB1: Interleukin-12 receptor subunit beta 1

IL12RB2: Interleukin-12 receptor subunit beta 2

IL1B: Interleukin-1 beta

IRF1: Interferon regulatory factor 1

IRF5: Interferon regulatory factor 5

IRF7: Interferon regulatory factor 7

IRF8: Interferon regulatory factor 8

ISG15: Interferon-stimulated gene 15

JUN: Jun proto-oncogene

KLF6: Kruppel-like factor 6

LGALS2: L-Galectin-2

LGR5: Leucine-rich repeat-containing G-protein coupled receptor 5

LRT: Lateral Rootless1

LST1: Leukocyte specific transcript 1

LYZ: Lysozyme

MTHFR: Methylenetetrahydrofolate reductase

MTRNR2L8: Mitochondrially encoded 16S ribosomal RNA

MX1: Myxovirus resistance protein 1

MX2: Myxovirus resistance protein 2

NLRP3: NOD-like receptor family pyrin domain containing 3

PDE4: Phosphodiesterase 4

PGE2: Prostaglandin E2

PPARG: Peroxisome proliferator-activated receptor gamma

PTGES: Prostaglandin E synthase

PTPN22: Protein tyrosine phosphatase non-receptor type 22

RAB2A: Ras-related protein Rab-2A

RHOC: Ras homolog family member C

RUNX3: Runt-related transcription factor 3

S100A4: S100 calcium-binding protein A4

S100A6: S100 calcium-binding protein A6

S100A8: S100 calcium-binding protein A8

S100A9: S100 calcium-binding protein A9

S100A10: S100 calcium-binding protein A10

SELL: Selectin L

SFRP2hi/DPP4: Secreted frizzled-related protein 2hi/Dipeptidyl peptidase 4

SPP1: Secreted phosphoprotein 1

STAT1: Signal Transducer and Activator of Transcription 1

STAT4: Signal Transducer and Activator of Transcription 4

TGF-β: Transforming Growth Factor-beta

TGFB2: Transforming Growth Factor-beta 2

TMEM176A: Transmembrane protein 176A

TMEM176B: Transmembrane protein 176B

TMSB10: Thymosin beta-10

TNFa: Tumor Necrosis Factor-alpha

TNIP1: TNFAIP3-interacting protein 1

TPT1: Tumor protein, translationally-controlled 1

TSPAN32: Tetraspanin 32

TYK2: Tyrosine kinase 2

VCAN: Versican

αSMA: Alpha Smooth Muscle Actin

Summary

Systemic sclerosis, or scleroderma (SSc), is a complex immune-mediated inflammatory disorder. SSc pathogenesis involves a triad of factors: immunological imbalance characterized by the presence of autoantibodies. pronounced vascular damage and extensive fibrosis of the skin and the internal organs. Clinically, SSc can be classified based on the extension of fibrosis. When it primarily affects the face and limbs, it is referred to as limited cutaneous SSc (lcSSc), while when it involves internal organs, it is defined as diffuse cutaneous SSc (dcSSc). Additionally, SSc can be classified by the autoantibodies generated by the patient, the predominant types being anti-topoisomerase (ATA+) and anti-centromere (ACA+), that correlate with dcSSc and lcSSC, respectively.

SSc has a clear genetic component and the role of common genetic variants in the susceptibility to SSc has been explored by using high throughput genotyping techniques such as genome-wide association studies (GWAS) arrays. The latest GWAS in the disease, published in 2019, identified 27 *loci* associated with SSc, discovering 13 new ones and providing new insights into the molecular pathways and specific cell types implicated in the pathogenesis of the disease. In addition, the mentioned large genomic study established solid grounds for comprehensive follow-up studies and data mining strategies. Consequently, the main aim of this doctoral thesis research was to further understand the biological mechanisms involved in SSc by applying novel analysis strategies on GWAS datasets and to generate and analyze the gene expression profiles of relevant cell subtypes at the single cell level.

A straightforward application of the identification of genetic risk factors and the estimation of their effects in large GWASs is the generation

of polygenic or genomic risk scores (PRS or GRS, respectively). This method enabled us to identify individuals at risk of developing the disease based on the presence of specific alleles in known disease-associated *loci*. In this doctoral thesis, we developed a 33 single nucleotide polymorphism (SNP) GRS that was proven to be able to differentiate between healthy individuals and SSc patients. Moreover, the generated GRS showed a relevant clinical management potential, as it could distinguish between individuals with SSc and those with two related immune-mediated disorders such as rheumatoid arthritis and Sjögren's syndrome, especially when genetic information was combined with immune cell count data.

On the other hand, two-sample Mendelian randomization methods allowed the scientific community to combine GWAS results for diseases and environmental risk factors to address the causality of risk factors on disease onset. High levels of body fat or obesity are known risk factors for numerous diseases, including immune-mediated diseases. Obesity is associated with a state of chronic low-level inflammation, in which adipocytes release pro-inflammatory cytokines. Therefore, this thesis included the study of the causal contribution of body fat distribution to the SSc. We utilized GWAS data from public repositories for anthropometric measures of body fat distribution, including body mass index (BMI), waist-to-hip ratio, and BMI adjusted for waist-to-hip ratio. However, our analyses did not reveal a causal relationship between SSc and any of these obesity-related anthropometric measures.

Finally, based on previous knowledge, monocytes were selected as targets cells to study in order to better understand the SSc pathogenesis. Monocytes are myeloid cells that circulate in the blood with various functions in innate immunity, ranging from direct action against threats to the activation of other cell types and chemotaxis. Therefore, we isolated

monocytes from peripheral blood in patients and controls and using single-cell transcriptome analysis (scRNA-seq), we identified aberrant gene expression profiles in SSc patients. Briefly, non-classical monocytes (ncMos) were found in higher proportions in SSc patients, and SSc ncMos also expressed increased levels of *PTGES* and interferon-mediated activation. *PTGES* encodes a prostaglandin E2 synthase, which was previously proposed as a therapeutic target in inflammation and may also be relevant for SSc patients. A SSc-related cluster of IRF7+ STAT1+ intermediate monocyte subset with an aberrant interferon response was also described. Additionally, we identified a M2-polarized population of classical monocytes that was depleted in patients. Considering that M2 macrophages are profibrotic, we hypothesized that these monocyte subset is being activated and migrating to tissues in SSc patients.

The results presented in this doctoral thesis signify a step forward in the genetic exploration of SSc from various perspectives. First, we employed previously published GWAS data to predict SSc onset based on genetic variants. Second, we utilized the same GWAS data to establish body fat distribution as a predisposing risk factor for the disease. Third, we employed cutting-edge techniques like scRNA-seq to describe circulating blood monocytes and their potential role in the disease.

Resumen

La esclerosis sistémica, o esclerodermia (SSc), es un trastorno inflamatorio mediado por el sistema inmunológico de etiología compleja. La patogénesis de la enfermedad involucra una tríada de factores: desequilibrio inmunológico caracterizado por la presencia de autoanticuerpos daño vascular pronunciado y fibrosis extensa de la piel y los órganos internos. Clínicamente, la SSc se puede clasificar según la extensión de la fibrosis. Cuando afecta principalmente la cara y las extremidades, se denomina esclerósis sistémica cutánea limitada (lcSSc), mientras que cuando involucra órganos internos, se define como esclerosis sistémica cutánea difusa (dcSSc). Además, la SSc se puede clasificar según los autoanticuerpos generados por el paciente, siendo los tipos predominantes la anti-topoisomerasa (ATA+) y los anticentrómero (ACA+), que se correlacionan con dcSSc y lcSSC, respectivamente.

La SSc es una enfermedad con un fuerte componente genético, estudiado en los últimos años gracias a técnicas de genotipado como los estudios de asociación de genoma completo (GWAS). El último GWAS en la enfermedad, publicado en 2019, identificó 27 *loci* asociados con la ES, descubriendo 13 nuevos y proporcionando nuevas perspectivas sobre las rutas moleculares y los efectos celulares específicos implicados en la patogénesis de la enfermedad. Adicionalmente, este gran estudio genómico estableció bases sólidas para estudios de seguimiento y estrategias de minería de datos. En consecuencia, el principal objetivo de esta tesis doctoral fue comprender mejor los mecanismos biológicos involucrados en la SSc mediante la aplicación de nuevas estrategias de análisis en conjuntos de datos de GWAS y generar y analizar los perfiles de expresión génica de subtipos celulares relevantes a nivel de células única.

Una aplicación directa de la identificación de factores de riesgo genéticos y la estimación de sus efectos en GWAS de grandes cohortes son los índices de riesgo poligénico o genómico (PRS o GRS, respectivamente). Esta técnica permite identificar a las personas en riesgo de desarrollar la enfermedad según la presencia de alelos específicos en los *loci* asociados a la enfermedad. En esta tesis doctoral, desarrollamos un GRS de 33 SNP que demostró ser capaz de diferenciar entre individuos sanos y pacientes con SSc. Además, el GRS generado mostró un potencial relevante en la gestión clínica, ya que podía distinguir entre individuos con SSc y aquellos con dos enfermedades autoinmunes relacionadas, como la artritis reumatoide y el síndrome de Sjögren, especialmente cuando se combinó información genética con datos de recuento de células inmunitarias.

Por otro lado, los métodos de aleatorización mendeliana de dos muestras permitieron a la comunidad científica combinar los resultados de GWAS para enfermedades y factores de riesgo ambientales para abordar la relación causal entre ellos. Los altos niveles de grasa corporal u obesidad son factores de riesgo conocidos para numerosas enfermedades, incluidas los transtornos mediados por el sistema inmunológico. La obesidad está asociada con un estado de inflamación crónica de bajo nivel, en el cual los adipocitos liberan citoquinas proinflamatorias. Por lo tanto, esta tesis incluye el estudio de la contribución causal de la distribución de grasa corporal a la SSc. Utilizamos datos de GWAS de repositorios públicos para medidas antropométricas de distribución de grasa corporal, incluyendo el índice de masa corporal (IMC), el ratio cintura-cadera y el IMC ajustado por el ratio cintura-cadera. Sin embargo, nuestros análisis no revelaron una relación causal entre la SSc y ninguna de estas medidas antropométricas relacionadas con la obesidad.

Finalmente, en base al conocimiento previo, los monocitos se seleccionaron como células objetivo para nuestro estudio, con el fin de comprender mejor la patogénesis de la SSc. Los monocitos son células mieloides que circulan en la sangre con diversas funciones en la inmunidad innata, que van desde la acción directa contra amenazas hasta la activación de otros tipos celulares y la quimiotaxis. Por lo tanto, aislamos monocitos de sangre periférica en pacientes y controles y, utilizando análisis de transcriptoma de células únicas (scRNA-seq), se identificaron perfiles de expresión génica aberrantes en pacientes con SSc. Se encontró una mayor proporción de monocitos no clásicos (ncMos) en pacientes con SSc, y estos también expresaron niveles aumentados de *PTGES* y activación mediada por interferón. *PTGES* codifica una prostaglandina E_2 sintasa, que previamente se ha propuesto como un objetivo terapéutico en la inflamación y que también puede ser relevante para los pacientes con SSc. También se describió un grupo celular relacionado con la SSc de un subconjunto de monocitos intermedios que expresan IRF7+ STAT1+ con una respuesta de interferón aberrante. Además, identificamos una población de monocitos clásicos polarizados a macrofagos M2 que estaba deplecionado en los pacientes. Dado que los macrófagos M2 son células profibróticas, planteamos la hipótesis de que este subconjunto de monocitos se activa y migra a los tejidos en pacientes con SSc.

Los resultados presentados en esta tesis doctoral representan un avance en la exploración genética de la SSc desde diversas perspectivas. En primer lugar, al utilizar datos de GWAS previamente publicados para predecir la aparición o distinguir la enfermedad de otras basado en variantes genéticas. En segundo lugar, al utilizar los mismos datos de GWAS para intentar dilucidar si la distribución de grasa corporal es un factor de riesgo para la enfermedad. En tercer lugar, al utilizar técnicas de vanguardia como

scRNA-seq para describir los monocitos de la sangre circulante y su papel potencial en la SSc.

Introduction

1. Systemic sclerosis is a chronic life-threatening disorder

The immune system comprises a highly complex network of cells, tissues and organs distributed throughout the body. It plays a critical role in protecting against diseases and pathogens and, it must be strictly regulated to function effectively. The ability of this system to respond appropriately to a threat is crucial for an individual's health and an excessive response can be as harmful as a deficiency (1,2).

Immune-mediated inflammatory diseases (IMIDs) constitute a group of heterogeneous disorders resulting from an overactive immune system and the loss of self-tolerance. IMIDs manifest across a spectrum of conditions affecting various organs, individuals of all ages and exhibiting a wide array of symptoms. The severity of each IMID can vary significantly, ranging from minor biochemical abnormalities to life-threatening situations, and they often have a chronic course (2). Consequently, IMIDs pose both clinical and economic challenges to society. For clinicians, distinguishing between IMIDs can be intricate, given the frequent overlap of clinical presentations and symptoms among these conditions. Therefore, an accurate diagnosis is crucial, particularly in the early stages, to initiate appropriate treatment. From a societal perspective, particularly within the healthcare system, the chronic nature of IMIDs requires the development of optimized strategies for cost-effective management of affected individuals. This is particularly important because patients with IMIDs often rely on the healthcare system for lifelong care and support (1).

The combined prevalence of IMIDs is high, affecting approximately 7-9% of the population, although this prevalence is not evenly distributed. IMIDs are more commonly observed in women than in men and exhibit variations among different ethnic groups (3,4). From an epidemiological perspective, there is significant variability in the incidence rates of various IMIDs. Some IMIDs, such as type 1 diabetes (T1D) or rheumatoid arthritis (RA) or occur more frequently, while others, like systemic lupus erythematosus (SLE) or systemic sclerosis (SSc) are considered rare diseases (5).

IMIDs can be classified into two main categories based on the extent of affected tissues: organ-specific diseases, such as T1D or multiple sclerosis (MS), and systemic diseases, including RA, SLE or SSc. Regardless of the affected tissue, IMIDs are typically characterized as complex and polygenic diseases. The complexity of these disorders arises from both genetic and environmental components, as the final phenotype is strongly influenced by external factors. Multiple genetic factors contribute to the onset of the disease; however, their individual contribution is modest, and it is the combination of these factors that leads to the dysregulation of the system and the disease. Nonetheless, it is worth noting that some IMIDs are monogenic or have monogenic forms as it happens in SLE, resulting from alterations in key genes (1,6).

SSc, also called scleroderma, is a complex chronic IMID. The hallmark characteristics of this disease include immune imbalance, vascular damage and alterations, and tissue fibrosis driven by excessive collagen deposition (7) (**Figure 1**). However, one challenging feature of SSc is the high heterogeneity between patients in their clinical manifestations, laboratory values and response to treatment.

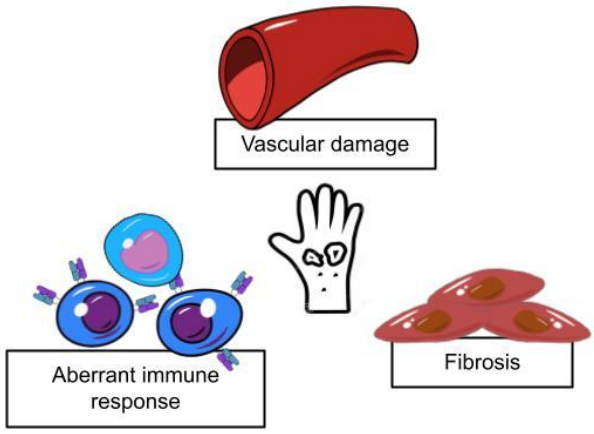


Figure 1. Hallmarks of the pathology of SSc.

Like many IMIDs, early stages of SSc often present with general symptoms such as fatigue, but many SSc patients debut with Raynaud's phenomenon. Raynaud's phenomenon is one of the primary vascular manifestations of the disease, characterized by fibrotic proliferation in small vessels, resulting in the fingers' tip turning white or blue due to impaired blood circulation and cyanosis (8). Not all individuals with Raynaud's phenomenon develop SSc, and the vascular patterns can be assessed by capillaroscopy and are used as a diagnostic tool in the early stages of the disease (9). Raynaud's phenomenon is not the sole vascular event associated with the disease; others include renal crisis and pulmonary arterial hypertension (PAH). Renal crisis is a severe complication that may necessitate renal transplantation, and alongside PAH are the leading causes of death in SSc (7).

The immune imbalance involvement in SSc mainly encompasses the overactivation of the innate and acquired immune responses. In this regard, this imbalance translates into the tissues through characteristic cellular infiltrates in patients, involving activated T cells and myeloid cells exhibiting

aberrant behavior (10–12). Furthermore, it has been described that some immune cell populations are elevated in the blood of SSc patients (12). All the above lead to the production of different autoantibodies and proinflammatory cytokines and chemokines, but also an interferon (IFN) signature that is characteristic of the disease (7).

Different organs and systems are affected in SSc, as is illustrated in **Figure 2**. Alongside vascular complications, the disease frequently affects organs like the lungs, gastroesophageal tract, and the heart (13). While gastroesophageal issues are the most prevalent among SSc patients, the most life-threatening complications involve the lungs, encompassing pulmonary hypertension, pulmonary fibrosis (PF), and interstitial lung disease (ILD) (13). ILD stands as the primary cause of mortality among patients, followed by SSc renal crisis. Digital ulcers and musculoskeletal problems are also frequently observed in these patients (13).

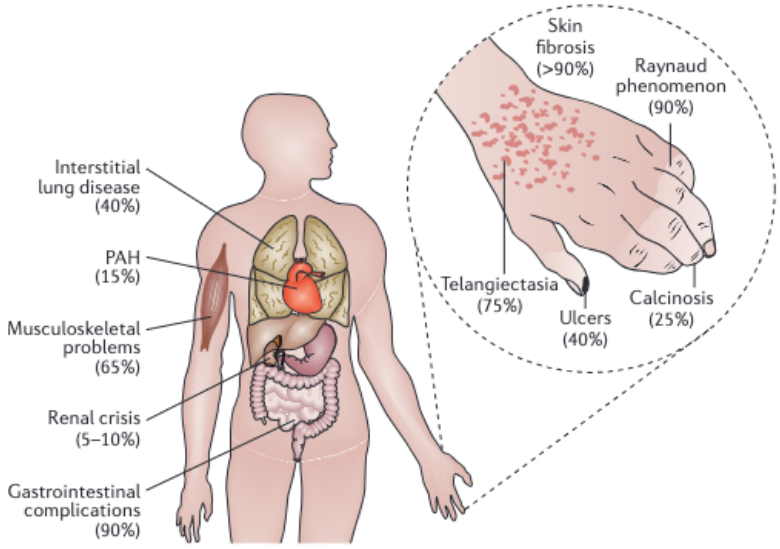


Figure 2. Different organs implicated in the pathology of systemic sclerosis (Extracted from Allanore *et al.*, Nat Rev Dis Primers, 2015).

Fibrosis in SSc typically commences in the skin, specifically in the distal fingers and toes before advancing proximally. Over time, an excess deposition of collagen leads to skin thickening and limited joint mobility, resulting in contractures affecting both large and small joints in some patients. After several years of disease progression, the skin can become tight, sclerotic, and occasionally atrophic. This fibrosis will extend to internal organs, affecting life function (7).

In terms of epidemiology, the incidence of SSc significantly varies by country. Globally, the estimation is 1 in 10,000 people (14). The peak of incidence appears around middle age, 50-60 years old. Gender-wise, there is a strong bias, with a higher incidence in women, with a 8:2 female to male ratio (15). However, the disease tends to be more severe in men, and their prognosis is generally poor (13).

Classifying SSc patients is a complex task due to the highly heterogeneous nature of the disease. As mentioned before, one of the defining characteristics of SSc is the extent of fibrosis, which serves as a key criterion for classification. Patients typically fall into one of two distinct groups: those with diffuse cutaneous systemic sclerosis (dcSSc), characterized by widespread fibrosis, and those with limited cutaneous systemic sclerosis (lcSSc), where fibrosis predominantly affects the hands, knees, and may or may not involve the face and neck. However, it is important to note that some patients may not fit neatly into either of these subtypes or may transition between them over time (14).

In addition to evaluating the progression of skin fibrosis, classification criteria also consider serological profiles. Approximately 95% of SSc patients test positive for antinuclear antibodies (ANA), which are normally mutually exclusive. There are three primary autoantibodies observed in SSc patients:

anticentromere antibodies (ACA), often associated with lcSSc; anti-topoisomerase I antibodies (ATA), and anti-RNA polymerase III antibodies (ARA), which are more commonly found in dcSSc patients. The serological profile is typically determined using indirect immunofluorescence, proving highly valuable for both diagnosis and classification (7,13,14).

2. Environmental risk factors

SSc is a complex disorder in which both genetic and environmental factors play an important role in the disease onset and development. Risk factors encompass a broad category of external agents, lifestyle and dietary factors influenced by sex, culture and social status, and can be classified into three primary categories: chemical, biological, and physical agents (16). Each category will be elaborated through in the following section.

Chemical agents constitute one of the most significant risk factors in SSc, with the first documented case dating back to 1914 (17). This category includes fractured silica crystal dust, organic solvents, asbestos, and various industrial agents (17–20). Typically, these aggressive agents come into contact with the skin or the lungs, two of the main organs/tissues affected in the disease. This relationship has been explored, identifying silica as a potential risk factor without a clear pathogenic mechanism (21). However, silica is known to affect T cells as an adjuvant, potentially leading to their overactivation and triggering autoimmune responses in predisposed individuals (21). Asbestos exposure may exert a similar effect and is also considered a risk factor. This exposure has been associated with various IMIDs (21), with animal models demonstrating the generation of ANAs following intratracheal injections of amphibole asbestos (22). Furthermore, asbestos exposure has been linked to alterations in cytokine levels in the blood, including increased levels of IL-17 and TGF-β (23–25). Organic solvents

are carbon-based substances capable of dissolving or dispersing other substances, which represent another significant risk factor in SSc. Research dating back to the mid-20th century has investigated the influence of organic solvents on SSc, revealing that individual exposure to these solvents nearly doubles the likelihood of developing the disease, with a higher risk observed in men (25). This risk factor has also been linked to the dcSSc subtype (19). A significant intoxication case to consider is the canola oil intoxication in Spain during the years 1980-1981 (26–28). The consumption of adulterated oil led to what was termed the Toxic Oil Syndrome, affecting thousands of people and resulting in the death of hundreds (26). Among the numerous and diverse symptoms exhibited by the affected individuals, some included skin thinning and other symptoms associated with SSc (28), primarily in women (29).

Biological agents constitute relevant risk factors for numerous IMIDs, being infectious agents one of the most important factors in this regard, due most IMIDs have been associated with them (30). The main hypothesis establishes that the association arises from the molecular resemblance between infectious agents and human proteins, potentially leading to the misidentification of self-epitopes as foreign invaders. When infections occur in individuals with a genetic predisposition, it can result in a breakdown of self-tolerance (30). For instance, antibodies against the Epstein-Barr virus (EBV) have been suggested to correlate with skin and joint manifestations in SLE (30), while both EBV and cytomegalovirus (CMV) have been linked to the onset of SSc (31,32). These associations seem to be due to the fact that both EBV and CMV can infect endothelial cells (ECs). Interestingly, in SSc, antibodies recognize a specific epitope from CMV, UL94, that induces apoptosis when it binds to the integrin-NAG-2 protein complex in ECs (31). There is further evidence of other infectious agents and their involvement in SSc, such as the increased presence of *Helicobacter pylori* in SSc patients (33,34).

Some physical agents, such as ultraviolet and ionizing radiation or electronic and magnetic fields, are capable of changing DNA structure, which causes cellular death and damage that could be a possible explanation for the onset of different IMIDs. Factors with a direct effect on the skin, such as ultraviolet light, have been linked to the development of MS and SLE (35,36). Nevertheless, no physical agents have been associated with SSc to date.

Lifestyle factors can greatly influence the susceptibility to IMIDs, and some of the most well-known include smoking, obesity, physical activity, educational level, and socioeconomic status (37–39). For instance, smoking is a common and widely recognized risk factor, established for other autoimmune diseases (37). However, it has been relatively underexplored in the context of SSc. Nevertheless, evidence suggests that smoking may increase the risk of fibrosis, vascular damage, and respiratory issues in SSc, and it is recommended that patients discontinue this habit (40).

Among lifestyle factors, obesity has emerged as a major public health issue in Western countries, with increasing prevalence across most nations. Co-occuring with the increase in obesity levels, the incidence of IMIDs has grown, raising suspicions about a correlation between the two phenomena. In fact, obesity has been characterized by a state of low-grade inflammation, marked by elevated levels of inflammatory markers such as IL-6 and C-reactive protein (CRP) in obese individuals (41). Adipose tissue is classified into two types: white (WAT) and brown adipose tissue, with WAT comprising the majority of adipose tissue in the body (42). WAT is primarily composed of adipocytes, along with a smaller fraction of other cells such as macrophages, which are more prevalent in females than males. Additionally, pre-adipocyte cells can differentiate into macrophages under the right stimuli, distinguishing them from macrophages derived from circulating monocytes (43). Adipocytes exhibit stimulatory actions, as demonstrated in culture

experiments that show an increased expression of adhesion molecules in ECs. Furthermore, adipocytes have the capability to activate adhesion and transmigration in circulating monocytes through the secretion of molecules, including chemokines like monocyte chemoattractant protein 1 (MCP-1) (44). All these factors contribute to the proinflammatory potential of adipocytes and adipose tissue. Diseases such as T1D and cardiovascular disease have been previously linked to obesity (41), however, this causal relationship with IMIDs is being recently explored (42,45,46).

3. The genetic architecture of SSc

In addition to exposure to environmental triggers, the development of a complex disease relies on the genetic predisposition of different individuals. In the case of SSc, the most relevant risk factor is still family history. Individuals with a relative who was previously diagnosed with SSc face a notably higher risk, reaching a 1.6%, compared to a 0.026% risk in the general population (47,48). Furthermore, shared genetic bases between IMIDs make it common for other IMIDs to present as comorbidities of SSc (49).

As a complex disease, SSc does not have a Mendelian inheritance pattern. Therefore, the polygenic component and environmental influences make it challenging to establish heritable patterns of the disease. Single nucleotide polymorphisms (SNPs) entail the alteration of a single pair of nucleotide bases within DNA. On average, these mutations occur approximately once every 1000 nucleotides, resulting in 4 to 5 million SNPs in a single individual. To date, research has identified over 80 million SNPs across various populations (50). A SNP must affect at least 1% of the population to be classified as such; otherwise, it falls under the category of rare variants or punctual mutations. The genetic component of polygenic

diseases consists in the accumulative effect of these SNPs, which are frequent in the population but have a low penetrance, differently from the rare mutations that impact Mendelian disease.

The role of SNPs in disease susceptibility can be addressed by performing genetic association studies. These studies compare the frequency of the different alleles between two groups of individuals, *i.e.* those affected by a particular condition (cases) and those who are not (controls). If an allele is significantly more frequent in cases than in controls, then the allele would have a risk effect. On the contrary, if an allele is significantly less frequent in cases than in controls, it would have a protective effect.

Genetic association studies are usually based on prior knowledge of the disease, meaning that they rely on a hypothesis. The selection of candidate genes and SNPs would be under the assumption of their impact in the disease due a known path or a previous association in a similar effect. Genome-wide association studies (GWAS) opened doors for comprehensive genome examination without the need to genotype all the known polymorphic positions of the genome, thanks to the phenomenon of linkage disequilibrium (LD). LD is the non-random association between genetic variants located close to each other on a chromosome, meaning they tend to be inherited together more often than expected by chance. In the GWAS techniques, the SNP genotyping panel is narrowed down by genotyping a limited number of SNPs and inferring those that are frequently inherited together due to the LD patterns observed in complete reference panels (51). These mean that GWAS are hypothesis free studies and make them the ground floor for further research based on their results.

To date, GWAS identified thousands of *loci* associated with complex diseases (52,53). The high number of tested SNPs generates a necessity of a more restrictive statistical significance threshold. The threshold is established at $p < 5x10^{-8}$ which corresponds to Bonferroni correction of one million independent tests, in this case SNPs (54). Furthermore, the possibility of false positives also increases, making it necessary to perform a replication step. Moreover, due to the number of SNPs that are inherited in the same haplotype blocks and are statistically indistinguishable, it is challenging to determine the causal variants within the associated region, often necessitating further fine mapping or functional studies.

Regarding the genetic component of SSc, the human leukocyte antigen (HLA) region, also known as the major histocompatibility complex (MHC) region, stands out as the first genetic association with the disease, exerting the most significant genetic influence on it. The HLA is highly polymorphic, in fact it is the most variable region in the human genome due to its function in antigen presentation. HLA class I has a role in the antigen presentation in every cell in the body, while HLA class II molecules play a crucial role in antigen presentation, involving various antigen-presenting cells such as mononuclear phagocytes, dendritic cells, endothelial cells and B cells. Consequently, this *locus* has a strong influence in IMIDs, and its association with human disease has been one of the main research issues in the past years (55).

HLA region have consistently shown the most significant associations with SSc susceptibility across various studies in different populations, including those involving candidate genes and large-scale genetic studies, as GWAS or, the Immunochip and GWASs (58, 60,61).

Within the HLA region, several classical alleles have been associated with SSc, being the most important *HLA-DRB1*11:04/HLA-DQA1*0501/HLA-DQB1*0301* haplotype (56–62). Our recent study, the most extensive HLA analysis in SSc using recent GWAS data, has confirmed the association of nine classical alleles with SSc, and unveiled that HLA class I also plays a substantial role in the disease's pathogenesis with the association of *HLA-B*08:01* (63). Importantly, the HLA region in SSc exhibits associations with specific patient subsets and distinct clinical features. For instance, *HLA-DQA1*02:01* is specifically linked to lcSSc, while *HLA-DQA1*05:01* is exclusively associated with dcSSc (63). Beyond the HLA region, there is a multitude of *loci* that have been linked to SSc in candidate gene studies or in GWAS (64). Over the past two decades, the number of *loci* associated with SSc has steadily increased, culminating in the identification of 27 disease-associated *loci* in the 2019 GWAS (65).

In this context, genes associated with the immune system have exhibited the strongest correlations. *IRF5*, which encodes the interferon regulatory factor 5, was the first non-HLA gene linked to SSc. Subsequent studies, including GWAS, reinforced this association. For instance, *IRF5* rs2004660 is a well-established SSc risk factor, and macrophages harboring this SNP have been shown to display heightened proinflammatory activity and polarization toward M1 macrophages. Several other *IRF5* variants have been linked to SSc (65,66), but the most intriguing one, rs4728142, appears to hold promise in prognosticating longer survival and preserved lung function, thanks to its role in decreasing *IRF5* expression (67). In addition to *IRF5*, other interferon regulatory factors (IRFs), including *IRF4*, *IRF5*, *IRF7*, and *IRF8* (**Table 1**), are associated with SSc. IRFs play a crucial role in regulating interferon transcription, orchestrating the activation of interferon genes following infection stimuli. Moreover, in SSc skin tissue,

IRF7 has been observed to be upregulated and active, with its deletion resulting in reduced profibrotic factors in SSc fibroblasts (68).

Similarly, genes associated with adaptive immune responses also play a role in SSc. *CD247*, among the first genes identified, is a prime example. This gene encodes the T-cell receptor T3 zeta chain, a key modulator of antigen-dependent T-cell activation following antigen recognition. The *CD247* rs2056626 variant has been associated with SSc from the early stages of genetic studies on the disease (66,69,70). Additionally, *STAT4* has been identified as an SSc risk factor, with two different SNPs, rs3821236 and rs4853458, being associated with the disease in the most recent GWAS (65). *STAT4*, is a transcription factor belonging to the STAT family, and plays a pivotal role in mediating the response to IL-12 signaling in lymphocytes and regulating T helper cell differentiation. Notably, several genes involved in the IL-12 signaling pathway, such as *TYK2* (rs2304256, rs34536443, rs12720356, rs35018800), *IL12A* (rs7758790, rs589446), *IL12RB1* (rs436857, rs2305743, rs8109496, rs11668601) and *IL12RB2* (rs3790566, rs924080, rs3790567), have been linked to SSc, encompassing components from the cytokine itself to its receptors (**Table 1.**). The *IL-12A* gene encodes the alpha subunit of the IL-12 cytokine, which forms a heterodimer and binds to the cell receptor. This receptor is encoded by two genes, *IL12RB1* and *IL12RB2*. Upon activation, the IL12RB2 protein provides a binding site for kinases like TYK2, which has also been associated with SSc (65,70–73).

Autophagy and apoptosis represent critical cellular processes involved in remodeling. These processes entail the degradation of organelles by lysosomes, posing a risk of exposing an individual to their own epitopes. Regarding genetic variations in SSc, some are associated with the disease and are implicated in these cellular processes. For instance, *ATG5*, which encodes a protein involved in autophagosome elongation, is linked to SSc,

specifically with rs9373839 (74). Several genes related to apoptosis have been associated with SSc, including *DNASE1L3*, which encodes a protein in the deoxyribonuclease 1 family responsible for DNA fragmentation during apoptosis. *DNASE1L3* rs35677470 has been associated with ACA+ patients in SSc (74,75). Lastly, *GSDMA* and *GSDMB* genes play crucial roles in pyroptosis, a form of cell death triggered by proinflammatory signals and associated with inflammation. Both genes have been linked to SSc (65,76).

One step forward in the study of genetic susceptibility in IMIDs has been the meta-analysis of GWAS (meta-GWAS). This kind of analysis allows combining data from different sources and investigating the shared genetic association, even including cohorts with relatively small size. With this strategy it was possible to increase the statistical power of SSc GWAS and to improve our capability to identify significant associations, with new genes associated with the disease, such as *IRF5-TNPO3 loci* or *DDX6* (65). The shared genetic background and susceptibility between IMIDs also benefited from this type of approach, showing that IMIDs shared up to half of their genome-wide significant associated-variants (77–80). Considering the shared genetic component of SSc with other IMIDs, SLE and RA stand out. Also, the most recent cross-disease meta-analysis between IMDs, including SSc, SLE, RA and idiopathic inflammatory myopathies, performed by our group, revealed five new *loci* between these diseases, *NAB1*, *KPNA4-ARL14*, *DGQK*, *LIMK1* and *PRR12*, some of them implicated in autoimmunity (81). Finally, it's noteworthy that meta-analyses can unveil shared genetic components among seemingly different diseases, as evidenced in the case of SSc and Crohn's disease (CD) (79). In a study conducted by our group, encompassing over 10,000 individuals with both diseases, 4 shared genetic *loci* were identified: *IL12RB2*, *IRF1/SLC22A5*, *STAT3*, and an intergenic locus at 6p21.31. These *loci* featured pleiotropic variants with opposite allelic effects in the diseases. Furthermore, an enrichment of the IL-12 family and

type I IFN signaling pathways was observed among the shared genetic risk *loci* for SSc and CD (79,82).

Table 1. Non-HLA genetic susceptibility *loci* associated with SSc. (Adapted from Ota, *Inflamm Regener*, 2021)

CHR	Gene	BP	SNP	Gene name	Study type	References
Innate immune response						
4	NFKB1	102527984	rs230534	Nuclear Factor Kappa B Subunit 1	meta-GWAS	(65)
5	TNIP1	151060536 - 151076171	rs4958881, rs2233287, rs3792783	TNFAIP3 Interacting Protein	GWAS, meta-GWAS,	(65,69,83,84)
6	IRF4	434464	rs9328192	Interferon Regulatory Factor 4	GWAS	(77)
6	TNFAIP3	137869500 - 137921300	rs5029929, rs2230926, rs6932056	Tumor necrosis factor Alpha Induced Protein 3	GWAS	(70,83,85,86)
7	IRF5	128933813 - 129077852	rs4728142, rs10488631, rs10488631, rs3757385, rs109542313, rs2004640, rs12537284, rs2280714	Interferon Regulatory Factor 5	GWAS	(66,69,70,74,77,83,87)
7	IRF5-TNPO3	129011368 - 129018785	rs36073657, rs12155080	Interferon Regulatory Factor 5, Transportin 3	meta-GWAS	(65)
7	TAP2	28808807-32743301	rs12538892, rs17500468	Transporter 2, ATP Binding Cassette Subfamily B Member	ImmunoChip	(74)
11	IRF7	589564-613208	rs1131665, rs4963128, rs702966	Interferon Regulatory Factor 7	CGA, meta-GWAS	(65,88)
16	IRF8	85938316-85985665	rs11642873, rs2280381, rs11117432, rs11644034, rs12711490, rs7202472, rs11117420	Interferon Regulatory Factor 8	GWAS, meta-GWAS,	(65,70,83,85,88–90)
Adaptive immune response						
1	TNFSF4	173243134 - 173364690	rs4916334, rs10798269, rs12039904	Tumor necrosis factor Superfamily Member 4	GWAS, meta-GWAS	(70,83,91)

1	CD247	167451088	rs2056626	T-Cell Receptor T3 Zeta Chain	GWAS, meta-GWAS	(65,66,89,92)
1	PTPN22	113834846	rs2476601	Protein Tyrosine Phosphatase Non-Receptor Type 22	GWAS	(77)
2	STAT4	191038032 - 191099907	rs7574865, rs3821236, rs4853458, rs10168266, rs3821236	Signal Transducer and Activator Of Transcription 4	GWAS, ImmunoChip, meta-GWAS	(65,69,70,87,89,90)
8	BLK	11486464-11491677	rs13277113, rs2736340	BLK Proto-Oncogene, Src Family Tyrosine Kinase	GWAS	(83)
15	CSK	74784926	rs1378942	C-Terminal Src Kinase	GWAS, GWAS follow-up	(65,83,91)
IL-12 Signaling Pathway and cytokines						
1	IL-12RB2	67294457-67356694	rs3790566, rs924080, rs3790567	Interleukin 12 Receptor Subunit Beta 2	GWAS, meta-GWAS	(65,72)
3	IL-12A	160015740	rs589446	Interleukin 12A	GWAS, ImmunoChip	(65)
19	TYK2	10352442-10364976	rs2304256, rs34536443, rs12720356, rs35018800	Tyrosine Kinase 2	ImmunoChip follow-up	(71)
19	IL-12RB1	18079534-18103711	rs436857, rs2305743, rs8109496, rs11668601	Interleukin 12 Receptor Subunit Beta 1	meta-GWAS	(65,73)
Apoptosis, Autophagy Pathway						
3	DNASEIL3	58197809	rs35677470	Deoxyribonuclease 1 Like 3	ImmunoChip, meta-GWAS	(74)
6	ATG5	106207742 - 106286165	rs9373839, rs633724	Autophagy Related 5	GWAS, ImmunoChip, meta-GWAS	(65,74,77,83)
6	NOTCH4	32222529	rs443198	Notch Receptor 4	GWAS	(88)
6	PRDM1	106129393	rs4134466	PR/SET Domain 1	GWAS	(90)

17	GSDMB	39907028	rs883770	Gasdermin B	meta-GWAS	(65)
17	GSDMA	39965640	rs3894194	Gasdermin A	GWAS	(76)
Vascular homeostasis and fibrosis						
3	PPARG	12218016	rs310746	Peroxisome Proliferator Activated Receptor Gamma	GWAS follow-up	(93)
Other						
3	NAB1	19066954	rs16832798	NGFI-A Binding Protein 1	meta-GWAS	(65)
3	POGLUT1-TIMMDC1-CD80-ARHGAP31	119397203	rs9884090	Protein O-Glucosyltransferase 1 Translocase of Inner Mitochondrial Membrane Domain Containing 1 Cluster of Differentiation 80 Rho GTPase Activating Protein 31	meta-GWAS	(65)
4	DGKQ	971891	rs11724804	Diacylglycerol Kinase Theta	meta-GWAS	(65)
11	DDX6	118768544	rs11217020	DEAD-Box Helicase 6	meta-GWAS	(65)
11	RAB2A-CHD7	618072	rs6598008	Member RAS Oncogene Family Chromodomain Helicase DNA Binding Protein 7	meta-GWAS	(65)
11	TSPAN32, CD81-AS1	2327289	rs2651804	Tetraspanin 32 Cluster of Differentiation 81 Antisense RNA 1	meta-GWAS	(65)
17	NUP85-GRB2	75228444	rs1005714	Nucleoporin 85 Growth Factor Receptor Bound Protein 2	meta-GWAS	(65)

5. Genomic strategies to deepen into the genetic and environmental risk factors of SSc.

One of the strengths of the GWAS is that they establish a firm basis for follow-up studies. Whether fine mapping specific loci or using the bulk of the metadata provided by the GWAS, the possibilities expand, especially due the public use of the data, with initiatives such as GWAS catalog (94). The information pertaining to a disease or a phenotypic trait, contained within the summary statistics of GWAS can be used for various techniques and approached in different ways. Using this data, we can delve into the underlying biology of the phenotype, estimate heritability, explore genetic correlations with other phenotypes, calculate the genetic risk of developing the disease, investigate therapeutic targets by identifying genes involved in disease susceptibility pathways, and explore the causal relationship between a risk factor and a final phenotype (95–97).

One of the most relevant applications for the discoveries in GWAS would be the capability to predict whether an individual is at risk of suffering from or developing a particular disease or condition, as well as the likelihood of this happening based on their genetic information, is a clinically crucial tool. Polygenic risk scores (PRS) or genomic risk scores (GRS) based on GWAS datasets have emerged to address this need. These are tools that, from their inception, aim to assess an individual's susceptibility to a specific disease based on the presence or absence of risk/protective alleles (**Figure 3**). The application of these methods would be a game changing situation in the clinic, especially in the early disease detection, making it possible to perform differential diagnosis and to initiate prevention therapies and interventions.

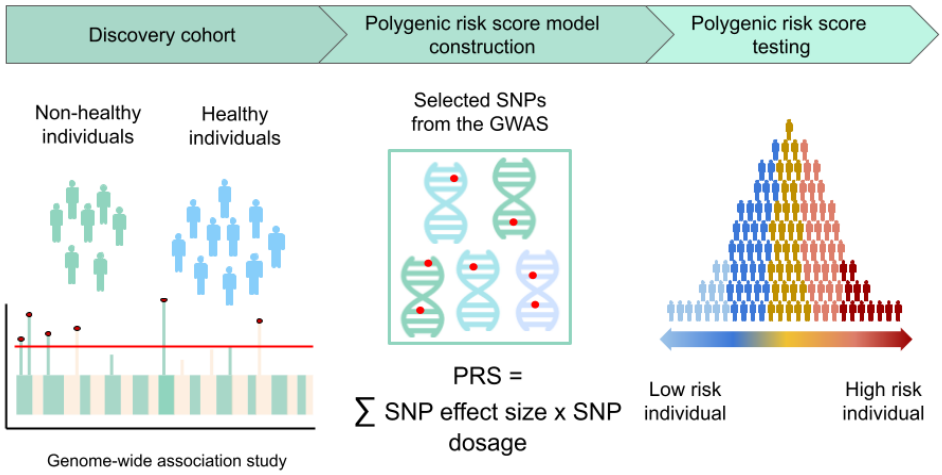


Figure 3. Polygenic risk score study design.

Furthermore, PRS has the potential to improve precision and personal medicine by predicting the evolution of the disease into the different subtypes in the early stages or predicting the possible comorbidities. However, it is necessary to distinguish between heritability explained and individual risk. The explained heritability is defined as the proportion of phenotypic variation in a population that can be explained by genetic variation (98).

A PRS is typically calculated as a weighted sum of the number of risk alleles carried by an individual. These risk alleles and their associated weights are determined based on the identified susceptibility loci and their measured effects identified through GWAS. It's worth noting that in some scenarios, a less stringent statistical significance threshold than that employed for genome-wide significance may be utilized to improve or estimate the overall predictability of the PRS. Nevertheless, this approach may come at the cost of reduced generalizability in certain contexts (99). PRS models can be combined with different parameters such as cell counts, sex, age, and multiple linked variants specific to the disease.

Although they are not usually integrated in daily clinical practice yet, PRS have been developed successfully in different IMIDs. For example, a recent study conducted by Khera *et al.* (100) has ignited a debate about the use of PRS in clinical practice. The authors developed a brilliant model for five common diseases: coronary artery disease (CAD), type 2 diabetes (T2D), inflammatory bowel disease (IBD), and atrial fibrillation. They found that, for many individuals, PRS risk variants were as significant as those associated with single variants influencing rare monogenic forms of diseases that are already routinely considered in clinical settings (100).

Additionally, a PRS model could enable the screening for several diseases by leveraging the shared genetic components, as observed among IMIDs or IMIDs. Recently, Knevel *et al*. (101) developed G-PROB (Genetic Probability tool) to calculate the probability of a patient developing various inflammatory arthritis-causing diseases, including RA, SLE, spondyloarthropathy, psoriatic arthritis, and gout. This tool was tested in clinical settings and exhibited good performance, increasing the accuracy of clinician diagnosis in 51% of cases after incorporating prediction information (101).

It's important to note that PRS models are not meant to replace physician diagnosis and intervention but rather to enhance their judgment and improve patient prognosis. These promising results suggest that personalized medicine could become more accessible. However, the challenge lies in the feasibility of genotyping every individual to incorporate this tool into clinical practice. Issues such as patient privacy and data interpretation also pose significant hurdles. Moreover, the primary concern arises from the population used to generate these models. Currently, 85% of GWAS have been conducted in the European population, which means that all subsequent studies, including PRS, will have a much lower predictive

ability in other genetic ancestries and will require further validation (102). Despite these advances, no PRS model had been developed for SSc before this doctoral thesis.

When it comes to leveraging GWAS data, Mendelian Randomization (MR) methods have gained popularity within the field of genetics. Randomized controlled trials are considered the gold standard for establishing scientific evidence regarding the causality of specific exposures on different outcomes or diseases. However, there are instances where conducting such trials may not be feasible or ethical. In such cases, MR studies have emerged as a powerful tool. This approach assesses the impact of a risk factor or exposure on a disease or outcome by employing genetic variants as instrumental variables (IVs) (103). The strength of MR methods stems from the inherent randomness and heritability of genes. They not only control for quantitative traits like urate levels or phospholipids in the blood, but also behavioral tendencies such as food intake or smoking. Therefore, MR allows us to infer causality between different events. These studies also address confounding factors more effectively than other approaches by utilizing genetic data as IVs, and they are less susceptible to issues like reverse causation compared to randomized controlled trials (104,105).

To conduct an MR study, genetic information for both the exposure and the outcome is required (**Figure 4**). Traditionally, this information was sourced from a single large cohort, making it challenging to find a well-matched population. However, with the advent of GWAS data and standardized methodologies, it has become possible to combine data from different cohorts in an approach known as two-sample MR (2SMR) (104).

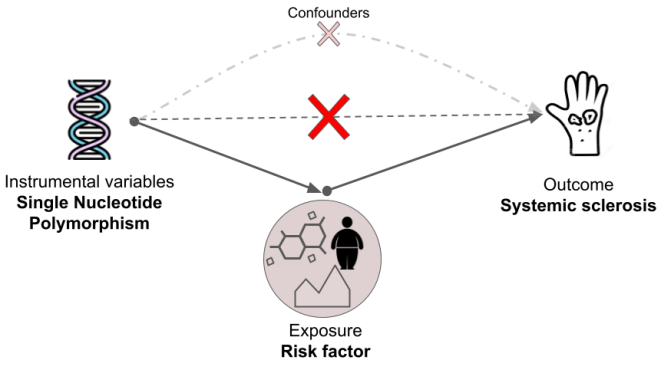


Figure 4. Principles of employing genetic variants as instrumental variables for estimating the causal impact of exposure factors on disease through Mendelian randomization.

For IVs to be considered valid in 2SMR studies, they must meet three fundamental assumptions. First, the variants must be associated with the risk factor in a GWAS. Second, the IVs should not be influenced by unmeasured confounders, as this would result in a direct association between the IV and the disease. Finally, the IVs should only affect the disease through the risk factor (104). To assess the validity of these assumptions, MR studies incorporate sensitivity analyses. One of the primary sources of invalid IVs is horizontal pleiotropy, where a genetic variant influences multiple related risk factors, introducing bias into the study. Detecting this type of pleiotropy can be challenging, but several methods have been developed to address it. One notable method is MR Egger regression, which allows for the presence of pleiotropic effects in all variants, as long as they are not directly proportional to the effects of the variants on the risk factor of interest (106).

MR studies have made a significant impact on clinical research in recent years. This approach has been widely applied to IMIDs, which are

influenced by both environmental factors and genetic susceptibility. In the case of RA, for example, MR studies have demonstrated the likelihood of obesity as a risk factor (45), as well as the risk of several on-site cancers (107) and variations in the gut microbiome (108). Similar research has been conducted for SLE, where shorter telomere length in leukocytes has been shown to be protective against the disease in both Western and Asian populations (109). Previous to this dissertation, there was only one MR study performed in SSc, which evaluated the effect of C-reactive protein levels on several diseases, including SSc, finding no causal relationship with it (110).

6. Cellular involvement in the SSc pathogenesis

SSc is a systemic disease in which multiple systems and cellular lineages show clear signs of deregulation and aberrant behavior (111). Therefore, considering the specific cellular context has proven vital to dissect the etiopathogenesis of the disease.

In this line, vascular damage is an early and prevalent issue in SSc, often marked by the Raynaud Phenomenon. Consequently, endothelial cells (ECs), which form the inner lining of blood vessels, play a crucial role in SSc pathophysiology as they influence cell survival, tissue modification processes (like angiogenesis and vasculogenesis) and interactions with surrounding cells, all contributing to SSc vasculopathy (112,113). The dysfunction of ECs, characterized by compromised junctions and elevated apoptotic markers in SSc patients, can activate myeloid cells such as dendritic cells and macrophages, subsequently triggering lymphoid cell activation (113). This immune response recruits tissue remodeling cells like fibroblasts and myofibroblasts, ultimately leading to organ dysfunction (114). The fibrotic process in SSc involves the accumulation of extracellular

matrix (ECM), comprising collagen, elastin, fibronectin, and glucosamine. This excess of ECM in affected tissues results also in organ dysfunction (115,116). Fibroblasts, initially quiescent cells, become activated due to various stimuli like cytokines, chemokines, TGF-β, reactive oxygen species, and mechanical stress. These activated fibroblasts exhibit distinct transcriptomic profiles and functions, contributing to SSc pathogenesis, with TGF-β playing a central role (117–119). In response to tissue injury or inflammation, fibroblasts can transform into myofibroblasts, characterized by the expression of α-smooth muscle actin (*αSMA*) and resistance to apoptosis (115,116). Interestingly, myofibroblasts may not always originate from fibroblasts but can also derive from adipocytes, perivascular pericytes, and the trans-differentiation of epithelial and vascular endothelial cells (116,120).

In SSc, an altered innate immune response is observed (121). For example, monocytes that are key players of the innate immune response are known to be deregulated in SSc. Monocytes can be classified into distinct 3 subgroups as they circulate in the bloodstream: classical monocytes (cMo) are those which transmigrate into the tissues across endothelium, responding to different signals; intermediate monocytes (iMo) have a longer halflife in blood than cMo and expressed APC phenotype; and non-classical monocytes (ncMo), also known as patrolling monocytes, they displaying pro-inflammatory and profibrotic properties (122). Higher proportions of ncMo are observed in SSc patients, particularly in severe cases with lung complications and IFN signaling sensitivity (123–125). These ncMo also display abnormal biophysical and pro-inflammatory properties in SSc patients (126).

Macrophages, derived from monocytes, play a major role in the affected tissue of patients with SSc. Initially, increased proportions of

profibrotic M2 macrophages were observed, but later studies revealed profiles with intermediate characteristics between proinflammatory M1 and profibrotic M2 macrophages. Notably, early-stage SSc presents a proinflammatory M1 profile, which transitions into an M1/M2 equilibrium in later stages (127,128). Additionally, a new subset of *CXCL4*-induced macrophages (M4) has been identified, although their specific role in the disease is unclear (129). Furthermore, a monocyte-derived dendritic cell population exclusive to affected SSc patient skin has been associated with a more severe skin disease (130).

The adaptive immune response in SSc includes altered behavior and activation of different T and B lymphocyte subsets. This involves a Th1/Th2 imbalance skewed towards Th2, while SSc is linked to proinflammatory Th17 cell activation and variations in their populations, including IFNγ+IL-17+Th17 in peripheral blood (131–133). Regulatory T cells (Tregs), which are essential to limit the immune response, are also a subject of extensive study in SSc, with investigations into compromised immunosuppressive activity (133).

T cell involvement extends to the skin, where Th22 subsets are elevated in affected tissues (131). Moreover, a specialized subset of T cells, T helper follicular cells expressing CXCL13+, actively contributes to B cell activation within SSc-affected tissue (134). B cells also participate in profibrotic crosstalk by secreting antibodies and cytokines and are detected in lesion sites in lungs and endothelial tissue, with differences in B cell subtype populations observed (135,136).

In summary, as is shown in **Figure 5**, the cellular involvement in SSc in a complex interplay between a large number of cells and several systems within the body that need to be accounted for.

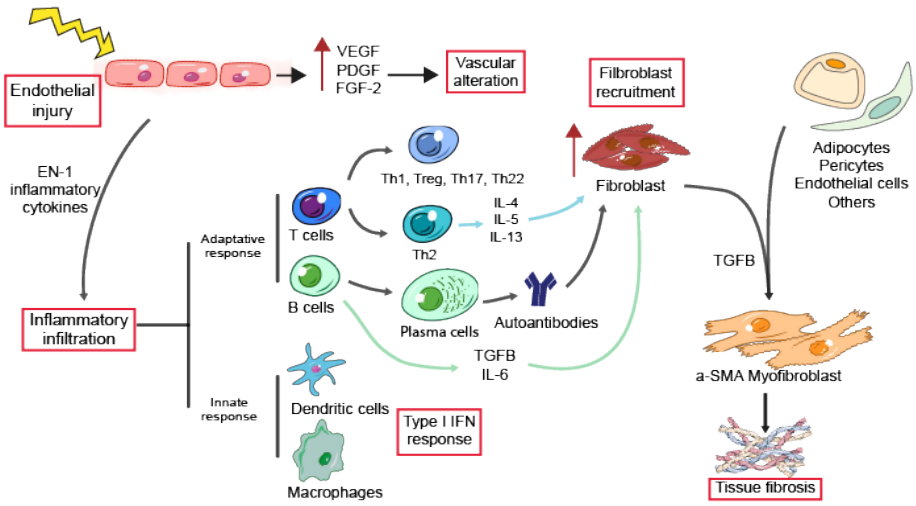


Figure 5. Cells and molecules implicated in the systemic sclerosis pathogenesis. (Adapted from Campochiaro and Allanore, *Arthritis Res Ther*, 2021)

7. Transcriptomic strategies to address the cell specific context in SSc

Gene expression is different in each cell or cluster of cells, and it is affected by numerous factors, within and outside the cells. Gene expression studies range from the analysis of the transcriptome of the tissues or selected cell subtypes, RNA sequencing (RNA-seq), to the analysis of individual cells, single cell RNA-seq (scRNA-seq).

The study of expression patterns in different tissues has been an essential line of research in SSc. Skin tissue is one of the most studied tissues in SSc due its importance in the disease. In this regard, studies in skin tissue in SSc revealed distinct gene expression patterns in the affected zones (137), associating a group of TGFβ-responsive genes to disease severity. Even more, the skin and peripheral blood mononuclear cells (PBMCs) of lcSSc patients with PAH showed gene markers associated with vascular damage

and inflammation (138). Non-affected skin tissue in patients is also a key point in the pathology and it is variable between patients (139). The gene expression profile of unaffected skin from lcSSc patients was indistinguishable from healthy skin and showed signs of immunity, inflammation, fibrosis and extracellular matrix turnover, among others (139).

Peripheral blood serves as a highly valuable source of information in disease, as numerous immune cells circulate it for migration and communication, and it contains several molecules that can be implemented as biomarkers. In terms of gene expression, a distinct inflammatory profile has been observed in lcSSc with PAH (140), along with an upregulation of endoplasmic reticulum stress and the unfolded protein response (141). Most recently, a large-scale RNA-seq study, in which our group participated, involved 162 SSc patients divided into a discovery and a replication cohort (142). Pathway analysis unveiled dysregulation in type I IFN, Toll-like receptor cascades, p53 suppressor activity, degranulation, and platelet activation. Remarkably, neutrophils were identified as the primary contributors to the expression of these genes (142). Another peripheral blood RNA-seq study, led by our team, integrated transcriptomic data with GWAS genotypes to analyze expression quantitative trait loci (eQTL) (143). The study revealed 49,123 validated cis-eQTLs from 4,539 SSc-associated SNPs from the GWAS, at p value < 10^{-5}. Among the 1,436 genes located within 1 Mb of these SNPs, 565 were found to have one or more eQTLs associated with SSc-related SNPs. Then, these SNPs were prioritized resulting in the identification of 233 candidates, of which 134 associated with key features of SSc and 105 exhibited differential expression in the blood cells, skin, or lung tissue of SSc patients. Transcription factor analysis identified enriched motifs for 24 transcription factors among SSc eQTLs, with five showing differential regulation in various tissues of SSc patients. Moreover, we

identified ten candidate genes that are amenable to targeting by approved medications for immune-mediated diseases. However, it's important to note that only three of these genes have undergone clinical trials in SSc patients (143). Another study, led by our team, integrated RNA-seq expression data with Capture Hi-C (high-throughput chromosome conformation capture) data for 8 SSc patients and 8 controls, focusing on CD14+ monocytes and CD4+ T lymphocytes (144). Seven known SSc signals in these cell types were confirmed (*NFKB1*, *CD247*, *STAT4, IRF8, DDX6, CSK, IKZF3*), along with the proposal of new candidate genes such as *CXCR5*, a gene involved in T helper cell differentiation into follicular Th cells (144). Interestingly, no differences in chromatin interactions were observed between patients and controls in any cell type. Another noteworthy finding was that at the *CD247* and *STAT4* loci, genes highly associated with the disease, chromatin interactions were exclusively observed in CD4+ cells (144).

In recent years, there has been a deeper exploration of the individual roles of cells within tissues in SSc. Within the skin, a gene expression microarray study described the role of M2 Macrophages and their distinctive pathways, involving differentiation, IFN activation, and tissue remodeling through extracellular matrix deposition in SSc patients (145).

Gene expression studies have witnessed an unprecedented advance in recent years with the advent of scRNA-seq, enabling the comprehensive analysis of transcriptomes at the single-cell level (146,147). In brief, scRNA-seq involves isolating and lysing individual cells from a dissociated tissue, followed by reverse transcription of the individual cell transcriptome to generate a cDNA library that is subsequently sequenced (146,148) (**Figure 6**). This technique continually evolves, encompassing molecular advancements and data analysis techniques, bridging emerging technologies with existing ones, such as integrating with data from GWAS or Assay for

Transposase-Accessible Chromatin using sequencing (ATAC-seq) (149–151).

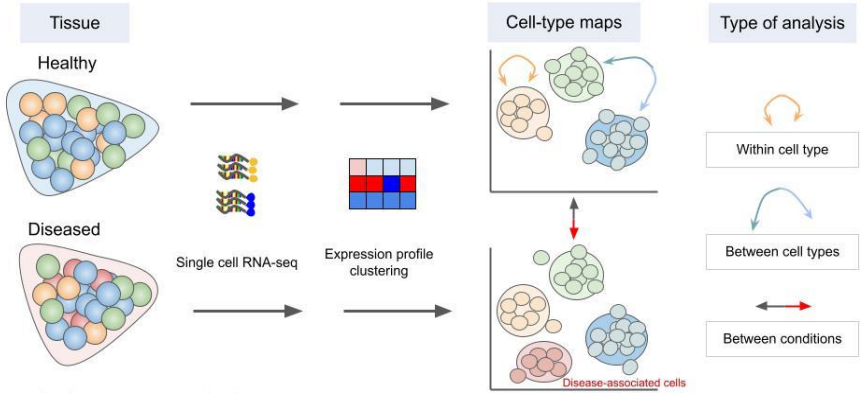


Figure 6. Single cell RNA-seq study design. (Adapted from Sandberg *et al*. Nat Methods. 2014)

The exploration of immune cells through scRNA-seq has expanded extensively in IMIDs (152–157), including SSc (134,158–162). As in previous gene expression studies, the skin remains a focal point of investigation. The application of scRNA-seq technologies lead to the identification of a distinctive SSc myofibroblast population, originating from *SFRP2*^{hi}/*DPP4*-expressing progenitor fibroblasts and regulated by a group of transcription factors, including *IRF7* (159). On the other hand, a large-scale scRNA-seq and scATAC-seq study (160), comprising 145 SSc patients and approximately 65,000 cells from PBMCs and skin biopsies, unveiled another fibroblast subset characterized by the expression of *LGR5* (160). A recent study focusing on skin and the serological subtypes suggested varying cell-autonomous interactions with autoantibodies in early and late-stage SSc (162). Intriguingly, it also observed distinct TGFβ responses in fibroblasts and smooth muscle cells in early ATA+ dcSSc skin samples, while in early ARA+ dcSSc, these responses occurred in the endothelial layer.

The role of myofibroblasts in the lungs of SSc patients with ILD has also been explored using scRNA-seq (158). The study found that SSc-ILD myofibroblasts in the lung exhibit significant phenotypic changes compared to controls, marked by increased expression of collagen and profibrotic factors (158). Moreover, in the lungs of SSc-ILD patients, scRNA-seq identified a distinct profibrotic macrophage population (161). This study also conducted a ATAC-seq, revealing that these profibrotic macrophages are characterized by altered chromatin accessibility of Secreted Phosphoprotein 1 (*SPP1*), and the involvement of numerous transcription factors such as *KLF6* (161).

Finally, scRNA-seq has been employed to investigate immune cells in SSc in three different studies until now (130,134,160). In skin tissue, Xue *et al.* (130) described a cluster of macrophages expressing *FCGR3A* associated with SSc. They also identified two subsets of myeloid cells: one consisting of monocyte-derived dendritic cells (DC) expressing *FCN-1* and *S100A8* and *S100A9*, and the other, a cluster of plasmacytoid DCs that was almost exclusive for dcSSc patients (130). Another scRNA-seq study in PBMCs revealed changes in immune cell composition associated with specific disease subtypes, disease duration, severity, and autoantibody status (160). Specifically, there was an increase in the ncMo population in SSc and DCs, the latter only in dcSSc, along with a decrease in naïve T cells (160). The most recent scRNA-seq study in SSc, which is part of this dissertation, was focused on the deep characterization of a specific cell subtype in peripheral blood *i.e.* CD14+ monocytes, and it will be discussed in detail in Chapter 3.

Objectives and justification

Recent studies have expanded our understanding of SSc; however, many aspects of the etiology and pathogenesis of the disease remain unknown. Therefore, our hypothesis is that identifying molecular and genetic traits at different levels in SSc patients will lead to improved treatment and, consequently, better prognosis for them. To identify these molecular traits, studies must be conducted at various levels: at the genomic level through the generation of a polygenic risk score to provide a genomic profile of the patients and Mendelian randomization studies to identify new environmental risk factors for the disease. At the transcriptomic level, this involves analyzing key cells in inflammation, thus understanding the mechanisms underlying its pathogenesis.

In this regard, we will contribute to the overall knowledge of SSc. The aim is to enhance clinical management, making it easier to treat and monitor these patients. Additionally, we seek to discover new therapeutic targets by identifying clinically relevant points within the genetic and metabolic framework of the cellular processes in this disease that may be prone to treatment with existing drugs.

Objectives

1. Generation of a genomic risk score (GRS) for SSc.
    a. Stratify and compare the subtypes of SSc, limited (lcSSc), and diffuse (dcSSc).
    b. Compare the predictive value of this score between SSc with other immune-mediated diseases.

2. Identify new environmental risk factors for SSc through a 2SMR study.
    a. Analyze the causal role of anthropometric traits associated with obesity in the development of SSc.

3. Identify specific cellular populations for SSc within the relevant cells for the disease using single-cell RNA sequencing (scRNA-seq).
    a. Identify specific cellular subpopulations for SSc in peripheral blood CD14+ monocytes.
    b. Characterize disease-specific gene expression profiles in different subtypes of CD14+ monocytes.
    c. Define differences in the transcriptome of CD14+ monocytes in patients with lcSSc and dcSSc.

Publications

Chapter 1. Genomic risk score impact on susceptibility to systemic sclerosis

Genomic risk score impact on susceptibility to systemic sclerosis

Lara Bossini-Castillo[1*], Gonzalo Villanueva-Martin[2*], Martin Kerick[2], Marialbert Acosta-Herrera[2], Elena López-Isac[2], Carmen Pilar Simeon[3], Norberto Ortego-Centeno[4], Shervin Assasi[5], International SSc Group[†], Australian Scleroderma Interest Group (ASIG)[¤], PRECISESADS Consortium[#], Flow Cytometry PRECISESADS study group[ᵏ], Nicolas Huzelmann[6], Armando Gabrielli[7], Jeska de Vries-Bouwstra[8], Yannick Allanore[9], Carmen Fonseca[10], Christopher P. Denton[10], Timothy Radstake[11], Marta E Alarcón-Riquelme[12], Lorenzo Beretta[13], Maureen Mayes[5], Javier Martín[2].

*These authors contributed equally.

[#¤†ᵏA] complete list of authors is provided in the supplementary material

Corresponding authors: Lara Bossini-Castillo PhD lbossinicastillo@ugr.es

Prof. Javier Martín javiermartin@ipb.csic.es

1.- Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Centro de Investigación Biomédica (CIBM). Granada. Spain.

2.- Institute of Parasitology and Biomedicine López-Neyra, IPBLN. Consejo Superior de Investigaciones Científicas (CSIC). Granada. Spain.

3.- Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain.

4.- Department of Internal Medicine, Clinic University Hospital, Granada, Spain.

5.- Division of Rheumatology and Clinical Immunogenetics, The University of Texas Health Science Center-Houston, TX, USA

6.- Department of Dermatology, University of Cologne, Cologne, Germany.

7.- Istituto di Clinica Medica Generale, Ematologia ed Immunologia Clinica. Università Politecnica delle Marche, Ancona, Italy.

8.- Service of Rheumatology, Leiden University Medical Center, Leiden, The Netherlands.

9.- Department of Rheumatology A, Cochin Hospital, INSERM U1016, Paris Descartes University, Paris, France.

10.- Centre for Rheumatology, Royal Free and University College Medical School, London, United Kingdom

11.- Department of Rheumatology & Clinical Immunology, Laboratory of Translational Immunology, department of Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.

12.- GENYO, Centre for Genomics and Oncological Research Pfizer, University of Granada, Andalusian Regional Government, PTS GRANADA, Granada, Spain

13.- Scleroderma Unit, Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy.

ABSTRACT

Objectives

Genomic Risk Scores (GRS) successfully demonstrated the ability of genetics to identify those individuals at high risk for complex traits including immune-mediated inflammatory diseases (IMIDs). We aimed to test the performance of GRS in the prediction of risk for systemic sclerosis (SSc) for the first time.

Methods

Allelic effects were obtained from the largest SSc GWAS to date (9,095 SSc and 17,584 healthy controls with European ancestry). The best-fitting GRS was identified under the additive model in an independent cohort that comprised 400 patients with SSc and 571 controls. Additionally, GRS for clinical subtypes (limited cutaneous SSc, and diffuse cutaneous SSc) and serological subtypes (anti-topoisomerase positive ATA+, and anti-centromere positive ACA+) were generated. We combined the estimated GRS with demographic and immunological parameters in a multivariate generalized linear model.

Results

The best-fitting SSc GRS included 33 SNPs and discriminated between patients with SSc and controls (AUC = 0.673). Moreover, the GRS differentiated between SSc and other IMIDs, such as rheumatoid arthritis and Sjögren syndrome. Finally, the combination of GRS with age and immune cell counts significantly increased the performance of the model (AUC = 0.787). While the SSc GRS was not able to discriminate between ATA+ and ACA+ patients (AUC < 0.5), the serological subtype GRS, which was based on the

allelic effects observed for the comparison between ACA+ and ATA+ patients, reached a AUC = 0.693.

Conclusions

GRS was successfully implemented in SSc. The model discriminated between patients with SSc and controls or other IMIDs, confirming the potential of GRS to support early and differential diagnosis for SSc.

Keywords

Systemic sclerosis, genomic risk score, genome-wide association study, immune mediated-inflammatory diseases.

Introduction

Complex diseases are a devastating consequence of usually unknown environmental factors and the combined effects of tens to thousands of genetic variants that are spread throughout the genome [1]. The advanced use of bioinformatic tools will provide a better understanding of the intricate network of multiple genetic effects that shapes the architecture of complex diseases [2].

Immune-mediated inflammatory diseases (IMIDs) comprise a variety of complex diseases characterized by the loss of self-tolerance, the maintenance of chronic inflammation and an aberrant immune response [3]. Genome-wide association studies (GWAS) have largely increased our understanding of the aetiology of complex diseases, providing new data about the genome and lighting the way to the identification of genes and pathways that contribute to disease susceptibility and prognosis. Many susceptibility loci have been discovered for IMIDs, and several are shared between diseases, adding a common genetic background to their overlapping clinical and immunological characteristics [4]. Additionally, GWAS findings have also confirmed that the contribution of each associated locus to disease risk is often small and has low predictive value [1].

To address complex disease susceptibility, three main components must be considered: genetics, environmental exposures, and lifestyle factors [1,4]. As for genetics, large cohorts have been genotyped in GWAS efforts, and hundreds of genetic risk factors have been identified [5]. However, GWAS data can be examined in various ways, moving forward to a more precise genetic profiling, its use for personalized medicine, and the identification of individuals with higher risk of displaying a specific phenotype [6]. Genomic risk scores (GRS) take into account disease heritability and the additive effect

of genetic polymorphisms, and they provide a disease risk score per individual to evaluate their relative risk to suffer a disease [7–9].

GRS are calculated essentially by combining the weighted effects of the risk alleles for each individual; these weighted effects are assigned depending on the strength of the association to the risk of disease — the effect size [7,10]. The identification of individuals with high risk or those prone to developing more aggressive phenotypes is a useful tool for personalized medicine and clinical management of patients. GRS have been successful in several diseases such as schizophrenia [11] and obesity [12]. This strategy had a great impact on cardiovascular diseases like coronary artery disease (CAD) [12–14], but also in IMIDs such as sarcoidosis [15], systemic lupus erythematosus (SLE) [16,17], and vitiligo [18] recently.

Systemic sclerosis (SSc) or scleroderma is a complex chronic autoimmune disease. It belongs to the group of IMIDs and it has one of the highest mortality rates among them [19]. SSc affects the connective tissue and shows complex and varied clinical manifestations. Raynaud's phenomenon and gastro-oesophageal reflux are two common onset symptoms, but they are not exclusive to SSc. Conversely, the disease can manifest in different ways, such as affectation of the skin —inflammatory skin disease, extensive fibrosis—, musculoskeletal inflammation, and vascular damage [20–22]. Furthermore, SSc also shows organ-specific manifestations, such as lung fibrosis, pulmonary arterial hypertension, renal failure and gastrointestinal complications. Notably, the involvement of the lungs, with pulmonary hypertension and/or pulmonary fibrosis, is the leading cause of death in SSc [19].

Patients with SSc can be classified into different subgroups according to clinical outcome: limited cutaneous scleroderma (lcSSc) or diffuse

cutaneous scleroderma (dcSSc), depending on how widespread fibrosis is [23]. On other hand, they can also be classified depending on their serological status, considering the presence of the mutually exclusive anti-centromere or anti-topoisomerase auto-antibodies (i.e. ACA+ or ATA+) [22,23].

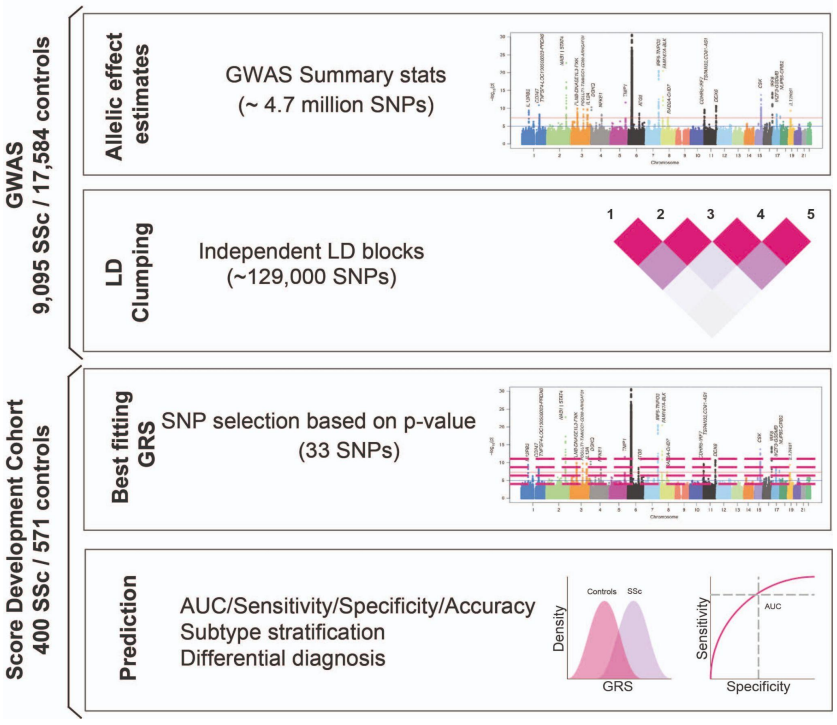


Figure 1 Overview of the study design. AUC, area under the receiver operating characteristic (ROC) curve; GRS, Genomic Risk Scores; GWAS, genome-wide association studies;SNPs, single nucleotide polymorphisms.

Since the first SSc GWAS in European populations was carried out ten years ago [24], our recently published meta-GWAS is the largest effort to decipher the genetic component of SSc [25]. In addition to the extensively known association of the human leukocyte antigen (HLA) region with the disease, twenty-seven non-HLA GWAS level associations and three suggestive loci were identified [25].

Considering the heterogeneity and variable prognosis of patients with SSc, GRS could be a powerful tool in clinical diagnosis to identify patients in the early stages of the disease and to differentiate them from patients with confounding diseases. By taking advantage of the summary statistics of this large meta-GWAS, we generated an accurate SSc GRS through the use of an independent and unique dataset comprising patients with SSc and with other IMIDs [3] (Figure 1). We generated subtype-specific GRS for the clinical and serological SSc subgroups of patients, and we tested the clinical implications of GRS in SSc. Finally, the GRS was complemented with additional demographic and immunological information.

Methods

GRS calculation

GRS was developed as implemented in PRSice-2 [26], using summary statistics and assuming an additive effect for the effective allele. Briefly, PRSice-2 calculated the product of the number of effect alleles per individual and the respective SNP weights. The score was averaged by the number of alleles included in the GRS per individual (--score avg). We used the minor allele frequency in the PRECISESADS cohort as the genotype for the samples with missing genotype. We applied a 10,000 permutation procedure to calculate the empirical P-value (--perm 10000).

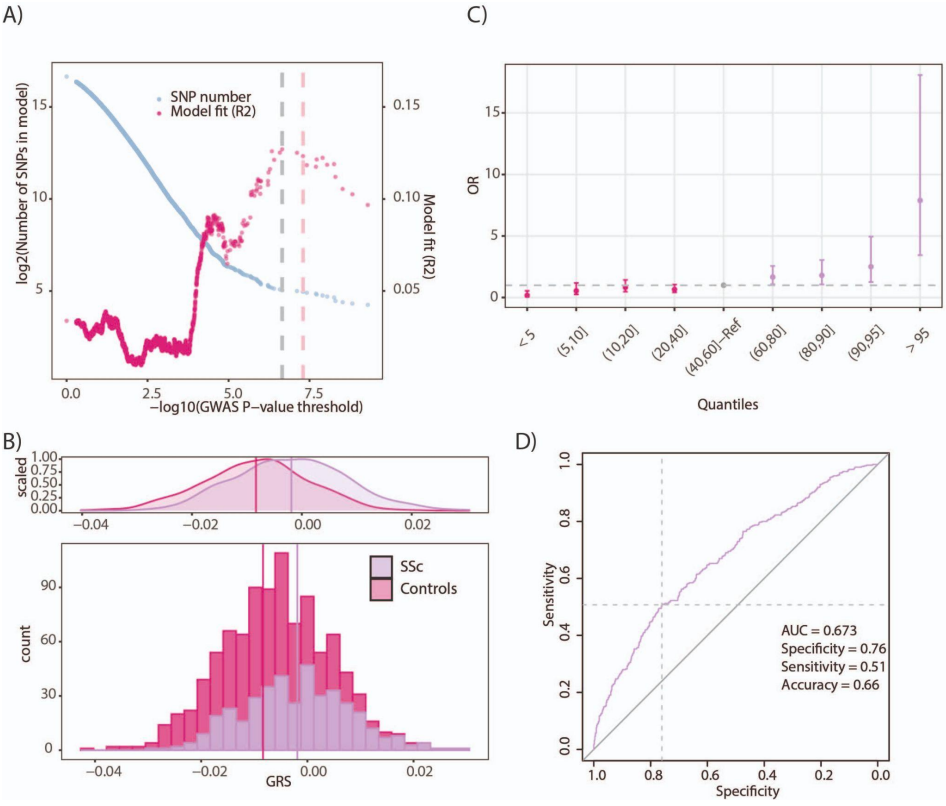


Figure 2 Systemic sclerosis Genomic Risk Scores (SSc GRS). (A) Identification of the best-fitting GRS in the score development cohort. Tested p value thresholds for the SNPs included in the GWAS summary statistics are presented in the x-axis. The number of SNPs included in the models corresponding to each p value threshold is shown on the left y-axis. Model fit (R2) is represented in the right y-axis. (B) Distribution of GRS for patients with SSc and healthy controls in the score development cohort. (C) Relative risk for individuals in different quantiles of the GRS distribution. (D) receiver operating characteristic (ROC) curve for the 33 SNP SSc GRS. AUC, area under the ROC curve; GWAS, genome-wideassociation studies; SNPs, single nucleotide polymorphisms.

PRSice-2 allowed us to fit different GRS models by selecting only the variants that passed a number of different p-value thresholds in the GWAS summary statistics (--bar-levels 5e-11, 5e-10, 5e-09, 5e-08, 5e-07, 5e-06, 5e-05, 0.0001, 0.001, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 1, but GRS calculated at all intermediate p-value thresholds, high resolution parameters, were calculated) using sex (female/male) as covariate. Therefore, the model fit is defined as: R2 of the full model (SSc case or control ~ GRS + Sex) - R2 of the null model (SSc case or control ~ Sex).

Multivariate model

In order to test if a combination of GRS with demographic factors and the counts of immune cell subpopulations in peripheral blood would improve the predictive value of our model, we divided our score development cohort into an initial set, comprising the non-Spanish individuals in the PRECISESADS study (n = 518), in which we developed a multivariate model and a testing set that comprised all the Spanish individuals in this study (n = 339).

First, we built several generalized linear models that included GRS and each demographic and immune parameter in Supplementary Table 1

individually, then we compared them to the null model that included only GRS and sex as covariates. Improvement over the null model was defined by a LRT (p-value < 0.05).

Secondly, we generated a multivariate model that included the 13 phenotypic variables that had been identified as informative in the previous step. Using leave-one-out prediction (i.e. including all variables but one in the model) and comparing to the full model, we calculated the contribution of all variables to the multivariate model. This model was applied to the testing set of individuals.

Details about the cohorts, LD clumping, GRS additive model, the model fitting analyses and the effects of including country of origin as covariates are shown in the Supplementary Methods section.

Results

- ***A 33-variant GRS discriminates between patients with SSc and controls***

We calculated GRS in an independent score development cohort comprising 400 patients with SSc and 571 healthy controls [27]. We observed that the best-fitting GRS (GRS R2 = 0.13; p-value = 1.27×10^{-17}; permutation p-value = 9.99×10^{-5}) included 33 independent SNPs that had a p-value < 2.215×10^{-7} (Figure 2A). Sex, which was included as a covariate, contributed very modestly to the explained variance (R2 = 0.01).

As expected, the SSc cases and controls showed significantly different GRS distributions (Figure 2B, control group mean = -8.35×10^{-3} and SSc group mean = -1.91×10^{-3}, t-test p-value < 2.2×10^{-16}). Reassuringly, individuals with GRS in the 95th percentile showed a 5-fold higher relative risk (OR = 7.89, 95%; CI = 3.44 - 18.08) than the reference quantile (40th-60th percentiles) (Figure 2C).

Reassuringly, the 33 variant GRS had a 67% chance of accurately predicting if an individual was a patient with SSc or an unaffected control (AUC = 0.673, 95% CI : 0.64-0.71, p-value = 3.90×10^{-23}, Figure 2D). We determined a best-fitting GRS threshold (GRS controls < -1.86×10^{-3} < GRS cases, details in Supplementary Methods) and reached a moderate discrimination between cases and controls (specificity = 0.76; sensitivity = 0.51; accuracy = 0.66, Figure 2D).

We observed that if the ROC curves were calculated separately for each country of origin, the AUC determined by the 33 variant GRS ranged from 0.60 to 0.75 (Supplementary Figure 2A). However, variability of the AUC did not correlate with either country longitude, latitude or distance to 1000 Genomes GBR and CEU populations (see Methods, Supplementary Figure 2B-

D).

- ***Subtype stratified SSc GWAS summary stats discriminate between clinical and serological subtypes***

The 33 variant GRS previously described distinguished between patients with SSc and healthy controls. However, SSc is a heterogeneous disease with both clinical and serological subtypes that influence the prognosis of the disease, and the prediction of these subtypes is a major clinical demand. The 33 SNP SSc GRS showed no predictive value for clinical subtypes (dcSSc vs. lcSSc AUC = 0.496, 95% CI: 0.40-0.59, p-value = 0.93**, Supplementary Figure 3**) and serological subtypes (ATA+ vs. ACA+ AUC = 0.464, 95% CI: 0.37-0.56, p-value = 0.45, **Supplementary Figure 3**). Furthermore, this SSc GRS was not able to predict the development of pulmonary fibrosis in patients with SSc (SSc with pulmonary fibrosis vs. SSc without pulmonary fibrosis AUC = 0.479, 95% CI: 0.38-0.57, p-value = 0.66, **Supplementary Figure 3**).

Therefore, we used the allelic effects obtained in the GWAS comparison between dcSSc and lcSSc and between ATA+ and ACA+ patients to build subtype specific GRS. The best-fitting GRS p-value threshold for the variants in the dcSSc versus lcSSc comparison, clinical subtype GRS, comprised up to 9,780 SNPs (SNP p-value threshold for the best- fitting dcSSc vs lcSSc GRS < 9.99x10-2, **Figure 3A**). This clinical subtype GRS was not limited to highly significant variants but it also included thousands of additional SNPs with very low significance. The GRS for the variants in the ATA+ versus ACA+ comparison, serological subtype GRS, required up to 35,058 SNPs (SNP p-value threshold for the best- fitting ATA+ vs ACA+ GRS < $3.48x10^{-1}$, **Figure 3A**). The clinical subtype GRS did not explain much of the phenotypic variance between dcSSc and lcSSc (R^2 = 0.053), while the

explained variance between them using the serological subtype GRS was comparable to the SSc GRS (R^2 = 0.115).

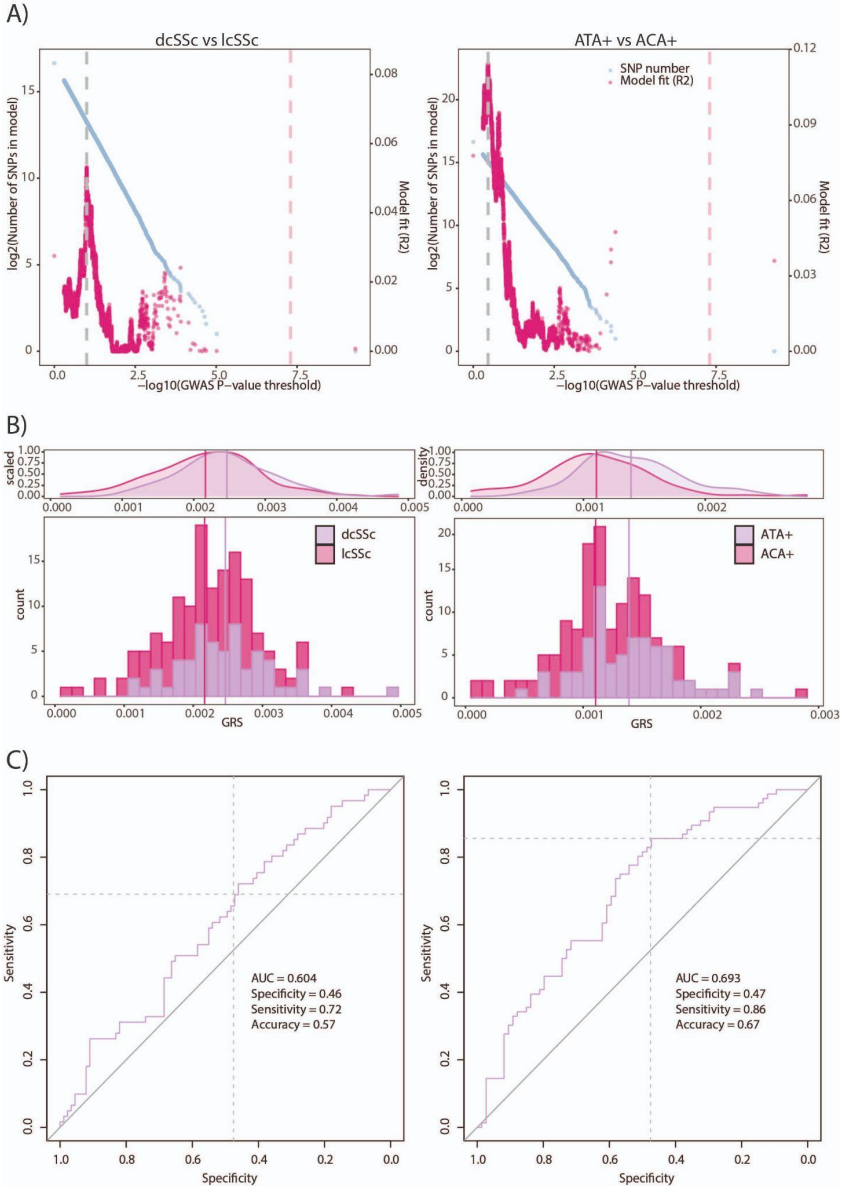


Figure 3 Characteristics of clinical subtype-specific Genomic Risk Scores (GRS) (left) and serological subtype-specific GRS (right). (A) Identification of the best-fitting GRS in the score development cohort. Tested p value thresholds for the SNPs

included in the GWAS summary statistics are presented in the x-axis. The number of SNPs included in the models corresponding to each p value threshold is shown on the left y-axis. Model fit (R2) is represented in the right y-axis. (B) Distribution of GRS for patients with systemic sclerosis (SSc) in each subtype group. (C) Receiver operating characteristic (ROC) curves for the 9 780 SNP clinical subtype-specific GRS and 35058 SNP serological subtype-specific GRS. AUC, area under the ROC curve; SNPs, single nucleotide polymorphisms.

In this context, the subtype-specific GRS distributions (mean dcSSc GRS = $2.46x10^{-3}$; mean lcSSc GRS = $2.16x10^{-3}$; t-test p-value = $1.21x10^{-2}$, **Figure 3B**), and AUC based on the clinical subtype GRS led to a modest classification of the patients into the dcSSc or lcSSc groups (AUC = 0.604, 95% CI: 0.51-0.70, p-value = $2.59x10^{-2}$, **Figure 3C**). However, the serological subtype GRS (comprising 35,058 SNPs) showed more distinctive GRS distributions between ATA+ and ACA+ patients (mean ATA+ GRS = $1.39x10^{-3}$ and mean ACA+ GRS = $1.11x10^{-3}$, t-test p-value = $1.12x10^{-4}$, **Figure 3B**), and best classification results for the ATA+ or ACA+ subgroups of patients (AUC = 0.693, 95% CI: 0.61-0.78, p-value = $7.58x10^{-6}$, **Figure 3C**).

Considering the clinical relevance of pulmonary fibrosis for the prognosis of patients with SSc, we tested the predictive value of both the clinical and the serological GRS on the development of lung fibrosis. Interestingly, we observed that the serological GRS was marginally able to discriminate between patients with and without lung fibrosis but the model did not reach statistical significance (AUC = 0.575, 95% CI: 0.48-0.67, p-value = 0.11, **Supplementary Figure 3**).

- ***GRS separates SSc from other IMIDs***

Considering the shared genetic component of IMIDs, the implementation of the proposed GRS might help to identify high risk

individuals not only for SSc but also for other immune-related traits. Regarding the accuracy of the 33 variant SSc GRS in other IMIDs, we observed that the SSc GRS was able to separate patients with rheumatoid arthritis (RA) (RA group mean = -4.46×10^{-3}; t-test p-value < 2.8×10^{-9}), Sjögren's syndrome SJS (SJS) group mean = -1.78×10^{-3}; t-test p-value < 3.54×10^{-6}) and SLE (SLE group mean = -3.67×10^{-3}; t-test p-value < 8.51×10^{-13}) from the non-affected individuals. However, as expected, the GRS differences between patients with RA, SJS and SLE and controls were less significant than between SSc cases and controls (**Figure 4A**). Furthermore, using the SSc GRS in these three additional IMIDs, the AUCs showed a modest predictive value (AUC RA = 0.608, 95% CI: 0.57-0.64, p-value = 6.58×10^{-9}; SJS = 0.590, 95% CI: 0.55-0.63, p-value = 1.58×10^{-6}; AUC SLE =0.623, 95% CI: 0.59-0.66, p-value = 3.94×10^{-12}, **Figure 4B**).

A key point towards GRS being implemented from bench-to-bedside is not only the ability to identify individuals at high risk of developing SSc in the general population, but also to help in the differential diagnosis between SSc and other IMIDs. In the pursuit of this objective, we tested the effectiveness of our SSc GRS to correctly classify between patients with SSc and those affected by other IMIDs. We report statistical differences between the GRS distributions for SSc and RA (t-test p-value < 3.78x10-4) or SJS (t-test p-value < 3.70x10-6), but only nominally significant differences in the case of SLE (t-test p-value < 1.37x10-2) (Figure 4A). These results were aligned with the predictive capacity of the GRS in the separation between patients with SSc and other IMIDs. The greatest AUC was observed for the classification of patients with SSc versus patients with SJS (SJS AUC = 0.585, 95% CI: 0.55-0.62, p-value = 2.22x10-5), and decreased in more closely related IMIDs, such as RA (AUC RA = 0.568, 95% CI: 0.53-0.61, p-value = 8.84x10-4) and, especially, SLE (SLE AUC = 0.553, 95% CI: 0.51-0.59, p-value = 1.19x10-2) (Figure 4C).

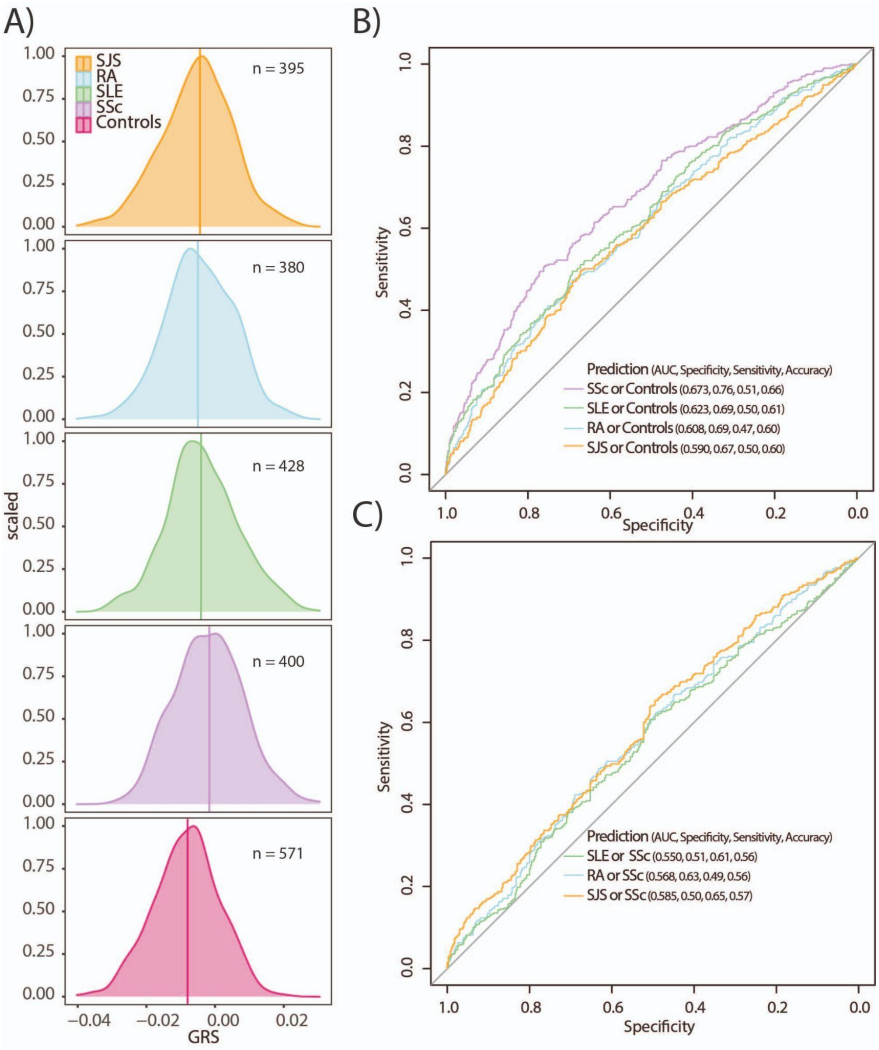


Figure 4 Impact of the 33 SNP systemic sclerosis (SSc) Genomic Risk Scores (GRS) on the differential classification with other immune-mediated inflammatory diseases IMIDs). (A) Distribution of GRS for healthy controls and patients with SSc, systemic lupus erythematosus (SLE), rheumatoid arthritis (RA) and Sjögren syndrome (SJS). (B) Receiver operating characteristic (ROC) curves for the predictive value of the SSc GRS to distinguish between patients with SSc, SLE, RA or SJS and healthy controls. (C) ROC curves for the predictive value of the SSc GRS to distinguish between patients with SLE, RA or SJS and patients with SSc. AUC, area under the ROC curves.

- ***Age and immune cell counts improve the prediction accuracy***

The score development cohort recruited in the PRECISESADS study was comprehensively phenotyped and allowed us to complement our GRS with additional demographic (age, sex) and immunological (immune cell counts in peripheral blood estimated using a large flow cytometry panel) parameters (163) (**Supplementary Table 1**). We divided our score development cohort into an initial set (n = 518), and a testing subgroup (n = 339). The initial set allowed us to test the relevance of the different parameters in a combined GRS and phenotypic model. On the other hand, the testing set confirmed these findings.

First, we identified the demographic and immunological parameters which improved the GRS model (LRT p-value < 0.05) (**Supplementary Table 2**). Twelve immune cell subtypes in peripheral blood showed a significant contribution to the model, but the most significant contribution among the phenotypic variables corresponded to age (LRT p-value = 3.47×10^{-20}, **Supplementary Table 2**).

When we combined only the informative variables into the same model, multivariate GLM, in addition to GRS and age, only 4 out of the 12 immune cell types remained as independently associated in the multivariate model: resting NK cells, M0 macrophages, activated dendritic cells and memory B cells (**Supplementary Table 3**). The contribution of sex to the model did not remain significant when considering all the independent variables together and GRS score distributions between male and female patients did not show significant information (t-test p-value = 0.24, **Supplementary Table 3**). Using leave-one-out prediction, we identified age as the most informative variable, followed by GRS (**Supplementary Table 4**). We observed that the contribution of GRS to the model was comparable to the

contribution of all significant parameters of immune cell count together (GRS LRT p-value = $2.59x10^{-12}$; GRS LRT p-value = $1.26x10^{-12}$, **Supplementary Table 4**).

The multivariate GLM described above (SSc status ~ GRS+Age+Memory B cells+Resting NK cells+M0 Macrophagues+Activated dendritic cells) greatly outperformed the GRS and sex only model both in the initial (AUC discovery = 0.847, 95% CI: 0.81-0.88, p-value = $1.10x10^{-90}$) and in the testing set (AUC = 0.787, 95% CI: 0.73-0.84, p-value = $1.31x10^{-24}$), as illustrated in **Figure 5**. Moreover, the multivariate GLM outperformed the models that did not include age, GRS or both (**Figure 5**).

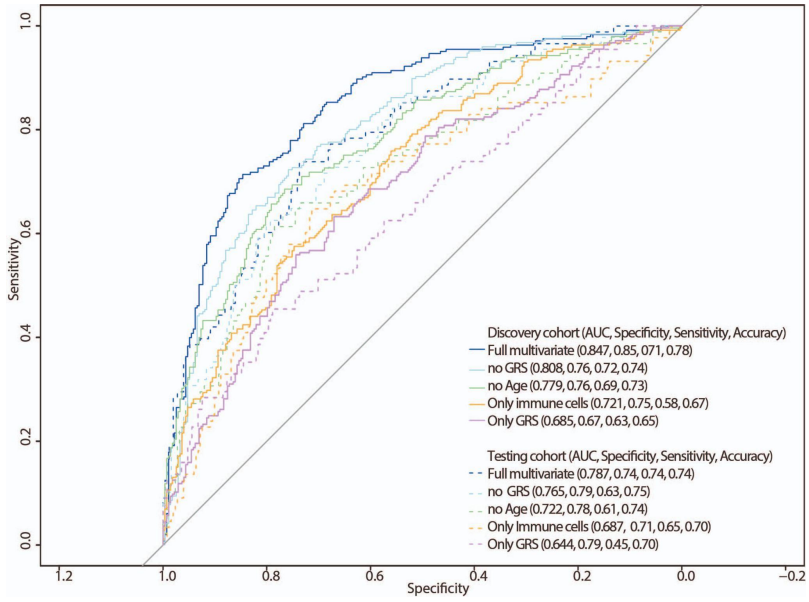


Figure 5 Receiver operating characteristic (ROC) curves for the predictive value of the multivariate generalised linear model (GLM), (SSc status ~GRS+Age+Memory B cells+Resting NK cells+M0 Macrophages+Activated dendritic cells) to distinguish between patients with SSc and healthy controls in the initial and replication cohorts depending on the variables included in the models. GRS, Genomic Risk Scores; NK cells, natural killer cells; SSc, systemic sclerosis.

Discussion

We generated a GRS based on the allelic effects identified in the largest GWAS in SSc to date (65). We obtained a predictive GRS model comprising 33 genetic polymorphisms, which allowed us to differentiate between SSc and controls in an independent SSc patient cohort (164). A serological subtype-specific GRS (based on the GWAS comparison between ATA+ and ACA+ SSc patients) showed the best predictive value to classify patients based on the presence of different autoantibodies. Furthermore, we demonstrated the accuracy of the model in the differentiation between SSc and other IMIDs, such as RA and SJS. Finally, we complemented the SSc GRS with demographic data and peripheral blood immune cell counts in a multivariate model which reached a very significant recall rate.

The reported SSc GRS showed good predictive value (AUC = 0.673), in line with the GRS developed for other IMIDs. For example, a similar AUC was reported for inflammatory bowel disease with a GRS based on the allelic effects observed for 12,882 cases and 132,532 healthy controls (AUC=0.72) (165) and in SLE (AUCs ranging 0.62 - 0.78 (166,167). Moreover, Stahl *et al*. (168) implemented a Bayesian inference model in a GWAS that comprised 5,485 cases of RA and 22,609 healthy controls, and the model explained 18% of the total variance, which is comparable to the variance explained by our model (R^2 = 0.13). We would like to note that the previously conducted GWAS comprised 9,095 SSc cases and 17,584 controls, and the SSc GRS was developed in an independent cohort of 400 patients with SSc and 571 non-affected controls recruited for the PRECISESADS project (164). Since sample size is key in the identification of reliable genetic association signals and in the accurate estimation of allelic effects in GWAS (99,100,169), the presented SSc GRS represents a robust model supported by substantial statistical power. Nevertheless, despite the promising results of the described SSc GRS, the

sensitivity and specificity of the model are still far from clinical use and it will require the addition of extra information and/or the development of well-powered phenotype-specific GWAS to identify cases with specific phenotypes with higher statistical power.

Furthermore, we consider that the SSc GRS is not heavily influenced by LD clumping, since we included only the top HLA SNP association in the GRS in order to avoid an overrepresentation of HLA polymorphisms without discarding completely the potential of this region in GRS. Nevertheless, it should be noted that all the samples included in the GWAS summary stats and in the score development cohorts for the SSc GRS had European ancestry (65,164) (**Supplementary Figure 1**). One of the major limitations of GRS implementation is the bias towards populations with a similar ethnic origin to the discovery sample, i.e., the GRS shows better accuracy in closely related populations (100,170). As we illustrated in **Supplementary Figure 2**, we found differences in the AUCs reached by the SSc GRS in the score development cohort depending on the origin of the individuals. Consequently, the performance of the SSc GRS in non-European or mixed populations should be taken with caution (100,171).

A possible confounding factor for GRS in IMIDs is the shared genetic and immunological component that makes diagnosis complex and a slow clinical process especially in the early stages of these diseases (78,172,173). As a clinical tool, a robust GRS improves early diagnosis and helps in differential diagnosis (169). Although, the accuracy of the SSc GRS in differentiating between SSc and other IMIDs is still far from clinical standards, the model was able to discriminate between SSc and RA in 56.8% of the cases, and between SSc and SJS in 58.5% of the cases (**Figure 4**). However, for SLE and SSc, which have a well-documented shared genetic component (172,174), it was not possible to reach an accuracy that allowed for case differentiation.

Taking into account the above, we consider that the reported GRS could enhance SSc diagnosis in the future and may contribute to personalized medicine, as a tool to assist physicians in the diagnosis of SSc.

In addition to comorbidities with other IMIDs, there is great variability in the disease course followed by patients with SSc, since their treatment and prognosis in the long term is very heterogeneous (175). Chen *et al.* (167), developed a GRS based on a GWAS analyzing patients with SLE with and without renal involvement, but this lupus nephritis-specific GRS did not outperform the SLE severity predictions achieved with a SLE GRS. Following a similar strategy, we generated two additional GRS based on the GWAS comparisons between clinical and serological subtypes in patients with SSc. Remarkably, we showed that the serological subtype-specific GRS was able to differentiate SSc cases within the serological subtypes (ACA+ or ATA+), which is a promising result in the use of GRS to predict prognosis in SSc (60). Regarding specific clinical outcomes, we focused on the use of GRS to predict lung fibrosis due to the disastrous effect of lung involvement on the survival of patients with SSc. We could not use SSc lung involvement GWAS data, but we observed that the serological subtype-specific GRS allowed us to correctly infer the existence of lung fibrosis on patients with SSc in 57.5% of cases (**Supplementary Figure 3**).

Finally, we explored the possibilities of combining GRS with demographic and immunological covariates. We found that, out of all the covariates tested, age and the relative abundance of different immune cell types proved to be informative and resulted in a higher sensitivity in the case/control classification. As expected, age was confirmed as a very relevant factor in our model. Age is known to influence SSc, since patients with SSc are often diagnosed in their midlife ages (176,177). On the other hand, sex was included as a covariate to calculate the best p-value threshold for the GRS and

in the multivariate model, but, in both cases, it was not very informative. This lack of significant contribution of sex to the GRS model was also reported previously in SLE (167). Therefore, these counterintuitive results for a known SSc risk factor (176), were likely due to the selection of a sex matched control population (**Supplementary Table 1**), which would rule out the relevance of this parameter. The immune cell types included in the multivariate GRS were also concordant with the known etiopathogenesis of the disease (14). Functional defects or genetic susceptibility variants located in relevant genes for dendritic cells, macrophages and B cells have been described in patients with SSc (64,74,178–180). T cell subtypes were relevant covariates in the model initially, but no T cell subset was selected for the multivariate model (**Supplementary Tables 2-4**). Considering the central role of T cells in SSc, we hypothesize that since we could not include the Th1, Th2 or Th17 fractions in the model, this effect might have been overlooked(64).

We have generated a GRS using a GWAS data set and a score development cohort in which training was carried out and empirical p-values for the GRS were obtained via permutation. Therefore, although both cohorts were independent, out-of-sample prediction has not been performed and it is a limitation of the present study. Consequently, our model and results should be considered as seminal work for future validation in additional cohorts of SSc patients.

In summary, we developed a GRS based on the largest GWAS in SSc, resulting in a sensitive model to differentiate between SSc cases and non-affected controls, but also to differentiate within the different SSc serological subtypes (ATA+ and ACA+). Additionally, the GRS was also useful to differentiate patients with SSc from those affected by RA and SJS. We have shown that the GRS strategy in SSc has great potential to contribute to the field. However, several limitations and challenges, such as non-European

ancestry or sample size must be overcome to implement this strategy in clinical management.

Key messages

What is already known about this subject?

-Systemic sclerosis (SSc) is a complex immune mediated disease (IMID) for which a genomic risk score has never been implemented.

What does this study add?

-A SSc GRS discriminates between patients with SSc and healthy controls with a remarkable predictive value.

-Clinical information, such as serologic subtype and immune cells counts, adds accuracy to the GRS model.

-The SSc GRS is capable of discriminating between SSc and other IMIDs.

How might this impact clinical practice or future developments?

- This SSc GRS is a promising tool to improve the diagnosis and prognosis of patients with SSc.

Competing interests

Lara Bossini-Castillo: none; Gonzalo Villanueva-Martin: none; Martin Kerick: none; Marialbert Acosta-Herrera: none; Elena López Isac: none; International SSc Group: none; PRECISESADS Consortium: none; Marta E Alarcón-Riquelme: none; Lorenzo Beretta: none; Javier Martín: none

Contributorship

Lara Bossini-Castillo: data analysis, manuscript drafting, revision and approval; Gonzalo Villanueva-Martin: data analysis, manuscript drafting,

revision and approval; Martin Kerick: interpretation of data, manuscript revision and approval; Marialbert Acosta-Herrera: interpretation of data, manuscript revision and approval; Elena López Isac: data interpretation, manuscript revision and approval; PRECISESADS Clinical Consortium: data acquisition, manuscript revision and approval; Marta E Alarcón-Riquelme: data interpretation, manuscript revision and approval; Lorenzo Beretta: data interpretation, manuscript revision and approval; Javier Martín: study design, manuscript drafting, revision and approval.

Funding info

This work was supported by the Spanish Ministry of Science and Innovation (grant ref. RTI2018101332-B-100), Red de Investigación en Inflamación y Enfermedades Reumáticas (RIER) from Instituto de Salud Carlos III (RD16/0012/0013), and EU/EFPIA/Innovative Medicines Initiative Joint Undertaking PRECISESADS grant no. 115565. LBC and MAH were funded by the Spanish Ministry of Science and Innovation through the *Juan de la Cierva* incorporation program (ref. IJC2018-038026-I and IJC2018-035131-I respectively). GVM was funded by the Spanish Ministry of Science and Innovation through the *Ayudas para contratos predoctorales para la formación de doctores 2019* program (ref. RTI2018-101332-B-I00).

Acknowledgments

This work is part of the PhD Thesis entitled *Deciphering the genetic basis of systemic sclerosis*, by GVM. We would like to thank Sofia Vargas and Gemma Robledo for their excellent technical assistance. We also appreciate the controls and the affected individuals who generously provided the samples for these studies.

Ethical approval information

An ethical protocol was prepared, consensus was reached across all partners, academic and industrial, translated into all participants' languages and approved by each of the local ethical committees of the clinical recruitment centers. The studies adhered to the standards set by the International Conference on Harmonization and Good Clinical Practice (ICH-GCP), and to the ethical principles that have their origin in the Declaration of Helsinki (2013). The protection of the confidentiality of records that could identify the included subjects is ensured as defined by the EU Directive 2001/20/EC and the applicable national and international requirements relating to data protection in each participating country. The CS study is registered with number NCT02890121, and the inception study with number NCT02890134 in ClinicalTrials.gov.

The study (PRECISESADS cross-sectional study) was approved by the following ethic committees: Comitato Etico Area 2 (Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano and University of Milan); approval no. 425bis Nov 19, 2014, and no. 671_2018 Sep 19, 2018; Klinikum der Universitaet zu Koeln, Cologne, Germany. Geschaftsstelle Ethikkommission; Pôle de pathologies rhumatismales systémiques et inflammatoires, Institut de Recherche Expérimentale et Clinique, Université catholique de Louvain, Brussels, Belgium. Comité d`Èthique Hospitalo-Facultaire; University of Szeged, Szeged, Hungary. Csongrad Megyei Kormanyhivatal; Hospital Clinic I Provicia, Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. Comité de Ética de Investigación Clínica del Hospital Clínic de Barcelona. Hospital Clinic del Barcelona; Servicio Andaluz de Salud, Hospital Universitario Reina Sofía Córdoba, Spain. Comité de Ética de la Investigación de Centro de Granada (CEI – Granada); Centro Hospitalar do Porto, Portugal. Comissao de ética para a Saude – CES do CHP; Centre Hospitalier Universitaire de Brest, Hospital de la

Cavale Blanche, Avenue Tanguy Prigent 29609, Brest, France. Comité de Protection des Personnes Ouest VI; Hôpitaux Universitaires de Genève, Switzerland. DEAS –Commission Cantonale d'éthique de la recherche Hôpitaux Universitaires de Genève; Biobanco del Sistema Sanitario Público de Andalucía, Granada, Spain; Katholieke Universiteit Leuven, Belgium. Commissie Medische Ethiek UZ KU Leuven /Onderzoek; Charite, Berlin, Germany. Ethikkommission; Medizinische Hochschule Hannover, Germany. Ethikkommission.

Data sharing statement

Raw data are property of the PRECISESADS consortium. Metadata and aggregated data are available upon request to the authors.

Patient and Public Involvement

None.

References

1 Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 2017;101:5–22.

2 Fugger L, Jensen LT, Rossjohn J. Challenges, Progress, and Prospects of Developing Therapies to Treat Autoimmune Diseases. Cell 2020;181:63–80.

3 Barturen G, Beretta L, Cervera R, et al. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. Nat Rev Rheumatol 2018;14:180.

4 Parkes M, Cortes A, van Heel DA, et al. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet 2013;14:661–73.

5 Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 2019;47:D1005–12.

6 Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet 2016;17:392–406.

7 Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet 2018;50:1219–24.

8 Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. Hum Mol Genet 2019;28:R133–42.

9 Choi SW, Mak TS-H, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc Published Online First: 24 July 2020. doi:10.1038/s41596-020-0353-1

10 Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. Nat Commun 2019;10:3328.

11 International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature 2009;460:748–52.

12 Khera AV, Chaffin M, Wade KH, et al. Polygenic Prediction of Weight

and Obesity Trajectories from Birth to Adulthood. Cell 2019;177:587–96.e9.

13 Inouye M, Abraham G, Nelson CP, et al. Genomic Risk Prediction of Coronary Artery Disease in 480,000 Adults: Implications for Primary Prevention. J Am Coll Cardiol 2018;72:1883–93.

14 Gladding PA, Legget M, Fatkin D, et al. Polygenic Risk Scores in Coronary Artery Disease and Atrial Fibrillation. Heart Lung Circ 2020;29:634–40.

15 Lareau CA, DeWeese CF, Adrianto I, et al. Polygenic risk assessment reveals pleiotropy between sarcoidosis and inflammatory disorders in the context of genetic ancestry. Genes Immun 2017;18:88–94.

16 Reid S, Alexsson A, Frodlund M, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. Ann Rheum Dis 2020;79:363–9.

17 Chen L, Wang Y-F, Liu L, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet Published Online First: 20 February 2020. doi:10.1093/hmg/ddaa030

18 Roberts GHL, Paul S, Yorgov D, et al. Family Clustering of Autoimmune Vitiligo Results Principally from Polygenic Inheritance of Common Risk Alleles. Am J Hum Genet 2019;105:364–72.

19 Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol 2012;24:165–70.

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

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

22 Denton CP, Khanna D. Systemic sclerosis. Lancet 2017;390:1685–99.

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

24 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.

25 López-Isac E, Acosta-Herrera M, Kerick M, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nature Communications. 2019;10. doi:10.1038/s41467-019-12760-y

26 Choi SW, O'Reilly P. PRSice 2: POLYGENIC RISK SCORE SOFTWARE (UPDATED) AND ITS APPLICATION TO CROSS-TRAIT ANALYSES. European Neuropsychopharmacology. 2019;29:S832. doi:10.1016/j.euroneuro.2017.08.092

27 Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases: Inception Cohort (PRECISESADSI). clinicaltrials.gov.
2016.https://clinicaltrials.gov/ct2/show/NCT02890134

28 Jamin C, Le Lann L, Alvarez-Errico D, et al. Multi-center harmonization of flow cytometers in the context of the European 'PRECISESADS' project. Autoimmun Rev 2016;15:1038–45.

29 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.

30 Stahl EA, Wegmann D, Trynka G, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet 2012;44:483–9.

31 Inshaw JRJ, Cutler AJ, Burren OS, et al. Approaches and advances in the genetic causes of autoimmune disease and their implications. Nat Immunol 2018;19:674–84.

32 Peterson RE, Kuchenbaecker K, Walters RK, et al. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell 2019;179:589–603.

33 Schork AJ, Anthony Schork M, Schork NJ. Genetic risks and clinical rewards. Nature Genetics. 2018;50:1210–1. doi:10.1038/s41588-018-0213-x

34 Márquez A, Kerick M, Zhernakova A, et al. Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. Genome Med 2018;10:97.

35 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.

36 Iaccarino L, Gatto M, Bettio S, et al. Overlap connective tissue disease syndromes. Autoimmun Rev 2013;12:363–73.

37 Morel PA, Chang HJ, Wilson JW, et al. Severe systemic sclerosis with anti-topoisomerase I Antibodies is associated with an HLA-DRw11 allele. Hum Immunol 1994;40:101–10.

38 Mayes MD, Lacey JV, Beebe-Dimmer J, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population. Arthritis & Rheumatism. 2003;48:2246–55. doi:10.1002/art.11073

39 Affandi AJ, Carvalheiro T, Ottria A, et al. Low RUNX3 expression alters dendritic cell function in patients with systemic sclerosis and contributes to enhanced fibrosis. Ann Rheum Dis 2019;78:1249–59.

40 van Bon L, Affandi AJ, Broen J, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. N Engl J Med 2014;370:433–43.

41 Skaug B, Khanna D, Swindell WR, et al. Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile. Ann Rheum Dis 2020;79:379–86.

42 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.

43 Bossini-Castillo L, López-Isac E, Mayes MD, et al. Genetics of systemic sclerosis. Semin Immunopathol 2015;37:443–51.

Chapter 2. The effect of body Fat Distribution on Systemic Sclerosis

The effect of body Fat Distribution on Systemic Sclerosis

Gonzalo Villanueva-Martin 1,*, Marialbert Acosta-Herrera 1, Martin Kerick 1, Elena López-Isac 1, Carmen P Simeon 2, Jose Luis Callejas 3, Shervin Assassi 4, Lorenzo Beretta 5, International SSc Group 6, Australian Scleroderma Interest Group (ASIG) 6, Yannick Allanore 7, Susan Proudman 8, Mandana Nikpour 8, Carmen Fonseca 9, Christopher P Denton 9, Timothy RDJ Radstake 10, Maureen D Mayes 4, Xia Jiang11,12,13, Javier Martin1,† and Lara Bossini-Castillo14,15,†,*

1. Department of Cell Biology and Immunology, Institute of Parasitology and Biomedicine López-Neyra, CSIC, Granada, Spain
2. Department of Internal Medicine, Valle de Hebrón Hospital, Barcelona, Spain.
3. Department of Internal Medicine, Hospital San Cecilio, Granada, Spain.
4. Department of Rheumatology, The University of Texas Health Science Center at Houston, Houston, Texas, USA.
5. Referral Center for Systemic Autoimmune Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico di Milano, Milan, Italy.
6. See supplementary notes.
7. Department of Rheumatology A, Hospital Cochin, Paris, Île-de-France, France.
8. The University of Melbourne at St. Vincent's Hospital, Melbourne, Victoria, Australia
9. Center for Rheumatology, Royal Free and University College Medical School, London, UK.
10. Department of Rheumatology and Clinical Immunology, University Medical Center Utrecht, Utrecht, The Netherlands.
11. Department of Clinical Neuroscience, Center for Molecular Medicine, Karolinska Institutet, Solna, Sweden.
12. Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, United States.
13. West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China.
14. Departamento de Genética e Instituto de Biotecnología, Universidad de Granada, Granada, Spain.
15. Instituto de Investigación Biosanitaria ibs.GRANADA, Granada, Spain.

† Authors share senior authorship.

● Correspondence: GVM: gvillanuevamartin@ipb.csic.es ; LBM: lbossinicastillo@ugr.es.

Abstract: Obesity contributes to a chronic proinflammatory state, which is a known risk factor to develop immune-mediated diseases. However, its role in systemic sclerosis (SSc) remains to be elucidated. Therefore, we conducted a two sample mendelian randomization (2SMR) study to analyze the effect of three body fat distribution parameters in SSc. As instrumental variables, we used the allele effects described for single nucleotide polymorphisms (SNPs) in different genome-wide association studies (GWAS) for SSc, body mass index (BMI), waist-to-hip ratio (WHR) and WHR adjusted for BMI (WHRadjBMI). We performed local (pHESS) and genome-wide (LDSC) genetic correlation analyses between each of the traits and SSc and we applied several MR methods (i.e. random-effects inverse variance-weight, MR-Egger regression, MR pleiotropy residual sum and outlier method and a multivariable model). Our results showed no genetic correlation or causal relationship between any of these traits and SSc. Nevertheless, we observed a negative causal association between WHRadjBMI and SSc, which might be due to the effect of gastrointestinal complications suffered by the majority of SSc patients. In conclusion, reverse causality might be a specially difficult confounding factor to define the effect of obesity in the onset of SSc.

Keywords: systemic sclerosis; mendelian-Randomization; obesity

Introduction

Systemic sclerosis (SSc) is an immune mediated disease (IMD), characterized by abnormal immunological activation, vascular damage and fibrosis of the skin [1,2]. SSc represents a major challenge for clinicians as it has a deep impact on the life quality and life expectancy of the affected patients [1]. Recent efforts in the study of the genetic factors that contribute to the onset and progression of SSc, such as several large scale genetic association studies and genome wide association studies (GWAS) [3], have contributed to identifying genetic susceptibility markers both in the Human Leukocyte Antigen (HLA) locus and outside this highly polymorphic region [4]. The largest GWAS to date comprised more than 9.000 patients with SSc, and allowed the identification of 19 non-HLA *loci* associated with the disease [3]. Moreover, recent studies have identified specific HLA-DQA1 alleles exclusively associated with different clinical subtypes of SSc [4]. Therefore, the number of relevant *loci* that have been firmly associated with this condition has remarkably increased over the last decade. Although the use of genetic risk factors to predict the risk of developing SSc was explored in a recent genomic risk score (GRS) [5], the involvement of these genetic risk factors in the disease pathogenesis and the affected biological pathways have not been fully established yet [6].

Despite the advances in the identification of the genetic factors contributing to the heritability of SSc, the complex nature of this disorder is an intrinsic obstacle to study the pathological mechanisms that lead to the disruption of the immune homeostasis and to the onset of fibrotic processes in affected individuals. Well-established environmental triggers for SSc are silica and solvents, which in extreme or long-term exposures are related to the disease development [7,8]. Moreover, demographic and clinical characteristics such as sex, age, ethnical origin, hormone levels, etc.

have been pointed out as risk factors for SSc [7,9]. But the roles of life-style and environmental triggers in the manifestation and prognosis of SSc are still elusive.

Mendelian Randomization (MR) uses SNPs as instrumental variants (IVs) in order to determine if they are acting on a disease or outcome through a risk factor or exposure [10,11]. The principle of the methods is that alleles are randomly distributed during gametogenesis, as well as their presence pre-exists the disease. These genetic facts mimic the random distribution of clinical trials and take away the causality of the disease on the variable, reducing confounding factors [12]. For a genetic variant to be considered as a IV, it is assumed that it is associated with the exposure. However, an IV cannot be associated with any confounding factor related to the risk factor or the outcome neither directly nor indirectly. Additionally, the effects of the IV on the outcome should only be mediated by the exposure [10]. Therefore, only when genetic polymorphisms which are relevant, independent and have a restricted effect on the outcome can be considered as IVs. In a classical MR study, the allele effects on the outcome and exposure are obtained from the same individuals [10,11]. However, detailed information for multiple traits is difficult to obtain in a large population. Two-sample MR (2SMR) methods allow us to combine the estimations of the IV allele effects relying only on GWAS summary statistics for the outcome and for the exposure from independent studies. The implementation of these methods has improved the statistical power to detect causal associations between risk factors and disease, which has shown promising results in several conditions [13].

Obesity-related diseases are becoming a public health issue in Western countries [14], since obesity rates are increasing due to unhealthy lifestyles. Obesity is defined by an excess of fat in the body and body fat

distribution can be measured by a variety of methods, for instance body mass index (BMI) and waist to hip ratio (WHR). BMI is the most common body fat proxy and it is the gold-standard for obesity. BMI is measured as the body weight normalized by height square (kg/m2) [15], and it is known that BMI > 25 kg/m2 is associated with an increased risk to suffer from chronic diseases such as cardiovascular disease, type II diabetes or specific cancers [16]. Nevertheless, BMI has certain limitations and anthropometric measures of abdominal obesity, such as WHR, seem to be better indicators of excessive fat mass [17]. Since WHR measures both visceral and gluteal fat, it stands out among other anthropometric traits [18]. If WHR is adjusted for BMI (WHRadjBMI), it is possible to obtain an anthropometric measure which is independent from the overall adiposity, and to combine the most standardized measure of obesity and the anthropometric measure that best captures the distribution of body fat [18,19]. Taking advantage of the publicly available GWAS results, MR approaches have been successful in identifying risk factors for IMDs, such as obesity-related traits [20,21]. The excess of fat has been associated with a low but persistent proinflammatory state that is believed to promote IMDs [14,22]. However, in the case of SSc, the relationship between body fat distribution and SSc remains to be explored.

Consequently, in order to analyze the effect of nutritional-status on SSc risk, we applied the novel 2SMR methods on the largest GWAS of SSc patients [3] with European ancestry and the biggest GWAS meta-analysis for fat distribution anthropometric measures to date [23].

Material and Methods

Instrumental variables

The study design of the 2SMR study for SSc and 3 obesity-related traits is summarized in Figure 1. The outcome instrumental variables (IV-outcome), i.e. the selected genetic variants and their effect sizes in SSc, were obtained from the largest SSc GWAS meta-analysis, which included 9,846 SSc patients and 18,333 healthy controls from 14 different cohorts with European ancestry [3]. Additionally, SNP effect sizes after stratification by sex, serological and clinical subtype as reported elsewhere [4] were also analyzed. Finally, we performed sex-specific analyses including only either the female or the male individuals from the different cohorts and following the previously described analysis framework [3].

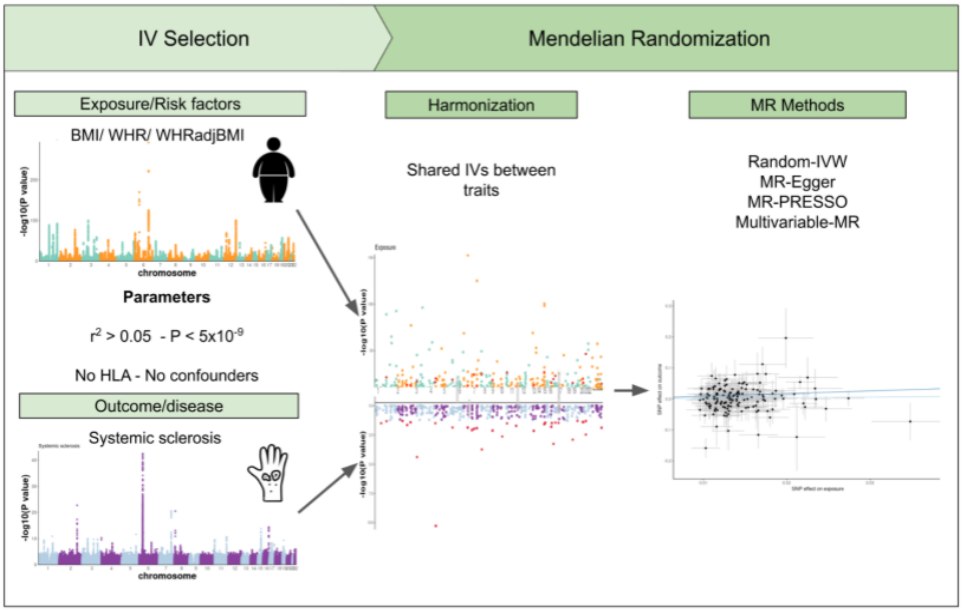


Figure 1. Schematic representation of the study design. The study is divided into several phases, i.e., selection of the instrumental variables for the

outcome and the exposures, data harmonization and generation of different Mendelian randomization models

In the case of the exposures, we obtained the IVs (IV-exposure) from a recent GWAS meta-analysis between the cohorts included in the Genetic Investigation of Anthropometric Traits consortium (GIANT) project and those recruited for the UK Biobank (UKBB) repository for different anthropometric measures [24]. We only the summary statistics comprising individuals with the European ancestry, which included 806,810 individuals and 27,381,302 SNPs for BMI, a classical obesity parameter, and for two parameters that assess body fat distribution, WHR comprising 697,734 individuals and 27,376,273 SNPs and WHRadjBMI covering 694,649 individuals and 27,375,636 SNPs [24]. None of the participants recruited in the SSc studies overlapped with the exposure GWASs to the best of our knowledge.

Genomic association analysis

Genetic correlation. To determine causality between obesity risk factors and SSc, we calculated the total genomic correlation between them. First, we performed an approximation implemented in the linkage disequilibrium regression score (LDSC) software [25]. Then, to study the contribution of specific regions (pairwise local genetic correlation), we used the methods supported in the ρ-HESS software [26]. Briefly, the ρ-HESS software splits the genome into 1,703 small regions through the chromosomes and uses LD matrices to create eigenvectors and to project the GWAS effect sizes. Then, local SNP-heritability per trait is calculated and, finally, genetic covariance between traits is estimated. We adjusted our significance thresholds for multiple testing, i.e. 1.1 x 10−3 (0.05/45) for LDSC and 2.9 x 10−5 (0.05/1,703) for ρ-HESS.

Mendelian randomization analysis. In order to assess if there was a causal relationship between body fat distribution and SSc or any of the stratified sets of patients, we performed a 2SMR study as implemented the R package "TwoSampleMR" [11]. Considering the complex linkage disequilibrium (LD) patterns and the strong genetic associations described in the HLA locus SSc [3,4,27], the extended HLA region (chromosome 6: 20,000,000 - 40,000,000 bp) was excluded from the MR analyses in order to prevent biases.

The selected IVs were based on the original independent signal analysis reported by Pulit et al. [24]. Briefly, the independent signals from results from the inverse variance meta-analysis ($P < 5 \times 10\text{-}9$) were identified by LD-based clumping ($r2 > 0.05$ and ± 5Mb). Secondary signals were also defined by conditional analyses ($P < 5 \times 10\text{-}9$) and locus LD-clumping. We extracted the association estimates for these SNPs or the best available proxy (according to the LD patterns observed in the UKBB cohort), which was present in the SSc dataset. The number of shared SNPs between SSc and the exposures reached 533, 247 and 262 for BMI, WHR, WHRadjBMI, respectively (Table A7).

Three gold-standard 2SMR methods were selected. A random-effects inverse variance-weight (IVW) approach, which pools the effects of each IV and balances to zero the global pleiotropy by assuming the validity or invalidity of all the SNPs [11]. A MR-Egger regression method [28], which is able to estimate causality even when all IVs are weak or invalid and to calculate horizontal pleiotropy. Although the previous methods are very robust for MR analysis, both of them have limitations to deal with outlier IVs. For that reason, we also applied the MR pleiotropy residual sum and outlier (MR-PRESSO) method [29]. The MR-PRESSO algorithm detects outlier IVs that exert horizontal pleiotropy in a multi-instrument

mendelian randomization analysis. Moreover, MR-PRESSO provides outlier-free causality estimates.

Additionally, to estimate the effect of the IVs controlling for their effect on other exposures, we performed a multivariable mendelian randomization analysis (MVMR) as implemented in the TwoSampleMR package [30]. This analysis included a set of unique LD-clumped IV-exposures for both BMI and WHR, which were regressed against SSc together, weighting for the inverse variance of SSc for these IVs.

False Discovery Rate (FDR) Benjamini & Hochberg correction was applied, and we considered $P < 0.05$ as significant [31].

Sensitivity analysis

The statistical power of our analyses was calculated using the algorithm described by Brion et al for MR studies [32]. Aiming to control for the effect of potential confounding factors, we removed from the MR analysis any the SNP with reported associations with known obesity-related confounding factors (Table A3) as reported by the GWAS catalog [33], SNPnexus [34] and ClinVar [35]. We studied the contribution of each SNP to the observed effects by carrying out a leave-one-out sensitivity analysis, as implemented in the "TwoSampleMR" package [11]. By these means, we observed that the exclusion of one SNP at a time did not affect the observed results.

Results

Leveraging mendelian randomization as a novel methodological strategy, we studied for the first time the causal contribution of body fat distribution to the risk of suffering from SSc (Figure 1). Here we used the GWAS summary statistics of the largest SSc meta-analysis [3] as an outcome and three obesity-related traits GWAS comprising thousands of European ancestry individuals as exposures.

Genomic correlation. Only the HLA locus harbours local genetic correlation between SSc and body fat distribution

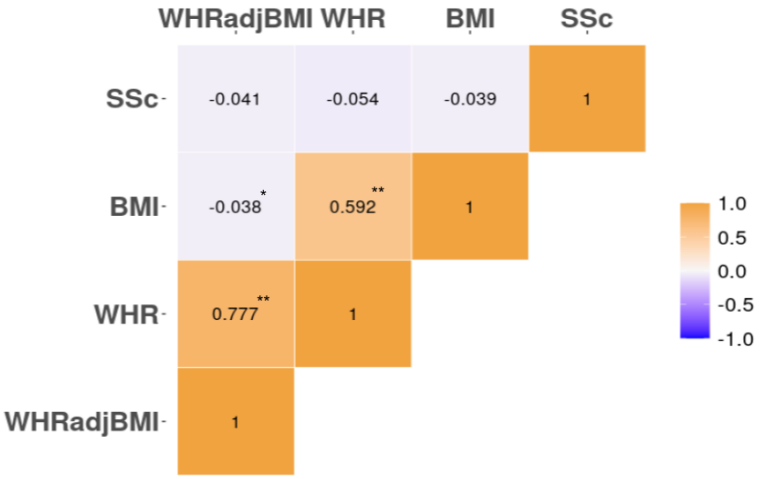


Figure 2. Pairwise global genetic correlation observed between the three obesity-related exposures and SSc. * = p > 0.05 (suggestive for statistical significance); ** = p > 0.00625 (Bonferroni-corrected).

At a genomic scale, we observed a strong genome-wide correlation between BMI and WHR (rg = 0.59, [95% CI -0.016 - 0.051]) and between WHR and WHRadjBMI (rg = 0.78, [95% CI -0.01 - 0.03]), but not between WHRadjBMI and BMI (rg = -4.02 x 10-2, [95% CI -0.016 - 0.049]), as

previously described [20] (Figure 2). However, our results showed no evidence of correlation between SSc and the three tested obesity-related traits (BMI rg = -0.039 [95% CI -0.033 - 0.102] ; WHR rg = -0.054, [95% CI -0.035 - 0.106]; WHRadjBMI rg = -0.041, [95% CI -0.04 - 0.122], all observed P > 0.05) (Figure 2).

Even when there is no correlation between traits at a genome-wide level, it is possible that the traits show local correlation at specific *loci*. To address this potential correlation, we performed a local genetic correlation analysis between BMI, WHR, WHRadjBMI and SSc (Figure A1). The local correlation observed in these regions reached rg = 8.5 x 10-4 and rg = 2.6 x 10-4 (Figure A1).

The analysis of the causal relationship between obesity-related traits and systemic sclerosis is limited by confounding factors.

Despite the limited genetic correlation found, we explored the possible causal relationship between body fat distribution and SSc. Considering the complex LD-patterns in the HLA-regions and the local genetic correlation found only in this locus, it was excluded from the following MR analyses. The available SSc dataset were powered enough to detect associations of 25% increased risk of SSc with BMI (99%), WHR (83%) and WHRadjBMI (92%) (Table A1), considering an explained phenotypic variance of 2.5-5% and the complete set comprising 28,179 individuals (34.9% cases). We were confident about the statistical power estimated for the largest subsets of patients, for instance, females (BMI power = 79%, WHR power = 82% and WHRadjBMI = 87%), lcSSc (BMI power = 94%, WHR power = 70% and WHRadjBMI = 81%) and ACA+ (BMI power = 83%, WHR power = 53% and WHRadjBMI = 65%). However, the analyses for the less frequent patient groups, i.e. males (BMI

Table 1. Association between genetically predicted obesity-related traits and risk of SSc. Analysis including index and secondary signals for the obesity-related traits and excluding the HLA region. BMI: body mass index, WHR: waist to hip ratio, WHRadjBMI: WHR adjusted for BMI, MR: Mendelian randomization, nSNPs: number of single nucleotide polymorphisms, OR: Odds Ratio, CI: confidence interval, p: p value, p adj: p value after FDR correction for multiple testing, IVW: inverse-variance weight, PRESSO: pleiotropy residual sum and outlier, NA: not applicable.

		Index and secondary SNPs ($p < 5 \times 10^{-9}$)				
	MR Approach	**nSNPs**	**OR (95% CI)**	***p***	***p* adj**	***p* for Heterogeneity or Pleiotropy**
BMI	MR-Egger	533	1.0575(0.6403–1.7466)	0.827	0.8273	0.6005
	Random-effects IVW		0.9326(0.7787–1.117)	0.449	0.4485	<0.001
	MR-PRESSO (1) *		0.943(0.7892–1.1269)	0.5189	NA	NA
WHR	MR-Egger	247	0.2698(0.0914–0.7965)	0.0185	0.0384	0.0519
	Random-effects IVW		0.7564(0.5567–1.0278)	0.0743	0.11145	<0.001
	MR-PRESSO (3) *		0.7809(0.5907–1.0324)	0.0838	NA	NA
WHRadjBMI	MR-Egger	262	0.4251(0.2014–0.8971)	0.0256	0.0384	0.1344
	Random-effects IVW		0.7269(0.5603–0.9431)	0.0163	0.0489	<0.001
	MR-PRESSO (1) *		0.77(0.6015–0.9857)	0.039	NA	NA

* Number of outlier SNPs detected by MR-PRESSO.

power = 30%, WHR power = 8% and WHRadjBMI = 10%) and ATA+ (BMI power = 14%, WHR power = 10% and WHRadjBMI = 12%) were clearly insufficient to identify true causal relationships (Table A1).

As reported in Table 1 and Table A2, classical MR methods showed no significant evidence of causality for BMI or WHR on SSc neither including only the index SNPs nor considering both the index SNPs and the secondary signals. The results for BMI under the random-effects IVW model showed a suggestive positive association with BMI, but this association did not reach statistical significance (OR under random-effects IVW = 1.15 [95% CI 0.67 - 1.98]) . Only a trend of negative association considering index and secondary signals was observed in the case of the random effects IVW model for WHR (Table 1). All the remaining models showed P > 0.05 and the ORs ranged 0.93 - 1.15 for BMI and 0.27 - 0.82 for WHR. In the case of WHRadjBMI (WHR after regressing out the effect of BMI), a negative association with SSc reached statistical significance in the three tested models (OR under random-effects IVW = 0.73 [95% CI 0.56 - 0.94], MR-Egger = 0.43 [95% CI 0.20-0.90], MR-PRESSO = 0.77 [95% CI 0.60-0.99]). These associations with WHRadjBMI remained negative in the analyses that included only index signals, but only the MR-Egger model was significant after multiple-testing correction (OR under MR-Egger = 0.69 [95% CI 0.51 - 0.93], (Table A2).

We carried out a sensitivity analysis, which implied the removal of SNPs associated with known obesity-related confounders (**Supplementary Table 3**), to address the effect of these confounders in the lack of significance for the BMI models and the negative relationships with WHR and WHRadjBMI. As shown in **Table 2** and **Supplementary Table 4**, the confounder-free models did not change the observed negative relationship and none of them reached a significant result after

Table 2. Association between genetically predicted obesity-related traits and risk of SSc. Analysis including index and secondary signals for the obesity-related traits and excluding the HLA region and known obesity-related confounder SNPs. BMI: body mass index, WHR: waist to hip ratio, WHRadjBMI: WHR adjusted for BMI, MR: Mendelian randomization, nSNPs: number of single nucleotide polymorphisms, OR: odds ratio, CI: confidence interval, p: p value, p adj: p value after FDR correction for multiple testing, IVW: inverse-variance weight, PRESSO: pleiotropy residual sum and outlier, NA: not applicable.

Index and Secondary SNPs ($p < 5 \times 10^{-9}$)

	MR Approach	nSNPs	OR (95% CI)	*p*	*p* adj	*p* for Heterogeneity or Pleiotropy
BMI	MR-Egger	483	1.422(0.721–2.803)	0.3103	0.3103	0.1769
	Random-effects IVW		0.909(0.741–1.115)	0.3598	0.3598	0.0011
	MR-PRESSO (1) *		0.922(0.754–1.128)	0.4288	NA	NA
WHR	MR-Egger	221	0.301(0.086–1.060)	0.0629	0.09435	0.1391
	Random-effects IVW		0.752(0.535–1.057)	0.1007	0.15105	< 0.001
	MR-PRESSO (2) *		0.764(0.559–1.044)	0.0927	NA	NA
WHRadjBMI	MR-Egger	237	0.335(0.137–0.819)	0.0172	0.0516	0.0772
	Random-effects IVW		0.716(0.534–0.961)	0.0261	0.0783	< 0.001
	MR-PRESSO (1) *		0.769(0.582–1.015)	0.0651	NA	NA

* Number of outlier SNPs detected by MR-PRESSO.

FDR correction. Although we observed effect size heterogeneity for the different genetic variants (**Supplementary Table 5**), the analyses of the intercept parameter in the MR-Egger models did not reveal any signs of horizontal pleiotropy and the effects were not affected by the removal of the outlier SNPs identified by the MR-PRESSO algorithm (**Tables 1-2**, **Supplementary Tables 2** and **Supplementary Table 4**). Furthermore, leave-one-out analyses did not highlight that these effects were influenced only by one variant (**Supplementary Figure 2**).

We decided to implement a MVMR model considering the significant associations observed for WHRadjBMI and the limitations of the univariate models to test for the combined influence of several exposures and to control for the effect of confounding factors. This analysis allowed us to directly test the association of BMI and WHR with SSc controlling for the effects of both parameters at the same time. As expected, the results of these analyses showed an effect for WHR (MVMR OR 0.80 [95% CI 0.57-1.13]) that was similar to the previously identified effect for WHRadjBMI (**Table 3**). Nevertheless, no significant association of BMI with SSc was revealed (MVMR OR 1.03 [95% CI 0.79-1.33]) (**Table 3**). These findings might point towards a negative or inexistent effect of WHR in SSc and, if any, a very modest risk effect for BMI.

Considering the well-known clinical and genetic differences between the SSc subsets of patients (181), we explored subset-specific effects for the selected exposures. Several associations remained significant in the stratified analyses, especially in the largest and more powerful subsets, such as lcSSc (**Supplementary Table 6**). However, the direction and magnitude of the exposure effects were consistent in all the subsets (**Supplementary Table 6**), which suggested an uniform effect, if any, in all the patients. There were no significant differences between the

Table 3. Multivariable MR (MVMR) model including BMI. WHR and risk of SSc. Analysis including index and secondary signals for the obesity-related traits and excluding the HLA region, with and without known obesity-related confounding SNPs. BMI: body mass index, WHR: waist to hip ratio, MR: Mendelian Randomization, nSNPs: number of Single Nucleotide Polymorphism, OR stand for Odd Ratio, CI: confidence interval, p: p value.

Before Confounder SNP Removal.					**After Confounder SNP Removal**				
Index and secondary SNPs ($p < 5 \times 10^{-9}$)					Index and secondary SNPs ($p < 5 \times 10^{-9}$)				
Outcome	**Exposure**	**nSNP**	**OR (95% CI)**	***p***	**Outcome**	**Exposure**	**nSNP**	**OR (95% CI)**	***p***
SSc	BMI	666	1.026(0.79–1.331)	0.849	SSc	BMI	610	1.027(0.760–1.387)	0.863
	WHR	666	0.804(0.573–1.128)	0.207		WHR	610	0.812(0.552–1.195)	0.291
Index SNPs ($p < 5 \times 10^{-9}$)					Index SNPs ($p < 5 \times 10^{-9}$)				
Outcome	**Exposure**	**nSNP**	**OR (95% CI)**	***p***	**Outcome**	**Exposure**	**nSNP**	**OR (95% CI)**	***p***
SSc	BMI	581	0.99(0.749–1.309)	0.946	SSc	BMI	524	1.013(0.726–1.412)	0.941
	WHR	581	0.876(0.607–1.263)	0.477		WHR	524	0.881(0.574–1.352)	0.561

models with and without the secondary signals (**Supplementary Table 6**). Moreover, taking into account the higher frequency of SSc in females (9 female: 1 male ratio) (13), we performed sex specific analyses too. In these analyses, we relied on female only and male only GWAS summary statistics for both SSc and the obesity-related risk factors. Once more, although the risk effect of BMI, WHR and WHRadjBMI seemed more evident in men, these effects did not reach statistical significance (**Supplementary Table 6**).

Discussion

This report addressed the risk effect of body fat distribution in SSc for the first time. We exhaustively exploited public GWAS summary statistics for both SSc and for anthropometric traits and the development of novel MR methods. We did not observe global genomic correlation between the outcome and any of the exposures. Moreover, local genetic correlation was only found in the HLA locus, a highly complex region. Different MR methods were then applied to identify possible causal relationships between the obesity traits and SSc. However, no significant risk causal effect of the exposures was found in this case.

Although our results do not support the causal relation between exposures and outcome, it should be noted that the statistical power of the SSc dataset is modest compared to similar studies performed to date in other IMDs, such as RA or IBD (182) (**Supplementary table 1**). SSc is a rare IMD and, despite the recent advances (65,66,74), the recruitment of large patient cohorts remains challenging. Therefore, future efforts to enlarge the size or to complement the available SSc GWAS information might help to identify causal risk factors.

We found that the effect of confounders might be more severe in the case of SSc than in other IMDs. Gastrointestinal involvement (GI), which affects more than 70% of the SSc patients (182), hinders food ingestion and patients are mostly thin (183). In fact, weight loss has been used as one of the SSc diagnostic markers (181). This direct effect of the onset symptoms in the exposures is known as reverse causality, and it is a remarkably difficult confounding factor to control for (184). Reverse causality might be the cause behind both the lack of significant risk effects of BMI in SSc and the reported negative relationship between WHR and SSc, which becomes

more evident when the effect of BMI is subtracted in the analysis of WHRadjBMI (**Tables 1-2**, **Supplementary Table 2** and **Supplementary Table 4**).

Bad diet habits and obesity are associated with an increased risk to suffer from IMDs such as RA and IBD (45,185,186). Higher BMI has been associated with increased risk to Crohn's disease (CD) and RA, but negative associations with BMI have been reported for ulcerative colitis (UC) and a recent study found reverse causality between WHR and RA (45,185,186). IMDs are often present as comorbidities and share altered molecular pathways, environmental triggers and genetic risk factors (187). Furthermore, the role of adipocytes in the activation of the immune system is prominent, especially due to the release of adipokines (188). Adipokines are molecules known to be involved in the "obesity-autoimmunity" relationship (42,189), such as lectins or cytokines, especially adiponectin, but also interleukins and tumor necrosis factor alpha (TNFα) (42). Interestingly, patients with SSc and a high BMI have been shown to have higher lectin levels than healthy controls (190) and it has been established that subcutaneous adipocytes can act as progenitor cells for fibroblasts (191,192). These fibroblasts may eventually transdifferentiate into myofibroblasts (193), activated profibrotic fibroblasts that are characteristic of the fibrotic lesions observed in SSc patients, and recent evidence has shown that the activation of adipocyte-derived mesenchymal cells from SSc skin biopsies to myofibroblasts is possible using soluble molecules present the skin microenvironment in SSc (194).

In order to rule out the role of obesity as a risk factor for SSc, body-fat distribution measures from the patients before the onset of GI or BMI matched case-control sets would be very valuable resources.

The negative association that is observed for WHR might be due to additional confounding factors that are inherent to SSc and that affect body fat distribution, for example, sex or lipid profiles (195). Remarkably, WHR is different in women than in men and there is a clear sex-bias in SSc (181). Therefore, we hypothesized that there could be a sex-specific association and performed stratified analyses with the female and male cohorts separately. Our results showed significant causal associations with SSc only in the females, but considering the statistical power differences and the similarity between the effect sizes, the lack of significance for the male group may be likely due to the reduced sample size (**Supplementary table 1**). The key role of sample size as a limitation of our study to identify weak risk effects was also clear in other stratified analyses, as we found consistent ORs for all the tested clinical subtypes of SSc patients but the models reached statistical significance only in the largest subsets (**Supplementary Table 6**).

In conclusion, this study found no significant evidence that supported the role of body-fat distribution as causal risk factor for SSc using 2SMR methods. Nevertheless, the current GWAS have a limited statistical power to identify modest contributions to SSc risk and the intrinsic nature of the SSc clinical complications might be acting as potential constraints in this study. Consequently, further analyses will be needed to rule out the role of obesity in the onset of SSc.

References

1. Radstake, T.R.D.J.; Gorlova, O.; Rueda, B.; Martin, J.-E.; Alizadeh, B.Z.; Palomino-Morales, R.; Coenen, M.J.; Vonk, M.C.; Voskuyl, A.E.; Schuerwegh, A.J.; et al. Genome-Wide Association Study of Systemic Sclerosis Identifies CD247 as a New Susceptibility Locus. Nat. Genet. 2010, 42, 426–429.

2. López-Isac, E.; Acosta-Herrera, M.; Kerick, M.; Assassi, S.; Satpathy, A.T.; Granja, J.; Mumbach, M.R.; Beretta, L.; Simeón, C.P.; Carreira, P.; et al. GWAS for Systemic Sclerosis Identifies Multiple Risk loci and Highlights Fibrotic and Vasculopathy Pathways. Nat. Commun. 2019, 10, 4955.

3. Acosta-Herrera, M.; Kerick, M.; Lopéz-Isac, E.; Assassi, S.; Beretta, L.; Simeón-Aznar, C.P.; Ortego-Centeno, N.; International SSc Group; Proudman, S.M.; Australian Scleroderma Interest Group (ASIG); 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, doi:10.1136/annrheumdis-2021-219884.

4. Bossini-Castillo, L.; Villanueva-Martin, G.; Kerick, M.; Acosta-Herrera, M.; López-Isac, E.; Simeón, C.P.; Ortego-Centeno, N.; Assassi, S.; International SSc Group; Australian Scleroderma Interest Group (ASIG); et al. Genomic Risk Score Impact on Susceptibility to Systemic Sclerosis. Ann. Rheum. Dis. 2021, 80, 118–127.

5. Bossini-Castillo, L.; López-Isac, E.; Mayes, M.D.; Martín, J. Genetics of Systemic Sclerosis. Semin. Immunopathol. 2015, 37, 443–451.

6. Barnes, J.; Mayes, M.D. Epidemiology of Systemic Sclerosis: Incidence, Prevalence, Survival, Risk Factors, Malignancy, and Environmental Triggers. Curr. Opin. Rheumatol. 2012, 24, 165–170.

7. Calderon, L.M.; Pope, J.E. Scleroderma Epidemiology Update. Curr. Opin. Rheumatol. 2021, 33, 122–127.

8. Allanore, Y.; Simms, R.; Distler, O.; Trojanowska, M.; Pope, J.; Denton, C.P.; Varga, J. Systemic Sclerosis. Nat Rev Dis Primers 2015, 1, 15002.

9. Burgess, S.; Butterworth, A.; Thompson, S.G. Mendelian Randomization Analysis with Multiple Genetic Variants Using Summarized Data. Genet. Epidemiol. 2013, 37, 658–665.

10. Bowden, J.; Davey Smith, G.; Burgess, S. Mendelian Randomization

with Invalid Instruments: Effect Estimation and Bias Detection through Egger Regression. Int. J. Epidemiol. 2015, 44, 512–525.

11. Smith, G.D.; Ebrahim, S. "Mendelian Randomization": Can Genetic Epidemiology Contribute to Understanding Environmental Determinants of Disease? Int. J. Epidemiol. 2003, 32, 1–22.

12. Hartwig, F.P.; Davies, N.M.; Hemani, G.; Davey Smith, G. Two-Sample Mendelian Randomization: Avoiding the Downsides of a Powerful, Widely Applicable but Potentially Fallible Technique. Int. J. Epidemiol. 2016, 45, 1717–1726.

13. Versini, M.; Jeandel, P.-Y.; Rosenthal, E.; Shoenfeld, Y. Obesity in Autoimmune Diseases: Not a Passive Bystander. Autoimmun. Rev. 2014, 13, 981–1000.

14. Borga, M.; West, J.; Bell, J.D.; Harvey, N.C.; Romu, T.; Heymsfield, S.B.; Dahlqvist Leinhard, O. Advanced Body Composition Assessment: From Body Mass Index to Body Composition Profiling. J. Investig. Med. 2018, 66, 1–9.

15. Huxley, R.; Mendis, S.; Zheleznyakov, E.; Reddy, S.; Chan, J. Body Mass Index, Waist Circumference and Waist:hip Ratio as Predictors of Cardiovascular Risk--a Review of the Literature. Eur. J. Clin. Nutr. 2010, 64, 16–22.

16. Cao, Q.; Yu, S.; Xiong, W.; Li, Y.; Li, H.; Li, J.; Li, F. Waist-Hip Ratio as a Predictor of Myocardial Infarction Risk: A Systematic Review and Meta-Analysis. Medicine 2018, 97, e11639.

17. Pulit, S.L.; Karaderi, T.; Lindgren, C.M. Sexual Dimorphisms in Genetic loci Linked to Body Fat Distribution. Biosci. Rep. 2017, 37, doi:10.1042/BSR20160184.

18. Randall, J.C.; Winkler, T.W.; Kutalik, Z.; Berndt, S.I.; Jackson, A.U.; Monda, K.L.; Kilpeläinen, T.O.; Esko, T.; Mägi, R.; Li, S.; et al. Sex-Stratified Genome-Wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic loci for Anthropometric Traits. PLoS Genet. 2013, 9, e1003500.

19. Tang, B.; Shi, H.; Alfredsson, L.; Klareskog, L.; Padyukov, L.; Jiang, X. Obesity-Related Traits and the Development of Rheumatoid Arthritis: Evidence From Genetic Data. Arthritis Rheumatol 2021, 73, 203–211.

20. Carreras-Torres, R.; Ibáñez-Sanz, G.; Obón-Santacana, M.; Duell, E.J.; Moreno, V. Identifying Environmental Risk Factors for Inflammatory Bowel

Diseases: A Mendelian Randomization Study. Sci. Rep. 2020, 10, 19273.

21. Alwarawrah, Y.; Kiernan, K.; MacIver, N.J. Changes in Nutritional Status Impact Immune Cell Metabolism and Function. Front. Immunol. 2018, 9, 1055.

22. Pulit, S.L. Summary-Level Data from Meta-Analysis of Fat Distribution Phenotypes in UK Biobank and Giant 2018.

23. Pulit, S.L.; Stoneman, C.; Morris, A.P.; Wood, A.R.; Glastonbury, C.A.; Tyrrell, J.; Yengo, L.; Ferreira, T.; Marouli, E.; Ji, Y.; et al. Meta-Analysis of Genome-Wide Association Studies for Body Fat Distribution in 694,649 Individuals of European Ancestry.

24. Bulik-Sullivan, B.K.; Loh, P.-R.; Finucane, H.K.; Ripke, S.; Yang, J.; Schizophrenia Working Group of the Psychiatric Genomics Consortium; Patterson, N.; Daly, M.J.; Price, A.L.; Neale, B.M. LD Score Regression Distinguishes Confounding from Polygenicity in Genome-Wide Association Studies. Nat. Genet. 2015, 47, 291–295.

25. Shi, H.; Mancuso, N.; Spendlove, S.; Pasaniuc, B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. Am. J. Hum. Genet. 2017, 101, 737–751.

26. Mayes, M.D.; Bossini-Castillo, L.; Gorlova, O.; Martin, J.E.; Zhou, X.; Chen, W.V.; Assassi, S.; Ying, J.; Tan, F.K.; Arnett, F.C.; et al. Immunochip Analysis Identifies Multiple Susceptibility loci for Systemic Sclerosis. Am. J. Hum. Genet. 2014, 94, 47–61.

27. Burgess, S.; Thompson, S.G. Interpreting Findings from Mendelian Randomization Using the MR-Egger Method. Eur. J. Epidemiol. 2017, 32, 377–389.

28. Verbanck, M.; Chen, C.-Y.; Neale, B.; Do, R. Publisher Correction: Detection of Widespread Horizontal Pleiotropy in Causal Relationships Inferred from Mendelian Randomization between Complex Traits and Diseases. Nat. Genet. 2018, 50, 1196.

29. Hemani, G.; Zheng, J.; Elsworth, B.; Wade, K.H.; Haberland, V.; Baird, D.; Laurin, C.; Burgess, S.; Bowden, J.; Langdon, R.; et al. The MR-Base Platform Supports Systematic Causal Inference across the Human Phenome. Elife 2018, 7, doi:10.7554/eLife.34408.

30. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. 1995, 57, 289–300.

31. Brion, M.-J.A.; Shakhbazov, K.; Visscher, P.M. Calculating Statistical Power in Mendelian Randomization Studies. Int. J. Epidemiol. 2013, 42, 1497–1501.

32. Buniello, A.; MacArthur, J.A.L.; Cerezo, M.; Harris, L.W.; Hayhurst, J.; Malangone, C.; McMahon, A.; Morales, J.; Mountjoy, E.; Sollis, E.; et al. The NHGRI-EBI GWAS Catalog of Published Genome-Wide Association Studies, Targeted Arrays and Summary Statistics 2019. Nucleic Acids Res. 2019, 47, D1005–D1012.

33. Oscanoa, J.; Sivapalan, L.; Gadaleta, E.; Dayem Ullah, A.Z.; Lemoine, N.R.; Chelala, C. SNPnexus: A Web Server for Functional Annotation of Human Genome Sequence Variation (2020 Update). Nucleic Acids Res. 2020, 48, W185–W192.

34. Landrum, M.J.; Lee, J.M.; Benson, M.; Brown, G.R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. Nucleic Acids Research 2018, 46, D1062–D1067.

35. Denton, C.P.; Khanna, D. Systemic Sclerosis. Lancet 2017, 390, 1685–1699.

36. Miller, J.B.; Gandhi, N.; Clarke, J.; McMahan, Z. Gastrointestinal Involvement in Systemic Sclerosis: An Update. J. Clin. Rheumatol. 2018, 24, 328–337.

37. Hughes, M.; Heal, C.; Siegert, E.; Hachulla, E.; Airó, P.; Riccardi, A.; Distler, O.; Matucci-Cerinic, M.; EUSTAR Collaborators Significant Weight Loss in Systemic Sclerosis: A Study from the EULAR Scleroderma Trials and Research (EUSTAR) Database. Ann. Rheum. Dis. 2020, 79, 1123–1125.

38. Davies, N.M.; Holmes, M.V.; Davey Smith, G. Reading Mendelian Randomisation Studies: A Guide, Glossary, and Checklist for Clinicians. BMJ 2018, 362, k601.

39. Julià, A.; Martínez-Mateu, S.H.; Domènech, E.; Cañete, J.D.; Ferrándiz, C.; Tornero, J.; Gisbert, J.P.; Fernández-Nebro, A.; Daudén, E.; Barreiro-de Acosta, M.; et al. Food Groups Associated with Immune-Mediated Inflammatory Diseases: A Mendelian Randomization and Disease Severity Study. Eur. J. Clin. Nutr. 2021, doi:10.1038/s41430-021-00913-6.

40. González-Serna, D.; Villanueva-Martin, G.; Acosta-Herrera, M.; Márquez, A.; Martín, J. Approaching Shared Pathophysiology in Immune-Mediated Diseases through Functional Genomics. Genes 2020, 11,

doi:10.3390/genes11121482.

41. Francisco, V.; Pino, J.; Campos-Cabaleiro, V.; Ruiz-Fernández, C.; Mera, A.; Gonzalez-Gay, M.A.; Gómez, R.; Gualillo, O. Obesity, Fat Mass and Immune System: Role for Leptin. Front. Physiol. 2018, 9, 640.

42. Żółkiewicz, J.; Stochmal, A.; Rudnicka, L. The Role of Adipokines in Systemic Sclerosis: A Missing Link? Arch. Dermatol. Res. 2019, 311, 251–263.

43. Iannone, F.; Praino, E.; Rotondo, C.; Natuzzi, D.; Bizzoca, R.; Lacarpia, N.; Fornaro, M.; Cacciapaglia, F. Body Mass Index and Adipokines/cytokines Dysregulation in Systemic Sclerosis. Clin. Exp. Immunol. 2021, doi:10.1111/cei.13651.

44. Driskell, R.R.; Lichtenberger, B.M.; Hoste, E.; Kretzschmar, K.; Simons, B.D.; Charalambous, M.; Ferron, S.R.; Herault, Y.; Pavlovic, G.; Ferguson-Smith, A.C.; et al. Distinct Fibroblast Lineages Determine Dermal Architecture in Skin Development and Repair. Nature 2013, 504, 277–281.

45. Salzer, M.C.; Lafzi, A.; Berenguer-Llergo, A.; Youssif, C.; Castellanos, A.; Solanas, G.; Peixoto, F.O.; Stephan-Otto Attolini, C.; Prats, N.; Aguilera, M.; et al. Identity Noise and Adipogenic Traits Characterize Dermal Fibroblast Aging. Cell 2018, 175, 1575–1590.e22.

46. Zhang, Z.; Shao, M.; Hepler, C.; Zi, Z.; Zhao, S.; An, Y.A.; Zhu, Y.; Ghaben, A.L.; Wang, M.-Y.; Li, N.; et al. Dermal Adipose Tissue Has High Plasticity and Undergoes Reversible Dedifferentiation in Mice. Journal of Clinical Investigation 2019, 129, 5327–5342.

47. Taki, Z.; Gostjeva, E.; Thilly, W.; Yaseen, B.; Lopez, H.; Mirza, M.; Hassuji, Z.; Vigneswaran, S.; Ahmed Abdi, B.; Hart, A.; et al. Pathogenic Activation of Mesenchymal Stem Cells Is Induced by the Disease Microenvironment in Systemic Sclerosis. Arthritis Rheumatol 2020, 72, 1361–1374.

Chapter 3. Non-classical circulating monocytes expressing high levels of microsomal prostaglandin E2 synthase-1 tag an aberrant IFN-response in systemic sclerosis

Non-classical circulating monocytes expressing high levels of *microsomal prostaglandin E2 synthase-1* tag an aberrant IFN-response in systemic sclerosis

Gonzalo Villanueva-Martin[1], Marialbert Acosta-Herrera[1,2], Elio G. Carmona[1,2], Martin Kerick[1], Norberto Ortego-Centeno[2,3], Jose Luis Callejas-Rubio[2], Norbert Mages[4], Sven Klages[4], Stefan Börno[4], Bernd Timmermann[4], Lara Bossini-Castillo[5,6*‡] and Javier Martin[1*‡].

1. Department of Cell Biology and Immunology, Institute of Parasitology and Biomedicine López-Neyra, CSIC, Granada, Spain.
2. Systemic Autoimmune Disease Unit, Hospital Clínico San Cecilio, Instituto de Investigación Biosanitaria Ibs. GRANADA, Granada, Spain.
3. Department of Medicine, University of Granada, Instituto de Investigación Biosanitaria Ibs. GRANADA, Granada, Spain.
4. Sequencing Core Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany.
5. Departamento de Genética e Instituto de Biotecnología, Centro de Investigación Biomédica (CIBM), Universidad de Granada, 18100 Granada, Spain.
6. Advanced Therapies and Biomedical Technologies (TEC-14), Instituto de Investigación Biosanitaria Ibs. GRANADA, 18016 Granada, Spain.

* These authors contributed equally.

‡Corresponding authors: Lara Bossini-Castillo (lbossinicastillo@ugr.es) and Javier Martin (javiermartin@ipb.csic.es).

Abstract

Systemic sclerosis (SSc) is a complex disease that affects the connective tissue, causing fibrosis. SSc patients show altered immune cell composition and activation in the peripheral blood (PB). PB monocytes (Mos) are recruited into tissues where they differentiate into macrophages, which are directly involved in fibrosis. To understand the role of CD14+ PB Mos in SSc, a single-cell transcriptome analysis (scRNA-seq) was conducted on 8 SSc patients and 8 controls. Using unsupervised clustering methods, CD14+ cells were assigned to 11 clusters, which added granularity to the known monocyte subsets: classical (cMos), intermediate (iMos) and non-classical Mos (ncMos) or type 2 dendritic cells. NcMos were significantly overrepresented in SSc patients and showed an active IFN-signature and increased expression levels of *PTGES*, in addition to monocyte motility and adhesion markers. We identified a SSc-related cluster of *IRF7*+ *STAT1*+ iMos with an aberrant IFN-response. Finally, a depletion of M2 polarised cMos in SSc was observed. Our results highlighted the potential of PB Mos as biomarkers for SSc and provided new possibilities for putative drug targets for modulating the innate immune response in SSc.

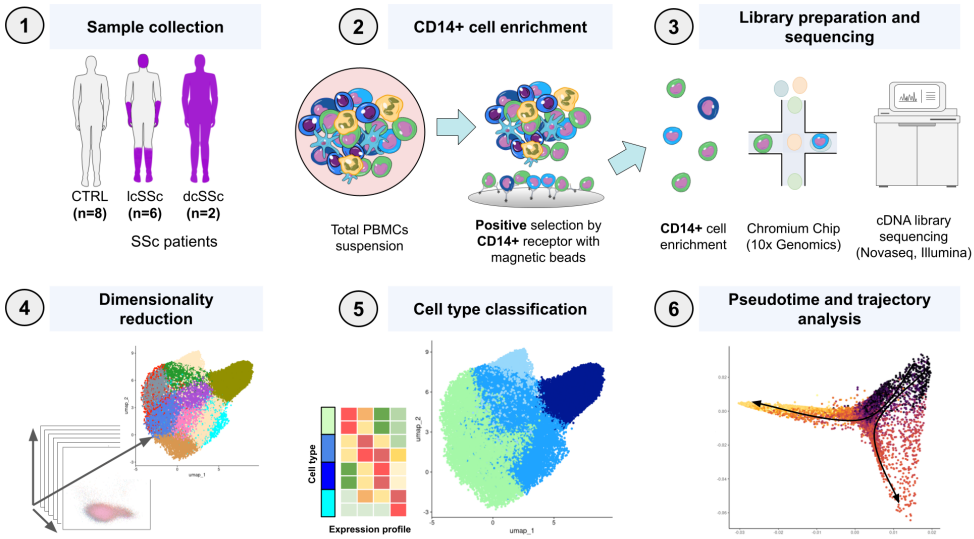
Keywords

Systemic sclerosis, single-cell transcriptome, scRNA-seq, monocyte, CD14

Highlights

- Non-classical monocytes in SSc were characterised by a high IFN-response signature and the upregulation of prostaglandin E2 synthesis.
- IRF7+ STAT1+ intermediate monocytes contributed especially to SSc.
- M2 polarised classical monocytes were depleted in the blood of SSc patients.

Graphical summary



1. Introduction

Systemic sclerosis (SSc) is a chronic life-threatening immune-mediated disease (IMD), which is characterised by an imbalanced immune response, endothelial damage and progressive fibrosis of the skin and internal organs [1]. Clinical manifestations among patients are highly heterogeneous, involving different extents of fibrosis, the appearance of autoantibodies against different nuclear structures and the onset of clinical complications or comorbidities [1]. SSc patients are classified into two major clinical subtypes: limited cutaneous SSc (lcSSc), if fibrosis is restricted to specific areas of the body (i.e. face and limbs), and diffuse cutaneous SSc (dcSSc), if fibrosis is generalised and affects mostly the torso and the proximal regions of the limbs [1]. From a geneticist point of view, SSc is classified as a complex disorder, as it is triggered by unknown environmental factors in genetically predisposed individuals [2]. Large genetic studies have contributed to establish 27 *loci* as firm genetic players in SSc susceptibility [3].

Both the adaptive and innate responses are chronically active and aberrant in SSc patients [4]. In this regard, innate myeloid cells that act as antigen presenting cells (APCs) [5] have been shown to be involved in pathological tissue scarring and fibrosis in SSc patients [6]. Moreover, alterations of the macrophage compartment have been suggested as essential drivers of connective tissue fibrosis in SSc [7]. Notably, in the early stages of SSc skin fibrosis, macrophages show a proinflammatory M1 and/or M2 concomitant profile that might progress towards a M1/M2 disequilibrium in later stages [8]. Furthermore, a new set of CXCL4-induced macrophages, which might be linked with profibrotic skills, has been identified in SSc patients. However, their exact role in the disease is yet to be defined [9].

Recently, the study of the molecular mechanisms leading to anomalous macrophage behaviour in SSc-affected tissues has reached an unprecedented level of detail thanks to the improvements of single cell transcriptome (scRNA-seq) technologies [6,10], which focused on the tissue-resident fibroblast, lymphocyte and macrophage populations [6,11,12]. Regarding macrophages, pioneer studies have identified a highly proliferative SSc-specific M2 macrophage subpopulation in the lungs distinguished by the expression of osteopontin (SPP1) [10], which induces profibrotic characteristics in the fibroblasts [13]. In the case of skin macrophages, scRNA-seq experiments of dcSSc skin have singled out a macrophage subpopulation characterised by the expression of high levels of Fcγ receptor IIIA (FCGR3A, also known as CD16) [6].

Interestingly, myeloid populations are not only a tissue-resident lineage, but also circulate in the peripheral blood as monocytes [4]. Due to the systemic nature of the disease, the monocyte compartment of SSc patients has abnormal biophysical properties and increased proportions of circulating inflammatory non-classical monocytes (ncMo) [14–17]. Moreover, SSc monocytes have been reported to increase their adhesion by reducing the expression of CD52 [18] and upregulating CCR3 [19] as a response to type I IFN. Moreover, circulating myeloid cells in the blood of SSc patients, especially those with a severe disease, have a gene expression profile that combines M1 and M2 surface markers [20–22].

Nevertheless, scRNA-seq technologies have not been applied to comprehensively characterise circulating CD14+ monocytes in the blood of SSc patients. Therefore, we will investigate the composition and the cell subtype-specific expression profiles of the monocyte compartment in SSc at the highest resolution by analysing the single cell transcriptomes of more

than 94,000 CD14+ peripheral blood cells from 8 patients affected with SSc and 8 controls.

2. Methods

2.1. Patient description

The study cohort consisted of 16 women. All individuals were of self-reported European ancestry and of similar age (average age SSc = 60; average age controls= 59), 8 of them were diagnosed with SSc and the remaining 8 were unaffected. Patients fulfilled the diagnostic criteria for the disease proposed by ACR [23] and were classified into limited cutaneous or diffuse cutaneous SSc according to the criteria proposed by LeRoy [24,25]. Clinical information of the patients, as well as their serological profile and drug treatment are shown in Supplementary Table 1. All participants were selected from the Hospital Universitario San Cecilio (Granada, Spain) by qualified staff and they signed a written consent before being enrolled in the study. All samples were irreversibly anonymised.

2.2. Cell suspension protocol

Thirty millilitres of PB of each participant were collected in EDTA tubes (Greiner #4550356) and processed for cryopreservation within 1 hour of extraction. CD14+ cells were isolated at the Instituto de Parasitología y Biomedicina López-Neyra (Granada, Spain). Next, plasma was separated from peripheral blood mononuclear cells (PBMCs) using Ficoll® Paque Plus (Merk #GE 17-1440-02) density gradients in Leucosep centrifuge tubes (Greiner #227290). Positive selection of CD14+ cells was performed using a magnetic bead kit (Stem Cell Easy Step, ref #17858) by following the protocol established by the manufacturer. The CD14+ cells accounted for

more than 86.5% of the cells in the samples as confirmed by flow-cytometry analyses and high CD14 mRNA expression was later confirmed at the single cell level (Supplementary Figure 1). Then, CD14+ cell suspensions were cryopreserved in 10% DMSO (Merk D2438) and 90% foetal bovine serum (FBS, Gibco #10082-147) medium and frozen in a -80ºC ultrafreezer at a controlled rate for at least 24 hours using a CoolCell container (Corning # 432000). Samples were kept in liquid nitrogen for long term storage.

2.3. Single cell RNA-sequencing library generation

To perform single cell whole transcriptome sequencing (scRNA-seq), we used the Next GEM technology by 10x Genomics. Samples were assayed following the manufacturer's instructions for the following kits: Chromium Next GEM Chip G Single Cell Kit (10xGenomics, PN-1000127) and Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1 (10xGenomics, PN-1000165_a). Subsequently, the generated cDNA libraries were sequenced using the NovaSeq 6000 platform (Illumina) with S2 and SP chemistry v1.5. The previously described settings allowed us to obtain an average of 85.85% of reads in cells, with an average of 31,847 reads per cell. The sequencing reads were aligned to the GRCh38 genome build and unique molecular identifiers (UMI) were processed by the 10x Cell Ranger Single Cell Software Suite (v3.0.0) using default parameters, with an average of 1,658 genes identified per cell.

2.4. Single cell RNA-sequencing data analysis

Cell Ranger results were imported into Scanpy (v1.8.2) (196) in Python (v3.8.1). All individuals passed the established quality filters. Any cell with less than 500 genes, more than 10% of reads mapping in mitochondrial genes or more than 30% of reads mapping in ribosomal genes was removed.

In addition, to avoid doublets, any cell with more than 3,000 detected genes was discarded. The *MTRNR2L8* gene was found to be aberrantly expressed in only one individual and was, therefore, excluded from the analysis. Altogether 94,525 CD14+ cells, and 22,637 genes passed the filters.

Normalisation by library size and logarithmic transformation was applied to the resulting UMIs for each cell that passed QC, using the scanpy tools and their default settings. Next, the scanpy cell cycle analysis tool was used, using a list of publicly available cell cycle-related genes [27], assigning to each cell a cell cycle phase. We used 5,000 highly variable genes (HVG) in our downstream analyses, which were selected at this point. To ensure that the results were not biassed by biological variation, the following parameters were regressed out using the scanpy regress_out function: number of UMIs per cell, proportion of reads in mitochondrial and ribosomal genes and cell cycle (S phase and G2M phase). Finally, the resulting data were scaled to unit variance and values exceeding standard deviation 10 were clipped.

Scaled data were then used to perform principal component analysis (PCA), and we used the first 20 PCs to perform a Batch Balanced k-Nearest Neighbour (BBkNN) integration graph, using the individuals as a correction key. This BBkNN graph was later used for embedding and visualisation with the Uniform Manifold Approximation and Projection (UMAP) algorithm [28], as well as for unsupervised clustering with the Leiden algorithm [29]. All the samples were properly integrated and similar clustering results were obtained using alternative algorithms (Supplementary Figures 2 and 3). The CellTypist package (v1.2.0) [30] was used to identify non-monocytic cells, mainly lymphoid cells, which were removed from the data. Additionally, apoptotic cells were identified on the basis of a panel of apoptosis markers

and discarded from further analysis. Finally, 94,525 cells remained in the dataset, on which the analyses described above were repeated.

After unsupervised clustering by the Leiden algorithm, 11 clusters were defined. Finally, the 11 clusters defined by the unsupervised Leiden clustering algorithm were assigned to known CD14+ cell subsets based on genetic markers from the literature (Supplementary Figure 4). However, each cluster was analysed individually to prevent interpretation bias based on previous immunological cell subset definitions.

2.4.1. Differential gene expression

In order to identify genes that can be used as cluster-specific marker genes and to analyse differential gene expression (DE) between cells in the same cluster but originated from different conditions, we applied the rank_genes_groups function implemented in scanpy. Scanpy then calculated differential expression for each gene and ranked them based on their Z-score and the underlying p-value. A Wilcoxon statistical test was applied for DE calculation and the Benjamini-Hochberg FDR (FDR < 0.1) strategy was applied as a correction method. Log2 fold changes were also calculated per group as implemented in the previously mentioned function.

2.4.2. Pathway enrichment analysis

The top 10% cluster marker genes and DE genes were considered for pathway enrichment analysis. The enrichment analysis was performed using the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome databases using R package EnrichR (v3.1.0). A p-value < 0.05 after FDR correction was established as a statistical significance threshold.

2.4.3. Trajectory analysis

To explore potential cell trajectories, we relied on the methods implemented in Monocle3 (v3.0) [31]. For this purpose, the object was first converted to a Seurat (v4.3.0) [32] object, using the SeuratData (v0.2.2) and SeuratDisk (v0.0.09020) packages. The sample was then downsampled to 9,000 cells, and then converted to a cell data set object with the Seurat function as.cell_data_set. The effect sizes of the raw counts were then estimated and the cluster_cell function was applied to perform the clustering and partitioning of the data in order to calculate the cell trajectories. Finally, the pseudotime branches were inferred with the learn_graph function. Based on the expression of genes related to the transdifferentiation of monocytes to macrophages (FCGR3A, CSF1R and RHOC), cluster 0 was chosen as the pseudotime root (this cluster corresponded to cluster 1 according to the monocle3 clustering).

We used the graph_test and find_gene_modules functions implemented in Monocle3 to identify modules of genes that changed with the pseudotime. We applied a multiple testing correction [33] on the results and, if a gene had a q value ≤0.05 and a Moran's I greater than 0.05 [34] (a measure of the degree of autocorrelation), it was considered to be significantly associated with the pseudotime trajectory.

To perform the diffusion mapping, an unsupervised dimensionality reduction analysis package Destiny (v3.12.0) [35] was used. Previously, the SingleCellExperiment package (v1.20.0) [36] was needed to adapt the Seurat object exported from Scanpy.

3. Results

3.1. Peripheral blood CD14+ monocytes show a IFN signature in SSc

In order to comprehensively characterise the pathological alterations of the transcriptome at the single cell level in PB monocytes of SSc patients, we analysed the transcriptome of 94,525 CD14+ cells. As shown in Supplementary Table 1**,** our study cohort was composed of 8 women affected by SSc (6 with lcSSc and 2 with dcSSc) and 8 non-affected women. Patients had several years of disease duration and all of them presented Raynaud's phenomenon and similar drug treatment. The majority of the recruited patients had gastrointestinal complications, but only some of them showed pulmonary involvement. The patients and controls were matched by ethnicity and age.

CD14+ monocytes represent a ~10% of the leukocytes in PB [37] and the number of isolated CD14+ cells per sample was consistent between the controls and the SSc patients, but also between patients with lcSSc and patients with dcSSc (Supplementary Table 2). We detected not only $CD14^{high}$ but also $CD14^{low}CD16^{high}$ cells, which corresponded to the non-classical monocyte population [38]. Therefore, the analysed cells showed a modest to high CD14 expression (Supplementary Figure 1), and an average of 1,658 genes per cell were detected.

After QC, we generated an integrated data set combining the SSc and control CD14+ monocyte transcriptomes. We observed that all samples were evenly distributed (Supplementary Figure 2), but each condition showed a distinct density pattern with qualitative differences between SSc and controls, as well as between lcSSc and dcSSc (Figure 1A). Moreover, the

comparison of the average gene expression between controls and SSc, lcSSc or dcSSc identified the overexpression in cases of 2,665, 2,640 and 1,057 genes, respectively (Figure 1B; Supplementary Figure 5A and 5B; Supplementary Tables 3-5).

The top DE genes (DEG) in SSc compared with controls included interferon response genes, such as *IFITM3*, *IRF1*, *IFITM2* and *IFI6*. Nevertheless, we also observed monocyte migration markers, for example *LGALS2*, and *TMSB10*. Additionally, antigen presentation molecules were remarkably DE, i.e. *HLA-A* and *HLA-DRB5* (Figure 1B; Supplementary Table 3). Furthermore, relevant SSc-associated transcription factors, as *STAT1* and *KLF6*, showed a significantly increased expression in patients with SSc (Figure 1B; Supplementary Table 3). In fact, pathway enrichment analysis highlighted that these SSc upregulated genes were enriched in several proinflammatory mechanisms such as: response to type I and type II IFN, Toll-like receptor signalling and Class I MHC-mediated antigen processing and presentation (Supplementary Table 6). It should also be noted that overexpressed genes in SSc included several key players of the innate immune response, the interferon-induced guanylate binding protein (GBP) family: *GBP1*, *GBP2*, *GBP3*, *GBP4* and *GBP5* (Supplementary Table 3) [39].

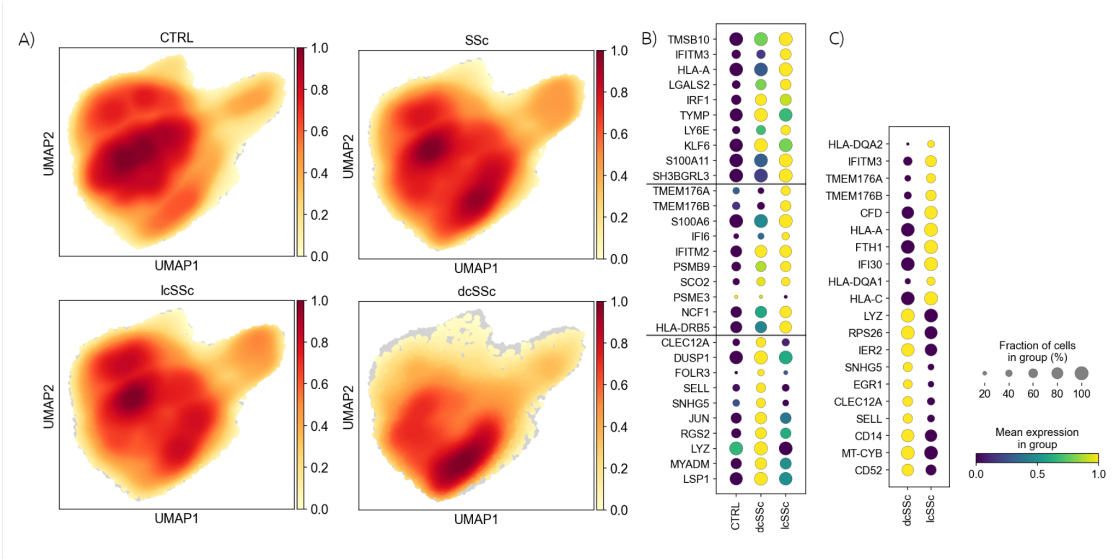


Figure 1. Cellular density and differential gene expression in SSc subtypes and controls. A) UMAP plots showing cellular density of CTRL, lcSSc, and dcSSc. Colour gradient indicates increasing density. B) Top 10 differentially expressed genes in SSc vs. CTRL, lcSSc vs. CTRL, and dcSSc vs. CTRL. C) Top 10 differentially expressed genes between lcSSc and dcSSc, and vice versa. Point size represents the fraction of cells per group expressing each gene, and colour represents the expression level of each gene in each group. SSc: Systemic Sclerosis; lcSSc: limited cutaneous SSc; dcSSc: diffuse cutaneous SSc; CTRL: controls.

The comparison between lcSSc patients and controls revealed that the general biological pathways enriched with DE genes were very similar to the observed trends for SSc (Supplementary Table 4). However, we detected subtype specific overexpression of genes encoding proteins of the complement cascade (such as, *CFD*) in lcSSc (Supplementary Table 4). On the contrary, patients with dcSSc showed increased levels of the monocyte activation related genes, such as *LYZ, CSF3R* (Supplementary Table 5).

Additionally, when lcSSc and dcSSc were compared, we identified 1,156 genes upregulated in lcSSc and 636 genes upregulated in dcSSc (Supplementary Table 7). Compared with dcSSc, the lcSSc subtype showed an overexpression of some interferon response genes (i.e. *S100A4*, *IFNGR2*), subtype-specific increased levels of negative regulators of dendritic cell differentiation (*TMEM176A* and *TMEM176B*) [40], and an enhanced antigen presentation profile with several HLA genes amongst the most DEG (Figure 1C; Supplementary Table 7). On the other hand, the top dcSSc upregulated genes compared to lcSSc were involved in monocyte migration (*SELL* [41], *CD52* [42], *VCAN* [43]) and in monocyte differentiation, *CLEC12A* [44] (Figure 1B-C, Supplementary Table 7). These DEGs were also enriched in interferon-related pathways and, additionally in intercellular communication, for example: immunoregulatory interactions between a lymphoid and a non-lymphoid cell (Supplementary Table 8). In addition, RUNX3-mediated immune response and migration were in the top enriched pathways (Supplementary Table 8).

3.2. Interferon activated non-classical CD16+ RHOC+ monocytes express high levels of microsomal prostaglandin E2 synthase-1 in SSc

Eleven cell clusters were defined on the basis of transcriptional similarity by implementing the community detection Leiden algorithm in an UMAP (Figure 2A) [45]. All the individuals contributed to all clusters (Figure 2B) and there was no cluster restricted to SSc or the SSc subtypes (Supplementary Figure 2). Notably, the implementation of the Louvain algorithm on a t -distributed stochastic neighbour embedding (t-SNE) visualisation resulted in similar clusters (Supplementary Figure 3).

Panels of known membrane surface markers allowed us to manually annotate clusters into the three major monocyte subsets: cMo (*CD14*$^{+}$/*SELL*, 5 clusters), iMo (*HLA-DRA*/*CD74*$^{+}$, 4 clusters) and ncMo (*FCGR3A*$^{+}$/*C1QA*$^{+}$, 1 cluster) (Figure 2C; Supplementary Figure 4). We also detected a DC2 population (*CD1C*$^{+}$ *FCER1A*$^{+}$/*CLEC10A*$^{+}$, 1 cluster).

We observed that the ncMo compartment (cluster 7) was significantly overrepresented in SSc patients, and especially in lcSSc cases (Figure 2D; Supplementary Figure 6). NcMo are CD16+ cells, which was consistent with cluster 7 showing the highest expression of the *CD16* encoding gene, i.e. *FCGR3A* (Figure 2D, Supplementary Figure 4).

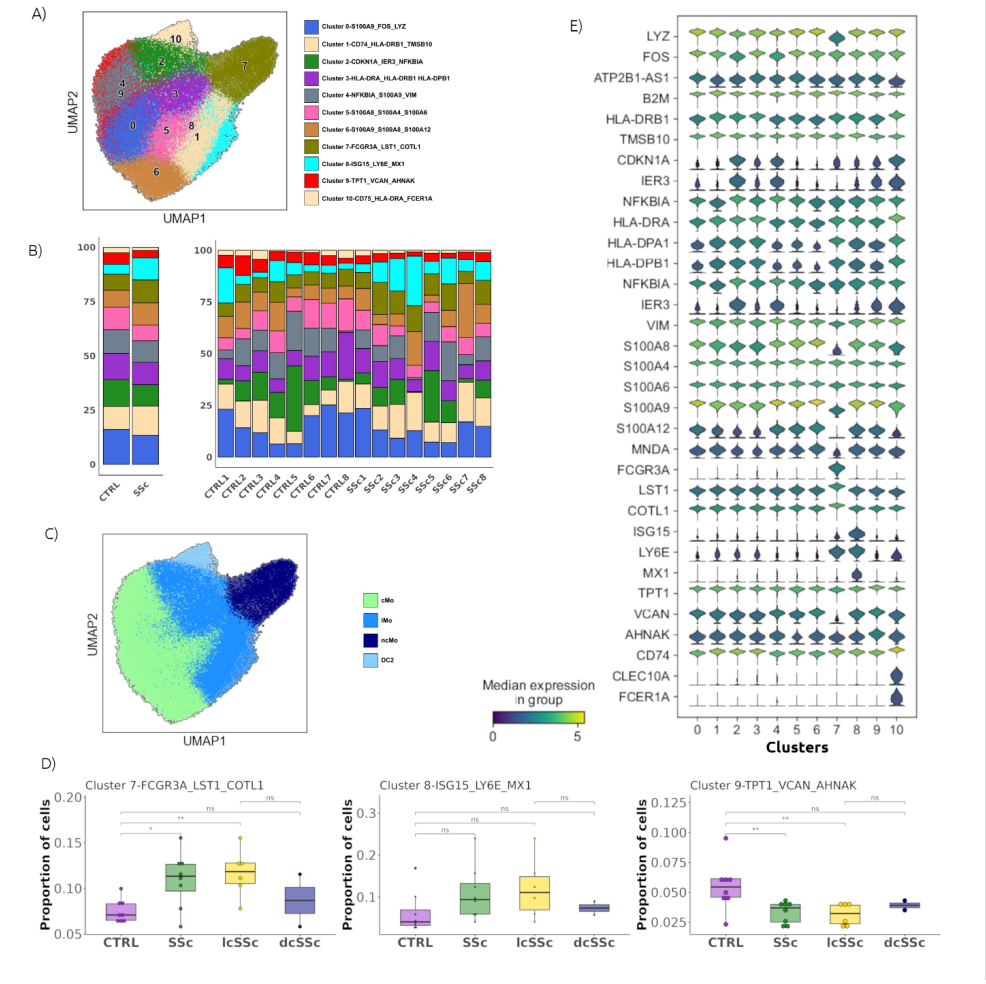


Figure 2. A) UMAP of the 11 CD14+ cell clusters from SSc and control samples, obtained using Leiden clustering and labelled from 0 to 10. B) Proportion of cells in each cluster by condition (CTRL or SSc) and by individual. Each cluster is represented by the same colour as in panel A. C) UMAP with the clusters classified and coloured according to the assigned cell type based on their expression of different marker genes. Clusters 0, 5, 4, 6, and 9 were classified as cMo; clusters 1, 2, 3, and 8 as iMo; cluster 7 as ncMo; and cluster 10 as DC2. D) Boxplots representing

the cell proportions for CTRL, SSc, lcSSc, and dcSSc (from left to right) in clusters 7, 8, and 9 . E) Violin plots of the top 3 DE genes of each cluster vs the rest, with colours representing the expression level in each group. cMo: classical monocytes; iMo: intermediate monocytes; ncMo: non-classical monocytes; DC2: dendritic cells type 2; DE: differential expression.

This ncMo cluster was characterised by high expression of *LST1*, which was also overexpressed in the SSc cells of this cluster (Figure 2D, Supplementary Table 9). *LST1* encodes a trans-membrane and soluble protein induced by immune response against bacteria and associated with the inhibition of lymphocyte proliferation [46] . Moreover, cell motility-related genes, such as *COTL1* [47] and *RHOC* [48], were also clear markers for this subset of monocytes (Figure 2D, Supplementary Table 9). It should be noted that *RHOC* was differentially expressed between SSc and controls, as well as being overexpressed in lcSSc cases compared with dcSSc patients. Although *RHOC* was a cluster marker (logFold change = 4.3) for ncMos (cluster 7), it was differentially expressed between SSc and controls in ncMo but also in the nearby subsets of antigen-presenting iMos (clusters 2 and 3) (Figure 3A; Supplementary Tables 3, 7, 9 and 10).

The interferon signature in ncMo was clear with several interferon induced genes, such as *IFITM3* and *IFITM2* in the top gene markers and differentially overexpressed in SSc patients (Supplementary Tables 9 and 10). The DEGs with the largest log fold changes (logFold change >1.5) included very promising *loci* related to SSc-associated fibrosis. For example, *PTGES*, which encodes an inducible microsomal enzyme that acts downstream from

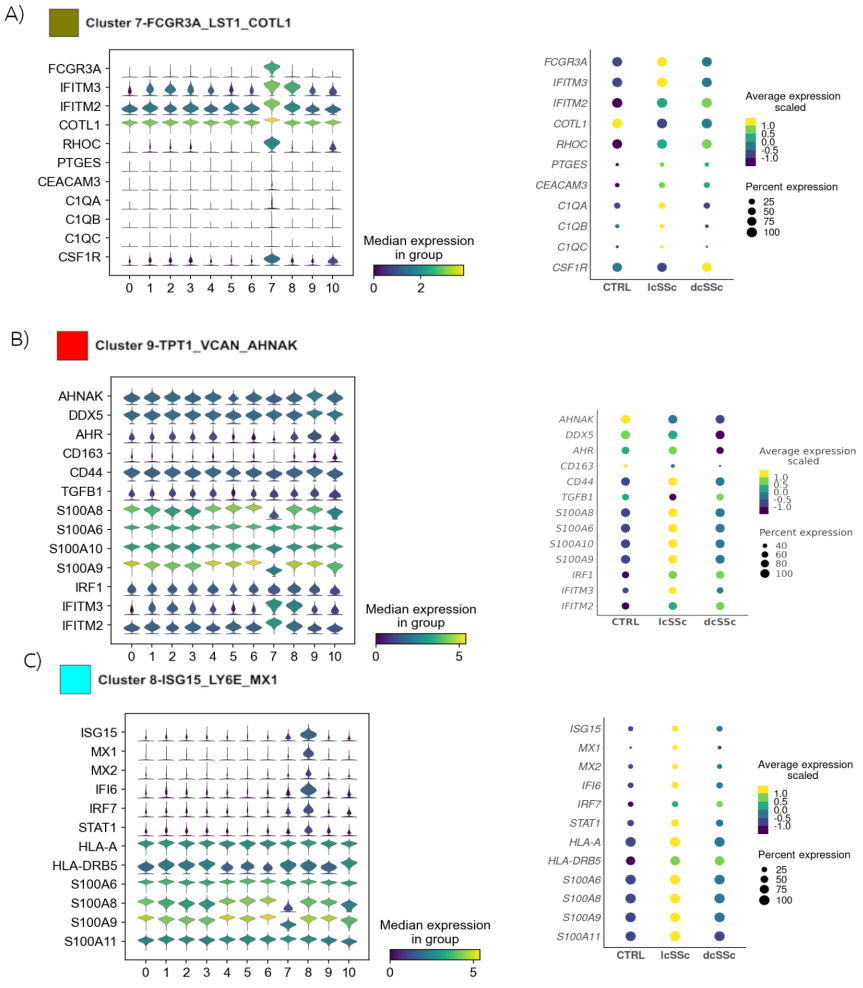


Figure 3. Violin plots and dotplots of gene expression in A) cluster 7, B) cluster 9 and C) cluster 8. The height of each violin indicates the cell proportion of each cluster and colours indicate expression levels and DE between controls and lcSSc and dcSSc is depicted in the dotpots.

cyclooxygenase-2 and catalyses the prostaglandin 2 (PGE2) synthesis [49] or *CEACAM3*, a cellular adhesion molecule [50], were exclusively DE in this cluster of ncMo (Figure 3A; Supplementary Table 10).Finally, several complement system genes, such as *C1QA*, *C1QB*, *C1QC* and *CSF1R*, showed the greatest fold change increases in the ncMo cluster compared with the rest of the CD14+ cells (Figure 3A; Supplementary Table 10).

3.3. Migration of M2 polarised monocytes is altered in SSc

We observed that cluster 9 was underrepresented in individuals affected by SSc, which was especially visible in lcSSc patients (Figure 2D). Cluster 9, a cMo subset, was characterised by a high expression of genes related to cell adhesion and migration, such as *VCAN*, *CD36*, *VIM* and *ITGB2* (Supplementary Table 9). Moreover, pathway enrichment analysis of the marker genes showed that the most relevant pathways also included cell surface interactions at vascular wall and several pathways related to extracellular matrix composition, (such as chondroitin sulfate/dermatan sulfate metabolism, diseases associated with glycosaminoglycan metabolism, etc.) and fibrosis (interleukin-4 and interleukin-13 signalling) [51] (Supplementary Table 11).

Remarkably, this cluster showed markers of monocyte activation (*AHNAK* [52], *DDX5* [53,54], and *CD44* [55]). Especially, we identified several markers of polarisation towards a M2 profibrotic phenotype [56], i.e. *AHR* [52], *TGFB1* and *CD163* [57] (Figure 3B and Supplementary Figure 7). Additionally, we observed some M1 markers among the cluster 9 markers (Supplementary Figure 7), such as *NLRP3* and *IL1B* (Supplementary Table 9) [58]. However, *NLRP3* had a higher expression in controls than in SSc (logFold change = -0,23) and *IL1B* was not differentially expressed

(Supplementary Table 9). Finally, the SSc cMos in this cluster showed a high overexpression of the S100A gene family (*S100A8*, *S100A6*, *S100A10*, *S100A9*) and other interferon-response genes (*IRF1*, *IFITM3*, *IFITM2*) (Figure 3B; Supplementary Table 10).

Then, we analysed the composition and characteristics of an iMo IFN signature-related cluster: cluster 8. SSc patients contributed more to this iMo cluster, which was marked by very high expression of genes related to IFN induction, such as *ISG15*, *MX1*, *MX2* and *IFI6* (Figures 2D and 3C). It should be noted that two master regulators of IFN-mediated immune activation, which have been previously involved in SSc pathogenesis and appeared DE in our comparison between all cells from SSc patients versus the control cells, *IRF7* and *STAT1*, marked exclusively this cluster (Supplementary Tables 4, 9 and 10). As expected by the cluster marker genes, DE analysis showed that the highest over-expression corresponded to MHC-I (*HLA-A*) and MHC-II (*HLA-DRB5*) genes and members of the S100A family (*S100A8*/*S100A9*, *S100A6* and *S100A11*), all markers of monocyte activation and inflammation (Figure 3C).

3.4. The expression of SSc genetic risk *loci* is altered in SSc CD14+ monocytes

Considering our success in establishing *IRF7* and *STAT1*, known SSc genetic risk factors, as cluster specific markers, we checked the expression of other known SSc risk *loci* [3].

We observed that *CSK*, *RAB2A*, *TSPAN32*, *GRB2*, *IL12RB1*, *IRF8*, *DDX6*, and *TNIP1* were DE in the comparison between all the SSc cells and all the control cells (Supplementary Table 3). All of them were upregulated in SSc, but only *CSK*, a kinase of the Src family that interacts with the immune-

related *PTPN22* locus, was significantly upregulated in lcSSc versus dcSSc (Supplementary Table 7). Remarkably, for some of these *loci* we were able to characterise cluster specific DEs.

We observed that the expression of *CSK* was increased in patient cMos (clusters 0, 4 and 5) and iMos (clusters 1, 2, 3 and 8), but not ncMos or DC2 (Supplementary Table 10). On the contrary, *RAB2A* locus, which encodes a Rab GTPase involved in intracellular vesicle trafficking, showed DE scattered in several monocyte clusters identified as either cMos (clusters 0 and 5), iMos (cluster 2) or ncMos (cluster 7).

As opposed to *IRF7*, which was a marker gene for clusters 7 and 8 but showed a generalised DE in several monocyte clusters, *IRF8* (also involved in transcriptional regulation via IFN) was a common marker for several clusters and it was upregulated in SSc in two cMo clusters (clusters 0 and 3) (Supplementary Tables 9 and 10). It should be noted that recent reports have identified a CD14+ cell-exclusive 3D chromatin interaction between a SSc-associated SNP, which was located in the vicinity of *IRF8* (rs11117420), and the promoter of this locus using Hi-C study in CD14+ monocytes obtained from SSc patient blood [59].

The *GRB2* locus was exclusively overexpressed in the SSc cells that belonged to cluster 0, and *ARHGAP31* and *TSPAN32* were significantly upregulated in SSc only in cluster 3 (Supplementary Table 10).

Notably, *IL12RB1*, which had been previously identified as a genetic risk locus for SSc [3,60], was highly expressed in the inflammatory SSc ncMos (Supplementary Table 9).

Finally, we observed that only one gene, *ANXA6*, located near the *TNIP1* SSc genetic susceptibility locus [61], showed a decreased expression

in SSc patients compared to controls in all the defined clusters (logFold change ranging 0.17-0.69) (Supplementary Tables 3-5). Additionally, *ANXA6* was a cluster marker for all the cMo clusters, except for cluster 5 (Supplementary Table 9). Interestingly, a decreased expression of *ANXA6* and a physical interaction between the *ANXA6* promoter and a nearby enhancer located in a *TNIP1* intron have been previously described in SSc CD4+ lymphocytes [62]. Moreover, the alleles of rs3792783, a SSc-associated SNP located in *TNIP1* [61], correlated both with the methylation status of the enhancer and with *ANXA6* expression [62]. Therefore, our findings support further investigation of the relevance of an altered expression of this locus in the context of CD14+ and especially in cMos in SSc patients.

3.5. IRF7+ STAT1+ intermediate monocytes show a distinctive IFN-response in SSc

Then, we explored the relationship between the different monocyte clusters and studied the existence of specific gene modules or cellular states that correlated with SSc. Consequently, we carried out a pseudotime analysis.

We observed that the monocyte clusters seemed to gradually differentiate from cMos towards either ncMos or DC2s with iMos acting as a crossroads (Figure 4A). The trajectory root was located in cMo clusters, such as *CD14*+, *SELL*+, and *CD36*+ (Supplementary Figure 4). However, the progression towards the endpoints relied on the expression of genes such as *FCGR3A* (*CD16*) and *FCER1A*, both markers of ncMos and DC2s, respectively (Figure 4B).

Moreover, the cMo clusters located at the start of the trajectory were clearly characterised by markers of an IFN-mediated response, i.e. *S100A12* (Figure 4B), that have already been addressed as SSc biomarkers above.

Nevertheless, the trajectory ended in the ncMo cluster and it was driven by the expression of two SSc-related markers: *RHOC* and *CDKN1C* (Supplementary Tables 3, 9 and 10). Both genes further suggested an overreactivity of ncMos in SSc, since the protein encoded by *RHOC* is key in the regulation of cell motility and *CDKN1C*, also known as *P57*, acts as a cyclin-dependent tumour suppressor.

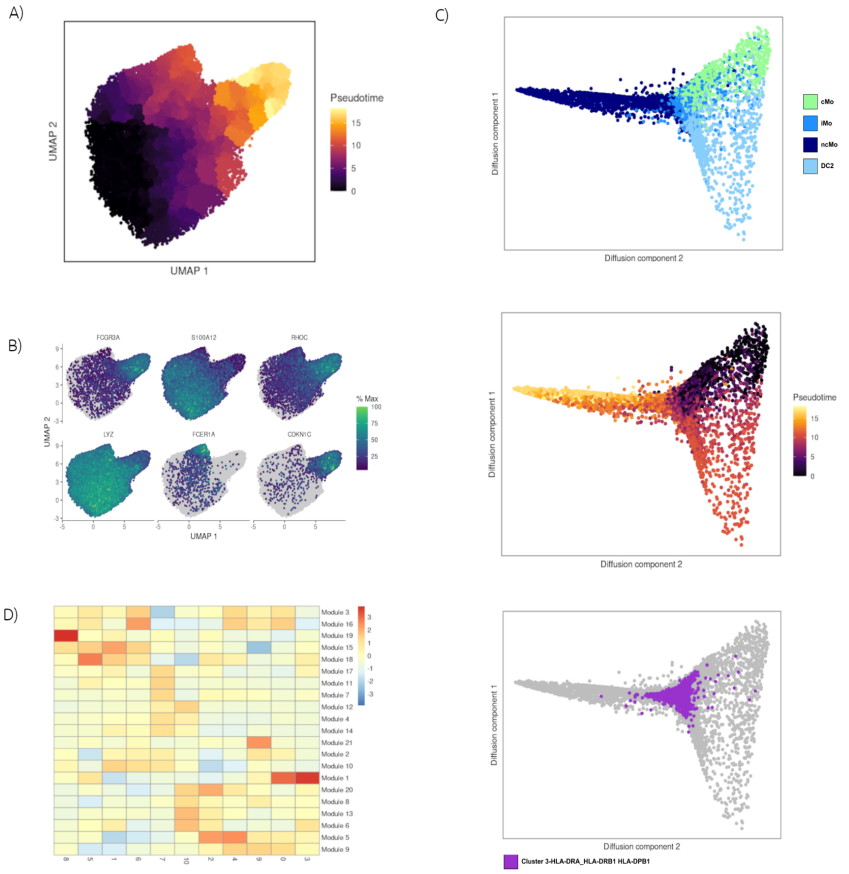


Figure 4. A) UMAP showing CD14+ cells from SSc and CTRLs coloured by pseudotime. B) Expression of the top 6 genes in the CD14+ pseudotime C) Diffusion maps coloured by cell type, pseudotime, and cluster 3 (from top to bottom). D) Heatmap of gene module expression per cluster. Numbers indicate clusters.

Then, we took into account that the estimated pseudotime seemed to have at least two different endpoints (Figure 4A) and decided to calculate a diffusion map to identify the branching points and connections between the different cell clusters. The diffusion map confirmed that ncMos and DC2s were established as two clearly different branches and, although the different iMos clusters were located between the cMos, and the ncMos or the DCs, cluster 3 appeared as an intersection between cMos and their polarisation to ncMos or DC2s (Figure 4B and C; Supplementary Figure 8). Although the transit from cMos to DC2s seemed gradual, the polarisation from cMos to ncMos looked linear with cluster 3 as a bottleneck (Figure 4C). Cluster 3 was characterised by a very high expression of HLA class II genes, which were also upregulated by the SSc cells in this cluster (Supplementary Tables 9 and 10).

Finally, we identified modules of co-expressed genes in the trajectory and we focused on those that showed a cluster specific pattern (Figure 4D). Module 1 was characteristic of the previously mentioned crossroads iMos cluster, cluster 3 (Figure 4D). This gene module was shaped by a variety of genes such as *MTHFR* (coding gene for the methylenetetrahydrofolate reductase enzyme) and *CELA2A* (which encodes a chymotrypsin like elastase) (Supplementary Table 12). But pathway enrichment suggested that this module might be correlated with signalling via *IL1R* (including *loci* such as *JUN*, *TGFB2*, *IL1RN*) and might be involved in the altered proportions of monocyte subsets in SSc.

On the contrary, module 19 was integrated by *ISG15* together with several proteins of the GBP family (*GBP1*, *GBP2*, *GBP3*) and IFN-induced genes (*IFI6*, *IFI44*, *IFI44L*) that were previously described as DEG in SSc (Supplementary Table 3 and 12). *ISG15* was a specific marker gene for cluster 8 (*IRF7*+ *STAT1*+ iMos), which was overrepresented in SSc and

showed the highest expression of module 19 (Figures 2D, 3C and 4D), which allowed us to restrict this particular IFN-response to a specific subset of SSc iMos.

4. Discussion

In specific pathogenic conditions, as in the SSc fibrotic tissue, macrophages are essential for the activation of profibrotic myofibroblasts [4,5]. However, still in peripheral blood circulation, SSc Mos show altered composition and expression profiles [16,63]. As opposed to affected tissue, blood is abundant and easily accessible, and it is often an appropriate biomarker for disease monitoring. Interestingly, single cell transcriptome analysis of PB immune cells and Mos has been fruitful to identify unique cell populations and disease activity-related profiles in IMDs [64,65]. Despite their central role in SSc pathogenesis, the circulating Mo compartment transcriptome had never been characterised at the single cell level before.

This study analysed the largest number of the circulating CD14+ cells (over 90,000 cells) in SSc patients compared to healthy controls. We prioritised identifying rare cell clusters and comprehensively characterising the differences between clusters over addressing interindividual variability. The reported findings provided valuable insights into CD14+ cellular heterogeneity and dynamics, and to identify disease markers in SSc. Nevertheless, the main limitation of the study is the number of studied individuals (8 SSc patients and 8 CTRLs) and further replication in larger independent cohorts are needed to validate the subtype-specific findings. Therefore, we consider that the comparisons comprising disease subtypes should be treated with caution, especially in the case of dcSSc.

Reassuringly, we observed an overrepresentation of inflammatory ncMos (cluster 7) (Figure 2D) as previously described [15,17]. Recently, Carvalheiro et al. described an increased frequency of *CXCL10*-producing ncMo in SSc and an elevated frequency of *CXCL8*-producing ncMos upon stimulation [66]. We observed that the ncMos in Cluster 7 exhibited a very high CXCL10 expression (logFold change = 1.3)

compared to other clusters (Supplementary Table 9) and that *CXCL8* expression was increased in SSc ncMos and iMo clusters as well (Supplementary Table 10).

Additionally, the identified ncMo cluster in our dataset showed a very high expression of known IFN-induced markers produced by myeloid cells in SSc fibrotic skin, such as *S1008A*/*S100A9*, which stimulate keratinocyte secretion of *CXCL2* and *CXCL3*, as well as *IL-6* (a known SSc hallmark) [67]. Consistent with a IFN-mediated effect in ncMos, SSc patient treatment with anifrolumab (a human monoclonal antibody against the interferon-α/β receptor subunit 1) in clinical trials correlated with a decreased *TGF-β* fibrosis and reduced expression levels of some of the ncMo markers observed in our study, such as *TGB1*, *CXCL10* and *B2M* [68] (Supplementary Tables 9-10). Therefore, the main role of the IFN-mediated ncMo activation might be to influence the cytokine profile of ncMos.

We also hypothesise that IFN might affect ncMo migration. Of particular significance is the upregulation of ncMo tissue migration markers, i.e. *CX3CR1* and *CEACAM3* [50,69] (Supplementary Tables 9-10). *CEACAM3* and several members of its family had been previously associated with SSc as well as correlated with interstitial lung disease, but these previous reports related *CEACAM3* to cMos [67]. Contrarily, our data showed that *CEACAM3* was an exclusive marker for ncMos, while *CEACAM4* was a marker for cMos and iMos clusters (Figure 3B, Supplementary Table 9 and 10). The transcriptomic signature for increased motility that we describe would match a recent study that investigated the biophysical properties of ncMos in SSc patients [16]. Matei et al. found that the ncMo of patients were pathologically more activated and exhibited biophysical characteristics that rendered them more prone to vascular migration and tissue infiltration [16]. While a mechanistic explanation for the IFN signal in SSc ncMos was out of the scope of our study, our findings are consistent with an imbalanced cytokine production and migration of ncMos in this disorder.

Additionally, we discovered that the SSc inflammatory ncMos show increased expression of prostaglandin E synthase, which is also known as mPGES-1 (microsomal Prostaglandin E Synthase-1) and encoded by *PTGES*, in a cluster-

specific fashion (Supplementary Table 9). Although the cytosolic prostaglandin E synthase gene (PTGES3) was a cluster marker for ncMos, only *PTGES* was overexpressed in SSc ncMos compared to control ncMos (Supplementary Table 10). Prostaglandin-2 injections are used as an effective treatment for Raynaud's phenomenon in SSc patients due to its effect as a vasodilator [70], but *PGE2* has a dual effect in inflammation. Depending on its association with different G-protein coupled *PGE2* receptor subtypes, PGE2 shows an anti-inflammatory and pro-resolving activity or it mediates proinflammatory non-resolving immune activation [71]. Remarkably, mPGES-1 is an inducible microsomal enzyme that has been associated with pathological overproduction of PGE2 [72]. We would like to highlight that fibroblast from *PTGES* null mice were resistant to the bleomycin skin fibrosis SSc model, and that *PTGES* has been involved in monocyte/macrophage activation via *PPARG* (a known SSc genetic risk factor [73]), after stimulation with *IL-17* [74,75]. Therefore, considering the recent advances in PTGES-specific inhibition, targeting this molecule specifically in SSc inflammatory ncMos might provide new drug targets for this disease.

Notably, a scRNA-seq analysis of SLE PBMCs reported that the Mo compartment showed the highest interferon-stimulated gene expression increase [76], concordantly with the large IFN-signature gene expression profile observed in our study. Interestingly, a SLE-specific cluster that was integrated by Mos expressing high levels of IFN-induced genes had similar cluster marker genes than a SSc-related cluster, the *ISG15*+*LY6E*+ iMo cluster (cluster 8) (Figure 2D, Supplementary Table 9). Moreover, this cluster showed relevant resemblance to a $C1q^{hi}$ monocyte cluster that was recently identified in a scRNA-seq study in PBMCs of patients with Behçet's disease (BD) [77]. Although we did not find a C1qhi cluster in SSc, the $C1q^{hi}$ monocytes in BD showed hybrid characteristics between the inflammatory SSc ncMos, which showed the largest expression of complement genes, and SSc-related iMos (cluster 8) (Supplementary Table 9). In BD, the *STAT1*-mediated response to IFNɣ was correlated with *IRF1* [78] in this cluster. Notably, *IRF1* was one of the most DEG in all the SSc monocyte clusters, especially in dcSSc patients, and a marker for cluster 8 (Supplementary 3-5, 10 and 9). In addition to *IRF1*, SSc iMos in cluster 8

showed an *IRF7* signal. Of note, genetic variants in the IRF7 locus were associated with SSc [79] and the *STAT1*/*IRF7* axis has also been implicated in fibroblast differentiation into myofibroblasts in SSc skin [11]. Therefore, we consider that our findings support that inflammatory ncMos and IFN-activated iMos in SSc have an aberrant response to IFN that might predispose them to a biassed macrophage polarisation.

As mentioned above, we observed several IFN-induced genes, such as several S100A family members or IFITM proteins, significantly DE in both clinical subtypes of SSc patients (Supplementary Tables 3-5). There is increasing evidence that suggests that high levels of these molecules might be associated with a deregulated monocyte proliferation and migration [67,80,81]. Considering that IFITM proteins, especially *IFITM3*, have been shown to be negatively regulated by mTOR inhibitors [82], our findings might support the emerging role of mTOR inhibition as a promising drug target for SSc [83] and particularly for lcSSc patients, who showed the highest *IFITM3* mRNA levels (Supplementary Table 8).

High levels of galectin-1 and galectin-3 (encoded by *LGALS1* and *LGALS3*, respectively) were previously reported in the sera of SSc patients [84], but for the first time, we found overexpression of *LGALS2* in SSc (Supplementary Tables 3-5 and 10). Remarkably, gal-2 is predominantly expressed in the gastrointestinal tract and it can bind to the surface of different immune lineages [85], including Mos and macrophages [86]. Nevertheless, unlike other galectins, gal-2 is expressed in immune cells only by the myeloid lineage [87]. Furthermore, gal-2 acts through a CD14/toll-like receptor (TLR)-4 pathway (a well-established SSc-related deregulated pathway in fibrotic skin and lung [88]) by altering Mo polarisation towards a proinflammatory phenotype [89]. Considering that a gal-2 antibody treatment has shown promising capacities of altering the polarisation of macrophages in a murine atherosclerosis model [90], these findings might also open new windows for treatment in SSc.

M2 macrophages are known to be increased in SSc skin [91] and to produce high levels of TGF-$\boldsymbol{\beta}$. TGF-$\boldsymbol{\beta}$ is a key profibrotic factor [92], which is known to

activate Mos more intensely in SSc than in healthy controls [93] and to polarise macrophages towards a profibrotic M2 phenotype [94]. Interestingly, we observed a cMo cluster (cluster 9) that showed markers of M2 polarisation, such as *CD163* and TGFB1, and which was depleted in patients with SSc (Figure 2D and Supplementary Figure 7). Furthermore, we observed that the top cluster 9 markers included *AHR* and *CD36* (Supplementary Figure 7, Supplementary Table 9), which might be informative of the role of these monocytes in the tissue. *AHR* has a key role in M1/M2 polarisation and is known to promote M2 polarisation and suppress M1 development [95]. *CD36* is a relevant apoptotic cell receptor and phagocytosis promoter that has been linked with an M2 phenotype and increased fibrosis [96]. Additionally, *IL1B*, a M1 marker that was also present in this cluster (Supplementary Table 9), has been shown to also mediate the activation of M2 macrophages in highly fibrotic skin tissue [97].

All the described findings connect cluster 9 with M2 polarised monocytes being actively recruited to affected connective tissue and are consistent with an altered M1/M2 balance in SSc blood with lower M1 polarisation levels in SSc. The underrepresentation of a highly activated cluster in an IMD might seem counterintuitive, but we hypothesise that it would be due to an increased migration of M2-polarised monocytes to SSc-affected tissue in patients. Remarkably, increased levels of monocyte migration markers, such as *VCAN* and *ITGB2*, were observed in this cluster (Figure 3B and Supplementary Table 9). *VCAN* (also known as versican) expression have been related with increased circulating Mo migration in SSc [98]. Besides, *ITGB2* has been identified as a SSc-associate monocyte gene and found to be upregulated in SSc skin macrophages [38].

Finally, SSc bleomycin mouse models showed that the modulation of M2 cytokine production by *PDE4* inhibition decreased skin fibrosis [99], and we observed high levels of expression of *PDE4* in cluster 9 and other cMos clusters (Supplementary Table 9). Therefore, we propose that specifically blocking the extravasation of the novel *TPT1*+ *VCAN*+ *AHNAK*+ cMo cluster into challenged tissue might benefit SSc patients.

5. Conclusions

In conclusion, we performed the most detailed characterisation of the CD14+ Mo compartment in SSc to date. We confirmed an overrepresentation of CD16+ ncMos at single cell level. Inflammatory SSc ncMos showed a high IFN-response signature and the upregulation of PGE_2 synthesis, monocyte adhesion markers and complement genes. We also identified an aberrant IFN-response in *IRF7*+ *STAT1*+ SSc iMos and, finally, we observed a depletion of M2 polarised cMos in SSc. These results reinforced the role of PB Mos as SSc biomarkers and provided new windows for clinical monitoring and drug targeting.

Acknowledgements

This research is part of the doctoral degree awarded to GV-M, within the Biomedicine program from the University of Granada entitled "Deciphering the genetic basis of systemic sclerosis". We would like to thank Sofia Vargas and Gemma Robledo for their excellent technical assistance and Gonzalo Borrego-Yaniz for his collaboration in the illustration of the cells for the figures. We also appreciate the controls and the affected individuals who generously provided the samples for these studies.

Funding

This work was supported by the grant P18-RT-4442 funded by *Consejería de Transformación Económica, Industria, Conocimiento y Universidades, Junta de Andalucía*. "Red de Investigación Cooperativa Orientada a Resultados en Salud'' (RICOR, RD21/0002/0039). LBC was supported by the Spanish Ministry of Science and Innovation through the *Juan de la Cierva Incorporación*' program (Grant ref. IJC2018-038026-I, funded by MCIN/AEI /10.13039/501100011033), which includes FEDER funds. MAH is a recipient of a Miguel Servet fellowship (CP21/00132) from the *Instituto de Salud Carlos III* (Spanish Ministry of Science and Innovation). GV-M was funded by the Grant PRE2019-087586 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF Investing in your future".

Figure legends

Figure 1. Cellular density and differential gene expression in SSc subtypes and controls. A) UMAP plots showing cellular density of CTRL, lcSSc, and dcSSc. Colour gradient indicates increasing density. B) Top 10 differentially expressed genes in SSc vs. CTRL, lcSSc vs. CTRL, and dcSSc vs. CTRL. C) Top 10 differentially expressed genes between lcSSc and dcSSc, and

vice versa. Point size represents the fraction of cells per group expressing each gene, and colour represents the expression level of each gene in each group. SSc: Systemic Sclerosis; lcSSc: limited cutaneous SSc; dcSSc: diffuse cutaneous SSc; CTRL: controls.

Figure 2. A) UMAP of the 11 CD14+ cell clusters from SSc and control samples, obtained using Leiden clustering and labelled from 0 to 10. B) Proportion of cells in each cluster by condition (CTRL or SSc) and by individual. Each cluster is represented by the same colour as in panel A. C) UMAP with the clusters classified and coloured according to the assigned cell type based on their expression of different marker genes. Clusters 0, 5, 4, 6, and 9 were classified as cMo; clusters 1, 2, 3, and 8 as iMo; cluster 7 as ncMo; and cluster 10 as DC2. D) Boxplots representing the cell proportions for CTRL, SSc, lcSSc, and dcSSc (from left to right) in clusters 7, 8, and 9 . E) Violin plots of the top 3 DE genes of each cluster vs the rest, with colours representing the expression level in each group. cMo: classical monocytes; iMo: intermediate monocytes; ncMo: non-classical monocytes; DC2: dendritic cells type 2; DE: differential expression.

Figure 3. Violin plots and dotplots of gene expression in A) cluster 7, B) cluster 9 and C) cluster 8. Colours indicate expression levels and DE between controls and lcSSc and dcSSc is depicted in the dotpots.

Figure 4. A) UMAP showing CD14+ cells from SSc and CTRLs coloured by pseudotime. B) Expression of the top 6 genes in the CD14+ pseudotime C) Diffusion maps coloured by cell type, pseudotime, and cluster 3 (from top to bottom). D) Heatmap of gene module expression per cluster. Numbers indicate clusters.

Supplementary Figure 1. A) Flow cytometry plot for CD14+ membrane protein labelling resulting from magnetic bead enrichment. B) Combined UMAP of SSc and CTRL for *CD14* expression.

Supplementary Figure 2. A) Combined UMAP of SSc and CTRL samples coloured by individual B) Combined UMAP of SSc and CTRL samples where cells are coloured based on their condition.

Supplementary Figure 3. T-Distributed Stochastic Neighbour Embedding (t-SNE) maps with A) clusters defined by the Leiden algorithm, and B) clusters calculated using the Louvain algorithm.

Supplementary Figure 4. Classification of the different clusters into cellular identities based on classical monocytic markers. A) UMAP coloured according to the clusters obtained by the Leiden algorithm; B) Dotplot of cMo, iMo, ncMo, and DC2 marker genes. The dot size represents the number of cells expressing that gene and the colour intensity represents the median expression level. cMo: classical monocytes; iMo: intermediate monocytes; ncMo: non-classical monocytes; DC2: dendritic cells type 2.

Supplementary Figure 5. A) Top 10 differentially expressed genes in SSc vs. CTRL, lcSSc vs. CTRL, and dcSSc vs. CTRL by individual. B) Top 10 differentially expressed genes between lcSSc and dcSSc, and vice versa by individual. Point size represents the fraction of cells per group expressing each gene, and colour represents the expression level of each gene in each individual.

Supplementary Figure 6. Combined boxplot and violin plot of the cellular composition of each cluster. The proportion of cells from CTRLs (purple) and SSc (green) is represented in each plot. * indicates FDR > 0.1.

Supplementary figure 7. M1 and M2 macrophage marker expression stratified by cluster. The size of the dot represents the percent of cells in each cluster expressing the marker and the color shows the average expression**.**

Supplementary Figure 8. Diffusion map colored by pseudotime and by cluster**.**

Contributors

GVM: data analysis, data interpretation, manuscript drafting, revision and approval; MAH: data interpretation, manuscript revision and approval; EGC: data analysis, manuscript revision and approval; MK: data interpretation, manuscript revision and approval; NOC: data acquisition, manuscript revision and approval; JLCR: data acquisition, manuscript revision and approval; NM: experiments and manuscript revision and approval; SK: data analysis; SB: data analysis, manuscript revision and approval; BT: manuscript revision and approval; LBC: study design, data analysis, manuscript drafting, revision and approval; JM: study design, data interpretation, manuscript drafting, revision and approval.

Competing interests

GVM: none; MAH: none; EGC: none; MK: none; NOC: none; JLCR: none; NM: none; SK: none; SB: none; BT: none; LBC: none; JM: none.

Ethical approval

An ethical protocol was prepared with consensus across all partners and was approved by the local ethical committee of the clinical recruitment centre. The study adhered to the standards set by the International Conference on Harmonization and Good Clinical Practice (ICH-GCP), and to

the ethical principles that have their origin in the Declaration of Helsinki (2013). The protection of the confidentiality of records that could identify the included subjects is ensured as defined by the EU Directive 2001/20/EC and the applicable national and international requirements relating to data protection in each participating country. This study was approved by the Ethical Committee "CEIM/CEI Provincial de Granada", and the CSIC Ethical Committee.

Data availability

scRNA-seq count tables: Files containing scRNA-seq UMI counts and meta-data for all samples in this will be available in Zenodo.

Code availability

All the codes used for processing and analyzing the data in this study were compiled into a single publicly available GitHub repository [https://github.com/gonv/scRNA-seq_SSc_CD14_nb].

Reference

[1] E.R. Volkmann, K. Andréasson, V. Smith, Systemic sclerosis, Lancet. 401 (2023) 304–318.

[2] L. Bossini-Castillo, E. López-Isac, M.D. Mayes, J. Martín, Genetics of systemic sclerosis, Semin. Immunopathol. 37 (2015) 443–451.

[3] E. López-Isac, M. Acosta-Herrera, M. Kerick, S. Assassi, A.T. Satpathy, J. Granja, M.R. Mumbach, L. Beretta, C.P. Simeón, P. Carreira, N. Ortego-Centeno, I. Castellvi, L. Bossini-Castillo, F.D. Carmona, G. Orozco, N. Hunzelmann, J.H.W. Distler, A. Franke, C. Lunardi, G. Moroncini, A. Gabrielli, J. de Vries-Bouwstra, C. Wijmenga, B.P.C. Koeleman, A. Nordin, L. Padyukov, A.M. Hoffmann-Vold, B. Lie, R. Ríos, J.L. Callejas, J.A. Vargas-Hitos, R. García-Portales, M.T. Camps, A. Fernández-Nebro, M.F. González-Escribano, F.J. García-Hernández, M.J. Castillo, M.A. Aguirre, I. Gómez-Gracia, B. Fernández-Gutiérrez, L. Rodríguez-Rodríguez, P. García de la Peña, E. Vicente, J.L. Andreu, M. Fernández de Castro, F.J. López-Longo, L. Martínez, Fonollosa, A. Guillén, G. Espinosa, C. Tolosa,

A. Pros, M. Rodríguez-Carballeira, F.J. Narváez, M. Rubio-Rivas, Ortiz-Santamaría, A.B. Madroñero, M.A. González-Gay, B. Díaz, L. Trapiella, A. Sousa, M.V. Egurbide, P. Fanlo-Mateo, L. Sáez-Comet, F. Díaz, Hernández, E. Beltrán, J.A. Román-Ivorra, E. Grau, J.J. Alegre-Sancho, M. Freire, F.J. Blanco-García, N. Oreiro, T. Witte, A. Kreuter, G. Riemekasten, P. Airó, C. Magro, A.E. Voskuyl, M.C. Vonk, R. Hesselstrand, S. Proudman, W. Stevens, M. Nikpour, J. Zochling, J. Sahhar, J. Roddy, P. Nash, K. Tymms, M. Rischmueller, S. Lester, T. Vyse, A.L. Herrick, J. Worthington, C.P. Denton, Y. Allanore, M.A. Brown, T.R.D.J. Radstake, C. Fonseca, H.Y. Chang, M.D. Mayes, J. Martin, GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways, Nat. Commun. 10 (2019) 1–14.

[4] A. Lescoat, V. Lecureur, J. Varga, Contribution of monocytes and macrophages to the pathogenesis of systemic sclerosis: recent insights and therapeutic implications, Curr. Opin. Rheumatol. 33 (2021) 463–470.

[5] G. Kania, M. Rudnik, O. Distler, Involvement of the myeloid cell compartment in fibrogenesis and systemic sclerosis, Nat. Rev. Rheumatol. 15 (2019) 288–302.

[6] D. Xue, T. Tabib, C. Morse, Y. Yang, R.T. Domsic, D. Khanna, R. Lafyatis, Expansion of Fcγ Receptor IIIa-Positive Macrophages, Ficolin 1-Positive Monocyte-Derived Dendritic Cells, and Plasmacytoid Dendritic Cells Associated With Severe Skin Disease in Systemic Sclerosis, Arthritis Rheumatol. 74 (2022) 329–341.

[7] R. Bhandari, M.S. Ball, V. Martyanov, D. Popovich, E. Schaafsma, S. Han, M. ElTanbouly, N.M. Orzechowski, M. Carns, E. Arroyo, K. Aren, M. Hinchcliff, M.L. Whitfield, P.A. Pioli, Profibrotic Activation of Human Macrophages in Systemic Sclerosis, Arthritis Rheumatol. 72 (2020) 1160–1169.

[8] B. Skaug, D. Khanna, W.R. Swindell, M.E. Hinchcliff, T.M. Frech, V.D. Steen, F.N. Hant, J.K. Gordon, A.A. Shah, L. Zhu, W. Jim Zheng, J.L. Browning, A.M.S. Barron, M. Wu, S. Visvanathan, P. Baum, J.M. Franks, M.L. Whitfield, V.K. Shanmugam, R.T. Domsic, F.V. Castelino, E.J. Bernstein, N. Wareing, M.A. Lyons, J. Ying, J. Charles, M.D. Mayes, S. Assassi, Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile, Annals of the Rheumatic Diseases. 79 (2020) 379–386. https://doi.org/10.1136/annrheumdis-2019-215894.

[9] M. van der Kroef, T. Carvalheiro, M. Rossato, F. de Wit, M. Cossu, E. Chouri, C.G.K. Wichers, C.P.J. Bekker, L. Beretta, N. Vazirpanah, E. Trombetta, T.R.D.J. Radstake, C. Angiolilli, CXCL4 triggers monocytes and macrophages to produce PDGF-BB, culminating in fibroblast activation: Implications for systemic sclerosis, J. Autoimmun. 111 (2020) 102444.

[10] A. Papazoglou, M. Huang, M. Bulik, A. Lafyatis, T. Tabib, C. Morse, J. Sembrat, M. Rojas, E. Valenzi, R. Lafyatis, Epigenetic Regulation of Profibrotic Macrophages in Systemic Sclerosis-Associated Interstitial Lung Disease, Arthritis Rheumatol. 74 (2022) 2003–2014.

[11] T. Tabib, M. Huang, N. Morse, A. Papazoglou, R. Behera, M. Jia, M. Bulik, D.E. Monier, P.V. Benos, W. Chen, R. Domsic, R. Lafyatis, Myofibroblast transcriptome indicates SFRP2hi fibroblast progenitors in systemic sclerosis skin, Nat. Commun. 12 (n.d.). https://doi.org/10.1038/s41467-021-24607-6.

[12] P.F. Alyxzandria M. Gaydosik Tracy Tabib Robyn Domsic Dinesh Khanna Robert Lafyatis, Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis, Ann. Rheum. Dis. 80 (n.d.) 1453–1460.

[13] X. Gao, G. Jia, A. Guttman, D.J. DePianto, K.B. Morshead, K.-H. Sun, N. Ramamoorthi, J.A. Vander Heiden, Z. Modrusan, P.J. Wolters, A. Jahreis, J.R. Arron, D. Khanna, T.R. Ramalingam, Osteopontin Links Myeloid Activation and Disease Progression in Systemic Sclerosis, Cell Rep Med. 1 (2020) 100140.

[14] L. Schneider, N.A. Marcondes, V. Hax, I.F. da Silva Moreira, C.Y. Ueda, R.R. Piovesan, R. Xavier, R. Chakr, Flow cytometry evaluation of CD14/CD16 monocyte subpopulations in systemic sclerosis patients: a cross sectional controlled study, Adv Rheumatol. 61 (2021) 27.

[15] A. Lescoat, V. Lecureur, M. Roussel, B.L. Sunnaram, A. Ballerie, G. Coiffier, S. Jouneau, O. Fardel, T. Fest, P. Jégo, CD16-positive circulating monocytes and fibrotic manifestations of systemic sclerosis, Clin. Rheumatol. 36 (2017) 1649–1654.

[16] A.-E. Matei, M. Kubánková, L. Xu, A.-H. Györfi, E. Boxberger, D. Soteriou, M. Papava, J. Prater, X. Hong, C. Bergmann, M. Kräter, G. Schett, J. Guck, J.H.W. Distler, Biophysical phenotyping of circulating immune cells identifies a distinct monocyte-driven signature in systemic sclerosis, Arthritis Rheumatol. (2022).

https://doi.org/10.1002/art.42394.

[17] C. Gur, S.-Y. Wang, F. Sheban, M. Zada, B. Li, F. Kharouf, H. Peleg, S. Aamar, A. Yalin, D. Kirschenbaum, Y. Braun-Moscovici, D.A. Jaitin, T. Meir-Salame, E. Hagai, B.K. Kragesteen, B. Avni, S. Grisariu, C. Bornstein, S. Shlomi-Loubaton, E. David, R. Shreberk-Hassidim, V. Molho-Pessach, D. Amar, T. Tzur, R. Kuint, M. Gross, O. Barboy, A. Moshe, L. Fellus-Alyagor, D. Hirsch, Y. Addadi, S. Erenfeld, M. Biton, T. Tzemach, A. Elazary, Y. Naparstek, R. Tzemach, A. Weiner, A. Giladi, A. Balbir-Gurman, I. Amit, LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma, Cell. 185 (2022) 1373–1388.e20.

[18] M. Rudnik, F. Rolski, S. Jordan, T. Mertelj, M. Stellato, O. Distler, P. Blyszczuk, G. Kania, Regulation of Monocyte Adhesion and Type I Interferon Signaling by CD52 in Patients With Systemic Sclerosis, Arthritis Rheumatol. 73 (2021) 1720–1730.

[19] R. Lee, C. Reese, B. Perry, J. Heywood, M. Bonner, M. Zemskova, R.M. Silver, S. Hoffman, E. Tourkina, Enhanced chemokine-receptor expression, function, and signaling in healthy African American and scleroderma-patient monocytes are regulated by caveolin-1, Fibrogenesis Tissue Repair. 8 (2015) 11.

[20] S. Soldano, A.C. Trombetta, P. Contini, V. Tomatis, B. Ruaro, R. Brizzolara, P. Montagna, A. Sulli, S. Paolino, C. Pizzorni, V. Smith, M. Cutolo, Increase in circulating cells coexpressing M1 and M2 macrophage surface markers in patients with systemic sclerosis, Ann. Rheum. Dis. 77 (2018) 1842–1845.

[21] A. Lescoat, A. Ballerie, S. Jouneau, O. Fardel, L. Vernhet, P. Jego, V. Lecureur, M1/M2 polarisation state of M-CSF blood-derived macrophages in systemic sclerosis, Ann. Rheum. Dis. 78 (2019) e127.

[22] A.C. Trombetta, S. Soldano, P. Contini, V. Tomatis, B. Ruaro, S. Paolino, R. Brizzolara, P. Montagna, A. Sulli, C. Pizzorni, V. Smith, M. Cutolo, A circulating cell population showing both M1 and M2 monocyte/macrophage surface markers characterizes systemic sclerosis patients with lung involvement, Respiratory Research. 19 (2018). https://doi.org/10.1186/s12931-018-0891-z.

[23] Preliminary criteria for the classification of systemic sclerosis (scleroderma). Subcommittee for scleroderma criteria of the American Rheumatism Association Diagnostic and Therapeutic Criteria

Committee, Arthritis Rheum. 23 (1980) 581–590.

[24] E.C. LeRoy, T.A. Medsger Jr, Criteria for the classification of early systemic sclerosis, J. Rheumatol. 28 (2001) 1573–1576.

[25] E.C. LeRoy, C. Black, R. Fleischmajer, S. Jablonska, T. Krieg, T.A. Medsger Jr, N. Rowell, F. Wollheim, Scleroderma (systemic sclerosis): classification, subsets and pathogenesis, J. Rheumatol. 15 (1988) 202–205.

[26] F.A. Wolf, P. Angerer, F.J. Theis, SCANPY: large-scale single-cell gene expression data analysis, Genome Biol. 19 (2018) 15.

[27] E.Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A.R. Bialas, N. Kamitaki, E.M. Martersteck, J.J. Trombetta, D.A. Weitz, J.R. Sanes, A.K. Shalek, A. Regev, S.A. McCarroll, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets, Cell. 161 (2015) 1202–1214.

[28] L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, (2018). https://doi.org/10.48550/ARXIV.1802.03426.

[29] V.A. Traag, L. Waltman, N.J. van Eck, From Louvain to Leiden: guaranteeing well-connected communities, Sci. Rep. 9 (2019) 5233.

[30] C. Domínguez Conde, C. Xu, L.B. Jarvis, D.B. Rainbow, S.B. Wells, T. Gomes, S.K. Howlett, O. Suchanek, K. Polanski, H.W. King, L. Mamanova, N. Huang, P.A. Szabo, L. Richardson, L. Bolt, E.S. Fasouli, K.T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z.K. Tuong, L. Campos, H.S. Mousa, E.J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D.K. Menon, O.A. Bayraktar, L.K. James, K.B. Meyer, N. Yosef, M.R. Clatworthy, P.A. Sims, D.L. Farber, K. Saeb-Parsy, J.L. Jones, S.A. Teichmann, Cross-tissue immune cell analysis reveals tissue-specific features in humans, Science. 376 (2022) eabl5197.

[31] J. Cao, M. Spielmann, X. Qiu, X. Huang, D.M. Ibrahim, A.J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F.J. Steemers, C. Trapnell, J. Shendure, The single-cell transcriptional landscape of mammalian organogenesis, Nature. 566 (2019) 496–502.

[32] Y. Hao, S. Hao, E. Andersen-Nissen, W.M. Mauck 3rd, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E.P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.M. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, R.

Satija, Integrated analysis of multimodal single-cell data, Cell. 184 (2021) 3573–3587.e29.

[33] J.D. Storey, The positive false discovery rate: a Bayesian interpretation and the q-value, Ann. Stat. 31 (2003) 2013–2035.

[34] P.A.P. Moran, Notes on continuous stochastic phenomena, Biometrika. 37 (1950) 17–23.

[35] P. Angerer, L. Haghverdi, M. Büttner, F.J. Theis, C. Marr, F. Buettner, destiny – diffusion maps for large-scale single-cell data in R, (n.d.). https://doi.org/10.1101/023309.

[36] R.A. Amezquita, V.J. Carey, L.N. Carpp, L. Geistlinger, A.T.L. Lun, F. Marini, K. Rue-Albrecht, D. Risso, C. Soneson, L. Waldron, H. Pagès, M. Smith, W. Huber, M. Morgan, R. Gottardo, S.C. Hicks, Orchestrating Single-Cell Analysis with Bioconductor, (n.d.). https://doi.org/10.1101/590562.

[37] C. Auffray, M.H. Sieweke, F. Geissmann, Blood monocytes: development, heterogeneity, and relationship with dendritic cells, Annu. Rev. Immunol. 27 (2009) 669–692.

[38] H.-K.M. Makinde, J.L.M. Dunn, G. Gadhvi, M. Carns, K. Aren, A.H. Chung, L.N. Muhammad, J. Song, C.M. Cuda, S. Dominguez, J.E. Pandolfino, J.E. Dematte D'Amico, G.S. Budinger, S. Assassi, T.M. Frech, D. Khanna, A. Shaeffer, H. Perlman, M. Hinchcliff, D.R. Winter, Three distinct transcriptional profiles of monocytes associate with disease activity in Scleroderma patients, Arthritis Rheumatol. 75 (2023) 595–608.

[39] K. Tretina, E.-S. Park, A. Maminska, J.D. MacMicking, Interferon-induced guanylate-binding proteins: Guardians of host defense in health and disease, J. Exp. Med. 216 (2019) 482–500.

[40] M. Hill, S. Russo, D. Olivera, M. Malcuori, G. Galliussi, M. Segovia, The intracellular cation channel TMEM176B as a dual immunoregulator, Front Cell Dev Biol. 10 (2022) 1038429.

[41] A. Ivetic, H.L. Hoskins Green, S.J. Hart, L-selectin: A Major Regulator of Leukocyte Adhesion, Migration and Signaling, Front. Immunol. 10 (2019) 1068.

[42] M. Rudnik, F. Rolski, S. Jordan, T. Mertelj, M. Stellato, O. Distler, P. Blyszczuk, G. Kania, Regulation of Monocyte Adhesion and

Type I Interferon Signaling by CD52 in Patients With Systemic Sclerosis, Arthritis Rheumatol. 73 (2021) 1720–1730.

[43] M. Frankenberger, T.P.J. Hofer, A. Marei, F. Dayyani, S. Schewe, C. Strasser, A. Aldraihim, F. Stanzel, R. Lang, R. Hoffmann, O. Prazeres da Costa, T. Buch, L. Ziegler-Heitbrock, Transcript profiling of CD16-positive monocytes reveals a unique molecular fingerprint, Eur. J. Immunol. 42 (2012) 957–974.

[44] A.S.J. Marshall, J.A. Willment, E. Pyz, K.M. Dennehy, D.M. Reid, P. Dri, S. Gordon, S.Y.C. Wong, G.D. Brown, Human MICL (CLEC12A) is differentially glycosylated and is down-regulated following cellular activation, Eur. J. Immunol. 36 (2006) 2159–2169.

[45] L. McInnes, J. Healy, J. Melville, UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, (2018). https://doi.org/10.48550/ARXIV.1802.03426.

[46] I. Rollinger-Holzinger, B. Eibl, M. Pauly, U. Griesser, F. Hentges, B. Auer, G. Pall, P. Schratzberger, D. Niederwieser, E.H. Weiss, H. Zwierzina, LST1: A Gene with Extensive Alternative Splicing and Immunomodulatory Function1, J. Immunol. 164 (2000) 3169–3176.

[47] P. Provost, J. Doucet, A. Stock, G. Gerisch, B. Samuelsson, O. Rådmark, Coactosin-like protein, a human F-actin-binding protein: critical role of lysine-75, Biochem. J. 359 (2001) 255–263.

[48] J.J. Bravo-Cordero, M. Oser, X. Chen, R. Eddy, L. Hodgson, J. Condeelis, A novel spatiotemporal RhoC activation pathway locally regulates cofilin activity at invadopodia, Curr. Biol. 21 (2011) 635–644.

[49] M.R. McCann, R. Monemdjou, P. Ghassemi-Kakroodi, H. Fahmi, G. Perez, S. Liu, X. Shi-Wen, S.K. Parapuram, F. Kojima, C.P. Denton, D.J. Abraham, J. Martel-Pelletier, L.J. Crofford, A. Leask, M. Kapoor, mPGES-1 null mice are resistant to bleomycin-induced skin fibrosis, Arthritis Res. Ther. 13 (2011) R6.

[50] K. Yokoyama, H. Mitoma, S. Kawano, Y. Yamauchi, Q. Wang, M. Ayano, Y. Kimoto, N. Ono, Y. Arinobu, K. Akashi, T. Horiuchi, H. Niiro, CEACAM 1, 3, 5 and 6 -positive classical monocytes correlate with interstitial lung disease in early systemic sclerosis, Front. Immunol. 13 (2022) 1016914.

[51] A. Mantovani, S.K. Biswas, M.R. Galdiero, A. Sica, M. Locati, Macrophage plasticity and polarization in tissue repair and

remodelling, J. Pathol. 229 (2013) 176–185.

[52] A. de Morrée, B. Flix, I. Bagaric, J. Wang, M. van den Boogaard, L. Grand Moursel, R.R. Frants, I. Illa, E. Gallardo, R. Toes, S.M. van der Maarel, Dysferlin regulates cell adhesion in human monocytes, J. Biol. Chem. 288 (2013) 14147–14157.

[53] M.C. Gingras, J.F. Margolin, Differential expression of multiple unexpected genes during U937 cell and macrophage differentiation detected by suppressive subtractive hybridization, Exp. Hematol. 28 (2000) 65–76.

[54] R.M. Nyamao, J. Wu, L. Yu, X. Xiao, F.-M. Zhang, Roles of DDX5 in the tumorigenesis, proliferation, differentiation, metastasis and pathway regulation of human malignancies, Biochim. Biophys. Acta Rev. Cancer. 1871 (2019) 85–98.

[55] J. Rachmilewitz, M.L. Tykocinski, Differential effects of chondroitin sulfates A and B on monocyte and B-cell activation: evidence for B-cell activation via a CD44-dependent pathway, Blood. 92 (1998) 223–229.

[56] A. Bhattacharjee, M. Shukla, V.P. Yakubenko, A. Mulya, S. Kundu, M.K. Cathcart, IL-4 and IL-13 employ discrete signaling pathways for target gene expression in alternatively activated monocytes/macrophages, Free Radic. Biol. Med. 54 (2013) 1–16.

[57] F. Zhang, H. Wang, X. Wang, G. Jiang, H. Liu, G. Zhang, H. Wang, R. Fang, X. Bu, S. Cai, J. Du, TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype, Oncotarget. 7 (2016) 52294–52306.

[58] Y. He, H. Hara, G. Núñez, Mechanism and Regulation of NLRP3 Inflammasome Activation, Trends Biochem. Sci. 41 (2016) 1012–1021.

[59] D. González-Serna, C. Shi, M. Kerick, J. Hankinson, J. Ding, A. McGovern, M. Tutino, G. Villanueva Martin, N. Ortego-Centeno, J.L. Callejas, J. Martin, G. Orozco, Functional genomics in primary T cells and monocytes identifies mechanisms by which genetic susceptibility loci influence systemic sclerosis risk, Arthritis Rheumatol. (2022). https://doi.org/10.1002/art.42396.

[60] E. López-Isac, L. Bossini-Castillo, S.G. Guerra, C. Denton, C. Fonseca, S. Assassi, X. Zhou, M.D. Mayes, C.P. Simeón, N. Ortego-Centeno, I. Castellví, P. Carreira, Spanish Scleroderma Group, O. Gorlova, L.

Beretta, A. Santaniello, C. Lunardi, R. Hesselstrand, A. Nordin, G. Riemekasten, T. Witte, N. Hunzelmann, A. Kreuter, J.H.W. Distler, A.E. Voskuyl, J. de Vries-Bouwstra, B.P. Koeleman, A. Herrick, J. Worthington, T.R.D.J. Radstake, J. Martin, Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus, Arthritis Rheumatol. 66 (2014) 3521–3523.

[61] E. López-Isac, European Scleroderma Group†, M. Acosta-Herrera, M. Kerick, S. Assassi, A.T. Satpathy, J. Granja, M.R. Mumbach, L. Beretta, C.P. Simeón, P. Carreira, N. Ortego-Centeno, I. Castellvi, L. Bossini-Castillo, F. David Carmona, G. Orozco, N. Hunzelmann, J.H.W. Distler, A. Franke, C. Lunardi, G. Moroncini, A. Gabrielli, J. de Vries-Bouwstra, C. Wijmenga, B.P.C. Koeleman, A. Nordin, L. Padyukov, A.-M. Hoffmann-Vold, B. Lie, S. Proudman, W. Stevens, M. Nikpour, T. Vyse, A.L. Herrick, J. Worthington, C.P. Denton, Y. Allanore, M.A. Brown, Timothy R D, C. Fonseca, H.Y. Chang, M.D. Mayes, J. Martin, Australian Scleroderma Interest Group (ASIG), GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways, Nature Communications. 10 (2019).

[62] T. Li, L. Ortiz-Fernández, E. Andrés-León, L. Ciudad, B.M. Javierre, E. López-Isac, A. Guillén-Del-Castillo, C.P. Simeón-Aznar, E. Ballestar, J. Martin, Epigenomics and transcriptomics of systemic sclerosis CD4+ T cells reveal long-range dysregulation of key inflammatory pathways mediated by disease-associated susceptibility loci, Genome Med. 12 (2020) 81.

[63] A. Lescoat, V. Lecureur, M. Roussel, B.L. Sunnaram, A. Ballerie, G. Coiffier, S. Jouneau, O. Fardel, T. Fest, P. Jégo, CD16-positive circulating monocytes and fibrotic manifestations of systemic sclerosis, Clin. Rheumatol. 36 (2017) 1649–1654.

[64] D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R.J. Rossi, C.-H. Chung, R.P. Huynh, P. Singh, W.F. Flynn, J.-A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M.A. Collet, A.L. Lucido, K. Palucka, P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J.F. Banchereau, Mapping systemic lupus erythematosus heterogeneity at the single-cell level, Nat. Immunol. 21 (2020) 1094–1106.

[65] W. Zheng, X. Wang, J. Liu, X. Yu, L. Li, H. Wang, J. Yu, X. Pei, C. Li, Z. Wang, M. Zhang, X. Zeng, F. Zhang, C. Wang, H. Chen, H.-Z. Chen, Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behçet's disease, Proc. Natl. Acad. Sci. U. S. A. 119

(2022) e2204289119.

[66] T. Carvalheiro, S. Horta, J.A.G. van Roon, M. Santiago, M.J. Salvador, H. Trindade, T.R.D.J. Radstake, J.A.P. da Silva, A. Paiva, Increased frequencies of circulating CXCL10-, CXCL8- and CCL4-producing monocytes and Siglec-3-expressing myeloid dendritic cells in systemic sclerosis patients, Inflamm. Res. 67 (2018) 169–177.

[67] J.S. Berkowitz, T. Tabib, H. Xiao, G.M. Sadej, D. Khanna, P. Fuschiotti, R.A. Lafyatis, J. Das, Cell-type-specific biomarkers of systemic sclerosis disease severity capture cell-intrinsic and cell-extrinsic circuits, Arthritis Rheumatol. (2023). https://doi.org/10.1002/art.42536.

[68] M. Wu, S. Assassi, Dysregulation of Type 1 Interferon Signaling in Systemic Sclerosis: a Promising Therapeutic Target?, Curr Treatm Opt Rheumatol. 7 (2021) 349–360.

[69] J. Yang, L. Zhang, C. Yu, X.-F. Yang, H. Wang, Monocyte and macrophage differentiation: circulation inflammatory monocyte as biomarker for inflammatory diseases, Biomark Res. 2 (2014) 1.

[70] J.J. Belch, R. Madhok, B. Shaw, P. Leiberman, R.D. Sturrock, C.D. Forbes, Double-blind trial of CL115,347, a transdermally absorbed prostaglandin E2 analogue, in treatment of Raynaud's phenomenon, Lancet. 1 (1985) 1180–1183.

[71] A. Koeberle, O. Werz, Perspective of microsomal prostaglandin E2 synthase-1 as drug target in inflammation-related disorders, Biochem. Pharmacol. 98 (2015) 1–15.

[72] A. Koeberle, O. Werz, Perspective of microsomal prostaglandin E2 synthase-1 as drug target in inflammation-related disorders, Biochem. Pharmacol. 98 (2015) 1–15.

[73] E. López-Isac, L. Bossini-Castillo, C.P. Simeon, M.V. Egurbide, J.J. Alegre-Sancho, J.L. Callejas, J.A. Roman-Ivorra, M. Freire, L. Beretta, A. Santaniello, P. Airó, C. Lunardi, N. Hunzelmann, G. Riemekasten, T. Witte, A. Kreuter, J.H.W. Distler, A.J. Schuerwegh, M.C. Vonk, A.E. Voskuyl, P.G. Shiels, J.M. van Laar, C. Fonseca, C. Denton, A. Herrick, J. Worthington, S. Assassi, B.P. Koeleman, M.D. Mayes, T.R.D.J. Radstake, J. Martin, N. Ortego-Centeno, R. Ríos, N. Navarrete, R.G. Portales, M.T. Camps, A. Fernández-Nebro, M.F. González-Escribano, J. Sánchez-Román, F.J. García-Hernández, M.J. Castillo, M.Á. Aguirre, I. Gómez-

Gracia, P. Carreira, B. Fernández-Gutiérrez, L. Rodríguez-Rodríguez, E. Vicente, J.L. Andreu, M.F. de Castro, P.G. de la Peña, F.J. López-Longo, L. Martinez, V. Fonollosa, C.P. Simeón, G. Espinosa, I. Castellví, C. Tolosa, A. Pros, M.R. Carballeira, F.J. Narváez, M.R. Rivas, V.O. Santamarí, M.Á. González-Gay, B. Díaz, L. Trapiella, A. Sousa, P.F. Mateo, L. Sáez-Comet, F. Díaz, V. Hernández, E. Beltrán, F.J. Blanco García, N. Oreiro, A genome-wide association study follow-up suggests a possible role for PPARG in systemic sclerosis susceptibility, Arthritis Res. Ther. 16 (2014) 2–9.

[74] F. Maione, G.M. Casillo, F. Raucci, A.J. Iqbal, N. Mascolo, The functional link between microsomal prostaglandin E synthase-1 (mPGES-1) and peroxisome proliferator-activated receptor γ (PPARγ) in the onset of inflammation, Pharmacol. Res. 157 (2020) 104807.

[75] F. Raucci, A. Saviano, G.M. Casillo, M. Guerra-Rodriguez, A.A. Mansour, M. Piccolo, M.G. Ferraro, E. Panza, V. Vellecco, C. Irace, F. Caso, R. Scarpa, N. Mascolo, M. Alfaifi, A.J. Iqbal, F. Maione, IL-17-induced inflammation modulates the mPGES-1/PPAR-γ pathway in monocytes/macrophages, Br. J. Pharmacol. 179 (2022) 1857–1873.

[76] D. Nehar-Belaid, S. Hong, R. Marches, G. Chen, M. Bolisetty, J. Baisch, L. Walters, M. Punaro, R.J. Rossi, C.-H. Chung, R.P. Huynh, P. Singh, W.F. Flynn, J.-A. Tabanor-Gayle, N. Kuchipudi, A. Mejias, M.A. Collet, A.L. Lucido, K. Palucka, P. Robson, S. Lakshminarayanan, O. Ramilo, T. Wright, V. Pascual, J.F. Banchereau, Mapping systemic lupus erythematosus heterogeneity at the single-cell level, Nat. Immunol. 21 (2020) 1094–1106.

[77] W. Zheng, X. Wang, J. Liu, X. Yu, L. Li, H. Wang, J. Yu, X. Pei, C. Li, Z. Wang, M. Zhang, X. Zeng, F. Zhang, C. Wang, H. Chen, H.-Z. Chen, Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behçet's disease, Proc. Natl. Acad. Sci. U. S. A. 119 (2022) e2204289119.

[78] W. Zheng, X. Wang, J. Liu, X. Yu, L. Li, H. Wang, J. Yu, X. Pei, C. Li, Z. Wang, M. Zhang, X. Zeng, F. Zhang, C. Wang, H. Chen, H.-Z. Chen, Single-cell analyses highlight the proinflammatory contribution of C1q-high monocytes to Behçet's disease, Proc. Natl. Acad. Sci. U. S. A. 119 (2022) e2204289119.

[79] F.D. Carmona, R. Gutala, C.P. Simeón, P. Carreira, N. Ortego-Centeno, E. Vicente-Rabaneda, F.J. García-Hernández, P. García de la Peña, M. Fernández-Castro, L. Martínez-Estupiñán, M.V. Egurbide, B.P. Tsao, P. Gourh, S.K. Agarwal, S. Assassi, M.D. Mayes, F.C. Arnett, F.K. Tan,

J. Martín, Spanish Scleroderma Group, Novel identification of the IRF7 region as an anticentromere autoantibody propensity locus in systemic sclerosis, Ann. Rheum. Dis. 71 (2012) 114–119.

[80] Y. Liu, R. Lu, W. Cui, Y. Pang, C. Liu, L. Cui, T. Qian, L. Quan, Y. Dai, Y. Jiao, Y. Pan, X. Ye, J. Shi, Z. Cheng, L. Fu, High IFITM3 expression predicts adverse prognosis in acute myeloid leukemia, Cancer Gene Ther. 27 (2020) 38–44.

[81] A.D. Hoffmann, S.E. Weinberg, S. Swaminathan, S. Chaudhuri, H.F. Mubarak, M.J. Schipma, C. Mao, X. Wang, L. El-Shennawy, N.K. Dashzeveg, J. Wei, P.J. Mehl, L.J. Shihadah, C.M. Wai, C. Ostiguin, Y. Jia, P. D'Amico, N.R. Wang, Y. Luo, A.R. Demonbreun, M.G. Ison, H. Liu, D. Fang, Unique molecular signatures sustained in circulating monocytes and regulatory T cells in Convalescent COVID-19 patients, bioRxiv. (2022). https://doi.org/10.1101/2022.03.26.485922.

[82] G. Shi, S. Ozog, B.E. Torbett, A.A. Compton, mTOR inhibitors lower an intrinsic barrier to virus infection mediated by IFITM3, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) E10069–E10078.

[83] D. Benfaremo, S. Svegliati, C. Paolini, S. Agarbati, G. Moroncini, Systemic Sclerosis: From Pathophysiology to Novel Therapeutic Approaches, Biomedicines. 10 (2022). https://doi.org/10.3390/biomedicines10010163.

[84] V. Sundblad, R.A. Gomez, J.C. Stupirski, P.F. Hockl, M.S. Pino, H. Laborde, G.A. Rabinovich, Circulating Galectin-1 and Galectin-3 in Sera From Patients With Systemic Sclerosis: Associations With Clinical Features and Treatment, Front. Pharmacol. 12 (2021) 650605.

[85] M.N. Negedu, C.A. Duckworth, L.-G. Yu, Galectin-2 in Health and Diseases, Int. J. Mol. Sci. 24 (2022). https://doi.org/10.3390/ijms24010341.

[86] D. Paclik, L. Werner, O. Guckelberger, B. Wiedenmann, A. Sturm, Galectins distinctively regulate central monocyte and macrophage function, Cell. Immunol. 271 (2011) 97–103.

[87] M. Uhlén, L. Fagerberg, B.M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C.A.-K. Szigyarto, J. Odeberg, D. Djureinovic, J.O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J.M. Schwenk, M. Hamsten, K.

von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, F. Pontén, Proteomics. Tissue-based map of the human proteome, Science. 347 (2015) 1260419.

[88] P. Laurent, V. Sisirak, E. Lazaro, C. Richez, P. Duffau, P. Blanco, M.-E. Truchetet, C. Contin-Bordes, Innate Immunity in Systemic Sclerosis Fibrosis: Recent Advances, Front. Immunol. 9 (2018) 1702.

[89] C. Yıldırım, D.Y.S. Vogel, M.R. Hollander, J.M. Baggen, R.D. Fontijn, S. Nieuwenhuis, A. Haverkamp, M.R. de Vries, P.H.A. Quax, J.J. Garcia-Vallejo, A.M. van der Laan, C.D. Dijkstra, T.C.T.M. van der Pouw Kraan, N. van Royen, A.J.G. Horrevoets, Galectin-2 induces a proinflammatory, anti-arteriogenic phenotype in monocytes and macrophages, PLoS One. 10 (2015) e0124347.

[90] J. Kane, M. Jansen, S. Hendrix, L.A. Bosmans, L. Beckers, C. van Tiel, M. Gijbels, N. Zelcer, C.J. de Vries, P. von Hundelshausen, M. Vervloet, E. Eringa, A.J. Horrevoets, N. van Royen, E. Lutgens, Anti-Galectin-2 Antibody Treatment Reduces Atherosclerotic Plaque Size and Alters Macrophage Polarity, Thromb. Haemost. 122 (2022) 1047–1057.

[91] N. Higashi-Kuwata, T. Makino, Y. Inoue, M. Takeya, H. Ihn, Alternatively activated macrophages (M2 macrophages) in the skin of patient with localized scleroderma, Exp. Dermatol. 18 (2009) 727–729.

[92] T. Rőszer, Understanding the Mysterious M2 Macrophage through Activation Markers and Effector Mechanisms, Mediators Inflamm. 2015 (2015) 816460.

[93] A. van Caam, J. Aarts, T. van Ee, E. Vitters, M. Koenders, F. van de Loo, P. van Lent, F. van den Hoogen, R. Thurlings, M.C. Vonk, P.M. van der Kraan, TGFβ-mediated expression of TGFβ-activating integrins in SSc monocytes: disturbed activation of latent TGFβ?, Arthritis Res. Ther. 22 (2020) 42.

[94] F. Zhang, H. Wang, X. Wang, G. Jiang, H. Liu, G. Zhang, H. Wang, R. Fang, X. Bu, S. Cai, J. Du, TGF-β induces M2-like macrophage polarization via SNAIL-mediated suppression of a pro-inflammatory phenotype, Oncotarget. 7 (2016) 52294–52306.

[95] X. Yang, H. Liu, T. Ye, C. Duan, P. Lv, X. Wu, J. Liu, K. Jiang, H. Lu, H. Yang, D. Xia, E. Peng, Z. Chen, K. Tang, Z. Ye, AhR activation attenuates calcium oxalate nephrocalcinosis by diminishing M1

macrophage polarization and promoting M2 macrophage polarization, Theranostics. 10 (2020) 12011–12025.

[96] S. Pennathur, K. Pasichnyk, N.M. Bahrami, L. Zeng, M. Febbraio, I. Yamaguchi, D.M. Okamura, The macrophage phagocytic receptor CD36 promotes fibrogenic pathways on removal of apoptotic cells during chronic kidney injury, Am. J. Pathol. 185 (2015) 2232–2245.

[97] P. Laurent, J. Lapoirie, D. Leleu, E. Levionnois, C. Grenier, B. Jurado-Mestre, E. Lazaro, P. Duffau, C. Richez, J. Seneschal, J.-L. Pellegrin, J. Constans, T. Schaeverbeke, I. Douchet, D. Duluc, T. Pradeu, C. Chizzolini, P. Blanco, M.-E. Truchetet, C. Contin-Bordes, Fédération Hospitalo-Universitaire ACRONIM and the Centre National de Référence des Maladies Auto-Immunes Systémiques Rares de l'Est et du Sud-Ouest (RESO), Interleukin-1β-Activated Microvascular Endothelial Cells Promote DC-SIGN-Positive Alternatively Activated Macrophages as a Mechanism of Skin Fibrosis in Systemic Sclerosis, Arthritis Rheumatol. 74 (2022) 1013–1026.

[98] A. Masuda, H. Yasuoka, T. Satoh, Y. Okazaki, Y. Yamaguchi, M. Kuwana, Versican is upregulated in circulating monocytes in patients with systemic sclerosis and amplifies a CCL2-mediated pathogenic loop, Arthritis Res. Ther. 15 (2013) R74.

[99] C. Maier, A. Ramming, C. Bergmann, R. Weinkam, N. Kittan, G. Schett, J.H.W. Distler, C. Beyer, Inhibition of phosphodiesterase 4 (PDE4) reduces dermal fibrosis by interfering with the release of interleukin-6 from M2 macrophages, Ann. Rheum. Dis. 76 (2017) 1133–1141

.

General discussion

To unravel the genetic architecture of an IMID researchers can benefit from both generating new data and the exploitation of the available data sources, followed by the combination of them. This dissertation followed both approaches in order to clarify some aspects in SSc onset and pathogenesis.

GWAS have marked a significant revolution in genetics ever since the publication of the first study 17 years ago (53). Over these nearly two decades, the genetic landscape of various diseases, including SSc, has greatly expanded (65). Moreover, the data obtained from GWAS have become a valuable resource for subsequent research. Finally, the genetic findings based on GWAS are gradually finding their way into daily clinical practice and what was mere science fiction a few years ago is now becoming a reality.

For example, there are novel proposals for the implementation of GRS or PRS in clinical practice in a controlled and step-by-step process as the G-PROB tool developed by Knevel *et al*. in 2020 (101). This risk prediction model enabled the prioritization between RA, SLE, spondyloarthropathy, psoriatic arthritis, and gout. Using the genetic information available prior to the patient's visit to the clinic, the G-PROB tool allowed the exclusion of at least one of the candidate diseases during clinical assessments and resulted in a confirmed diagnosis in 64% of cases (101). This model was a proof of concept and represented a significant step towards PRS use in rheumatic diseases and a promising advance in the steps that the GWAS strategy opened years ago for precision medicine.

Genetic risk prediction models should aspire to integrate into clinical practice, encompassing patients' genetic data before their appointments

through screening models similar to those already established for monogenic and hereditary diseases. These tools would facilitate shorter diagnostic timelines, providing physicians with the necessary time to devise effective therapies and providing patients with relief from the uncertainty associated with undiagnosed conditions. In the case of SSc, a recent review by Volkmann *et al*. (7) highlighted that patients with SSc receive a definitive diagnosis in 65.9% of cases at 5 years and 72.7% at 10 years, relying on indicators such as Raynaud's phenomenon, disease-specific autoantibodies, and capillaroscopy patterns.

Therefore, we considered taking advantage of the privileged resources of our team, who led the largest SSc GWAS to date, and generating the first SSc GRS as part of this dissertation. Then, we analyzed thousands of models and finally selected a GRS model that included 33 highly associated SNPs. Moreover, we developed subtype-specific models that considered the allele effects of thousands of SNPs on the different serological and clinical subtypes of patients (197). We would like to highlight that our GRS model was able to distinguish between SSc and other IMIDs, not only between SSc and unaffected individuals, which is a major request from clinicians as differential diagnosis between IMIDs is key for them. For this reason, we think that we set the basis for a new discrimination tool that could help in differential diagnosis in the early stages of SSc.

Our results also proved that biochemical analyses and demographic data could significantly enhance the accuracy of predictive models. The incorporation of a multivariable model that included cell counts for various immune cell types and age resulted in a substantial enhancement of accuracy, increasing the AUC to encouraging values (AUC_{GRS} = 0.69, $AUC_{discovery}$ = 0.84). Additionally, considering the results of the scRNA-seq analysis in CD14+ monocytes included in this thesis, we hypothesize that our

model might improve if levels of the altered populations were included in the model similarly to inclusion of M0 macrophages, which were included as part of the immune cell counts. It is worth noting that the ncMos $CD14^{low}CD16^{high}$ monocyte population was previously shown to be elevated in SSc patients and confirmed in our study (12,198). Our research has underscored the significant role of this monocyte subpopulation in the inflammation observed in SSc patients, potentially adding crucial elements to a prospective model, including the consideration of *PTGES* expression as a SSc-related marker. These biological or cellular factors, in conjunction with traditional risk factors, may enrich prediction models, thereby contributing to enhanced diagnostic accuracy.

Nonetheless, for the seamless integration of PRS into clinical practice, we must equip clinicians with the requisite training to interpret and communicate the results to patients (98,171). Additionally, we must not overlook the broader population's need for data interpretation and privacy, given that PRS necessitates a substantial volume of genetic information. In this context, comprehending the concept of high and low-risk individuals and how clinicians can therapeutically intervene remains a complex challenge (98). In this regard, models as ours, should be brought closer to physicians in a collaborative effort. Not just to receive feedback and improve it, but to integrate it into their clinical practice more easily.

Finally, the extensive catalog of GWAS studies published in recent years is predominantly based on individuals of European ancestry, rendering PRS models inadequately applicable to other populations with limited predictive value outside this demographic (102,199,200). In our own GRS model, there are differences between the predictive value between the cohorts, with an AUC range from 0.60 to 0.75. Consequently, there is an urgent need to revise existing PRS models or develop new ones that account

for these disparities (102,199–202). In this regard, recently Zhang *et al* (202) published a new approach in this direction, CT-SLEB, an algorithm trained with multiple sources of non-European population cohorts, to cover this large range of the population. Another approach involves adjusting GWAS for a polygenic score, as recently published by Campos et al. (201), to enhance statistical power for discovery across all ancestral groups. Regarding SSc, a new GWAS is being carried out including larger non-European cohorts at the moment. Several trans-ethnic studies have been conducted in SSc previously, such as the most recent study in Turkish and Iranian populations (203), led by our group, and the study conducted in Japanese and European populations by Terao *et al.* (90). Trans-ethnic studies have also been performed in other IMIDs using GWAS data, as seen in RA, resulting in the identification of 124 *loci*, 36 of which were novel (204). Therefore, a project should be planned in the disease with a trans-ethnic approach, boosting the statistical power with largest cohort and taking advantage of natural differences in LD across ethnically diverse populations. Our predictive model should serve as a baseline for future PRSs in the disease that encompass or are adjusted for different ethnicities.

As described in the introduction of this dissertation, despite the advances in the field of genetics, the known environmental risk factors contributing to the development of SSc are limited and primarily revolve around exposure to various chemicals (176). Environmental factors related to lifestyle have not been thoroughly explored, with little to no new contributions in recent years. Therefore, with the second publication included in this dissertation, our aim was to broaden the knowledge about the risk factors affecting the disease, focusing on an extremely important social issue—obesity—and using cutting-edge statistical techniques like MR.

Obesity as a determining factor in SSc was previously studied but within the context of SSc-ILD. Nagy *et al.* (205) investigated the effect of overweight conditions on SSc-ILD patients and pulmonary deterioration. They observed that those with a lower BMI (BMI < 25 kg/m2) experienced a significant increase in pulmonary deterioration compared to those with higher BMI values. In their publication, the authors questioned why being overweight appeared to be a protective factor against SSc-ILD (205). Although it has been demonstrated that obesity is a risk factor for IMDs (45), our results did not seem to dismiss this counterintuitive protective effect in SSc. However, due the nature of the disease, it is difficult to discern the effect of the SNPs in SSc patients, as happens in other IMDs.

It should be highlighted that GI occurs in 90% of SSc patients, according to recent studies (206). Furthermore, despite guidelines published by leading medical organizations related to SSc, such as The European League Against Rheumatism (EULAR) Scleroderma Trials and Research Group and the United Kingdom Scleroderma Study Group (UKSSG) (207,208), rheumatologists continue to face challenges in its treatment (206). It has been confirmed that SSc patients tend to have lower body mass and even malnutrition issues due to GI involvement (209). We hypothesize that due to these complications of SSc, the proinflammatory role of obesity that was observed in other IMIDs would not cause an increased risk in SSc as obesity would hardly develop in SSc patients and the risk variants, even if present, will not exert their risk effects. Even though we did not establish a causal relationship between body mass distribution and SSc, the relevance of adipocytes in inflammation has been known for over two decades due to their release of cytokines and adipokines (41). Hence, we cannot rule out the role that adipocytes may play in the disease.

In spite of this negative result, we think that the scientific community should not cease efforts to study different environmental risk factors and their causal relationship with SSc. As recently, taking advantage of 2SMR approach, our group successfully established a potential causal relationship between genetically predicted longer telomeres in leukocytes and the onset of SSc (210). Thanks to the availability of public GWAS data in repositories like the GWAS catalog (94), we can continue to make strides in understanding the risk factors impacting SSc, covering a largely uncharted territory to date. Factors such as vitamin or metabolite levels, smoking or hormone production are subjects of investigation in 2SMR studies and are now publicly available for research.

Genetic predisposition in combination with environmental risk factors affect the individual outcome, including the generalized immune imbalance in SSc. A number of transcriptomic and DNA conformation studies showed cell type-specific effects in different immune lineages in the disease. For example, Hi-ChIP results (65) or the Hi-C capture study (144) confirmed T cell and monocyte specific 3D DNA contacts. However, the majority of the cell-specific analyses in SSc have focused on studying the transcriptomic profiles of different subsets.

In the case of monocytes, previous transcriptomic studies investigated their role in SSc both in Europeans (142) and in African Americans (211). Beretta *et al.* (142) found a dysregulation of several crucial pathways, as type I IFN or Toll like receptor (TLR) cascade; Neutrophil showed as the major contributor to the gene expression in their whole blood study. However, no specific subpopulations were identified, or biomarkers established. Allen *et al.* (211) results contrast with previous methylation and eQTLs studies, as they found modest differences between cases and controls. They concluded that these contracts may be due to variation in DNA

methylation and gene expression across cell types, individuals of different genetic clinical and environmental backgrounds (211). In both cases, only cMo and ncMo were targeted, leaving aside iMo, as it is difficult to capture them. With scRNA-seq, which allows for greater granularity in analysis, we were able to confirm the over-representation of ncMos and to identify and characterize SSc-related clusters of iMos. Those population were *IRF7*+*STAT1*+iMos or *HLA-DRA*+*HLA-DRB1*+iMos, which might contribute to SSc pathogenesis through different paths, such as controlling the differentiation to ncMo, as it could be inferred in the trajectory analyses (Figure 4D of Chapter 3) (198). In this regard, we took a first step in demonstrating iMo may have a role in SSc.

Previously published scRNA-seq studies in SSc focused on affected tissues as crucial to the disease as the skin and lungs. Thanks to these studies, specific cellular populations were defined in patients, suggesting a pathogenic cellular microenvironment (159–161). Immune cells are part of this microenvironment, with aberrant behavior of myeloid cells and some lymphocyte populations (130,212). Blood sampling is much less invasive than a lung or skin biopsy, is less stressful for the patient, requires no recovery time, and can be done during a routine visit to the physician. In the blood, immune cells belonging to both the innate and adaptive immune systems can be sampled while migrating to the affected tissues. Therefore, identifying the abnormal behavior of these cells in blood can lead us to develop new treatments and therapeutic targets. Despite these advantages and the fact that SSc is a systemic disorder, little research using scRNA-seq has been done on peripheral blood circulating immune populations (213).

Before the publication of the third manuscript included in this dissertation, only one study had conducted an exploration of PBMCs with inconclusive results (160), and another study screened the single

transcriptome of 837 monocytes, in which they found a inflammatory gene module of ncMo (213). With more than 90,000 analyzed monocyte cells from patients and controls, we conducted the most comprehensive characterization of CD14+ Mos in SSc so far. We were able to define a specific profile of these cells in patients, differentiate between different subpopulations, and identify key genes in their pro-inflammatory profile. For instance, several publications had previously identified the S100A protein family as a potential biomarker for the disease (214–216). In our scRNA-seq study, we have also observed the overexpression of genes encoding some of these proteins in various monocyte populations, reinforcing the validity of scRNA-seq for both discovering new cell populations and confirming data obtained through other techniques.

One of the key findings in the third publication of this dissertation was the over-expression of *PTGES* in ncMo among SSc patients (198). *PTGES* encodes Prostaglandin E Synthase, an enzyme within the Prostaglandin E2 (PGE_2) pathway, suggesting its potential as a biomarker for the disease, which should be validated in a larger cohort. This result also prompts questions about the role of prostaglandins in SSc. While prostaglandins, with their vasodilating effects, are effective in treating Raynaud's phenomenon and digital ulcers (217), other prostaglandin-derived lipids have been identified with potential implications in the disease (218). Given that PGE_2 is a known inflammation mediator, and *PTGES* has been proposed as an anti-inflammatory target (219), we believe that both molecules are worth exploring in the context of SSc. Consequently, PGE_2 might potentially contribute to ncMo-mediated inflammation and *PTGES* might serve as a promising biomarker.

Additionally, our study serves as a bridge that connects circulating cells with the altered profiles of tissue resident immune cells. We identified

two cellular clusters with this function, one with a cellular destination to M2 monocytes and another to type 2 dendritic cells. Both cell types are essential in the inflammatory response and cellular activation, and both have been found to be altered in the skin of SSc patients (128,130).

Adaptive immune responses play a crucial role in SSc pathogenesis. CD4+ T cells, which are central to this response, exert influence through their activation and immune regulation, mediated by factors like the T cell receptor (TCR) and the release of cytokines. Although previous research has highlighted the imbalances and polarization of CD4+ T cell subtypes (220), the interaction between adaptive and innate immune response and their impact on SSc remains to be fully understood, with the HLA acting as a central element connecting them. In SLE, Guo *et al.* (221) analyzed CD4+ T cells and established correlations between chromatin accessibility and disease severity. They also identified transcriptional dysfunction within Treg cells, particularly within a Treg subgroup exhibiting exhaustion-like properties (221). Similarly, in the case of RA, Argyriou *et al.* (222) conducted scRNA-seq on synovial fluid CD4+ T cells and identified the G-protein coupled receptor 56 (*GPR56*) as a marker that distinguishes T peripheral helper cells in ACPA+ and ACPA- RA patients. In this regard, our group conducted a scRNA-seq study on CD4+ T cells, with the same cohort as in the scRNA-seq in monocytes (198) included in this thesis. The CD4+ study is designed to delve into the characterization of CD4+ subpopulations within PBMCs from SSc patients, mirroring the approaches used in the aforementioned studies.

The incidence of SSc exhibits a gender bias, as reflected in the prevalence ratio between women and men, which stands at 8:2 (15). The implementation of sex stratification was a clear limitation of the studies included in this doctoral thesis. However, it's noteworthy that sex was not

among the covariates that contributed the most to the PRS model in this dissertation, despite the evidence indicating such a bias. Nevertheless, our scRNA-seq study cohort included only women and we could not address if sex-based differences were observed at the transcriptomic level. In any case, we consider that high-resolution transcriptomic studies, such as scRNA-seq, offer a promising avenue to elucidate the underlying reasons for sex bias. Such studies require fewer individuals while providing a high level of detail. Peters *et al.* (223) outline studies of sex bias in three phases, the first of which pertains to sex and gender differences at the disease's general level (risk, disease presentation, etc.), the second at the mechanistic level, and finally, the practical implications. In the case of SSc, due to the prevalence ratio, conducting large-scale research can be a complex challenge to achieve the necessary statistical power for reliable results. However, transcriptomic studies can fill the gap and bring us closer to the second point mentioned by Peters and Woodward, providing insights into the sex-specific mechanisms and metabolic pathways involved.

Finally, an important step that should be taken in the future to make the most out of the findings and dataset generated in this thesis would be the integration of the various layers of 'omics' that we explored. A good example of future possibilities would be the application of single-cell disease relevance score (scDRS), a tool that links PRS with scRNA-seq data, identifying cells that express high-risk genes based on GWAS data (150). Another possible framework would include the implementation of CellPhonedb and NitcheNet, which are tools aimed at investigating communication between different cell types using information from genes encoding receptors and soluble molecules (224,225). These are just a few examples of integration tools emerging to bridge the gaps between data from individual studies.

This dissertation comprised multiple approaches to shed light into the molecular causes of SSc. The establishment of the GRS model has provided a foundation for predicting disease risk considering genetic risk markers. Although our results couldn't establish a causal relationship between obesity traits and SSc, we contributed to the understanding of risk factors in SSc. Lastly, through the characterization of monocytes in SSc, we took a significant step toward defining the role that these cells play in the disease's pathogenesis and their possible use as disease biomarkers. Therefore, through these diverse approaches, we contributed to advancing the knowledge of the disease and laying the groundwork for future research.

Conclusions

1. A sensitive genomic risk score capable of distinguishing between patients with systemic sclerosis and unaffected individuals, between different serological subtypes of systemic sclerosis and between systemic sclerosis patients and individuals affected by rheumatoid arthritis and Sjögren's syndrome was developed.

2. No significant evidence indicating that body fat distribution was a causal risk factor for systemic sclerosis was found using a two-sample Mendelian randomization strategy.

3. Single cell transcriptome analysis revealed that non-classical monocytes are overrepresented in the blood of systemic sclerosis patients, exhibit an inflammatory and interferon signal and are characterized by the overexpression of *PTGES*, which encodes an enzyme in the prostaglandin E_2 metabolism.

4. A population of intermediate monocytes characterized by an aberrant response to interferon and marked by the expression of *IRF7* and *STAT1* was found in the blood of systemic sclerosis patients.

5. A depletion of classical monocytes polarized into M2 macrophages was observed in the blood of systemic sclerosis patients.

Conclusiones

1. Se desarrolló una puntuación de riesgo genómico capaz de distinguir sensiblemente entre pacientes con esclerosis sistémica y personas no afectadas, entre diferentes subtipos serológicos de esclerosis sistémica y entre pacientes con esclerosis sistémica e individuos afectados por artritis reumatoide y síndrome de Sjögren.

2. No se encontraron pruebas significativas que indiquen que la distribución de grasa corporal sea un factor de riesgo causal para la esclerosis sistémica utilizando una estrategia de aleatorización mendeliana de dos muestras.

3. Un análisis del transcriptoma de células individuales reveló que los monocitos no clásicos están sobrerrepresentados en la sangre de los pacientes con esclerosis sistémica, muestran una señal inflamatoria e interferón y se caracterizan por la sobreexpresión de *PTGES*, que codifica una enzima en el metabolismo de la prostaglandina E2.

4. Se encontró una población de monocitos intermedios caracterizados por una respuesta aberrante al interferón y marcados por la expresión de *IRF7* y *STAT1* en la sangre de pacientes con esclerosis sistémica.

5. Se observó una disminución de los monocitos clásicos polarizados hacia macrófagos M2 en la sangre de pacientes con esclerosis sistémica.

References

1. Pisetsky DS. Pathogenesis of autoimmune disease. Nat Rev Nephrol. 2023 Aug;19(8):509–24.

2. Theofilopoulos AN, Kono DH, Baccala R. The multiple pathways to autoimmunity. Nat Immunol. 2017 Jun 20;18(7):716–24.

3. Ngo ST, Steyn FJ, McCombe PA. Gender differences in autoimmune disease. Front Neuroendocrinol. 2014;35(3):347–69.

4. Moroni L, Bianchi I, Lleo A. Geoepidemiology, gender and autoimmune disease. Autoimmun Rev. 2012 May;11(6–7):A386-92.

5. Cooper GS, Bynum MLK, Somers EC. Recent insights in the epidemiology of autoimmune diseases: improved prevalence estimates and understanding of clustering of diseases. J Autoimmun. 2009 Oct 9;33(3–4):197–207.

6. Cheng MH, Anderson MS. Monogenic autoimmunity. Annu Rev Immunol. 2012 Jan 6;30:393–427.

7. Volkmann ER, Andréasson K, Smith V. Systemic sclerosis. Lancet. 2023 Jan 28;401(10373):304–18.

8. Wigley FM, Flavahan NA. Raynaud's Phenomenon. N Engl J Med. 2016 Aug 11;375(6):556–65.

9. Smith V, Herrick AL, Ingegnoli F, Damjanov N, De Angelis R, Denton CP, et al. Standardisation of nailfold capillaroscopy for the assessment of patients with Raynaud's phenomenon and systemic sclerosis. Autoimmun Rev. 2020 Mar;19(3):102458.

10. O'Reilly S, Hügle T, van Laar JM. T cells in systemic sclerosis: a reappraisal. Rheumatology . 2012 Sep;51(9):1540–9.

11. Pillai S. T and B lymphocytes in fibrosis and systemic sclerosis. Curr Opin Rheumatol. 2019 Nov;31(6):576–81.

12. Lescoat A, Lecureur V, Varga J. Contribution of monocytes and macrophages to the pathogenesis of systemic sclerosis: recent insights and therapeutic implications. Curr Opin Rheumatol. 2021 Nov 1;33(6):463–70.

13. Allanore Y, Simms R, Distler O, Trojanowska M, Pope J, Denton CP, et al. Systemic sclerosis. Nat Rev Dis Primers. 2015 Apr;1:15002.

14. Denton CP, Khanna D. Systemic sclerosis. Lancet. 2017;390(10103):1685–99.

15. De Angelis R, Giuggioli D, Bajocchi G, Dagna L, Zanframundo G, Foti R, et al. Sex-related Differences in Systemic Sclerosis: A Multicenter Cross-sectional Study From the National Registry of the Italian Society for Rheumatology. J Rheumatol. 2022 Feb;49(2):176–85.

16. Marie I, Gehanno J-F. Environmental risk factors of systemic sclerosis. Semin Immunopathol. 2015 Sep;37(5):463–73.

17. Bramwell B. Sclerodermia, Diffuse. Trans Med Chir Soc Edinb. 1914;33:152–70.

18. Otsuki T, Maeda M, Murakami S, Hayashi H, Miura Y, Kusaka M, et al. Immunological effects of silica and asbestos. Cell Mol Immunol. 2007 Aug;4(4):261–8.

19. Marie I, Menard J-F, Duval-Modeste A-B, Joly P, Dominique S, Bravard P, et al. Association of occupational exposure with features of systemic sclerosis. J Am Acad Dermatol. 2015 Mar;72(3):456–64.

20. Lee S, Hayashi H, Maeda M, Chen Y, Matsuzaki H, Takei-Kumagai N, et al. Environmental factors producing autoimmune dysregulation--chronic activation of T cells caused by silica exposure. Immunobiology. 2012 Jul;217(7):743–8.

21. Marie I, Gehanno J-F, Bubenheim M, Duval-Modeste A-B, Joly P, Dominique S, et al. Prospective study to evaluate the association between systemic sclerosis and occupational exposure and review of the literature. Autoimmun Rev. 2014 Feb;13(2):151–6.

22. Ferro A, Zebedeo CN, Davis C, Ng KW, Pfau JC. Amphibole, but not chrysotile, asbestos induces anti-nuclear autoantibodies and IL-17 in C57BL/6 mice. J Immunotoxicol. 2014 Jul-Sep;11(3):283–90.

23. Salazar KD, Copeland CB, Luebke RW. Effects of Libby amphibole asbestos exposure on two models of arthritis in the Lewis rat. J Toxicol Environ Health A. 2012;75(6):351–65.

24. Furuzawa-Carballeda J, Vargas-Rojas MI, Cabral AR. Autoimmune inflammation from the Th17 perspective. Autoimmun Rev. 2007

Jan;6(3):169–75.

25. Kettaneh A, Al Moufti O, Tiev KP, Chayet C, Tolédano C, Fabre B, et al. Occupational exposure to solvents and gender-related risk of systemic sclerosis: a metaanalysis of case-control studies. J Rheumatol. 2007 Jan;34(1):97–103.

26. Martinez-Tello FJ, Navas-Palacios JJ, Ricoy JR, Gil-Mart?n R, Conde-Zurita JM, Colina-Ruiz Delgado F, et al. Pathology of a new Toxic Syndrome caused by ingestion of adulterated oil in Spain. Virchows Arch A Path Anat and Histol. 1982;397(3):261–85.

27. Patterson R, Germolec D. Review article toxic oil syndrome: Review of immune aspects of the disease. J Immunotoxicol. 2005 Jan;2(1):51–8.

28. Kaufman LD, Izquierdo Martinez M, Serrano JM, Gomez-Reino JJ. 12-year followup study of epidemic Spanish toxic oil syndrome. J Rheumatol. 1995 Feb;22(2):282–8.

29. de la Paz MP, Philen RM, Borda IA. Toxic oil syndrome: The perspective after 20 years. Epidemiol Rev. 2001 Jan 1;23(2):231–47.

30. Kivity S, Agmon-Levin N, Blank M, Shoenfeld Y. Infections and autoimmunity--friends or foes? Trends Immunol. 2009 Aug;30(8):409–14.

31. Lunardi C, Bason C, Navone R, Millo E, Damonte G, Corrocher R, et al. Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells. Nat Med. 2000 Oct;6(10):1183–6.

32. Arvia R, Zakrzewska K, Giovannelli L, Ristori S, Frediani E, Del Rosso M, et al. Parvovirus B19 induces cellular senescence in human dermal fibroblasts: putative role in systemic sclerosis-associated fibrosis. Rheumatology . 2022 Aug 30;61(9):3864–74.

33. Yazawa N, Fujimoto M, Kikuchi K, Kubo M, Ihn H, Sato S, et al. High seroprevalence of Helicobacter pylori infection in patients with systemic sclerosis: association with esophageal involvement. J Rheumatol. 1998 Apr;25(4):650–3.

34. Radić M, Martinović Kaliterna D, Bonacin D, Morović Vergles J, Radić J. Correlation between Helicobacter pylori infection and systemic sclerosis activity. Rheumatology . 2010 Sep;49(9):1784–5.

35. Miller FW, Alfredsson L, Costenbader KH, Kamen DL, Nelson LM, Norris JM, et al. Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop. J Autoimmun. 2012 Dec;39(4):259–71.

36. Kaul A, Gordon C, Crow MK, Touma Z, Urowitz MB, van Vollenhoven R, et al. Systemic lupus erythematosus. Nat Rev Dis Primers. 2016 Jun 16;2:16039.

37. Wieczorek M, Gwinnutt JM, Ransay-Colle M, Balanescu A, Bischoff-Ferrari H, Boonen A, et al. Smoking, alcohol consumption and disease-specific outcomes in rheumatic and musculoskeletal diseases (RMDs): systematic reviews informing the 2021 EULAR recommendations for lifestyle improvements in people with RMDs. RMD Open [Internet]. 2022 Mar;8(1). Available from: http://dx.doi.org/10.1136/rmdopen-2021-002170

38. Sharif K, Watad A, Bragazzi NL, Lichtbroun M, Amital H, Shoenfeld Y. Physical activity and autoimmune diseases: Get moving and manage the disease. Autoimmun Rev. 2018 Jan;17(1):53–72.

39. Chen J, Liao S, Pang W, Guo F, Yang L, Liu H-F, et al. Life factors acting on systemic lupus erythematosus. Front Immunol. 2022 Sep 15;13:986239.

40. Zhang Y-J, Zhang L, Huang X-L, Duan Y, Yang L-J, Wang J. Association between cigarette smoking and impaired clinical symptoms in systemic sclerosis: A review. Cell Immunol. 2017 Aug;318:1–7.

41. Fantuzzi G. Adipose tissue, adipokines, and inflammation. J Allergy Clin Immunol. 2005 May;115(5):911–9; quiz 920.

42. Versini M, Jeandel PY, Rosenthal E, Shoenfeld Y. Obesity in autoimmune diseases: Not a passive bystander. Autoimmun Rev. 2014;13(9):981–1000.

43. Arai S, Maehara N, Iwamura Y, Honda S-I, Nakashima K, Kai T, et al. Obesity-associated autoantibody production requires AIM to retain the immunoglobulin M immune complex on follicular dendritic cells. Cell Rep. 2013 Apr 25;3(4):1187–98.

44. Patterson CC, Dahlquist GG, Gyürüs E, Green A, Soltész G, EURODIAB Study Group. Incidence trends for childhood type 1 diabetes in Europe during 1989-2003 and predicted new cases 2005-20: a multicentre

prospective registration study. Lancet. 2009 Jun 13;373(9680):2027–33.

45. Tang B, Shi H, Alfredsson L, Klareskog L, Padyukov L, Jiang X. Obesity-Related Traits and the Development of Rheumatoid Arthritis: Evidence From Genetic Data. Arthritis Rheumatol. 2021 Feb;73(2):203–11.

46. Wang M, Huang S, Lin X, Wen C, He Z, Huang L. The Causal Relationship Between Blood Lipids and Systemic Lupus Erythematosus Risk: A Bidirectional Two-Sample Mendelian Randomization Study. Front Genet. 2022 Apr 13;13:858653.

47. Arnett FC, Cho M, Chatterjee S, Aguilar MB, Reveille JD, Mayes MD. Familial occurrence frequencies and relative risks for systemic sclerosis (scleroderma) in three United States cohorts. Arthritis Rheum. 2001 Jun;44(6):1359–62.

48. Englert H, Small-McMahon J, Chambers P, O'Connor H, Davis K, Manolios N, et al. Familial risk estimation in systemic sclerosis. Aust N Z J Med. 1999 Feb;29(1):36–41.

49. Harroud A, Hafler DA. Common genetic factors among autoimmune diseases. Science. 2023 May 5;380(6644):485–90.

50. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

51. Raychaudhuri S, Rich SS. Autoimmunity: insights from human genomics. Curr Opin Immunol. 2012 Oct;24(5):513–5.

52. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet. 2017 Jul;101(1):5–22.

53. Abdellaoui A, Yengo L, Verweij KJH, Visscher PM. 15 years of GWAS discovery: Realizing the promise. Am J Hum Genet [Internet]. 2023 Jan 5; Available from: http://dx.doi.org/10.1016/j.ajhg.2022.12.011

54. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. 2012 Dec 27;8(12):e1002822.

55. Matzaraki V, Kumar V, Wijmenga C, Zhernakova A. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. Genome Biol. 2017 Apr 27;18(1):76.

56. Arnett FC, Gourh P, Shete S, Ahn CW, Honey RE, Agarwal SK, 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 May;69(5):822–7.

57. Loubière LS, Lambert NC, Madeleine MM, Porter AJ, Mullarkey ME, Pang JM, et al. HLA allelic variants encoding DR11 in diffuse and limited systemic sclerosis in Caucasian women. Rheumatology . 2005 Mar;44(3):318–22.

58. Vlachoyiannopoulos PG, Dafni UG, Pakas I, Spyropoulou-Vlachou M, Stavropoulos-Giokas C, Moutsopoulos HM. Systemic scleroderma in Greece: low mortality and strong linkage with HLA-DRB1*1104 allele. Ann Rheum Dis. 2000 May;59(5):359–67.

59. Vargas-Alarcón G, Granados J, Ibañez de Kasep G, Alcocer-Varela J, Alarcón-Segovia D. Association of HLA-DR5 (DR11) with systemic sclerosis (scleroderma) in Mexican patients. Clin Exp Rheumatol. 1995 Jan-Feb;13(1):11–6.

60. Morel PA, Chang HJ, Wilson JW, Conte C, Saidman SL, Bray JD, et al. Severe systemic sclerosis with anti-topoisomerase I Antibodies is associated with an HLA-DRw11 allele. Hum Immunol. 1994 Jun;40(2):101–10.

61. Rands AL, Whyte J, Cox B, Hall ND, McHugh NJ. MHC class II associations with autoantibody and T cell immune responses to the scleroderma autoantigen topoisomerase I. J Autoimmun. 2000 Dec;15(4):451–8.

62. Beretta L, Rueda B, Marchini M, Santaniello A, Simeon CP, Fonollosa V, et al. Analysis of Class II human leucocyte antigens in Italian and Spanish systemic sclerosis. Rheumatology (Oxford). 2012 Jan 1;51(1):52–9.

63. Acosta-Herrera M, Kerick M, Lopéz-Isac E, Assassi S, Beretta L, Simeón-Aznar CP, 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 Aug;80(8):1040–7.

64. Bossini-Castillo L, López-Isac E, Mayes MD, Martín J. Genetics of systemic sclerosis. Semin Immunopathol. 2015 Sep;37(5):443–51.

65. López-Isac E, Acosta-Herrera M, Kerick M, Assassi S, Satpathy AT, Granja J, et al. GWAS for systemic sclerosis identifies multiple risk loci and

highlights fibrotic and vasculopathy pathways. Nat Commun. 2019 Oct 31;10(1):4955.

66. Radstake TRDJ, Gorlova O, Rueda B, Martin J-E, Alizadeh BZ, Palomino-Morales R, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. Nat Genet. 2010 May;42(5):426–9.

67. Sharif R, Mayes MD, Tan FK, Gorlova OY, Hummers LK, Shah AA, et al. IRF5 polymorphism predicts prognosis in patients with systemic sclerosis. Ann Rheum Dis. 2012 Jul;71(7):1197–202.

68. Wu M, Skaug B, Bi X, Mills T, Salazar G, Zhou X, et al. Interferon regulatory factor 7 (IRF7) represents a link between inflammation and fibrosis in the pathogenesis of systemic sclerosis. Ann Rheum Dis. 2019;78(11):1583–91.

69. Allanore Y, Saad M, Dieudé P, Avouac J, Distler JHW, Amouyel P, et al. Genome-Wide scan identifies TNIP1, PSORS1C1, and RHOB as novel risk loci for systemic sclerosis. PLoS Genet [Internet]. 2011;7(7). Available from: http://dx.doi.org/10.1371/journal.pgen.1002091

70. Martin J-E, Broen JC, Carmona FD, Teruel M, Simeon CP, Vonk MC, et al. Identification of CSK as a systemic sclerosis genetic risk factor through Genome Wide Association Study follow-up. Hum Mol Genet. 2012 Jun 15;21(12):2825–35.

71. López-Isac E, Campillo-Davo D, Bossini-Castillo L, Guerra SG, Assassi S, Simeón CP, et al. Influence of TYK2 in systemic sclerosis susceptibility: a new locus in the IL-12 pathway. Ann Rheum Dis. 2016 Aug;75(8):1521–6.

72. Bossini-Castillo L, Martin JE, Broen J, Gorlova O, Simeón CP, Beretta L, 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(4):926–33.

73. López-Isac E, Bossini-Castillo L, Guerra SG, Denton C, Fonseca C, Assassi S, et al. Identification of IL12RB1 as a novel systemic sclerosis susceptibility locus. Arthritis Rheumatol. 2014 Dec;66(12):3521–3.

74. Mayes MD, Bossini-Castillo L, Gorlova O, Martin JE, Zhou X, Chen WV, et al. Immunochip analysis identifies multiple susceptibility loci for systemic sclerosis. Am J Hum Genet. 2014;94(1):47–61.

75. Zochling J, Newell F, Charlesworth JC, Leo P, Stankovich J, Cortes A, et al. An Immunochip-based interrogation of scleroderma susceptibility variants identifies a novel association at DNASE1L3. Arthritis Res Ther. 2014 Oct 21;16(5):438.

76. Moreno-Moral A, Bagnati M, Koturan S, Ko J-H, Fonseca C, Harmston N, et al. Changes in macrophage transcriptome associate with systemic sclerosis and mediate contribution to disease risk. Ann Rheum Dis. 2018 Apr;77(4):596–601.

77. López-Isac E, Martín J-E, Assassi S, Simeón CP, Carreira P, Ortego-Centeno N, et al. Brief Report: IRF4 Newly Identified as a Common Susceptibility Locus for Systemic Sclerosis and Rheumatoid Arthritis in a Cross-Disease Meta-Analysis of Genome-Wide Association Studies. Arthritis Rheumatol. 2016 Sep;68(9):2338–44.

78. Márquez A, Kerick M, Zhernakova A, Gutierrez-Achury J, Chen WM, Onengut-Gumuscu S, et al. Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations. Genome Med [Internet]. 2018 Dec 20;10(1). Available from: http://dx.doi.org/10.1186/s13073-018-0604-8

79. González-Serna D, Ochoa E, López-Isac E, Julià A, Degenhardt F, Ortego-Centeno N, et al. A cross-disease meta-GWAS identifies four new susceptibility loci shared between systemic sclerosis and Crohn's disease. Sci Rep. 2020 Feb 5;10(1):1862.

80. Márquez A, Martín J. Genetic overlap between type 1 diabetes and other autoimmune diseases. Semin Immunopathol. 2022 Jan;44(1):81–97.

81. Acosta-Herrera M, Kerick M, González-Serna D, Wijmenga C, Franke A, Gregersen PK, et al. Genome-wide meta-analysis reveals shared new loci in systemic seropositive rheumatic diseases. Ann Rheum Dis. 2019;78(3):311–9.

82. Ortíz-Fernández L, Martín J, Alarcón-Riquelme ME. A summary on the genetics of systemic lupus erythematosus, rheumatoid arthritis, systemic sclerosis, and Sjögren's syndrome. Clin Rev Allergy Immunol. 2023 Jun;64(3):392–411.

83. Martin J-E, Assassi S, Diaz-Gallo L-M, Broen JC, Simeon CP, Castellvi I, et al. A systemic sclerosis and systemic lupus erythematosus pan-meta-GWAS reveals new shared susceptibility loci. Hum Mol Genet. 2013 Oct 1;22(19):4021–9.

84. Bossini-Castillo L, Martin JE, Broen J, Simeon CP, Beretta L, Gorlova OY, et al. Confirmation of*TNIP1*but not*RHOB*and*PSORS1C1*as systemic sclerosis risk factors in a large independent replication study. Ann Rheum Dis. 2013 Apr;72(4):602–7.

85. Terao C, Ohmura K, Kawaguchi Y, Nishimoto T, Kawasaki A, Takehara K, et al. PLD4 as a novel susceptibility gene for systemic sclerosis in a Japanese population. Arthritis Rheum. 2013 Feb;65(2):472–80.

86. Gorlova OY, Li Y, Gorlov I, Ying J, Chen WV, Assassi S, et al. Gene-level association analysis of systemic sclerosis: A comparison of African-Americans and White populations. PLoS One. 2018 Jan 2;13(1):e0189498.

87. Xu Y, Wang W, Tian Y, Liu J, Yang R. Polymorphisms in STAT4 and IRF5 increase the risk of systemic sclerosis: a meta-analysis. Int J Dermatol. 2016 Apr;55(4):408–16.

88. Carmona FD, Gutala R, Simeón CP, Carreira P, Ortego-Centeno N, Vicente-Rabaneda E, et al. Novel identification of the *IRF7* region as an anticentromere autoantibody propensity locus in systemic sclerosis. Ann Rheum Dis. 2012 Jan;71(1):114–9.

89. Gorlova O, Martin J-E, Rueda B, Koeleman BPC, Ying J, Teruel M, et al. Identification of novel genetic markers associated with clinical phenotypes of systemic sclerosis through a genome-wide association strategy. PLoS Genet. 2011 Jul 14;7(7):e1002178.

90. Terao C, Kawaguchi T, Dieude P, Varga J, Kuwana M, Hudson M, et al. Transethnic meta-Analysis identifies GSDMA and PRDM1 as susceptibility genes to systemic sclerosis. Ann Rheum Dis. 2017;76(6):1150–8.

91. Bossini-Castillo L, Broen JCA, Simeon CP, Beretta L, Vonk MC, Ortego-Centeno N, et al. A replication study confirms the association of *TNFSF4 (OX40L)* polymorphisms with systemic sclerosis in a large European cohort. Ann Rheum Dis. 2011 Apr;70(4):638–41.

92. Dieude P, Boileau C, Guedj M, Avouac J, Ruiz B, Hachulla E, et al. Independent replication establishes the CD247 gene as a genetic systemic sclerosis susceptibility factor. Ann Rheum Dis. 2011 Sep 1;70(9):1695–6.

93. López-Isac E, Bossini-Castillo L, Simeon CP, Egurbide MV, Alegre-Sancho JJ, Callejas JL, et al. A genome-wide association study follow-up suggests

a possible role for PPARG in systemic sclerosis susceptibility. Arthritis Res Ther. 2014;16(1):2–9.

94. Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. Nucleic Acids Res. 2023 Jan 6;51(D1):D977–85.

95. Uffelmann E, Huang QQ, Munung NS, de Vries J, Okada Y, Martin AR, et al. Genome-wide association studies. Nature Reviews Methods Primers. 2021 Aug 26;1(1):1–21.

96. Gutierrez-Arcelus M, Rich SS, Raychaudhuri S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. Nat Rev Genet. 2016 Mar;17(3):160–74.

97. Ding J, Frantzeskos A, Orozco G. Functional genomics in autoimmune diseases. Hum Mol Genet. 2020;00(00):1–7.

98. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. Nat Rev Genet. 2018 Sep;19(9):581–90.

99. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. Nat Rev Genet. 2016;17(7):392–406.

100. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219–24.

101. Knevel R, le Cessie S, Terao CC, Slowikowski K, Cui J, Huizinga TWJ, et al. Using genetics to prioritize diagnoses for rheumatology outpatients with inflammatory arthritis. Sci Transl Med [Internet]. 2020 May 27;12(545). Available from: http://dx.doi.org/10.1126/scitranslmed.aay1548

102. Kachuri L, Chatterjee N, Hirbo J, Schaid DJ, Martin I, Kullo IJ, et al. Principles and methods for transferring polygenic risk scores across global populations. Nat Rev Genet [Internet]. 2023 Aug 24; Available from: http://dx.doi.org/10.1038/s41576-023-00637-2

103. Boehm FJ, Zhou X. Statistical methods for Mendelian randomization in genome-wide association studies: A review. Comput Struct Biotechnol J. 2022 May 14;20:2338–51.

104. Burgess S, Scott RA, Timpson NJ, Smith GD, Thompson SG. Using published data in Mendelian randomization: A blueprint for efficient identification of causal risk factors. Eur J Epidemiol. 2015;30(7):543–52.

105. Zhao SS, MacKie SL, Zheng J. Why clinicians should know about Mendelian randomization. Rheumatology. 2021;60(4):1577–9.

106. Bowden J, Smith GD, Burgess S. Mendelian randomization with invalid instruments: Effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512–25.

107. Yuan S, Chen J, Ruan X, Vithayathil M, Kar S, Li X, et al. Rheumatoid arthritis and risk of site-specific cancers: Mendelian randomization study in European and East Asian populations. Arthritis Res Ther. 2022 Dec 13;24(1):270.

108. Lee YH. Causal association of gut microbiome on the risk of rheumatoid arthritis: a Mendelian randomisation study. Ann Rheum Dis. 2022 Jan;81(1):e3.

109. Wang X-F, Xu W-J, Wang F-F, Leng R, Yang X-K, Ling H-Z, et al. Telomere Length and Development of Systemic Lupus Erythematosus: A Mendelian Randomization Study. Arthritis Rheumatol. 2022 Dec;74(12):1984–90.

110. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, et al. Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. PLoS Med. 2016 Jun;13(6):e1001976.

111. Truchetet ME, Brembilla NC, Chizzolini C. Current Concepts on the Pathogenesis of Systemic Sclerosis. Clin Rev Allergy Immunol. 2023 Jun;64(3):262–83.

112. Mostmans Y, Cutolo M, Giddelo C, Decuman S, Melsens K, Declercq H, et al. The role of endothelial cells in the vasculopathy of systemic sclerosis: A systematic review. Autoimmun Rev. 2017 Aug;16(8):774–86.

113. Matucci-Cerinic M, Kahaleh B, Wigley FM. Review: evidence that systemic sclerosis is a vascular disease. Arthritis Rheum. 2013 Aug;65(8):1953–62.

114. Bhattacharyya S, Wei J, Varga J. Understanding fibrosis in systemic sclerosis: shifting paradigms, emerging opportunities. Nat Rev Rheumatol. 2011 Oct 25;8(1):42–54.

115. Varga J, Abraham D. Systemic sclerosis: a prototypic multisystem fibrotic disorder. J Clin Invest. 2007 Mar;117(3):557–67.

116. Korman B. Evolving insights into the cellular and molecular pathogenesis of fibrosis in systemic sclerosis. Transl Res. 2019 Jul;209:77–89.

117. Tsou P-S, Varga J, O'Reilly S. Advances in epigenetics in systemic sclerosis: molecular mechanisms and therapeutic potential. Nat Rev Rheumatol. 2021 Oct;17(10):596–607.

118. Sfikakis PP, McCune BK, Tsokos M, Aroni K, Vayiopoulos G, Tsokos GC. Immunohistological demonstration of transforming growth factor-beta isoforms in the skin of patients with systemic sclerosis. Clin Immunol Immunopathol. 1993 Nov;69(2):199–204.

119. Lafyatis R. Transforming growth factor β- At the centre of systemic sclerosis. Nat Rev Rheumatol. 2014;10(12):706–19.

120. Van Caam A, Vonk M, Van Den Hoogen F, Van Lent P, Van Der Kraan P. Unraveling SSc pathophysiology; The myofibroblast. Front Immunol. 2018;9(NOV):1–22.

121. Kerick M, Acosta-Herrera M, Simeón-Aznar CP, Callejas JL, Assassi S, International SSc Group, et al. Complement component C4 structural variation and quantitative traits contribute to sex-biased vulnerability in systemic sclerosis. NPJ Genom Med. 2022 Oct 5;7(1):57.

122. Narasimhan PB, Marcovecchio P, Hamers AAJ, Hedrick CC. Nonclassical Monocytes in Health and Disease. Annu Rev Immunol. 2019;37(1):439–56.

123. Lescoat A, Lecureur V, Roussel M, Sunnaram BL, Ballerie A, Coiffier G, et al. CD16-positive circulating monocytes and fibrotic manifestations of systemic sclerosis. Clin Rheumatol. 2017 Jul;36(7):1649–54.

124. Schneider L, Marcondes NA, Hax V, da Silva Moreira IF, Ueda CY, Piovesan RR, et al. Flow cytometry evaluation of CD14/CD16 monocyte subpopulations in systemic sclerosis patients: a cross sectional controlled study. Adv Rheumatol. 2021 May 22;61(1):27.

125. Skaug B, Assassi S. Type I interferon dysregulation in Systemic Sclerosis. Cytokine. 2020 Aug;132:154635.

126. Matei A-E, Kubánková M, Xu L, Györfi A-H, Boxberger E, Soteriou D, et al. Biophysical phenotyping of circulating immune cells identifies a distinct monocyte-driven signature in systemic sclerosis. Arthritis Rheumatol [Internet]. 2022 Oct 25; Available from: http://dx.doi.org/10.1002/art.42394

127. Soldano S, Trombetta AC, Contini P, Tomatis V, Ruaro B, Brizzolara R, et al. Increase in circulating cells coexpressing M1 and M2 macrophage surface markers in patients with systemic sclerosis. Ann Rheum Dis. 2018 Dec;77(12):1842–5.

128. Trombetta AC, Soldano S, Contini P, Tomatis V, Ruaro B, Paolino S, et al. A circulating cell population showing both M1 and M2 monocyte/macrophage surface markers characterizes systemic sclerosis patients with lung involvement. Respir Res. 2018 Sep 24;19(1):186.

129. van der Kroef M, Carvalheiro T, Rossato M, de Wit F, Cossu M, Chouri E, et al. CXCL4 triggers monocytes and macrophages to produce PDGF-BB, culminating in fibroblast activation: Implications for systemic sclerosis. J Autoimmun. 2020 Jul;111:102444.

130. Xue D, Tabib T, Morse C, Yang Y, Domsic RT, Khanna D, et al. Expansion of Fcγ Receptor IIIa-Positive Macrophages, Ficolin 1-Positive Monocyte-Derived Dendritic Cells, and Plasmacytoid Dendritic Cells Associated With Severe Skin Disease in Systemic Sclerosis. Arthritis Rheumatol. 2022 Feb;74(2):329–41.

131. Mathian A, Parizot C, Dorgham K, Trad S, Arnaud L, Larsen M, et al. Activated and resting regulatory T cell exhaustion concurs with high levels of interleukin-22 expression in systemic sclerosis lesions. Ann Rheum Dis. 2012 Jul;71(7):1227–34.

132. Frantz C, Auffray C, Avouac J, Allanore Y. Regulatory T Cells in Systemic Sclerosis. Front Immunol. 2018 Oct 15;9:2356.

133. Jin W, Zheng Y, Zhu P. T cell abnormalities in systemic sclerosis. Autoimmun Rev. 2022 Nov;21(11):103185.

134. Gaydosik AM, Tabib T, Domsic R, Khanna D, Lafyatis R, Fuschiotti P. Single-cell transcriptome analysis identifies skin-specific T-cell responses in systemic sclerosis. Ann Rheum Dis. 2021

Nov;80(11):1453–60.

135. Beesley CF, Goldman NR, Taher TE, Denton CP, Abraham DJ, Mageed RA, et al. Dysregulated B cell function and disease pathogenesis in systemic sclerosis. Front Immunol. 2022;13:999008.

136. Fox DA, Lundy SK, Whitfield ML, Berrocal V, Campbell P, Rasmussen S, et al. Lymphocyte subset abnormalities in early diffuse cutaneous systemic sclerosis. Arthritis Res Ther. 2021 Jan 6;23(1):10.

137. Milano A, Pendergrass SA, Sargent JL, George LK, McCalmont TH, Connolly MK, et al. Molecular subsets in the gene expression signatures of scleroderma skin. PLoS One. 2008 Jul 16;3(7):e2696.

138. Sargent JL, Milano A, Bhattacharyya S, Varga J, Connolly MK, Chang HY, et al. A TGFbeta-responsive gene signature is associated with a subset of diffuse scleroderma with increased disease severity. J Invest Dermatol. 2010 Mar;130(3):694–705.

139. Derrett-Smith EC, Martyanov V, Chighizola CB, Moinzadeh P, Campochiaro C, Khan K, et al. Limited cutaneous systemic sclerosis skin demonstrates distinct molecular subsets separated by a cardiovascular development gene expression signature. Arthritis Res Ther. 2017 Jul 4;19(1):156.

140. Pendergrass SA, Hayes E, Farina G, Lemaire R, Farber HW, Whitfield ML, et al. Limited systemic sclerosis patients with pulmonary arterial hypertension show biomarkers of inflammation and vascular injury. PLoS One. 2010 Aug 17;5(8):e12106.

141. Lenna S, Farina AG, Martyanov V, Christmann RB, Wood TA, Farber HW, et al. Increased expression of endoplasmic reticulum stress and unfolded protein response genes in peripheral blood mononuclear cells from patients with limited cutaneous systemic sclerosis and pulmonary arterial hypertension. Arthritis Rheum. 2013 May;65(5):1357–66.

142. Beretta L, Barturen G, Vigone B, Bellocchi C, Hunzelmann N, De Langhe E, et al. Genome-wide whole blood transcriptome profiling in a large European cohort of systemic sclerosis patients. Ann Rheum Dis. 2020 Sep;79(9):1218–26.

143. Kerick M, González-Serna D, Carnero-Montoro E, Teruel M, Acosta-Herrera M, Makowska Z, et al. Expression Quantitative Trait Locus Analysis in Systemic Sclerosis Identifies New Candidate Genes Associated With Multiple Aspects of Disease Pathology. Arthritis

Rheumatol. 2021 Jul;73(7):1288–300.

144. González-Serna D, Shi C, Kerick M, Hankinson J, Ding J, McGovern A, et al. Identification of Mechanisms by Which Genetic Susceptibility Loci Influence Systemic Sclerosis Risk Using Functional Genomics in Primary T Cells and Monocytes. Arthritis Rheumatol. 2023 Jun;75(6):1007–20.

145. Mahoney JM, Taroni J, Martyanov V, Wood TA, Greene CS, Pioli PA, et al. Systems level analysis of systemic sclerosis shows a network of immune and profibrotic pathways connected with genetic polymorphisms. PLoS Comput Biol. 2015 Jan;11(1):e1004005.

146. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol. 2019 Jun 19;15(6):e8746.

147. Efremova M, Vento-Tormo R, Park JE, Teichmann SA, James KR. Immunology in the Era of Single-Cell Technologies. Annu Rev Immunol. 2020;38:727–57.

148. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature. 2012 Sep 20;489(7416):391–9.

149. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM 3rd, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun 13;177(7):1888-1902.e21.

150. Zhang MJ, Hou K, Dey KK, Sakaue S, Jagadeesh KA, Weinand K, et al. Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nat Genet. 2022 Oct;54(10):1572–80.

151. Kharchenko PV, Silberstein L, Scadden DT. Bayesian approach to single-cell differential expression analysis. Nat Methods. 2014 Jul;11(7):740–2.

152. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018 Jan;18(1):35–45.

153. Stewart BJ, Ferdinand JR, Clatworthy MR. Using single-cell technologies to map the human immune system - implications for nephrology. Nat Rev Nephrol. 2020 Feb;16(2):112–28.

154. Cerosaletti K, Barahmand-Pour-Whitman F, Yang J, DeBerg HA, Dufort MJ, Murray SA, et al. Single-Cell RNA Sequencing Reveals Expanded Clones of Islet Antigen-Reactive CD4 T Cells in Peripheral

Blood of Subjects with Type 1 Diabetes. J Immunol. 2017 Jul 1;199(1):323–35.

155. Sharma S, Pettus J, Gottschalk M, Abe B, Gottlieb P, Teyton L. Single-Cell Analysis of CD4 T Cells in Type 1 Diabetes: From Mouse to Man, How to Perform Mechanistic Studies. Diabetes. 2019 Oct;68(10):1886–91.

156. Stephenson W, Donlin LT, Butler A, Rozo C, Bracken B, Rashidfarrokhi A, et al. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. Nat Commun. 2018 Feb 23;9(1):791.

157. Der E, Suryawanshi H, Buyon J, Tuschl T, Putterman C. Single-cell RNA sequencing for the study of lupus nephritis. Lupus Sci Med. 2019 Jun 4;6(1):e000329.

158. Valenzi E, Bulik M, Tabib T, Morse C, Sembrat J, Trejo Bittar H, et al. Single-cell analysis reveals fibroblast heterogeneity and myofibroblasts in systemic sclerosis-associated interstitial lung disease. Ann Rheum Dis. 2019 Oct;78(10):1379–87.

159. Tabib T, Huang M, Morse N, Papazoglou A, Behera R, Jia M, et al. Myofibroblast transcriptome indicates SFRP2 fibroblast progenitors in systemic sclerosis skin. Nat Commun. 2021 Jul 19;12(1):4384.

160. Gur C, Wang S-Y, Sheban F, Zada M, Li B, Kharouf F, et al. LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma. Cell. 2022 Apr 14;185(8):1373-1388.e20.

161. Papazoglou A, Huang M, Bulik M, Lafyatis A, Tabib T, Morse C, et al. Epigenetic regulation of profibrotic macrophages in systemic sclerosis-associated interstitial lung disease. Arthritis Rheumatol. 2022 Dec;74(12):2003–14.

162. Clark KEN, Xu S, Attah M, Ong VH, Buckley CD, Denton CP. Single-cell analysis reveals key differences between early-stage and late-stage systemic sclerosis skin across autoantibody subgroups. Ann Rheum Dis [Internet]. 2023 Aug 14; Available from: http://dx.doi.org/10.1136/ard-2023-224184

163. Jamin C, Le Lann L, Alvarez-Errico D, Barbarroja N, Cantaert T, Ducreux J, et al. Multi-center harmonization of flow cytometers in the context of the European “PRECISESADS” project. Autoimmun Rev. 2016 Nov;15(11):1038–45.

164. Molecular Reclassification to Find Clinically Useful Biomarkers for Systemic Autoimmune Diseases: Inception Cohort (PRECISESADSI) [Internet]. clinicaltrials.gov. 2016. Available from: https://clinicaltrials.gov/ct2/show/NCT02890134

165. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet. 2015 Sep;47(9):979–86.

166. Reid S, Alexsson A, Frodlund M, Morris D, Sandling JK, Bolin K, et al. High genetic risk score is associated with early disease onset, damage accrual and decreased survival in systemic lupus erythematosus. Ann Rheum Dis. 2020 Mar;79(3):363–9.

167. Chen L, Wang Y-F, Liu L, Bielowka A, Ahmed R, Zhang H, et al. Genome-wide assessment of genetic risk for systemic lupus erythematosus and disease severity. Hum Mol Genet [Internet]. 2020 Feb 20; Available from: http://dx.doi.org/10.1093/hmg/ddaa030

168. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nat Genet. 2012 Mar 25;44(5):483–9.

169. Inshaw JRJ, Cutler AJ, Burren OS, Stefana MI, Todd JA. Approaches and advances in the genetic causes of autoimmune disease and their implications. Nat Immunol. 2018 Jul;19(7):674–84.

170. Peterson RE, Kuchenbaecker K, Walters RK, Chen C-Y, Popejoy AB, Periyasamy S, et al. Genome-wide Association Studies in Ancestrally Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. Cell. 2019 Oct 17;179(3):589–603.

171. Schork AJ, Schork MA, Schork NJ. Genetic risks and clinical rewards. Nat Genet. 2018 Sep;50(9):1210–1.

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

173. Iaccarino L, Gatto M, Bettio S, Caso F, Rampudda M, Zen M, et al. Overlap connective tissue disease syndromes. Autoimmun Rev. 2013 Jan;12(3):363–73.

174. Barturen G, Beretta L, Cervera R, Van Vollenhoven R, Alarcón-

Riquelme ME. Moving towards a molecular taxonomy of autoimmune rheumatic diseases. Nat Rev Rheumatol. 2018 Feb 21;14(3):180.

175. Gabrielli A, Avvedimento EV, Krieg T. Scleroderma. N Engl J Med. 2009 May 7;360(19):1989–2003.

176. Barnes J, Mayes MD. Epidemiology of systemic sclerosis: incidence, prevalence, survival, risk factors, malignancy, and environmental triggers. Curr Opin Rheumatol. 2012 Mar;24(2):165–70.

177. Mayes MD, Lacey JV, Beebe-Dimmer J, Gillespie BW, Cooper B, Laing TJ, et al. Prevalence, incidence, survival, and disease characteristics of systemic sclerosis in a large US population [Internet]. Vol. 48, Arthritis & Rheumatism. 2003. p. 2246–55. Available from: http://dx.doi.org/10.1002/art.11073

178. Affandi AJ, Carvalheiro T, Ottria A, Broen JC, Bossini-Castillo L, Tieland RG, et al. Low RUNX3 expression alters dendritic cell function in patients with systemic sclerosis and contributes to enhanced fibrosis. Ann Rheum Dis. 2019 Sep;78(9):1249–59.

179. van Bon L, Affandi AJ, Broen J, Christmann RB, Marijnissen RJ, Stawski L, et al. Proteome-wide analysis and CXCL4 as a biomarker in systemic sclerosis. N Engl J Med. 2014 Jan 30;370(5):433–43.

180. Skaug B, Khanna D, Swindell WR, Hinchcliff ME, Frech TM, Steen VD, et al. Global skin gene expression analysis of early diffuse cutaneous systemic sclerosis shows a prominent innate and adaptive inflammatory profile. Ann Rheum Dis. 2020 Mar;79(3):379–86.

181. Denton CP, Khanna D. Systemic sclerosis. Lancet. 2017 Oct;390(10103):1685–99.

182. Miller JB, Gandhi N, Clarke J, McMahan Z. Gastrointestinal Involvement in Systemic Sclerosis: An Update. J Clin Rheumatol. 2018 Sep;24(6):328–37.

183. Hughes M, Heal C, Siegert E, Hachulla E, Airó P, Riccardi A, et al. Significant weight loss in systemic sclerosis: a study from the EULAR Scleroderma Trials and Research (EUSTAR) database. Ann Rheum Dis. 2020 Aug;79(8):1123–5.

184. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018 Jul 12;362:k601.

185. Carreras-Torres R, Ibáñez-Sanz G, Obón-Santacana M, Duell EJ, Moreno V. Identifying environmental risk factors for inflammatory bowel diseases: a Mendelian randomization study. Sci Rep. 2020 Nov 6;10(1):19273.

186. Julià A, Martínez-Mateu SH, Domènech E, Cañete JD, Ferrándiz C, Tornero J, et al. Food groups associated with immune-mediated inflammatory diseases: a Mendelian randomization and disease severity study. Eur J Clin Nutr [Internet]. 2021 Apr 23; Available from: http://dx.doi.org/10.1038/s41430-021-00913-6

187. González-Serna D, Villanueva-Martin G, Acosta-Herrera M, Márquez A, Martín J. Approaching Shared Pathophysiology in Immune-Mediated Diseases through Functional Genomics. Genes [Internet]. 2020 Dec 9;11(12). Available from: http://dx.doi.org/10.3390/genes11121482

188. Francisco V, Pino J, Campos-Cabaleiro V, Ruiz-Fernández C, Mera A, Gonzalez-Gay MA, et al. Obesity, Fat Mass and Immune System: Role for Leptin. Front Physiol. 2018 Jun 1;9:640.

189. Żółkiewicz J, Stochmal A, Rudnicka L. The role of adipokines in systemic sclerosis: a missing link? Arch Dermatol Res. 2019 May;311(4):251–63.

190. Iannone F, Praino E, Rotondo C, Natuzzi D, Bizzoca R, Lacarpia N, et al. Body mass index and adipokines/cytokines dysregulation in Systemic Sclerosis. Clin Exp Immunol [Internet]. 2021 Aug 6; Available from: http://dx.doi.org/10.1111/cei.13651

191. Driskell RR, Lichtenberger BM, Hoste E, Kretzschmar K, Simons BD, Charalambous M, et al. Distinct fibroblast lineages determine dermal architecture in skin development and repair. Nature. 2013 Dec 12;504(7479):277–81.

192. Salzer MC, Lafzi A, Berenguer-Llergo A, Youssif C, Castellanos A, Solanas G, et al. Identity Noise and Adipogenic Traits Characterize Dermal Fibroblast Aging. Cell. 2018 Nov 29;175(6):1575-1590.e22.

193. Zhang Z, Shao M, Hepler C, Zi Z, Zhao S, An YA, et al. Dermal adipose tissue has high plasticity and undergoes reversible dedifferentiation in mice [Internet]. Vol. 129, Journal of Clinical Investigation. 2019. p. 5327–42. Available from: http://dx.doi.org/10.1172/jci130239

194. Taki Z, Gostjeva E, Thilly W, Yaseen B, Lopez H, Mirza M, et al. Pathogenic Activation of Mesenchymal Stem Cells Is Induced by the

Disease Microenvironment in Systemic Sclerosis. Arthritis Rheumatol. 2020 Aug;72(8):1361–74.

195. Huxley R, Mendis S, Zheleznyakov E, Reddy S, Chan J. Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk--a review of the literature. Eur J Clin Nutr. 2010 Jan;64(1):16–22.

196. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018 Feb 6;19(1):15.

197. Bossini-Castillo L, Villanueva-Martin G, Kerick M, Acosta-Herrera M, López-Isac E, Simeón CP, et al. Genomic Risk Score impact on susceptibility to systemic sclerosis. Ann Rheum Dis. 2021 Jan;80(1):118–27.

198. Villanueva-Martin G, Acosta-Herrera M, Carmona EG, Kerick M, Ortego-Centeno N, Callejas-Rubio JL, et al. Non-classical circulating monocytes expressing high levels of microsomal prostaglandin E2 synthase-1 tag an aberrant IFN-response in systemic sclerosis. J Autoimmun. 2023 Aug 24;140:103097.

199. Márquez-Luna C, Loh P-R, South Asian Type 2 Diabetes (SAT2D) Consortium, SIGMA Type 2 Diabetes Consortium, Price AL. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol. 2017 Dec;41(8):811–23.

200. Zhao Z, Fritsche LG, Smith JA, Mukherjee B, Lee S. The construction of cross-population polygenic risk scores using transfer learning. Am J Hum Genet. 2022 Nov 3;109(11):1998–2008.

201. Campos AI, Namba S, Lin S-C, Nam K, Sidorenko J, Wang H, et al. Boosting the power of genome-wide association studies within and across ancestries by using polygenic scores. Nat Genet. 2023 Oct;55(10):1769–76.

202. Zhang H, Zhan J, Jin J, Zhang J, Lu W, Zhao R, et al. A new method for multiancestry polygenic prediction improves performance across diverse populations. Nat Genet. 2023 Oct;55(10):1757–68.

203. González-Serna D, López-Isac E, Yilmaz N, Gharibdoost F, Jamshidi A, Kavosi H, et al. Analysis of the genetic component of systemic sclerosis in Iranian and Turkish populations through a genome-wide association study. Rheumatology. 2019;58(2):289–98.

204. Ishigaki K, Sakaue S, Terao C, Luo Y, Sonehara K, Yamaguchi K, et al. Multi-ancestry genome-wide association analyses identify novel genetic mechanisms in rheumatoid arthritis. Nat Genet. 2022 Nov;54(11):1640–51.

205. Nagy A, Palmer E, Polivka L, Eszes N, Vincze K, Barczi E, et al. Treatment and systemic sclerosis interstitial lung disease outcome: The overweight paradox. Biomedicines. 2022 Feb 13;10(2):434.

206. Kaniecki T, Abdi T, McMahan ZH. A practical approach to the evaluation and management of gastrointestinal symptoms in patients with systemic sclerosis. Best Pract Res Clin Rheumatol. 2021 Sep;35(3):101666.

207. Kowal-Bielecka O, Fransen J, Avouac J, Becker M, Kulak A, Allanore Y, et al. Update of EULAR recommendations for the treatment of systemic sclerosis. Ann Rheum Dis. 2017 Aug;76(8):1327–39.

208. Hansi N, Thoua N, Carulli M, Chakravarty K, Lal S, Smyth A, et al. Consensus best practice pathway of the UK scleroderma study group: gastrointestinal manifestations of systemic sclerosis. Clin Exp Rheumatol. 2014 Nov 5;32(6 Suppl 86):S-214-21.

209. Radić M, Kolak E, Đogaš H, Gelemanović A, Bučan Nenadić D, Vučković M, et al. Body composition parameters in systemic sclerosis-a systematic review and meta-analysis. Rheumatology [Internet]. 2023 Aug 30; Available from: https://academic.oup.com/rheumatology/advance-article/doi/10.1093/rheumatology/kead418/7255900

210. Rodriguez-Martin I, Villanueva-Martin G, Guillen-Del-Castillo A, Ortego-Centeno N, Callejas JL, Simeón-Aznar CP, et al. Contribution of Telomere Length to Systemic Sclerosis Onset: A Mendelian Randomization Study. Int J Mol Sci. 2023 Oct 25;24(21):15589.

211. Allen PC, Smith S, Wilson RC, Wirth JR, Wilson NH, Baker Frost D, et al. Distinct genome-wide DNA methylation and gene expression signatures in classical monocytes from African American patients with systemic sclerosis. Clin Epigenetics. 2023 Feb 17;15(1):25.

212. Padilla CM, Valenzi E, Tabib T, Nazari B, Sembrat J, Rojas M, et al. Increased CD8+ tissue resident memory T cells, regulatory T cells, and activated natural killer cells in systemic sclerosis lungs. Rheumatology [Internet]. 2023 Jun 13; Available from:

http://dx.doi.org/10.1093/rheumatology/kead273

213. Kobayashi S, Nagafuchi Y, Okubo M, Sugimori Y, Shirai H, Hatano H, et al. Integrated bulk and single-cell RNA-sequencing identified disease-relevant monocytes and a gene network module underlying systemic sclerosis. J Autoimmun. 2021 Jan;116:102547.

214. Balanescu P, Balanescu E, Baicus C, Balanescu A. S100A6, calumenin and cytohesin 2 as biomarkers for cutaneous involvement in systemic sclerosis patients: A case control study. J Pers Med. 2021 May 2;11(5):368.

215. Denton CP, Xu S, Zhang F, Maclean RH, Clark KEN, Borchert S, et al. Clinical and pathogenic significance of S100A4 overexpression in systemic sclerosis. Ann Rheum Dis. 2023 Sep;82(9):1205–17.

216. Švec X, Štorkánová H, Trinh-Minh T, Tran MC, Štorkánová L, Hulejová H, et al. S100A4 neutralizing monoclonal antibody 6B12 counteracts the established experimental skin fibrosis induced by bleomycin. Rheumatology [Internet]. 2023 Jun 14; Available from: http://dx.doi.org/10.1093/rheumatology/kead295

217. Belch JJ, Madhok R, Shaw B, Leiberman P, Sturrock RD, Forbes CD. Double-blind trial of CL115,347, a transdermally absorbed prostaglandin E2 analogue, in treatment of Raynaud's phenomenon. Lancet. 1985 May 25;1(8439):1180–3.

218. Gogulska Z, Smolenska Z, Turyn J, Mika A, Zdrojewski Z. Lipid Alterations in Systemic Sclerosis. Front Mol Biosci. 2021 Dec 21;8:761721.

219. Koeberle A, Werz O. Perspective of microsomal prostaglandin E2 synthase-1 as drug target in inflammation-related disorders. Biochem Pharmacol. 2015 Nov 1;98(1):1–15.

220. Liu M, Wu W, Sun X, Yang J, Xu J, Fu W, et al. New insights into CD4(+) T cell abnormalities in systemic sclerosis. Cytokine Growth Factor Rev. 2016 Apr;28:31–6.

221. Guo C, Liu Q, Zong D, Zhang W, Zuo Z, Yu Q, et al. Single-cell transcriptome profiling and chromatin accessibility reveal an exhausted regulatory CD4+ T cell subset in systemic lupus erythematosus. Cell Rep. 2022 Nov 8;41(6):111606.

222. Argyriou A, Wadsworth MH 2nd, Lendvai A, Christensen SM,

Hensvold AH, Gerstner C, et al. Single cell sequencing identifies clonally expanded synovial CD4+ TPH cells expressing GPR56 in rheumatoid arthritis. Nat Commun. 2022 Jul 13;13(1):4046.

223. Peters SAE, Woodward M. A roadmap for sex- and gender-disaggregated health research. BMC Med. 2023 Sep 13;21(1):354.

224. Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell-cell communication from combined expression of multi-subunit ligand-receptor complexes. Nat Protoc. 2020 Apr;15(4):1484–506.

225. Browaeys R, Saelens W, Saeys Y. NicheNet: modeling intercellular communication by linking ligands to target genes. Nat Methods. 2020 Feb;17(2):159–62.

The heart of science is an essential balance between two seemingly contradictory attitudes—an openness to new ideas, no matter how bizarre or counterintuitive, and the most ruthlessly skeptical scrutiny of all ideas, old and new.

Carl Sagan