# *Single-Cell RNA Sequencing Reveals Impaired Regulatory T Cell Function and a Pro-inflammatory Th17 Profile in Systemic Sclerosis*

Gonzalo Villanueva-Martin[1,2,3,‡], Gonzalo Borrego-Yaniz[1], Marialbert Acosta-Herrera[1,4], José Luis Callejas-Rubio[4,5], Norberto Ortego[5], Norbert Mages[6], Stefan Börno[6], Maria Gutierrez-Arcelus[2,3], Javier Martin[1,*] and Lara Bossini-Castillo[7,8,*].

1. Department of Cell Biology and Immunology, Institute of Parasitology and Biomedicine López-Neyra, CSIC, Granada, Spain.
2. Division of Immunology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA
3. Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, MA, USA
4. Systemic Autoimmune Disease Unit, Hospital Clínico San Cecilio, Instituto de Investigación Biosanitaria Ibs. GRANADA, Granada, Spain.
5. Department of Medicine, University of Granada, Granada, Spain.
6. Sequencing Core Facility, Max Planck Institute for Molecular Genetics, 14195 Berlin, Germany.
7. Departamento de Genética e Instituto de Biotecnología, Centro de Investigación Biomédica (CIBM), Universidad de Granada, 18100 Granada, Spain.
8. Advanced Therapies and Biomedical Technologies (TEC-14), Instituto de Investigación Biosanitaria Ibs. GRANADA, 18016 Granada, Spain.

* These authors contributed equally.
‡Corresponding author: Gonzalo Villanueva-Martin (gonzalo.villanuevamartin@childrens.harvard.edu).

# Abstract

**Objective:** Systemic sclerosis (SSc) is a complex autoimmune disease characterized by immune dysregulation, fibrosis, and vascular abnormalities. This study introduces a novel single-cell RNA sequencing (scRNA-seq) approach to comprehensively profile the CD4+ T cell subpopulations in SSc patients, offering unprecedented insights into the cellular and molecular underpinnings of the disease.
**Methods:** We employed scRNA-seq to analyze over 80,000 CD4+ T cells from 8 SSc patients and 8 healthy controls, identifying significant alterations in the composition and gene expression profiles of several T CD4+ cell subsets.

**Results:** Key findings include the overexpression of *FCRL3* in regulatory T cells (Tregs), which suggests a reduced suppressive function in SSc, contributing to the disease's immune imbalance. Additionally, we identified a pro-inflammatory and glucocorticoid-resistant profile in Th17 lymphocytes, marked by the overexpression of *MDR1*, reinforcing the chronic inflammatory state in SSc. A marked reduction in *AREG*+ naive T cells, crucial for Treg function, further indicates compromised immune regulation. We also observed an increased population of highly activated Th2 cells highlighting the pro-inflammatory environment in SSc. Moreover, significant clonal expansion in central memory T cells was noted, but no significant biases in V(D)J gene usage or epitope recognition were detected between SSc patients and controls.

**Conclusion:** Our findings advance the understanding of SSc pathogenesis by revealing specific immune dysregulations at a single-cell level, particularly impaired Treg function and a skewed Th17 and Th2 response. These insights could pave the way for targeted therapeutic strategies, representing a significant step forward in autoimmune disease research.

# Introduction

Systemic sclerosis or scleroderma (SSc), is a chronic immune-mediated inflammatory disease (IMID). The pathogenesis of SSc involves an intricate interplay of immune cell dysfunction, vascular damage, and progressive skin fibrosis [1]. Additionally, SSc predominantly affects middle-aged women [1].

Clinical and genetic evidence underscores a significant disruption of the immune system in the development of SSc, with a particularly notable involvement of lymphocyte activation and T cell subsets [2,3]. Among these, CD4+ T cells play a pivotal role not only as regulators of the immune response but also as effectors contributing to inflammation and tissue damage [2,3]. Additionally, the association of more than 30 genetic loci with SSc [4], many of which are related to T cell activation, highlights the importance of studying the CD4+ T cell lineage to better understand SSc susceptibility and to develop more targeted therapies.

Systemic immune deregulation is a SSc hallmark, but no increase in absolute cell numbers has been observed to date in the different subsets of circulating CD4+ helper T cells (Th) [5]. However, an imbalance in Th1 and Th2 cytokine production is well-established in SSc patients [5]. Furthermore, a skewed differentiation of Th cells towards the proinflammatory Th17 phenotype, as opposed to regulatory T cells (Tregs) that control the immune response, has been recognized as a feature of SSc [6,7]. Th17 cells have been implicated in the pathogenesis of SSc, particularly in vascular damage [5]. Reports on Th17 abundance in SSc peripheral blood are inconsistent [8,9]. Despite this controversy, imbalanced Th17 and Treg activity is considered critical in SSc pathogenesis. Altered proportions of Tregs in the PB and affected tissues of SSc patients and aberrant SSc-related gene expression have been reported, further supporting the relevance of this imbalance [5,10].

Additional scarce T cell subsets have also shown characteristic alterations in SSc [5]. For instance, the Th22 subset, known for its profibrotic role in tissue repair, wound healing, and autoimmunity, was overrepresented in SSc-affected skin tissues [11]. Moreover, a SSc-specific recirculating T helper follicular cell (Thf) subset, defined by high *CXCL13* expression (CXCL13+Thf), was identified in the skin tissue of SSc patients [12] suggesting that the study of scarce CD4+ T cell subsets in SSc might shed light on the etiopathogenesis of this IMID.

A key element in T cell biology is the T cell receptor (TCR), which is a highly polymorphic and antigen-specific membrane receptor required for T cell activation [13]. Lower diversity in the TCR repertoire of IMID patients is frequently observed [14], and it was recently described in SSc patients in comparison to healthy donors [15]. However, little is known about the TCR repertoire in specific CD4+ T cell subtypes and its association with SSc-related aberrant activation and behavior.

Considering the above evidence supporting an essential role of circulating CD4+ T cells in SSc, we performed the first comprehensive analysis of PB CD4+ T cells in SSc at the single-cell level to shed light into the cellular heterogeneity of these critical compartments for the disease, and their value as disease biomarkers. We characterized both the transcriptome and TCR repertoire of more than 80,000 CD4+ cells and established SSc-specific cellular composition and gene expression patterns.

# Materials and methods

## Patient description

This study included a cohort of 16 female participants, all with self-reported European ancestry. The average age was 60 years old for SSc patients and 59 for the unaffected controls. Among the participants, 8 were diagnosed with SSc according to the diagnostic criteria proposed by the ACR [16], and they were further classified, according to the criteria proposed by LeRoy, into limited cutaneous (lcSSc) or diffuse cutaneous SSc (dSSc) [17,18]. Participants were recruited at Hospital Universitario San Cecilio, Granada, Spain, with informed consent. Clinical information, including anonymized serological profiles and drug treatment of the patients, is shown in **Supplementary Table 1**.

## Cell suspension protocol

PB mononuclear cells (PBMCs) were separated as previously described [19]. Using a magnetic bead kit (Stem Cell Easy Step, ref #17858) CD4+ cells were negatively selected, following manufacturer's protocol. Flow cytometry and mRNA expression confirmed CD4+ cell purity (>99.6%, **Supplementary Figure 1**). Subsequently, CD4+ cell suspensions were cryopreserved as described elsewhere [19].

## Single cell RNA-sequencing and T cell receptor library generation

For single-cell whole transcriptome sequencing (scRNA-seq), the Next GEM technology provided by 10x Genomics was selected. Following the manufacturer's instructions, the samples were processed using the Chromium Next GEM Chip G Single Cell Kit (10xGenomics, PN-1000127 for gene expression and PN-1000005 for V(D)J profiling) 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 on the NovaSeq 6000 platform (Illumina) using S2 and SP chemistry v1.5. For the gene expression, these settings allowed an average 74.32% of reads assigned to cells, with an average of 4,372 UMIs per cell. For the TCR immune profiling, an average of 70.71% of reads in cells was observed. The

sequencing reads were aligned to the GRCh38 genome build, and the unique molecular identifiers (UMI) were processed using the 10x Cell Ranger Single Cell Software Suite (v3.0.0) with default parameters, resulting in an average of 1,357 genes identified per cell.

## Single cell RNA-sequencing data analysis

The outcomes from Cell Ranger were imported into *Scanpy* (v1.8.2) [20] within a Python environment (v3.8.1). All samples successfully met the established quality control (QC) filters. Finally, 81,714 CD4+ T cells and 21,583 genes passed all the QC filters.

A Batch Balanced k-Nearest Neighbour (BBkNN) integration graph was calculated using the first 20 principal components and clustering was performed employing the Leiden algorithm. The validity of the integration was supported by consistent clustering outcomes observed through alternative algorithms and the analysis of the UMAP representations of each individual after integration (**Supplementary Figure 2**).

Afterwards, unsupervised clustering by the Leiden algorithm returned 12 clusters, which were assigned to known CD4+ cell subtypes based on recent markers described in scRNA-seq studies in CD4+ cells [21] (**Supplementary Figure 3**). However, each cluster was analyzed as a distinct entity, independently of the previously assigned cell subtype.

We performed differential gene expression using the Wilcoxon statistical method. Additionally, Pathway enrichment analysis was calculated using the R package EnrichR (v3.1.0). Statistical significance was defined after applying the Benjamini-Hochberg False Discovery Rate correction method. We applied Partition-based Graph Abstraction to test the statistical strength of the connectivity between the clusters. To identify cell differentiation trajectories, we used the *Monocle3* algorithm [22] .

Finally, *Scirpy* (v1.0) was used for the analysis of TCR clonotypes and clonotype networks, clonal expansion, normalized Shannon entropy, TCR repertoire, preferential usage of V(D)J genes and epitope analysis.

Details of the scRNA-seq analyses are found in the **Additional Information** file.

# Results

## CD4+ T cells from SSc patients exhibit IFN-mediated late activation profiles

CD4+ T lymphocytes were extracted from the PB of 8 women, including 6 patients with lcSSc and 2 patients with dcSSc, alongside 8 undiagnosed women matched by age and ethnicity, as detailed in **Supplementary Table 1**.

Following QC, a final set of 81,637 CD4+ high quality scRNA-seq transcriptomes of individual cells from both SSc and unaffected individuals were selected for analysis. Although the number of cells was balanced between the patients with SSc and in the CTRLs (~4,100 per SSc patients and ~4,000 for CTRLs, **Supplementary Table 2**), the cell density on a UMAP plot was not uniform across conditions and lcSSc and dcSSc patients had clearly distinct patterns (**Figure 1A**). We hypothesize that these differences in cell density reflect variations in T cell activation between SSc patients and controls, as well as between the SSc subtypes.

DEG analysis between the SSc and control groups revealed 1,661 genes with increased expression in SSc and 2,298 genes with decreased expression in SSc (**Supplementary Table 3**). The top DEG in SSc were shared between both lcSSc and dcSSc (**Figure 1B**), but we also observed several subtype-specific markers (**Figure 1C**, **Supplementary Tables 4-6**). We observed marked profiles of late IFN-mediated CD4+ T cell activation, such as *IFITM2*, which was one of the most overexpressed genes among SSc patients (**Figure 1D**, **Supplementary Table 3**). S100 family genes were highly expressed in SSc (**Supplementary Table** 3). Interestingly, *TNFAIP3* stands out as the second most downregulated in SSc T cells, similar to Systemic Lupus Erythematosus (SLE) -related T cell dysfunction [23].

Regarding the comparison between the clinical subtypes, lcSSc and dcSSc, and controls, the results revealed no additional significant DEG (**Supplementary Tables 4 - 5**). Nevertheless, in the DE analysis between dcSSc and lcSSc, we observed a significant overexpression of *IFITM3* and *CCL5* in the lcSSc subset (**Supplementary Table 6**).

## The overexpression of *FCRL3* supports a reduced suppressive role of regulatory T cells in systemic sclerosis

We identified 12 distinct CD4+ T cell clusters and observed none of them was exclusively integrated by neither SSc nor CTRL cells (**Figure 2A**). Considering the different cluster marker genes (**Figure 2B** and **Supplementary Table 7**), each cluster was classified according to their responsiveness to activation, proliferative capacity and effector potential into [21]: naive T cells (Tn), central memory T cells (Tcm), effector memory T cells (Tem), re-expressing *CD45RA* effector memory T cells (Temra) (also known as CD4+ cytotoxic T Lymphocytes, CD4+ CTL), and regulatory T cells (Tregs).

As a result of this classification (**Figure 2C**), 5 clusters were identified as Tn, encompassing 57.3% of the total cells (**Supplementary Figure 3A-B**). These clusters exhibited the highest expression of *CCR7*, a hallmark Tn marker. Four clusters (30.2% of the cells) were characterized by the combined expression of *CREM*, *ITGB1*, and *KLRB1* among others and were classified as Tcm (**Supplementary Figure 3A-B**). Cluster 7 (6.1% of the cells) was identified as Tem, due to a high expression of *CCL5* and *GZMA* (**Supplementary Figure 3A-B**). Cluster 10 (5.1% of all cells) showed a high *FOXP3* expression, and it was identified as Tregs (**Supplementary Figure 3A-B**). Lastly, cluster 11 (1.2% of cells) was classified as Temra/CD4+ CTL, with elevated expression of *GZMH* and *GZMB*. We observed no statistically significant differences in the representation of the different T cell subtypes neither between cases and controls, nor between the different subtypes and controls (**Supplementary Figure 3C**). However, we found significant differences in three clusters, 3, 4 and 9 (**Figure 2C-D**).

Trajectory analysis and PAGA visualization supported the cell type annotation (**Figure 2E-F**). The PAGA analysis revealed a strong association among clusters annotated as Tn, as well as those annotated as Tcm, whereas single-cluster cell types appeared isolated in the diagram (**Figure 2E**). The diffusion map showed two main different branches originating from a common previous state (cluster 0). One branch was particularly differentiated, identified as Temra cells (cluster 11), which exhibited the highest pseudotime score, while the other branch contained a mixture of Tcm and Tem cells (**Figure 2F**).

As mentioned before, cluster 10 was identified as Treg cells. In this case, we observed that Tregs showed a high expression of a very relevant co-inhibitory molecule for Tregs, *TIGIT* (**Figure 2B**), which is known to be elevated in SSc Tregs [24]. *FCRL3*, a Treg marker, was significantly overexpressed in SSc Tregs (**Supplementary Figure 4**, **Supplementary Table 9**). The *FCRL3 locus* encodes the Fc Receptor Like 3 protein, an immunoglobulin receptor whose stimulation has been linked to the suppression of Treg activity [25]. All these findings suggest a defective suppressive role of Tregs in SSc patients that is consistent with the SSc-related immune imbalance.

## Th17 proinflammatory lymphocytes show a pro-inflammatory and glucocorticoid-resistant MDR1+ profile in systemic sclerosis

SSc patients showed higher proportions of cluster 3 cells (**Figure 2D**). This subset was classified as Th17 cells (**Supplementary Table 7** and **Supplementary Figure 4A-B**). Moreover, we found that the Th17 cluster had a strong proinflammatory profile based on the expression of high levels of *KLRB1/CD161* and *CCL20* (**Supplementary Table 7**). Furthermore, previously reported markers of pathogenic glucocorticoid-resistant Th17 activation were also increased in this subset, for example, *LTK*, *ABCB1* (also known as MDR1 or P-glycoprotein), *CTSH*, and *PTPN13* (**Supplementary Table 7**).

Additionally, the SSc Th17 lymphocytes showed the upregulation of immune activation genes such as *IFITM1*, *IFITM2* and *LTB* (**Supplementary Table 7**).

## Low levels of *AREG*+ naive T cells might contribute to a deficient regulatory T cell response in SSc patients

Cluster 4, a specific cluster of Tn cells, presented a reduced number of cells in SSc patients (**Figure 2C**). It is important to note that the most differentially overexpressed cluster marker gene for this cluster, in terms of log fold-change, was *AREG* (**Supplementary Table 8**). This gene encodes amphiregulin, a member of the epidermal growth factor (EGF) family of proteins. Moreover, this gene is

downregulated in SSc, and especially in the cells in this cluster. Interestingly, amphiregulin has been demonstrated to enhance the suppressive action of Tregs when binding to the EGFR receptor [26], which reinforces the findings that suggested a diminished Treg inhibition in patients.

## SSc patients show an increased and highly activated population of inflammatory T helper 2 cells

The Tcm compartment is known to play a relevant role in IMIDs [27], and the DE analysis between SSc and CTRLs in this subset revealed a consistent activation, motility, and inflammatory profile in SSc due to the increased expression of genes such as: *IFITM1*, *IFITM2* (**Supplementary Table 8**).

One of the clusters within this group, cluster 9, was found to be underrepresented in patients (**Figure 2C**). Interestingly, patient cells seemed to show a biased profile towards effector memory characteristics as they overexpressed *CCL5* and MHC class II genes, *i.e*. *HLA-DRB1* and *HLA-DPB1* (**Supplementary Table 9)**.

Therefore, we decided to further investigate the composition of this CD4+ T cell subtype. To accomplish this, clusters 3, 5, 6 and 9 (comprising a total of 24,708 cells) were selected and re-clusterized. As result, 9 sub-clusters with a homogenous distribution between SSc and unaffected controls were identified (**Figure 3A**). Subsequently, employing a panel of markers for Th subphenotypes, we classified these clusters as Th1, Th2, Th17, Th22, and follicular helper T cells (Thf) (**Figure 3B**).

The analysis of the cellular composition within the identified cell types unveiled significant differences between patients and controls. In SSc, there was an increase in Th2 cells and a decrease in the Th22 subset (**Figure 3C-D**). The Th2 cluster markers included classical markers such as *GATA3* and *IL4R*, along with *XBP1*, another Th2 marker [28]. Th2 cells also exhibited motility markers such as *VIM* and *ANXA1*, and activation markers like *SELL*, *CCR4*, and *CD48*/*SLAMF2*

(**Supplementary Table 10**). Therefore, we confirmed a highly activated SSc-specific expression profile in circulating Th2 cells.

## Tcm cell clones are differentially expanded in systemic sclerosis CD4+ T cells

To assess T-cell receptor usage, an immunological profiling of cells was conducted for both patients and controls. This profile aimed to examine differences in terms of clonal expansion (an increment in the abundance of a specific clonotype) and V(D)J gene usage.

There was no observable global difference in T cell expansion between SSc and controls in the clonotype expansion (**Figure 4A**). Considering the cell subtypes, the most differentiated subtypes i.e. Tcm, Tem and Temra subtypes showed an increased expansion (**Figure 4A**), but differences between SSc and controls were only significant in the Tcm subtype. The clonal expansion analysis by cluster revealed significant increases in clusters 3, 5, 7, 8, 9, and especially in cluster 11 (FDR < 0.05) (**Figure 4B**). However, differences in clonal expansion between SSc and controls only appeared in clusters 3 and 5 (FDR < 0.05). Nevertheless, Shannon diversity measures did not significantly differ between SSc and controls in any of the CD4+ T cell clusters or subtypes. The V(D)J analysis showed no bias in the usage of these genes between SSc patients and controls (**Figure 4C**), either between SSc subtypes. As for the epitope analysis, of 81,637 cells, only ~9,000 cells were matching successfully with an epitope in the VDJdb [29]. More than 20 different epitopes were identified for all the cells, but no relevant differences in the epitope recognition were observed between SSc patients and the non-affected CTRLs (**Figure 4D**).

# Discussion

While scRNA-seq studies in SSc have focused on affected tissues, circulating CD4+ cells in peripheral blood remained understudied [12,26,30,31]. In fact, our findings are complementary, though not always overlapping, with those reported in the study of CD4+ T cells in the skin of SSc patients. Gaydosik et al. identified a population of Tfh cells with high *CXCR3* expression in the skin, highlighting their potential role in tissue-specific immune responses in SSc [12]. In contrast, in our PB analysis, we did not detect Tfh cells as a distinct cluster, but we observed a reduction in the overall Tfh compartment in SSc patients compared to controls. These findings underscore the need for integrative studies that combine tissue and blood analyses to better understand the dynamics and functional roles of CD4+ T cells in SSc pathogenesis.

The Treg proportion in the PB of SSc patients remained controversial, with some studies suggesting an increased presence of these cells [32,33], while others indicate a decrease in the skin but not in the blood [34]. Our data matched Klein et al. with no difference in Treg counts, but reduced suppressive function as in previous reports. [33,34]. We found that SSc Tregs expressed high levels of *FCRL3* gene, which had been shown to reduce the suppressive capacity of these cells [25]. Argawal et al.[25] found that stimulating FCRL3 inhibited the ability of Tregs to suppress the proliferation of both CD4+ and CD8+ cells. Additionally, it promoted a Th17-like phenotype by increasing IL-17 production in these cells without decreasing the protein expression of FOXP3 [25]. Moreover, the activation of FCRL3 with specific FCRL3-Ab mimics the effect of suppressive molecules on the regulatory activity of Tregs, such as TLR2. Thus, this receptor appears to modulate Treg regulation, promoting a more proinflammatory state in these cells [25]. We consider that our results might be the basis for future studies to analyze the value of this molecule as a biomarker for SSc.

In addition to an abnormal intrinsic behavior of SSc Tregs, their function might be also impaired by alterations in other CD4+ subsets that interact with them. For example, we observed that Tn cells expressing amphiregulin (*AREG*), a molecule that enhances the immunosuppressive activity of Tregs [26,35] were less frequent in SSc patients. This molecule is implicated in multiple physiological pathways,

primarily in tissue repair and inflammation resolution [36], particularly in the type 2 inflammatory response [35]. In light of these findings, we hypothesize that decreased *AREG* levels in Tn cells may impair Treg function and promote immune dysregulation in SSc.

In our data, Th17 cells are included in the only cluster overrepresented in SSc patients. Cluster markers analysis showed high expression levels of *MDR1*, also known as Multidrug Resistance Protein 1 or P-glycoprotein. *MDR1*+Th17 cells seem to produce a mixed cytokine profile between Th1 and Th17 (IL-17A, IL-17F, IL-22 and IFN) and their resistance to glucocorticoids produces a selection of these cells under treatment as Ramesh et al. showed using *in vitro* studies [37]. In RA, elevated MDR1 expression has been observed in patients with highly active disease [38,39], while in CD the disruption of the homeostasis in the gene expression promotes disease progression [39,40]. Furthermore, we found additional pro-inflammatory markers in Th17 cells, such as *KLRB1*/*CD161* and *CCL20.* Reports on *CD161* showed that the main expression of this gene is produced in residing Tcm cells [41]. Once *CD161*+Th17 cells migrate to the tissue, they contribute to the inflammation process by stimulating the production of TNF and IFN-γ [41]. Their presence was also detected in Crohn's Disease mediating the destructive inflammation of the gut tissue [42] or in the synovial infiltrate in RA [43]. Therefore, we propose that this particular subset of Th17 cells should be considered in the context of SSc as they might play a key role in patient prognosis.

One limitation of our study is the imbalance in the number of patients with lcSSc and dcSSc, which warrants caution when interpreting comparisons between these subtypes. The smaller sample size of dcSSc patients may limit the generalizability of subtype-specific findings. Additionally, the varying treatments and comorbidities (i.e. SS), as detailed in **Supplementary Table 1**, are a potential source of heterogeneity. Despite these challenges, we did not observe that any patient or group of patients deviated significantly from the rest.

TCR analysis showed increased clonal expansion in Tcm from SSc, but no other significant differences. The lack of clear differences between SSc patients and controls may be attributed to the strong heterogeneity of the patients, the sample

size, or the limitations of the tools used. Nevertheless, further analyses are required to confirm the role of TCR diversity in SSc.

In conclusion, scRNA-seq analysis of CD4+ T cells in the PB of SSc patients revealed a disease-specific expression profile in Tregs. Additionally, we observed that proinflammatory Th17 cells in SSc expressed markers indicative of pathogenic behavior and drug resistance. Our findings underscore the value of single-cell transcriptome analysis of specific cell subtypes in PB to unravel the complexity of IMIDs such as SSc.

# Acknowledgements

We would like to thank Sofia Vargas and Gemma Robledo for their excellent technical assistance, and also to Sven Klages for his excellent bioinformatic assistance. 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). GVM was funded by the Grant PRE2019-087586 funded by MCIN/AEI/10.13039/501100011033 and by “ESF Investing in your future”. GBY’s contract is part of the grant PREP2022-000712, funded by the MCIN/AEI/10.13039/501100011033 and the ESF+. MAH is recipient of a Miguel Servet fellowship (CP21/00132) from Instituto de Salud Carlos III (ISCIII).


# Contributors

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

# Competing interests

GVM: none; GBY: none; MAH: none; JLCR: none; NOC: none; NM: none; SB: none; MGA: none; JM: none; LBC: 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 center. 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 (0371/2020)”, and the CSIC Ethical Committee (026/2020).

# Data availability

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

# 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_CD4_nb].

# Figures and figure legends

## Graphical abstract

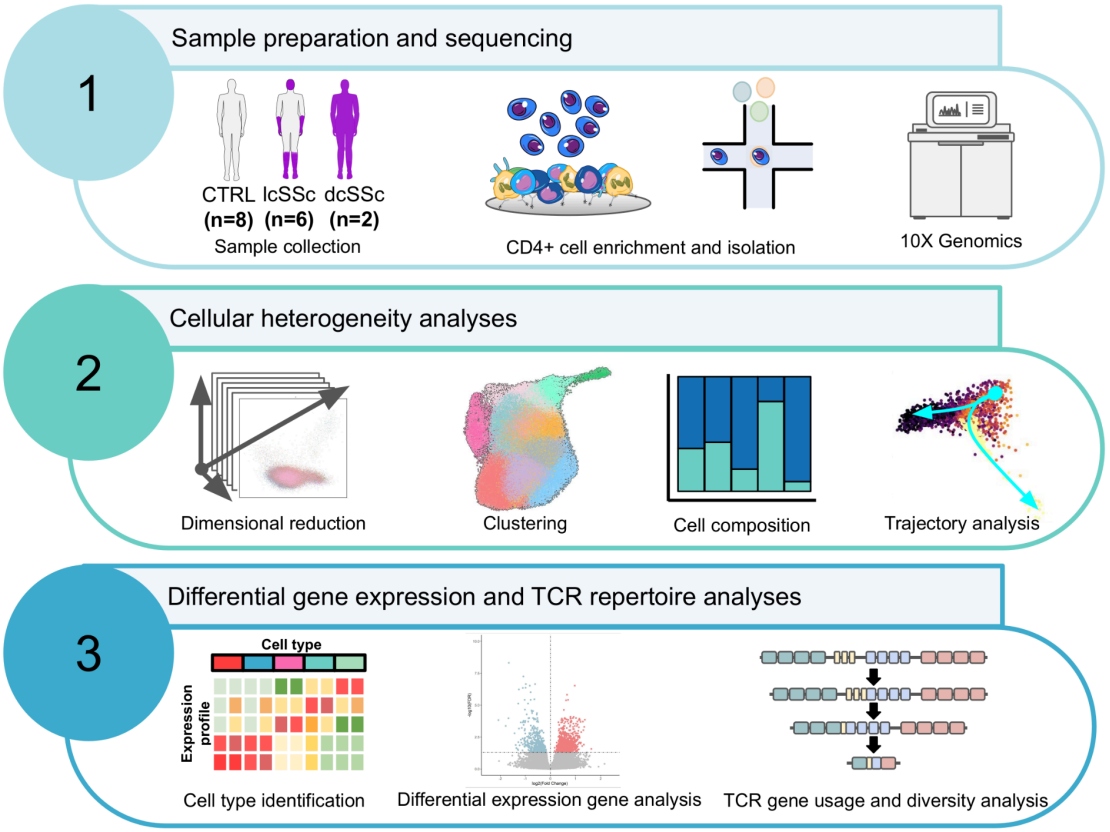


**Figure 1.** Single-cell analysis of CD4+ T cells in systemic sclerosis (SSc) and controls (CTRLs). **(a)** UMAP visualization of cell density by condition. The UMAP plot shows the distribution and density of CD4+ T cells across different conditions SSc, lcSSc (limited cutaneous SSc), dcSSc (diffuse cutaneous SSc), and controls. **(b)** Expression of the top 10 upregulated genes in SSc, lcSSc, and dcSSc versus controls. Dotplot displaying the expression levels of the top 10 upregulated genes in SSc, lcSSc, and dcSSc compared to controls. **(c)** Expression of the top 10 upregulated genes in dcSSc versus lcSSc and vice versa. **(d)** Expression of late activation markers in CD4+ T cells in SSc versus controls (as defined by Soskic et al., Nat Gen 2021).

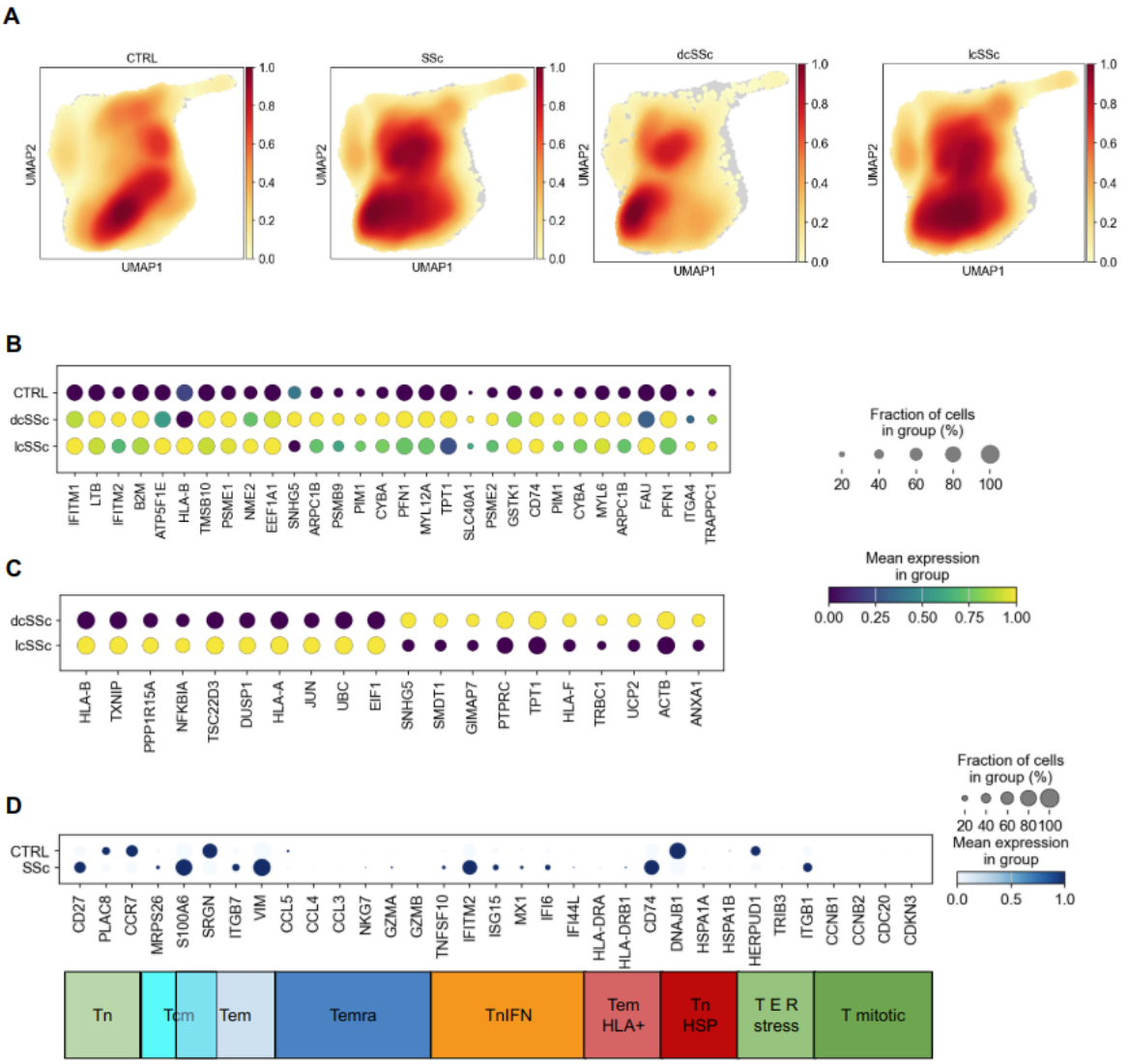


**Figure 2.** Clustering and characterization of CD4+ T cell populations in systemic sclerosis (SSc). **(a)** UMAP visualization of clusters and composition analysis. UMAP plot showing 12 clusters (labeled from 0 to 11 by size) identified in the CD4+ T cell population. The composition analysis illustrates the contribution of each individual to each cluster. The top three marker genes for each cluster are indicated. **(b)** Dot plot of marker genes for each cluster. Dot plot displaying the expression of marker genes for the 12 clusters. Each dot represents the percentage of cells expressing the gene, and the color indicates the average expression level. **(c)** UMAP visualization and composition analysis by cell type annotation. UMAP plot and composition analysis showing the distribution of cell types: Tn (naive T cells), Tcm (central memory T cells), Tem (effector memory T cells), Temra (effector memory T cells re-expressing CD45+), and Tregs (regulatory T cells). **(d)** Box plots of cluster composition differences between SSc patients and controls. Box plots highlighting the composition differences in three clusters that show significant variation between SSc patients and controls. *: $P_{FDR} <= 0.05$; **: $P_{FDR} <= 0.001$ **(e)** PAGA connectivity map. PAGA (Partition-based Graph Abstraction) map where each node represents a cluster, and the thickness of the lines indicates the strength of connections between clusters. **(f)** Diffusion map colored by pseudotime, cell type annotation, and cluster. Diffusion map visualization colored by (left to right): pseudotime, cell type annotation, and cluster identity, providing insights into the developmental trajectories and relationships between cell types and clusters.

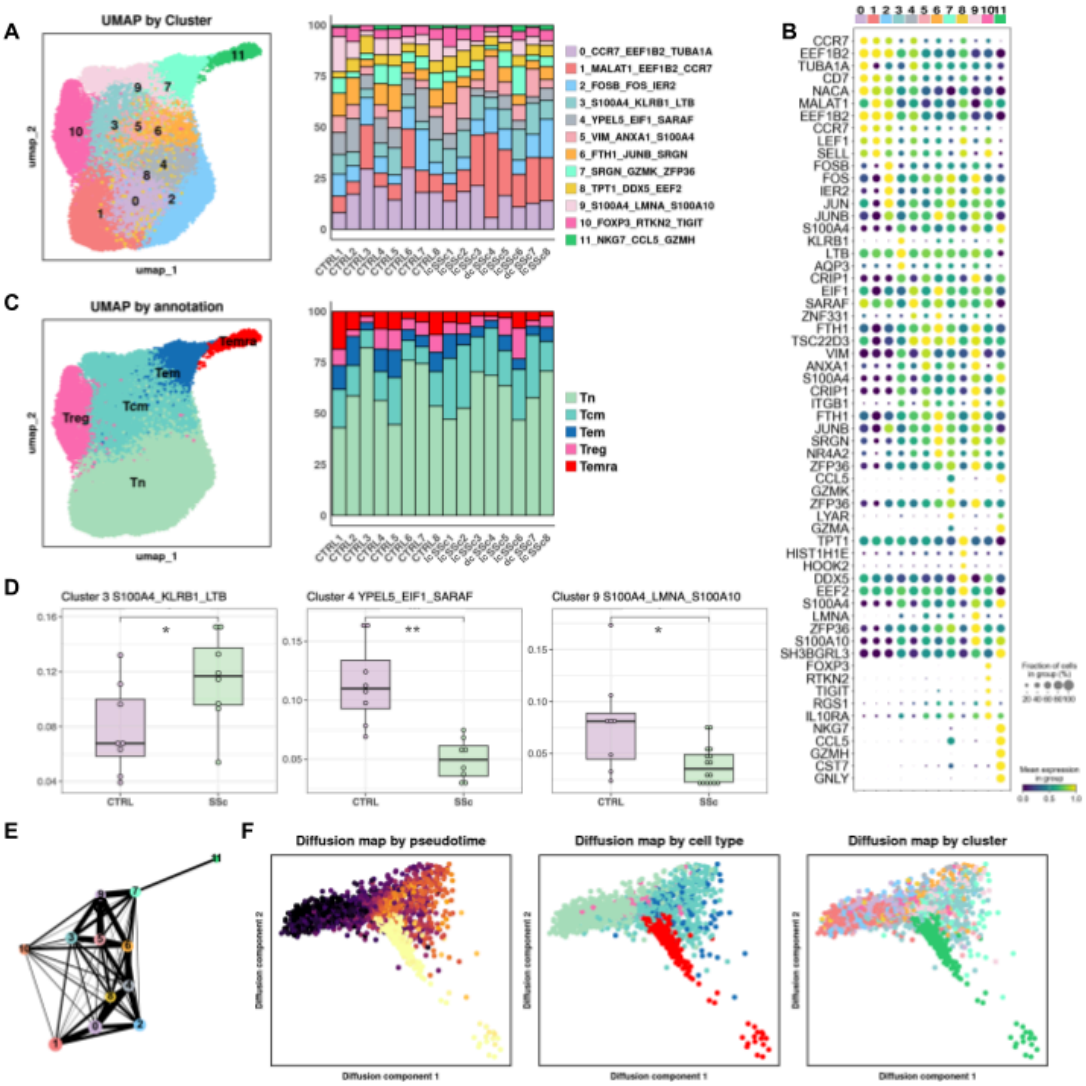

A
UMAP by Cluster
umap_1
umap_2
0_CCR7_EEF1B2_TUBA1A
1_MALAT1_EEF1B2_CCR7
2_FOSB_FOS_IER2
3_S100A4_KLRB1_LTB
4_YPEL5_EIF1_SARAF
5_VIM_ANXA1_S100A4
6_FTH1_JUNB_SRGN
7_SRGN_GZMK_ZFP36
8_TPT1_DDX5_EEF2
9_S100A4_LMNA_S100A10
10_FOXP3_RTKN2_TIGIT
11_NKG7_CCL5_GZMH
B
C
UMAP by annotation
Tn
Tcm
Tem
Treg
Temra
D
Cluster 3 S100A4_KLRB1_LTB
Cluster 4 YPEL5_EIF1_SARAF
Cluster 9 S100A4_LMNA_S100A10
CTRL
SSc
E
F
Diffusion map by pseudotime
Diffusion map by cell type
Diffusion map by cluster
Diffusion component 1
Diffusion component 2

**Figure 3.** Subclustering and characterization of Tcm (central memory T cell) populations in systemic sclerosis (SSc). **(a)** UMAP visualization of Tcm subclustering. Top UMAP: Colored by unsupervised clusters identified through subclustering of Tcm cells. Bottom UMAP: Colored by cell type annotation based on T helper gene markers. **(b)** T helper marker gene panel. Dot plot displaying the expression of T helper marker genes used to annotate the different clusters. The markers are based on previous studies and include genes specific to Th1, Th2, Th17, Tfh (follicular helper T cells), and other T helper subsets. **(c)** Cell type contributions stratified by condition (CTRL or SSc). Stacked bar plot showing the contribution of each cell type (annotated in panel B) to the Tcm population, stratified by condition (CTRL or SSc) and color-coded by cell type. **(d)** Contributions of CTRLs and SSc patients stratified by clinical subtype (dcSSc and lcSSc). Stacked bar plot illustrating the contribution of different Tcm cell subtypes in CTRLs, dcSSc (diffuse cutaneous SSc), and lcSSc (limited cutaneous SSc) patients.

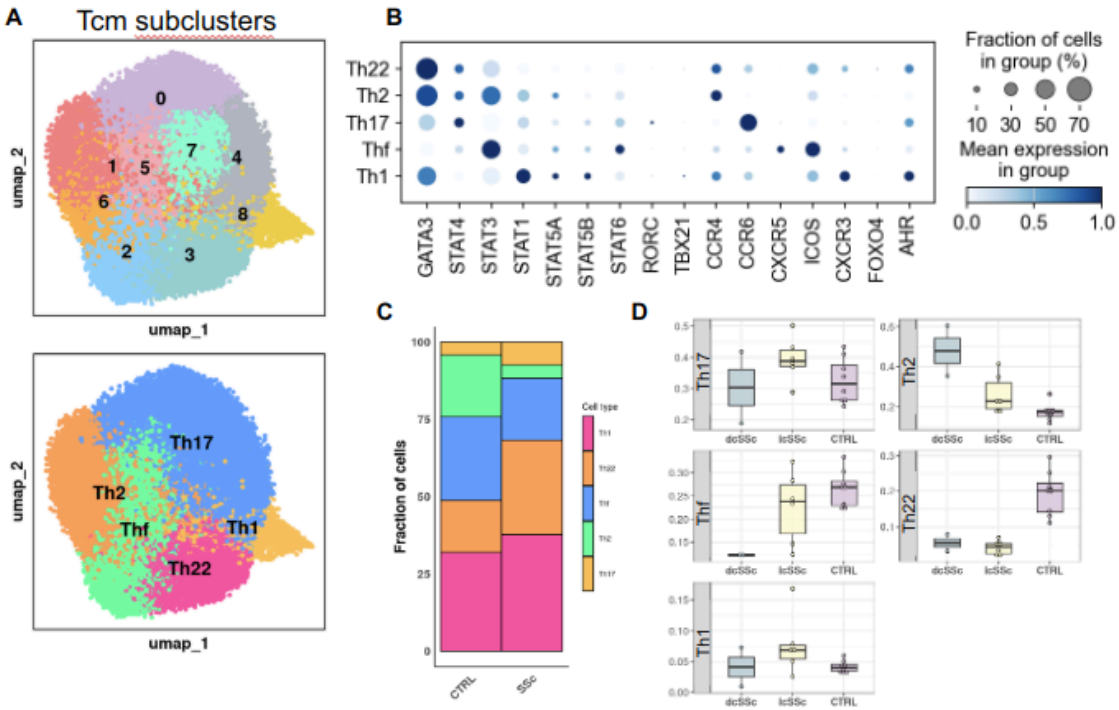
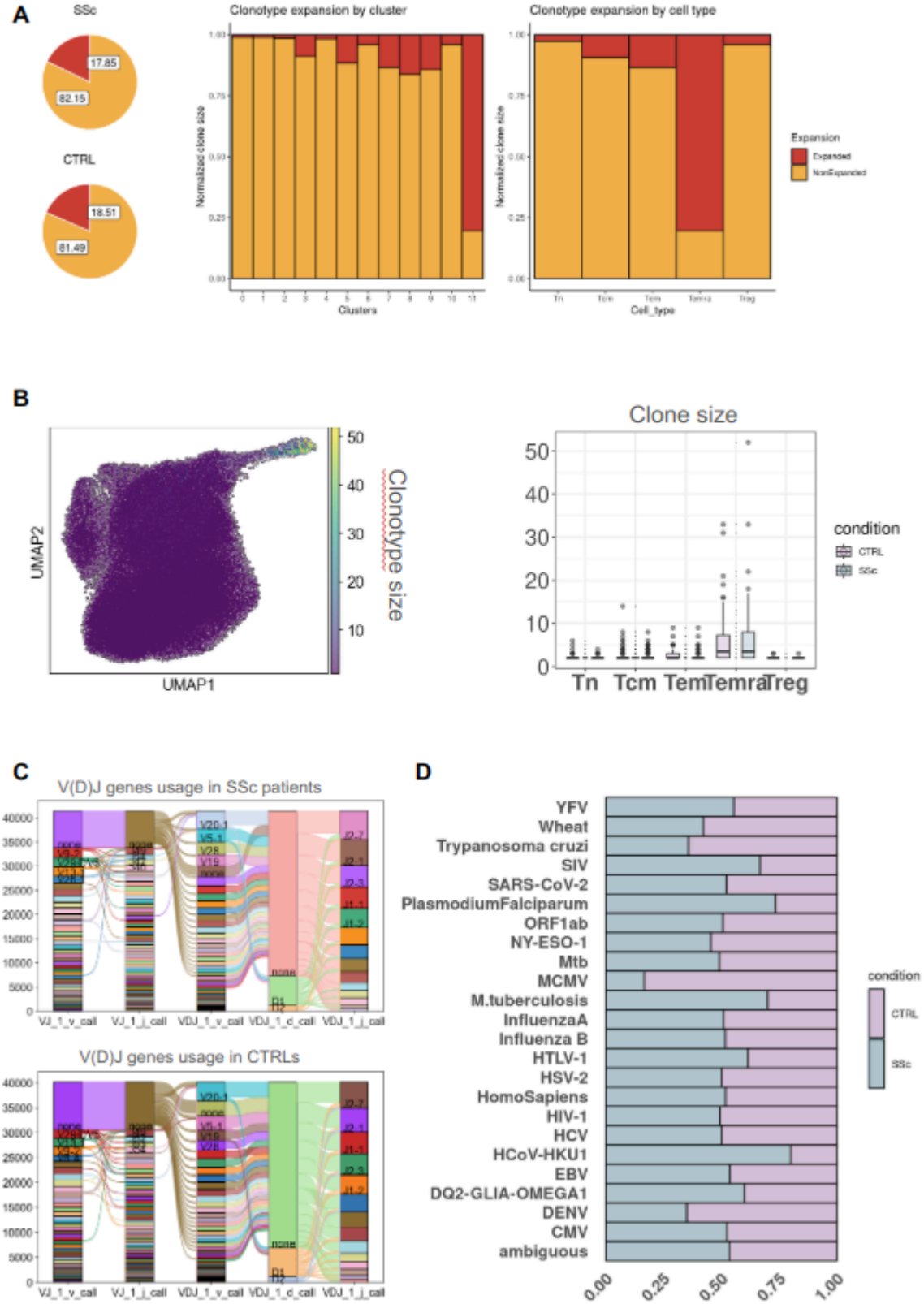


**Figure 4.** Clonotype expansion and TCR repertoire analysis in CD4+ T cells. **(a)** Clonotype expansion analysis. Left: Visualization of clonotype expansion across all cells for SSc and CTRLs. Middle: Comparison of clonotype expansion between SSc and CTRLs by cluster. Right: Comparison of clonotype expansion between SSc and CTRLs by cell type. **(b)** UMAP visualization and clonotype size distribution. Left: UMAP plot colored by clonotype size, showing the distribution of expanded clonotypes across the CD4+ T cell population. Right: Box plot displaying the distribution of clonotype sizes across different cell types. **(c)** VDJ gene usage analysis. Ribbon plots illustrating the usage of VDJ genes in the TCR repertoire: Top panel: VDJ gene usage in SSc patients. Bottom panel: VDJ gene usage in CTRLs. **(d)** Epitope analysis. Visualization of epitopes identified within the single-cell subset, highlighting potential antigen-specific responses in SSc and CTRLs.


A
SSc
17.85
82.15
CTRL
18.51
81.49
Clonotype expansion by cluster
Normalized clone size
Clusters
Clonotype expansion by cell type
Cell_type
Expansion
Expanded
NonExpanded
B
UMAP2
UMAP1
Clonotype size
Clone size
condition
CTRL
SSc
Tn
Tcm
Tem
Temra
Treg
C
V(D)J genes usage in SSc patients
V(D)J genes usage in CTRLs
VJ_1_v_call
VJ_1_j_call
VDJ_1_v_call
VDJ_1_d_call
VDJ_1_j_call
D
YFV
Wheat
Trypanosoma cruzi
SIV
SARS-CoV-2
PlasmodiumFalciparum
ORF1ab
NY-ESO-1
Mtb
MCMV
M.tuberculosis
InfluenzaA
Influenza B
HTLV-1
HSV-2
HomoSapiens
HIV-1
HCV
HCoV-HKU1
EBV
DQ2-GLIA-OMEGA1
DENV
CMV
ambiguous
condition
CTRL
SSc

# References


1. Volkmann ER, Andréasson K, Smith V. Systemic sclerosis. Lancet 2023;401:304–18.

2. Thoreau B, Chaigne B, Mouthon L. Role of B-Cell in the Pathogenesis of Systemic Sclerosis. Front. Immunol. 2022;13:933468.

3. 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;28:31–6.

4. López-Isac E, European Scleroderma Group†, Acosta-Herrera M, Kerick M, Assassi S, Satpathy AT, et al. GWAS for systemic sclerosis identifies multiple risk loci and highlights fibrotic and vasculopathy pathways. Nature Communications2019;10.

5. Jin W, Zheng Y, Zhu P. T cell abnormalities in systemic sclerosis. Autoimmun. Rev. 2022;21:103185.

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

7. Wei L, Abraham D, Ong V. The Yin and Yang of IL-17 in Systemic Sclerosis. Front. Immunol. 2022;13:885609.

8. Ito T, Tamura N, Okuda S, Tada K, Matsushita M, Yamaji K, et al. Elevated serum levels of soluble CD146 in patients with systemic sclerosis. Clin. Rheumatol. 2017;36:119–24.

9. Guo RH, Cheng H, Zhang XY, Yu Z, Wang GH, Hao SY, et al. Changes in peripheral T-lymphocyte subsets and serum cytokines in patients with systemic sclerosis. Front. Pharmacol. 2022;13:986199.

10. Kobayashi S, Nagafuchi Y, Shoda H, Fujio K. The Pathophysiological Roles of Regulatory T Cells in the Early Phase of Systemic Sclerosis. Front. Immunol. 2022;13:900638.

11. Truchetet ME, Brembilla NC, Montanari E, Allanore Y, Chizzolini C. Increased frequency of circulating Th22 in addition to Th17 and Th2 lymphocytes in systemic sclerosis: association with interstitial lung disease. Arthritis Res. Ther. 2011;13:R166.

12. 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;80:1453–60.

13. Irac SE, Soon MSF, Borcherding N, Tuong ZK. Single-cell immune repertoire analysis. Nat. Methods 2024;21:777–92.

14. Shen Y, Voigt A, Leng X, Rodriguez AA, Nguyen CQ. A current and future perspective on T cell receptor repertoire profiling. Front. Genet.

2023;14:1159109.

15. Servaas NH, Zaaraoui-Boutahar F, Wichers CGK, Ottria A, Chouri E, Affandi AJ, et al. Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in systemic sclerosis. J. Autoimmun. 2021;117:102574.

16. 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. 1980;23:581–90.

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

18. LeRoy EC, Black C, Fleischmajer R, Jablonska S, Krieg T, Medsger TA Jr, et al. Scleroderma (systemic sclerosis): classification, subsets and pathogenesis. J. Rheumatol. 1988;15:202–5.

19. 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;140:103097.

20. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19:15.

21. Soskic B, Cano-Gamez E, Smyth DJ, Ambridge K, Ke Z, Matte JC, et al. Immune disease risk variants regulate gene expression dynamics during CD4 T cell activation. Nat. Genet. 2022;54:817–26.

22. Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature 2019;566:496–502.

23. Zhao H, Wang L, Luo H, Li QZ, Zuo X. TNFAIP3 downregulation mediated by histone modification contributes to T-cell dysfunction in systemic lupus erythematosus. Rheumatology 2017;56:835–43.

24. Fleury M, Belkina AC, Proctor EA, Zammitti C, Simms RW, Lauffenburger DA, et al. Increased Expression and Modulated Regulatory Activity of Coinhibitory Receptors PD-1, TIGIT, and TIM-3 in Lymphocytes From Patients With Systemic Sclerosis. Arthritis Rheumatol 2018;70:566–77.

25. Agarwal S, Kraus Z, Dement-Brown J, Alabi O, Starost K, Tolnay M. Human Fc Receptor-like 3 Inhibits Regulatory T Cell Function and Binds Secretory IgA. Cell Rep. 2020;30:1292–9.e3.

26. Zaiss DMW, van Loosdregt J, Gorlani A, Bekker CPJ, Gröne A, Sibilia M, et al. Amphiregulin enhances regulatory T cell-suppressive function via the epidermal growth factor receptor. Immunity 2013;38:275–84.

27. Jaigirdar SA, MacLeod MKL. Development and Function of Protective and Pathologic Memory CD4 T Cells. Front. Immunol. 2015;6:456.

28. Pramanik J, Chen X, Kar G, Henriksson J, Gomes T, Park JE, et al. Genome-wide analyses reveal the IRE1a-XBP1 pathway promotes T helper cell differentiation by resolving secretory stress and accelerating proliferation. Genome Med. 2018;10:76.

29. Bagaev DV, Vroomans RMA, Samir J, Stervbo U, Rius C, Dolton G, et al. VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium. Nucleic Acids Res. 2020;48:D1057–62.

30. Gur C, Wang SY, Sheban F, Zada M, Li B, Kharouf F, et al. LGR5 expressing skin fibroblasts define a major cellular hub perturbed in scleroderma. Cell 2022;185:1373–88.e20.

31. Valenzi E, Tabib T, Papazoglou A, Sembrat J, Trejo Bittar HE, Rojas M, et al. Disparate Interferon Signaling and Shared Aberrant Basaloid Cells in Single-Cell Profiling of Idiopathic Pulmonary Fibrosis and Systemic Sclerosis-Associated Interstitial Lung Disease. Front. Immunol. 2021;12:595811.

32. Antiga E, Quaglino P, Bellandi S, Volpi W, Del Bianco E, Comessatti A, et al. Regulatory T cells in the skin lesions and blood of patients with systemic sclerosis and morphoea. Br. J. Dermatol. 2010;162:1056–63.

33. Radstake TRDJ, van Bon L, Broen J, Wenink M, Santegoets K, Deng Y, et al. Increased frequency and compromised function of T regulatory cells in systemic sclerosis (SSc) is related to a diminished CD69 and TGFbeta expression. PLoS One 2009;4:e5981.

34. Klein S, Kretz CC, Ruland V, Stumpf C, Haust M, Hartschuh W, et al. Reduction of regulatory T cells in skin lesions but not in peripheral blood of patients with systemic scleroderma. Ann. Rheum. Dis. 2011;70:1475–81.

35. Zaiss DMW, Gause WC, Osborne LC, Artis D. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. Immunity 2015;42:216–26.

36. Arruda LCM, Clave E, Douay C, Lima-Júnior JR, Slavov SN, Malmegrim KCR, et al. CMV-specific clones may lead to reduced TCR diversity and relapse in systemic sclerosis patients treated with AHSCT. Rheumatology (Oxford) 2020;59:e38–40.

37. Ramesh R, Kozhaya L, McKevitt K, Djuretic IM, Carlson TJ, Quintero MA, et al. Pro-inflammatory human Th17 cells selectively express P-glycoprotein and are refractory to glucocorticoids. J. Exp. Med. 2014;211:89–104.

38. Cao W, Kayama H, Chen ML, Delmas A, Sun A, Kim SY, et al. The Xenobiotic Transporter Mdr1 Enforces T Cell Homeostasis in the Presence of Intestinal Bile Acids. Immunity 2017;47:1182–96.e10.

39. Bossennec M, Di Roio A, Caux C, Ménétrier-Caux C. MDR1 in immunity: friend

or foe? Oncoimmunology 2018;7:e1499388.

40. Brinar M, Cukovic-Cavka S, Bozina N, Ravic KG, Markos P, Ladic A, et al. MDR1 polymorphisms are associated with inflammatory bowel disease in a cohort of Croatian IBD patients. BMC Gastroenterol. 2013;13:57.

41. Truong KL, Schlickeiser S, Vogt K, Boës D, Stanko K, Appelt C, et al. Killer-like receptors and GPR56 progressive expression defines cytokine production of human CD4 memory T cells. Nat. Commun. 2019;10:2263.

42. Kleinschek MA, Boniface K, Sadekova S, Grein J, Murphy EE, Turner SP, et al. Circulating and gut-resident human Th17 cells express CD161 and promote intestinal inflammation. J. Exp. Med. 2009;206:525–34.

43. Nistala K, Adams S, Cambrook H, Ursu S, Olivito B, de Jager W, et al. Th17 plasticity in human autoimmune arthritis is driven by the inflammatory environment. Proc. Natl. Acad. Sci. U. S. A. 2010;107:14751–6.