Journal of Autoimmunity 117 (2021) 102574

Contents lists available at ScienceDirect

Journal of Autoimmunity

journal homepage: www.elsevier.com/locate/jautimm

Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven $CD4^+$ and $CD8^+$ T-cell clusters in systemic sclerosis

N.H. Servaas [a,b], F. Zaaraoui-Boutahar [c], C.G.K. Wichers [a,b], A. Ottria [a,b], E. Chouri [a,b], A. J. Affandi [a,b], S. Silva-Cardoso [a,b], M. van der Kroef [a,b], T. Carvalheiro [a,b], F. van Wijk [b], T.R.D. J. Radstake [a,b], A.C. Andeweg [c], A. Pandit [a,b,*]

[a] *Department of Rheumatology & Clinical Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands*
[b] *Center for Translational Immunology, University Medical Center Utrecht, Utrecht University, Utrecht, the Netherlands*
[c] *Department of Viroscience, Erasmus Medical Center Rotterdam, Rotterdam, the Netherlands*

ARTICLE INFO

Keywords:
T-cell receptor
Repertoire sequencing
Systemic sclerosis
Autoimmunity
Longitudinal immunosequencing

ABSTRACT

The T-cell receptor (TCR) is a highly polymorphic surface receptor that allows T-cells to recognize antigenic peptides presented on the major histocompatibility complex (MHC). Changes in the TCR repertoire have been observed in several autoimmune conditions, and these changes are suggested to predispose autoimmunity. Multiple lines of evidence have implied an important role for T-cells in the pathogenesis of Systemic Sclerosis (SSc), a complex autoimmune disease. One of the major questions regarding the roles of T-cells is whether expansion and activation of T-cells observed in the diseases pathogenesis is antigen driven.

To investigate the temporal TCR repertoire dynamics in SSc, we performed high-throughput sequencing of $CD4^+$ and $CD8^+$ TCRβ chains on longitudinal samples obtained from four SSc patients collected over a minimum of two years. Repertoire overlap analysis revealed that samples taken from the same individual over time shared a high number of TCRβ sequences, indicating a clear temporal persistence of the TCRβ repertoire in $CD4^+$ as well as $CD8^+$ T-cells. Moreover, the TCRβs that were found with a high frequency at one time point were also found with a high frequency at the other time points (even after almost four years), showing that frequencies of dominant TCRβs are largely consistent over time. We also show that TCRβ generation probability and observed TCR frequency are not related in SSc samples, showing that clonal expansion and persistence of TCRβs is caused by antigenic selection rather than convergent recombination. Moreover, we demonstrate that TCRβ diversity is lower in $CD4^+$ and $CD8^+$ T-cells from SSc patients compared with memory T-cells from healthy individuals, as SSc TCRβ repertoires are largely dominated by clonally expanded persistent TCRβ sequences. Lastly, using "Grouping of Lymphocyte Interactions by Paratope Hotspots" (GLIPH2), we identify clusters of TCRβ sequences with homologous sequences that potentially recognize the same antigens and contain TCRβs that are persist in SSc patients.

In conclusion, our results show that $CD4^+$ and $CD8^+$ T-cells are highly persistent in SSc patients over time, and this persistence is likely a result from antigenic selection. Moreover, persistent TCRs form high similarity clusters with other (non-)persistent sequences that potentially recognize the same epitopes. These data provide evidence for an antigen driven expansion of $CD4^+$/$CD8^+$ T-cells in SSc.

1. Introduction

Systemic Sclerosis (SSc) is a complex chronic autoimmune disease, characterized by vascular abnormalities and widespread fibrosis affecting the skin and internal organs [1]. Although the pathogenic mechanisms underlying SSc remain largely unknown, multiple lines of evidence imply an important role for $CD4^+$ and $CD8^+$ T-cells in the progression of the disease. Activated T-cells infiltrate the skin of SSc patients already in the early phase of the disease [2,3]. These infiltrating T-cells can cross-talk with fibroblasts, inducing fibroblast activation and apoptosis through secretion of pro-inflammatory cytokines and fas/fas ligand engagement [4,5]. Moreover, T-cells isolated from SSc patients

* Corresponding author. University Medical Center Utrecht, KC.02.046.0, Lundlaan 6, 3583EA, Utrecht, the Netherlands.
E-mail address: a.pandit@umcutrecht.nl (A. Pandit).

https://doi.org/10.1016/j.jaut.2020.102574
Received 8 September 2020; Received in revised form 10 November 2020; Accepted 13 November 2020
Available online 8 December 2020
0896-8411/© 2020 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

undergo clonal expansion when cultured together with autologous fibroblasts, suggesting that auto-antigens presented by SSc fibroblasts can induce auto-reactive T-cell responses [6]. Apart from skin, peripheral blood T-cells from SSc patients also exhibit signs of activation and express activation markers, including IL-2R, HLA-DR, and CD29 [7–9], and secrete pro-inflammatory and pro-fibrotic factors [10–12].

The T-cell receptor (TCR) is a highly polymorphic surface receptor that allows T-cells to recognize antigenic peptides presented on the major histocompatibility complex (MHC) [13]. $CD4^{+}$ T-cells recognize peptides presented on the MHC class II complex, while $CD8^{+}$ T-cells recognize peptides presented on the MHC class I complex. Classical TCRs are heterodimers consisting of a paired α- and β-chain. These chains making up the TCR are generated through somatic recombination of V (variable), D (diversity) and J (joining) gene segments accompanied by pseudorandom insertions and deletions of nucleic acids at their joining regions [14], thereby giving rise to an enormously diverse TCR repertoire in every individual. By this process of VDJ recombination, the small set of genes that encode the TCR can be used to create over 10^{15} potential TCR clonotypes [13,15]. Previous estimates of number of unique T-cells in a human range from 10^{6} to 10^{11} [16–18], meaning that every individual only carries a small fraction of the potential repertoire.

High throughput sequencing of TCR repertoires is emerging as a valuable tool to unravel the exact role of T-cells in autoimmune diseases. The TCR repertoire has been proposed to serve as diagnostic biomarker for various autoimmune diseases, and recent studies have identified disease-associated TCR sequences in autoimmune diseases including autoimmune encephalomyelitis (AE), systemic lupus erythematosus (SLE), and rheumatoid arthritis (RA) [19–21]. Moreover, changes in T-cell repertoire diversity have been suggested to predispose the pathological manifestations in RA patients [22]. Prior studies examining the TCR repertoire in SSc have shown that there is an oligoclonal expansion of T-cells in the skin, lungs and blood of SSc patients [23–25], suggesting that expanded T-cells are involved in the disease pathogenesis. However, there are limitations to the results of previous studies that have examined the TCR repertoire in SSc patients. These include: a) lack of the use of high-throughput techniques; b) consideration of either $CD4^{+}$ or $CD8^{+}$ or unsorted T-cell populations; and c) the study of T-cells obtained only from a single time point thereby providing a static snapshot of the TCR repertoire in SSc.

Two major hypotheses have been postulated to explain mechanism of the expansion of T-cells in the context of autoimmunity [26,27]. The first hypothesis states that T-cells might expand non-specifically or by chance (bystander activation) due to chronic inflammation observed in autoimmune patients [27–29]. In this case, proliferation of T-cells is induced through non-specific activation in the presence of TLR ligands and cytokines during an immune response. Due to inherent biases in the V(D)J recombination process, some TCRβ sequences are more prevalent as they have a high generation probability [30]. As a result of this bias, during bystander activation, naïve T-cell clones with TCRβs that have high generation probabilities have a larger chance of being at the site of action due to their increased prevalence, and therefore have a higher chance to expand. In this case, expansion is a result of chance. The second hypothesis states that clonal expansion in autoimmunity is driven by chronic, and specific responses to antigens that selectively skew the TCR repertoire [26]. Here expansion is driven by antigen specific selection rather than chance. It remains to be unraveled which of these two mechanisms contributes to the activation and expansion of autoreactive T-cells in SSc.

To better understand T-cell responses in SSc pathogenesis, here we investigate the temporal TCR repertoire dynamics in SSc patients. We performed high-throughput sequencing of TCRβ chains of sorted $CD4^{+}$ and $CD8^{+}$ non-naive T-cells isolated from longitudinal samples from four SSc patients collected over a minimum of two years.

2. Materials and methods

2.1. Sample collection

Whole heparinized blood samples from SSc patients were obtained from the University Medical Center Utrecht. SSc patients were classified according to the ACR/EULAR criteria [31]. This study was conducted in accordance with the Declaration of Helsinki and was performed with approval of the Institutional Review Board of the University Medical Centre Utrecht, The Netherlands. The medical ethics committee of the UMC Utrecht approved the study (METC nr. 13–697). All participants enrolled in the study signed informed consent, and patient samples were anonymized upon collection. Peripheral blood mononuclear cells (PBMCs) were isolated by density gradient centrifugation on Ficoll-PlaqueTM Plus (GE Healthcare, Uppsala, Sweden). pDCs, mDCs, B-cells and monocytes were first depleted by magnetic bead sorting using the autoMACs Pro Separator (Miltenyi Biotec, Bergisch Gladbach, Germany) according to the manufacturer's protocol. The remaining peripheral blood lymphocytes (PBLs) were resuspended in freezing medium (80% FCS, (Sigma-Aldrich, Saint Louis, Missouri, USA)/20% DMSO, (Sigma-Aldrich)) and stored in sterile cryovials in liquid nitrogen (−196 °C) until further use. From all patients PBLs were collected at baseline (T0), at least one year after inclusion (T1, ranging from 12 to 19 months) and at least two years after inclusion (T2, ranging from 24 to 46 months) (Fig. 1a). High-resolution HLA typing was performed on DNA obtained from PBMCs from one time point for each patient by next generation sequencing.

2.2. T-cell sorting

PBLs were thawed in RPMI 1640 (Gibco, Thermo Fisher Scientific, Waltham, Massachusetts, USA) supplemented with 20% FCS (Sigma-Aldrich), and washed with PBS. Subsequently, the cells were resuspended in FACS buffer (PBS supplemented with 1% BSA (Sigma-Aldrich) and 0.1% sodium azide (Sigma-Aldrich)) and stained using the following antibodies: CD3-AF700 (clone UCHT1, Biolegend, San Diego, California, USA), CD4-BV711 (clone OKT4, Sony Biotechnology, San Jose, California, USA), CD8-V500 (clone RPA-T8, BD Bioscience, San Jose, California, USA), CD56-PE-CF594 (clone B159, BD Bioscience), CD16-BV785 (clone 3G8, Sony Biotechnology), CD14-PerCP-Cy5.5 (clone HCD14, Biolegend), HLADR-BV421 (clone L243, Biolegend), CD45RO-PE-Cy7 (clone L243, BD Bioscience), CD27-APC-eFluor780 (clone O323, eBioscience, San Diego, California, USA). Multiparametric flow cytometry sorting of non-naive $CD4^{+}$ T-cells ($CD3^{+}CD4^{+}CD45RO^{+/-}CD27^{+/-}$) and non-naive $CD8^{+}$ T-cells ($CD3^{+}CD8^{+}CD45RO^{+/-}CD27^{+/-}$) was performed on the BD FACSAria II (BD Bioscience), according to the gating strategy described in Fig. 1b. After sorting, cells were washed in PBS, lysed in TRIzol™ Reagent (Invitrogen, Carlsbad, California, USA), and stored at −20 °C until RNA isolation.

2.3. RNA isolation and TCR sequencing

RNA was isolated using the RNeasy Mini Kit (Qiagen, Venlo, Netherlands) by adding ethanol to the upper aqueous phase of processed TRIzol samples and transferring directly to the RNeasy spin columns. TCR amplification was performed according to the protocol published by Mamedov *et al.* [32]. Primer and barcode sequences are provided in Supplementary Table 1. Briefly, cDNA was generated by RACE using a primer directed to the TCRβ constant region. Thirteen nucleotide long unique molecular identifiers (UMIs) were incorporated during cDNA synthesis. Subsequently, two-stage semi-nested PCR amplification was performed including a size selection/agarose gel purification step after the first PCR. To minimize cross-sample contamination, 5-nucleotide sample specific barcodes were introduced at two steps during the library preparation process [32]. Resulting TCR amplicons were subjected to high-throughput sequencing using the Ovation Low Complexity

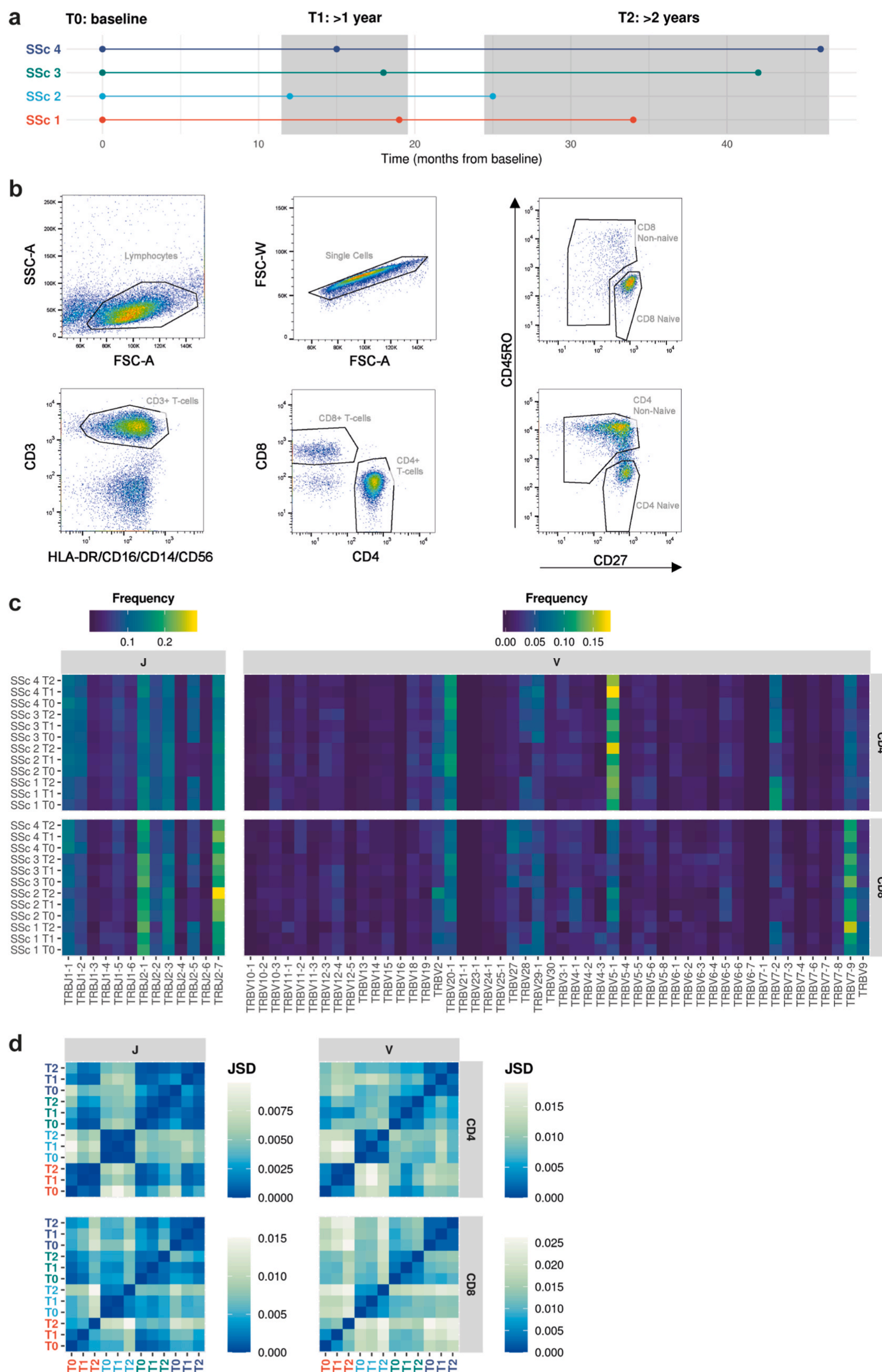

Fig. 1. High-throughput TCRβ sequencing of SSc patients over time shows temporal persistence of Vβ and Jβ gene segment usage. **a)** Experimental design of longitudinal sampling. **b)** Gating strategy for FACS sorting of $CD4^+$ non-naïve and $CD8^+$ non-naïve T-cells for PBL samples. **c)** Heatmap showing the frequency of Vβ and Jβ gene usage across different samples. **d)** Heatmap of Jensen-Shannon divergence of Vβ and Jβ gene usage between different samples. Lower divergence indicates similar Vβ/Jβ gene usage.

Sequencing System kit (NuGEN, San Carlos, California, USA) according to the manufacturer's instructions, and the Illumina MiSeq system (Illumina, San Diego, California, USA), using indexed paired-end 300 cycle runs.

2.4. TCR repertoire analysis

Raw paired-end reads were assembled using Paired-End reAd mergeR (PEAR) [33]. Sample specific barcode correction was performed using the 'umi_group_ec' command from the Recover T Cell Receptor (RTCR) pipeline [34], allowing zero mismatches in the barcode seed sequence for UMI detection (sample specific barcodes are provided in Supplementary Table 1). This strict barcode selection resulted in about 50% loss of reads, but ensured that there was minimal cross-sample contamination. Subsequently, barcode sequence reads having the same UMI were collapsed into consensus sequences using the RTCR pipeline to accurately recover TCRβ sequences. Downstream data analysis of TCRβ repertoires was performed using the tcR R package [35].

Healthy longitudinal sequencing data [36] was obtained from the immuneACCESS portal of Adaptive Biotechnologies repository at: https://clients.adaptivebiotech.com/pub/healthy-adult-time-course-TCRB. Healthy data for validation of the diversity analysis was obtained from the immuneACCESS repository for the datasets from Emerson *et al.* [37], Rowe *et al.* [38], Tourino *et al.* [39], Lindau *et al.* [40], Soto *et al.* [41], Savola *et al.* [42], and De Neuter *et al.* [43]. The processed healthy data from Wang *et al.* [44] was obtained from the supplementary data provided in their publication.

2.5. Statistical analyses

Statistical analyses were performed using R version 3.4.1 [45], and figures were produced using the R package ggplot2 [46]. Generation probabilities of TCRβ amino acid sequences were computed using the generative model of V(D)J recombination implemented by OLGA (Optimized Likelihood estimate of immunoGlobulin Amino-acid sequences) [47], using the default parameters. Diversity estimates were calculated by sample-size-based rarefaction and extrapolation using the R package iNEXT (iNterpolation/EXTrapolation) [48]. Clustering analysis was performed using the GLIPH2 [49] webtool (http://50.255.35.37:8080/). Significant clusters were considered based on the following parameters: number of samples = 3, number of CDR3 ≥ 3, vb_score<0.05, length_score<0.05. After filtering for significance, clusters were ordered based on final_score obtained from GLIPH2. Network graphs of clusters were produced using the R package igraph [50]. Unless indicated otherwise, analysis of differences was performed using Student *t*-test. For multiple group comparisons, one-way anova was used. P-values <0.05 were considered statistically significant.

2.6. Availability of data

The TCRβ sequencing data presented in this study have been deposited in NCBI's Gene Expression Omnibus (GEO) database under GEO: GSE156980. Both raw data and processed data are available.

3. Results

3.1. High-throughput TCR sequencing of SSc patients

To investigate the TCR repertoire dynamics in SSc, we performed high-throughput sequencing (HTS) of TCRβ chains on longitudinal samples obtained from four SSc patients collected over a minimum of two years. The clinical characteristics and HLA haplotypes of the SSc patients are included in Table 1. Among the SSc patients included, two were limited SSc (lcSSc), and two were diffuse SSc patients (dcSSc). From all patients PBLs were collected at baseline (T0), at least one year after inclusion (T1, ranging from 12 to 19 months) and at least two years after inclusion (T2, ranging from 24 to 46 months) (Fig. 1a). Frozen PBL samples were subjected to FACS sorting and sorted non-naive CD4$^+$ T-cells (CD3$^+$CD4$^+$CD45RO$^{+/-}$CD27$^{+/-}$) and non-naive CD8$^+$ T-cells (CD3$^+$CD8$^+$CD45RO$^{+/-}$CD27$^{+/-}$) were used for TCRβ sequencing (Fig. 1b). Sample specific barcodes and UMIs to barcode individual mRNA molecules were used to accurately recover TCRβ sequences using the RTCR pipeline [34]. We produced a total of 906 448 and 125 962 TCRβ UMI corrected amino acid (AA) sequence reads for CD4$^+$ and CD8$^+$ T-cells respectively. The number of UMI corrected AA sequence reads per sample was on average 75 000 for CD4$^+$ T-cells and on average 10 500 for CD8$^+$ T-cells (details see Supplementary Table 2).

3.2. Frequency of Vβ and Jβ gene segment usage indicates temporal persistence of TCRβ sequences in SSc patients

We first assessed the frequency of Vβ and Jβ gene segment usage in SSc patients over time (Fig. 1c). The most frequently used Vβ segments across all samples were V20–1, V5-1 and V7-9, for both CD4$^+$ and CD8$^+$ T-cells. When looking at Jβ segment usage, J2-7, J2-1 and J2-3 were most frequently observed. In previous studies, these Vβ (V20–1, V5-1 and V7-9) and Jβ genes (J2-7, J2-1 and J2-3) were also identified as the most frequently used genes in both healthy and diseased individuals [51,52], reflecting known intrinsic biases in the V-D-J rearrangement process [53]. Additionally, Vβ2, which we identified with a relatively

Table 1
Clinical characteristics of patients included. Time from T0 is indicated in months after the first sample was taken. Abbreviations: dcSSc = diffuse SSc, lcSSc = limited SSc, M = male, F = female, ILD = interstitial lung disease (1 = yes, 0 = no), ANA = anti-nuclear antibodies (1 = yes, 0 = no), mRSS = modified rodman skin score, unk = unkown.

Patient	Time	Months	Subset	Age	Sex	ILD	ANA	mRSS	HLA-A	HLA-B	HLA-C	HLA-DRB1	HLA-DQB1	HLA-DPB1
SSc 1	T0	0	dcSSc	65	F	1	1	11	*11:01 *31:01	*55:01 *56:01	*01:02 *03:03	*08:03 *11:01	*03:01 *03:01	*02:01 *04:01
	T1	19	dcSSc			1	1	11						
	T2	34	dcSSc			1	1	11						
SSc 2	T0	0	dcSSc	31	M	0	1	16	*23:01 *33:03	*08:01 *15:10	*03:04 *03:04	*08:06 *13:04	*03:01 *03:19	*01:01 *131:01
	T1	12	dcSSc			0	1	13						
	T2	25	dcSSc			0	1	13						
SSc 3	T0	0	lcSSc	46	F	1	1	4	*01:01 *02:01	*07:02 *37:01	*06:02 *07:02	*15:01 unk	*06:02 *06: unk	*03:01 *05:01
	T1	18	lcSSc			1	1	4						
	T2	42	lcSSc			1	1	8						
SSc 4	T0	0	lcSSc	49	F	0	1	2	*01:01 *24:02	*08:01 *14:02	*02:02 *07:01	*01:01 *03:01	*02:01 *05:01	*04:01 *04:02
	T1	15	lcSSc			0	1	2						
	T2	46	lcSSc			0	1	2						

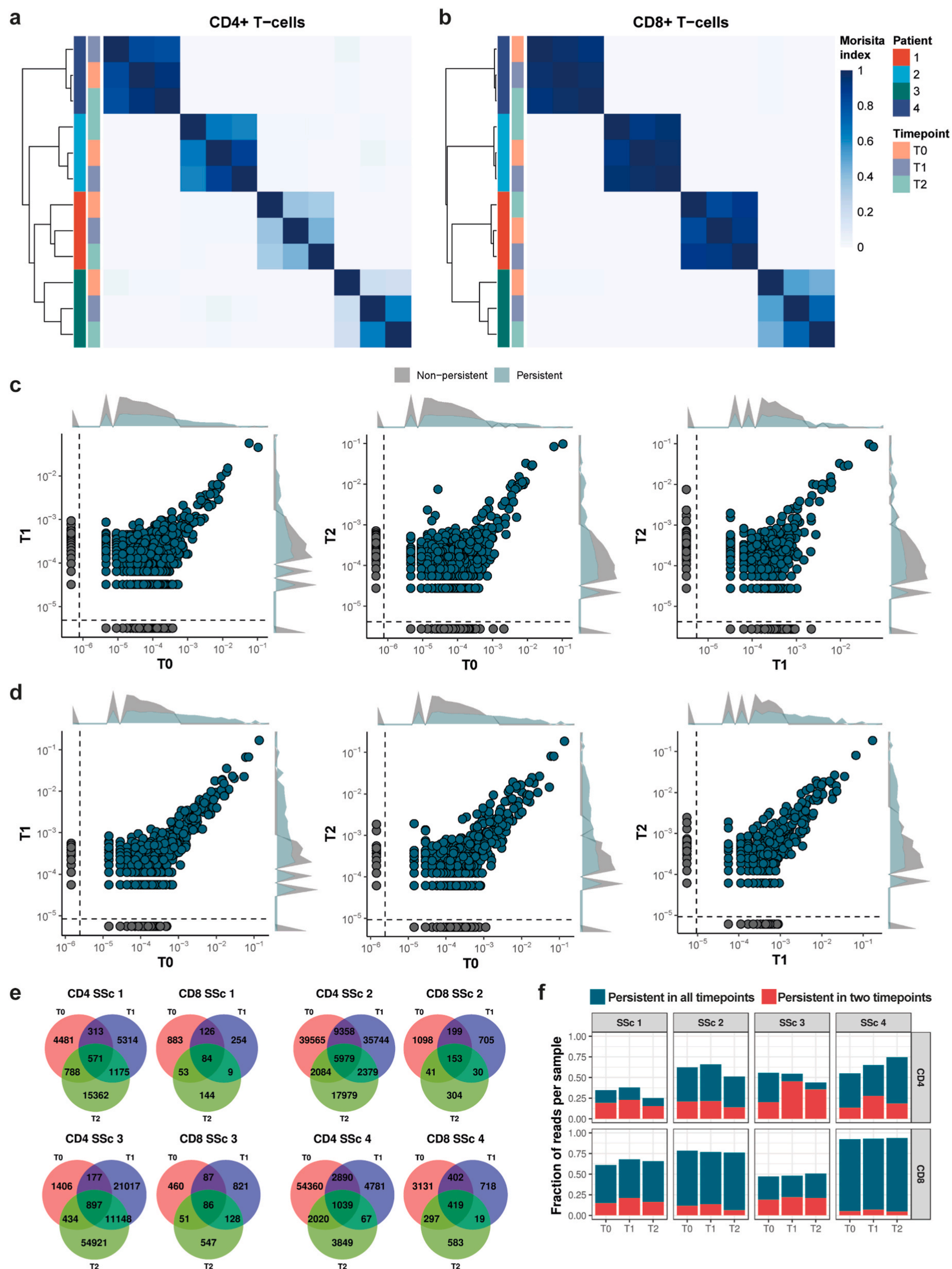

Fig. 2. SSc TCRβ repertoires are highly stable over time and are dominated by persistent TCRβ sequences. **a)** Heatmap showing high overlap (dark blue, morisita index) of TCR sequences within individuals and limited overlap (light blue) of TCR clones between individuals for CD4$^+$ T-cells and **b)** CD8$^+$ T-cells. **c)** Overlap between T-cell sequences within SSc patient 4 over time for CD4$^+$ T-cells and **d)** CD8$^+$ T-cells. Each dot on the scatter plots indicates a single TCR sequence. Axes denote frequency of sequences. Persistent sequences are shown in blue and non-persistent sequences in grey. Frequency was calculated based on the total reads for each sample. Density curves indicate the distribution of sequences across the samples. **e)** Venn diagram showing the exact number of persistent and non-persistent TCR sequences within each patient for CD4 and CD8 T-cells. **f)** Boxplots showing the fraction of total reads per sample (y-axis) across the different samples. Majority of reads is coming from sequences persistent across all time points (blue) or occurring in at least two time points (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

high frequency in CD4$^+$ and CD8$^+$ T-cells, was previously found to be one of the most frequent Vβ chains in peripheral blood T-cells in SSc patients in another study [54], showing that disease specific patterns are also present. Furthermore, we also observed individual specific patterns of Vβ and Jβ segments. As an example, SSc patient 2 displayed a lower frequency of V-28, V7-2 and J2-5 in both CD4$^+$ and CD8$^+$ T-cells across all time points compared to the other patients (Fig. 1c).

In order to quantify the relative similarity of Vβ and Jβ gene segment usage across all samples, we calculated the Jensen-Shannon divergence (JSD) between them (Fig. 1d). The JSD ranges from 0 to 1. A JSD of 0 indicates identical Vβ and Jβ segment usage, while a JSD of 1 indicates that the Vβ and Jβ segment usage is distinct between two samples. When comparing the JSD between samples taken from the same patient, we observed that the use of Vβ and Jβ segments for both CD4$^+$ and CD8$^+$ T-cells was rather consistent over time, while samples taken from different patients displayed a higher divergence amongst each other (Fig. 1d). Moreover, when comparing the differences in JSD between Vβ and Jβ segment usage, we observed that Vβ usage (Fig. 1d, right panels) was more variable then Jβ segment usage (Fig. 1d, left panels) across different individuals. This difference in variability between Vβ and Jβ usage is to be expected since the TCRβ locus has more Vβ than Jβ gene segments (according to the ImMunoGeneTics (IMGT) database [55]), resulting in a greater potential variability for Vβ segment usage. Overall this analysis shows that, Vβ and Jβ gene segment usage is largely persistent within SSc patients over time.

3.3. SSc TCRβ repertoires are highly stable over time

Apart from examining persistence in the use of Vβ and Jβ segments, we wanted to further examine the persistence of full CDR3 amino acid sequences within SSc patients over time. In order to quantify the overlap in TCRβ repertoire between different samples, Morisita's overlap index was calculated to intersect amino acid CDR3 sequences. This index ranges from 0 (no overlapping sequences) to 1 (identical repertoires). Overlap analysis revealed that samples taken from the same patient shared a high number of sequences, indicating a clear temporal persistence of the repertoire within patients, while overlap was extremely limited between samples taken from different patients. This pattern was consistent over all time points, for both CD4$^+$ and CD8$^+$ T-cells (Fig. 2a and Fig. 2b respectively). These results demonstrate that the TCRβ repertoire in SSc patients is highly unique and stable over long periods of time, also at the level of exact CDR3 amino acid sequences.

Next we examined whether the frequencies of TCRβ sequences were also consistent over time. We observed that TCRβ sequences that were found with a high frequency at one time point were also found with a high frequency at the other time points collected from the same patient, for both CD4$^+$ and CD8$^+$ T-cells (Fig. 2c and d respectively). This shows that the frequencies of highly abundant TCRβs within SSc patients are largely consistent over long periods of time, even after almost four years of follow-up for SSc patient 4. Persistence of dominant TCRβ sequences was observed for all the SSc patients included in our study, for both CD4$^+$ and CD8$^+$ T-cells (Supplementary Figure S1). The exact number of TCRβ amino acid sequences overlapping between samples taken from the same patient are shown in Fig. 2e. Although the absolute number of sequences that are overlapping within patients over time (i.e. persistent sequences) are low, they make up a substantial part of the samples in terms of abundance, based on UMI corrected reads, as shown in Fig. 2f. These results clearly demonstrate that SSc repertoires are largely dominated by persistent sequences.

In order to investigate whether persistent TCRβ sequences have any known antigen specificity, we queried the sequences that were persistently present in all three time points for every patient in VDJdb (a curated database of TCR sequences with known antigen specificities) [56]. The results of this analysis are shown in Table 2. In this table we show the hits for peptides presented on MHC II for CD4$^+$ T-cells and peptides presented on MHC I for CD8$^+$ T-cells, matching with the HLA haplotype of the patient from which the sequences were obtained.

For persistent TCRβs from CD8$^+$ T-cells, we mainly found associations with peptides related to viral antigens including EBV, CMV, influenza and HIV (Table 2). For SSc patient 2, we identified one TCRβ sequence (CASSRLAGGTDTQYF) associated with both CMV and HIV-1. This TCRβ sequence most likely represents a hit against CMV in patient 2, as all the patients included in our study are known to be HIV-1 negative. For persistent TCRβs from CD4$^+$ T-cells, we identified two persistent sequences (CASSLEETQYF and CASSLGGEETQYF) associated with CMV. No hits against TCRβ sequences associated with human autoantigens where identified in VDJdb. However, the vast majority of the records present in VDJdb are based on studies of viral and cancer epitopes, while TCRβ sequences associated with autoantigens remain understudied.

Overall, these results reveal a clear temporal persistence of clonally expanded CD4$^+$ and CD8$^+$ T-cells in SSc patients, and we show that the TCRβ repertoire in SSc patients is highly stable over time.

Table 2

Epitope specificity of persistent TCRβs (occurring in all three time points within a patient), according to VDJdb. For CD4$^+$ TCRβ sequences, specificities for peptides presented on MHC II are given and for CD8$^+$ TCRβ sequences, specificities for peptides presented on MHC I are given. Only results for HLA molecules matching the HLA haplotype of the specific patients are given. Sequences indicated with an asterisk (*) are identical.

CD4 - MHC II					
	CDR3	MHC A	MHC B	Epitope	Species
SSc 3	CASSLEETQYF	HLA-DRA*01:01	HLA-DRB1*15	pp65	CMV
SSc 3	CASSLGGEETQYF	HLA-DRA*01:01	HLA-DRB1*15	pp65	CMV
CD8 - MHC I					
	CDR3	**MHC A**	**MHC B**	**Epitope**	**Species**
SSc 1	CASSWGQGSNYGYTF	HLA-A*11:01	B2M	EBNA3B	EBV
SSc 2	CASSLGQAYEQYF	HLA-B*08:01	B2M	EBNA3A	EBV
SSc 2	CASSPGQGEGYEQYF	HLA-B*08:01	B2M	BZLF1	EBV
SSc 2	CASSPGTGEGYEQYF	HLA-B*08:01	B2M	BZLF1	EBV
SSc 2	CASSRLAGGTDTQYF*	HLA-B*08:01	B2M	Gag	HIV-1
SSc 2	CASSRLAGGTDTQYF*	HLA-B*08:01	B2M	IE1	CMV
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	BMLF1	EBV
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	M	Influenza
SSc 3	CSARDRTGNGYTF*	HLA-A*02:01	B2M	BMLF1	EBV
SSc 3	CASSPTDTQYF	HLA-A*02	B2M	pp65	CMV
SSc 4	CASSVGQAYEQYF	HLA-B*08:01	B2M	EBNA3A	EBV

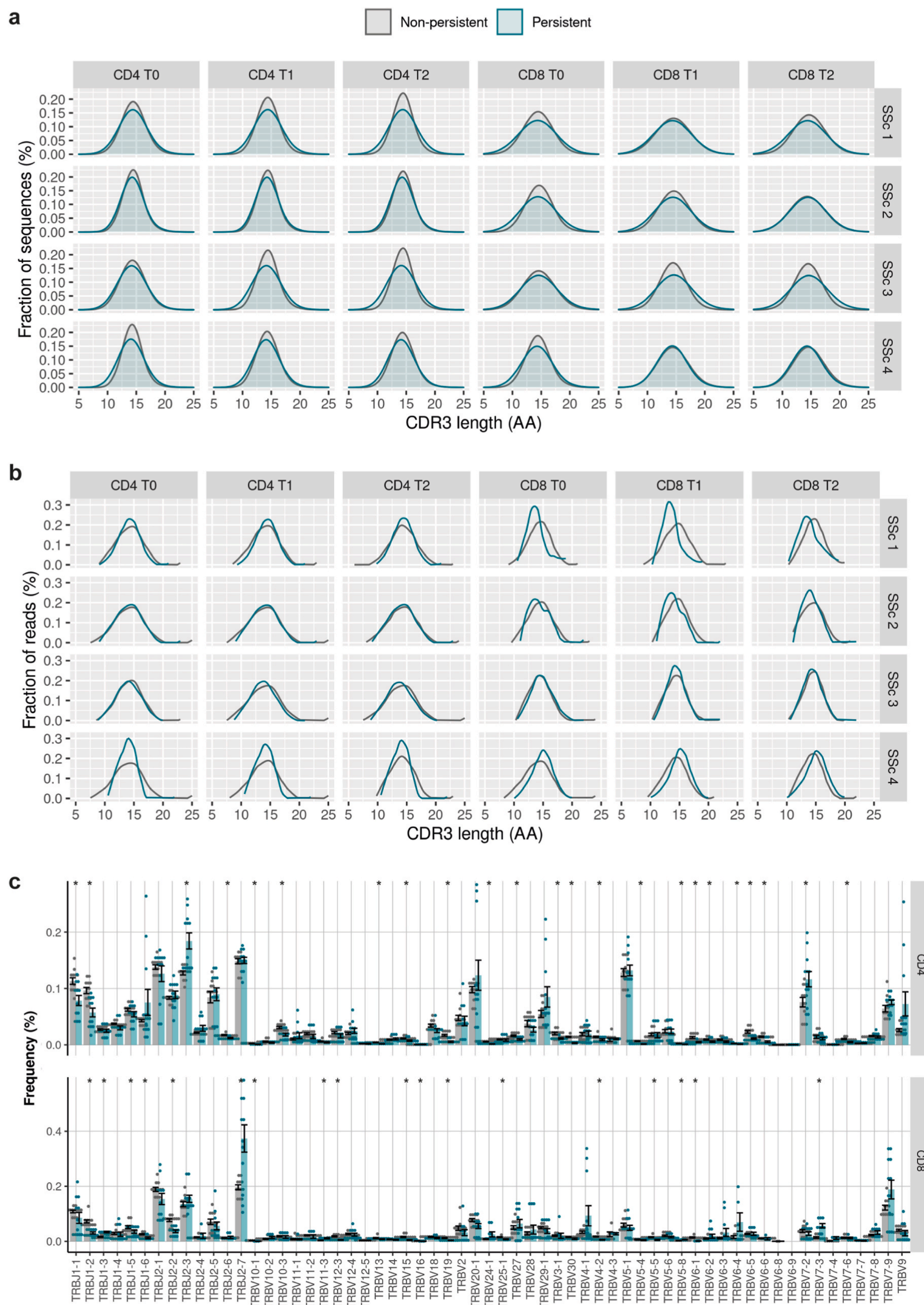

Fig. 3. Comparison of TCRβ characteristics between persistent and non-persistent sequences. **a)** Density plots showing the distribution of lengths of TCR amino acid sequences (x-axis), based on incidence. Distribution of lengths of persistent (blue) and non-persistent (grey) TCR sequences are shown. **b)** Line plots showing the distribution of lengths of TCR amino acid sequences (x-axis), based on abundance. **c)** Bar plot showing the frequency (y-axis) of TCR Vβ and Jβ gene segments (x-axis) of persistent sequences (blue) compared to non-persistent (grey) sequences. Every dot represents one sample. T-tests were performed to test if the gene usage was significantly different between persistent and non-persistent sequences (* = $p < 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

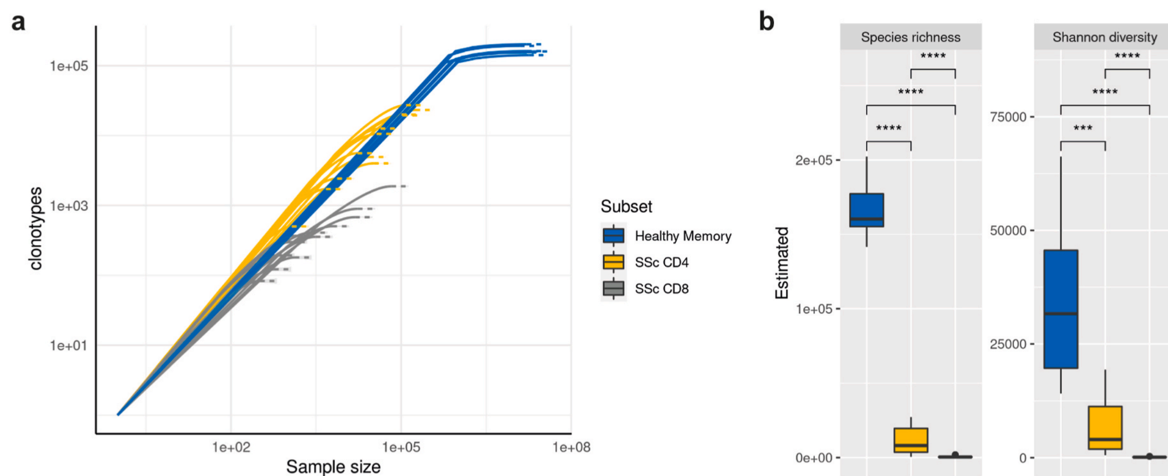

Fig. 4. SSc TCRβ repertoires have a decreased diversity compared to healthy memory repertoires. **a)** Sample based rarefaction and extrapolation curves. Solid lines depict observed data, dashed line depict extrapolated data. Calculated for all healthy memory samples (blue), SSc CD4 (yellow) and SSc CD8 samples (grey). Every line represents one sample. **b)** Boxplots show median of asymptotic diversity estimates calculated from the rarefied and extrapolated data shown in G. ***$p < 0.001$, ****$p < 0.0001$ (one way anova). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

3.4. Persistent clones display common features across SSc patients

Previously, TCRs in the context of autoimmune disease have been associated with certain characteristics such as shorter length and a bias in Vβ/Jβ-gene segment usage [21,39,57]. To further investigate the potential involvement of persistent TCRβs identified in SSc patients in autoimmune responses, we computed the distribution of lengths of all the TCRβ amino acid sequences and compared the lengths of persistent and non-persistent TCRβs. The lengths were calculated based on both incidence (without weighing sequences by their abundance) and abundance (also taking into account the frequency of the sequences). A Gaussian distribution of lengths was observed for both persistent and non-persistent sequences, in $CD4^+$ as well as $CD8^+$ T-cells, when looking at incidence (Fig. 3a), and distribution of lengths of sequences was similar between persistent and non-persistent sequences in all samples (two sample Kolmogorov-Smirnov tests >0.05 for all comparisons). When comparing the distribution of CDR3 lengths of persistent and non-persistent sequences based on abundance, we again observed no significant differences in the distributions neither for $CD4^+$ nor $CD8^+$ T-cells (two sample Kolmogorov-Smirnov tests >0.05 for all comparisons, Fig. 3b). Although the CDR3 length distributions where not significantly different between persistent and non-persistent sequences, In SSc patient 1 and 2, more shorter length sequences were observed in the persistent TCRβs, while in SSc patient 4, longer sequences were found (Fig. 3b). However, this skewness in lengths is mainly caused by the expansion of few dominant TCRβ sequences.

Next to comparing lengths, we also compared the frequencies of Vβ and Jβ segment usage between persistent and non-persistent TCR sequences to see if there was any preferential usage (Fig. 3c). Although most differences observed were small, we identified various Vβ and Jβ gene segments that had either higher or lower frequencies in persistent sequences across all SSc patients. As an example, TRBJ1-2 had a significantly lower frequency in persistent sequences in $CD4^+$ as well as $CD8^+$ T-cells across all SSc patients, while the frequency of TRBV7-2 and TRBV7-3 was higher in persistent sequences in $CD4^+$ and $CD8^+$ T-cells respectively (Fig. 3c). This analysis demonstrates that, although the number of exact sequences that are shared between SSc patients is low, TCRβ sequences that are persistently present in SSc patients over time show similarities in terms of Vβ and Jβ usage. Moreover, these are significantly different from non-persistent sequences, showing that persistent sequences display preferential usage of Vβ and Jβ segments across SSc patients. Given that similar TCR sequences are thought to be involved in T-cell responses to similar antigens [58–61], the preferential segment Vβ and Jβ segment usage across SSc patients over time might reflect chronic immune responses against antigens that are commonly present across patients.

3.5. SSc TCRβ repertoires are less diverse than healthy memory repertoires

The persistence of highly abundant TCRs is not necessarily unique to autoimmune repertoires and has in fact previously been observed in healthy individuals [36]. Therefore, we also compared our SSc data to a public dataset of longitudinal TCRβ sequences from healthy donors [36]. To investigate whether TCRβs in SSc patient repertoires are aberrantly expanded compared to healthy repertoires, we compared the Shannon diversities of healthy memory repertoires to those of SSc. As the samples differed in their sizes, to compare the diversity between samples we performed rarefaction analysis. The healthy control dataset from Chu *et al.* included here did not contain singletons (sequences represented by one read). We therefore performed rarefaction analysis and estimated Shannon diversity in our SSc patient data by including and excluding singletons. In Fig. 4a, we show the estimated Shannon diversity based on the rarefied and extrapolated data excluding singletons. The estimated species richness and Shannon diversity was significantly lower in SSc $CD4^+$ and $CD8^+$ T-cell repertoires as compared to the healthy memory repertoire (p-value <0.05, Fig. 4b). To estimate the effect of excluding singletons on the diversity, the same analysis was also performed including the SSc singletons. In this analysis, the estimated Shannon diversity of TCRβ sequences obtained from SSc $CD8^+$ T-cells were still significantly lower than the diversity of healthy memory cells, while the significant difference in diversity with the $CD4^+$ T-cells was no longer observed (Supplementary Figure S2a-b).

Additionally, next to the decreased diversity of SSc repertoires versus healthy repertoires, we also observed a decreased diversity of the SSc $CD8^+$ repertoire versus the SSc $CD4^+$ repertoire. Given the fact that there was a substantial difference in sequencing depth between $CD4^+$ and $CD8^+$ T-cells in our analysis (Supplementary Table 2), we wanted to confirm the results from the rarefaction analysis in a separate subsampling analysis. We repeated the diversity analysis for 100 random subsamples obtained from SSc $CD4^+$ T-cells and compared these results to the results obtained from SSc $CD8^+$ T-cells. In all subsamples analyzed, $CD8^+$ T-cells had a significantly lower diversity than the $CD4^+$ T-cells (Supplementary Figure S2c), corroborating with the results from the rarefaction analysis from the original/complete data.

As the healthy control dataset from Chu *et al.* contains some technical

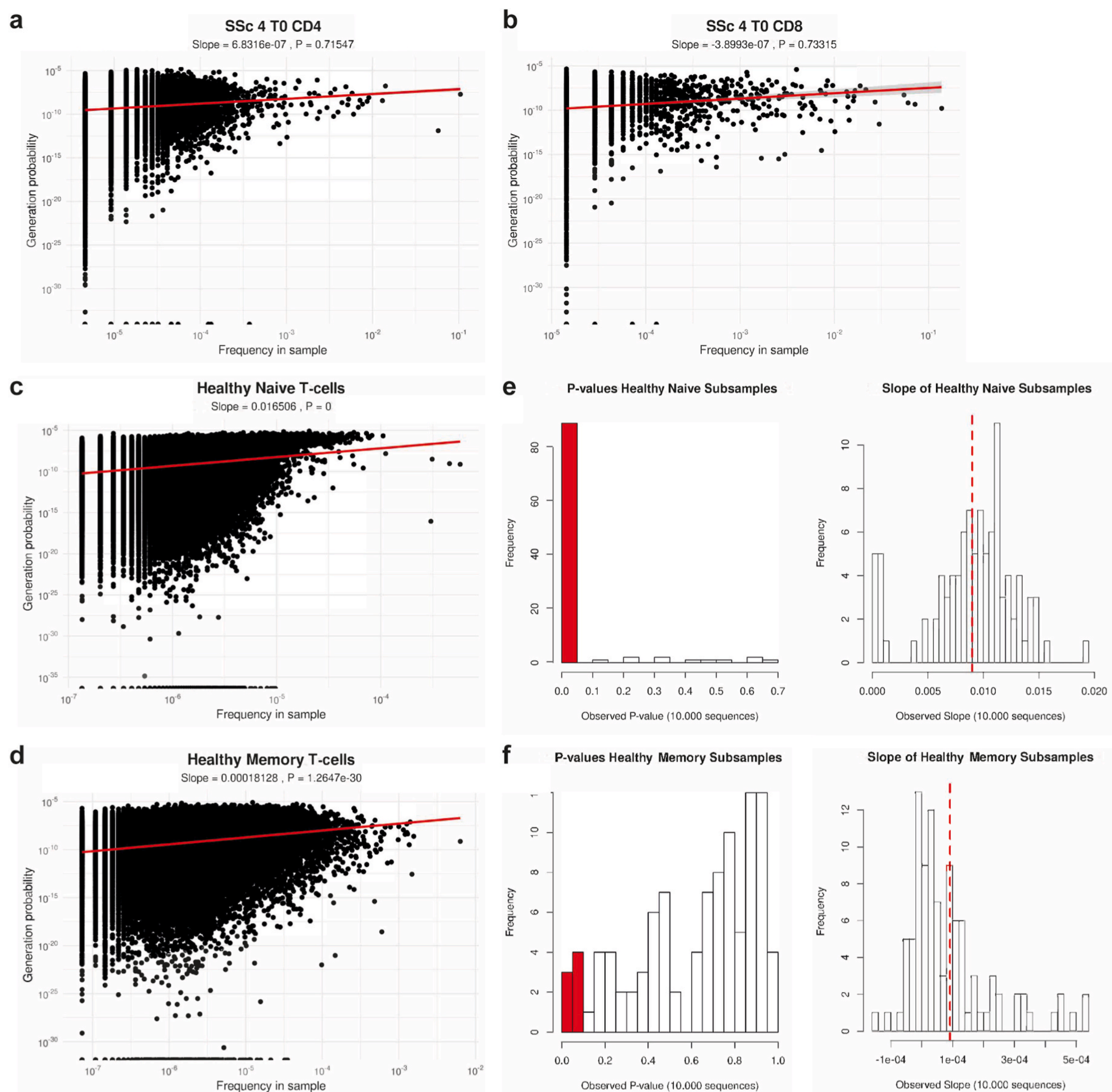

Fig. 5. Frequencies of dominant TCRβ sequences are not driven by generation probabilities in SSc patients. **a)** Linear regression plot between frequency (x-axis) and generation probability (y-axis) of one representative plot of taken from SSc CD4^{+} T-cells (patient 4, time point 0). Red line indicates the linear regression model fit. **b)** Linear regression plot between frequency and generation probability of one representative plot of taken from SSc CD8^{+} T-cells (patient 4, time point 0). **c)** Linear regression plot between frequency and generation probability of one representative plot of taken from healthy naïve T-cells (donor 1, time point 4). **d)** Linear regression plot between frequency and generation probability of one representative plot of taken from healthy memory T-cells (donor 1, time point 4). **e)** Histograms showing the distribution of p-values (left) and slopes (right) for linear regression between frequency and generation probability (computed using OLGA) of subsamples of healthy donor naïve T-cell samples (donor 1, time point 4). Distribution over 100 different subsamples is shown. Red bars in the p-value histogram indicates p-value ≤ 0.05. Red dashed line in the slope histogram indicates the mean slope for the 100 subsamples. **f)** Distribution of p-values (left) and slopes (right) for linear regression between frequency and generation probability of subsamples of healthy donor memory T-cell samples (donor 1, time point 4). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

differences from our SSc dataset (e.g. DNA versus RNA, no distinction between CD4^{+}/CD8^{+} T-cells, exclusion of singletons), we compared the diversity observed in SSc repertoires versus healthy donors in eight additional datasets. The eight additional datasets included TCRβ repertoires from sorted CD4^{+}/CD8^{+} T-cells isolated from healthy individuals (see Supplementary Table 3). We compared Shannon diversity, estimated by rarefaction analysis, for all eight datasets with our SSc dataset (Supplementary Figure 3). For datasets excluding singletons, we also excluded singletons from the SSc data. We were able to validate the decreased diversity observed in SSc patients for both CD4^{+} and CD8^{+} T-cells for the vast majority of the datasets analyzed. In more detail, in six out of seven datasets containing CD4^{+} T-cells, Shannon diversity was significantly lower in SSc as compared to CD4^{+} T-cells from healthy individuals. For all five datasets containing CD8^{+} T-cells, Shannon diversity was significantly lower in SSc as compared to CD8^{+} T-cells from healthy individuals.

Overall, these results demonstrate that the TCRβ repertoire diversity is lower in SSc patients compared to healthy individuals. This provides evidence for a skewed, clonally expanded repertoire in SSc, potentially due to chronic antigen driven T-cell responses.

3.6. *Persistent frequency of dominant TCR sequences is driven by antigenic selection rather than bystander activation*

The hypothesis of 1) bystander activation due to chronic inflammation and 2) clonal expansion driven by the chronic presence of antigens

Table 3
Number of clusters identified by GLIPH2 within SSc patients. All sequences from all time points from each patient (for the two cell types) were used as input for GLIPH2. The total clusters column represents the total number of clusters identified by GLIPH2. The significant clusters column represents the number of significant clusters (consisting of sequences from all three time points, at least three unique CDR3s, length score <0.05, and Vβ score <0.05).

Cell type	Patient	Total clusters	Significant clusters
CD4	SSc 1	7800	204
	SSc 2	50365	3735
	SSc 3	34882	554
	SSc 4	22674	550
CD8	SSc 1	335	2
	SSc 2	534	23
	SSc 3	453	22
	SSc 4	1500	106

skewing the TCR repertoire have been proposed. To test the hypothesis of bystander activation, we investigated whether easy to generate TCRβ sequences (having high generation probabilities) are present at high frequencies in SSc patients. To this extend, we calculated the generation probabilities (pgens) of the TCRβs identified in SSc patients using OLGA [47]. We then used linear regression to model the relationship between TCRβ frequency and pgen. Using this analysis, we show that in SSc CD4$^+$ and CD8$^+$ T-cells, TCRβ frequency and pgen are not related (p-value >0.05, Fig. 5a and Fig. 5b for CD4$^+$ and CD8$^+$ T-cells respectively). This indicates that T-cells persistent in SSc patients have not expanded because of random bystander effects.

For naive T-cells isolated from healthy individuals, we found a significant positive correlation between the pgen and abundance of TCRβs (p-value <0.05, Fig. 5c). For memory T-cells isolated from the same healthy individuals, we also observed a significant positive correlation between pgen and abundance of TCRβs (p-value <0.05, Fig. 5d). Notably, the slope for memory T-cells was lower than that observed for the naïve T-cells (0.00018 versus 0.0165, respectively). From this analysis we show that for naïve and memory TCRβs obtained from healthy individuals, pgen and abundance are positively correlated, whereas in non-naive SSc T-cells no relationship between pgen and abundance is observed. However, as the samples obtained from healthy individuals were sequenced more deeply than our SSc samples, this observed difference in correlation might be confounded by sequencing depth. Therefore, we repeated the linear regression analysis for 100 random subsamples obtained from the healthy dataset and compared these results to the results obtained from SSc samples. Upon subsampling of the naïve healthy T-cells, 89% of the slopes observed the in linear regression were significantly different from zero (linear regression p-value <0.05 and slope >0, as indicated by the red bars in Fig. 5e), confirming that indeed there is a positive relationship between pgen and abundance for TCRβs obtained from healthy naïve T-cells. However, when looking at healthy memory T-cells, we did not confirm the correlation between frequency and pgen that was observed in the full sample (p-value <0.05 in 7% of subsamples, Fig. 5f), showing that in healthy memory T-cells there is no clear correlation between TCRβ pgen and abundance. Thus, subsampling analysis shows that in both healthy memory T-cells and SSc non-naïve CD4$^+$ and CD8$^+$ T-cells, TCRβ pgen and abundance are not correlated. These results suggest that the more abundant TCRβs in SSc repertoires are likely there because of selection, similar to what is observed in healthy memory repertoires, rather than bystander activation of naïve T-cells that are not antigen specific.

3.7. Clusters of similar TCRβ sequences can be identified in SSc patients over time

In order to identify TCRβs in SSc patients that are potentially involved in chronic autoimmune responses, we used "Grouping of Lymphocyte Interactions by Paratope Hotspots" (GLIPH2), that employs sequence similarity and motif analysis to group TCR sequences that potentially recognize the same epitope [49]. To screen for antigen specific TCRβ clusters, we used all sequences obtained from every time point for each individual SSc patient as an input for GLIPH2. In order to exclude false positives, for each patient we considered the clusters that contained sequences from all time points, had at least three unique CDR3s, had similar CDR3 lengths (length score <0.05), and shared similar Vβ-gene frequency distributions (Vβ score <0.05). The number of clusters that were obtained for every patient for CD4$^+$ and CD8$^+$ T-cells are given in Table 3.

Significant clusters were identified in all patients, either based on global similarity (CDR3 sequence differing by maximum one amino acid) or local similarity (shared motif within CDR3 amino acid region). All clusters identified by GLIPH2 are given in Supplementary Table 4. In Fig. 6a, an example network of TCRβ sequences based on clustering analysis by GLIPH2 in CD4$^+$ T-cells is given. Red and purple nodes within this network represent TCRβs that are persistently present in three or two time points respectively, while blue shaded nodes represent sequences are found only in one time point. Nodes are connected when they are part of the same cluster as identified by GLIPH2. Global and local motifs are indicated in green and orange respectively. A "%" sign within the motif indicates a variable amino acid across the sequences in which that particular motif is present.

Within the network, persistent TCRβs share motifs with other persistent and non-persistent TCRβs sequences. This shows that persistent TCRβ sequences cluster together with other, similar TCRβ sequences, potentially representing groups of T-cells responding to the same antigen. The top 10 clusters, based on final cluster score outputted by GLIPH2, and their corresponding motifs for this network are given in Fig. 6b. Some clusters also shared TCRβ sequences with other clusters within a patient, showing that clusters also display convergence between each other (Fig. 6c, top 3 clusters are shown). Similar results were obtained for CD8$^+$ T-cells, where we also identified many clusters of persistent and non-persistent TCRβs sharing motifs within the CDR3 region (Fig. 6d–f). These results demonstrate that, apart from the presence of individual persistent clonally expanded T-cells, clusters of T-cells with potentially similar specificities are present within SSc patients over time.

Lastly, we performed an overlap analysis of all motifs from significant clusters identified by GLIPH2 between the different SSc patients (Fig. 6g and h). We did not observe any motifs for CD4$^+$ T-cells or for CD8$^+$ T-cells that overlap between all four patients. For the CD4$^+$ T-cells, there were 11 motifs that were identified in clusters from three out of four SSc patients (Fig. 6g). These represent groups of T-cells that are likely to recognize the same or highly similar antigens across SSc patients, which could potentially be involved in SSc pathogenesis. We also performed the GLIPH2 analysis on the data from memory T-cells for healthy donors obtained from Chu *et al.* [36]. For this analysis, the same parameters as for SSc were used, however, as for healthy donor 1 and healthy donor 2, memory T-cells were obtained from only two time points, for these donors we also considered clusters that had at least two unique CDR3s to be significant. The number of significant clusters that were obtained for the memory T-cells obtained from healthy donors is given in Supplementary Figure 4a. In order to determine whether the 11 motifs that we found overlapping in CD4$^+$ TCRβ clusters between SSc patients were specific for SSc, we overlapped them with the motifs obtained from the healthy donor memory T-cells. Of the 11 motifs, 2 (S%TTDT and S%DRAYE) were found to be uniquely present in SSc patients (Supplementary Figure 4b). In order to investigate whether these motifs have been associated to known antigens, we queried them against the VDJdb motif database, which contains several curated CDR3 motifs specific to different epitopes [56]. No hits were identified for these motifs.

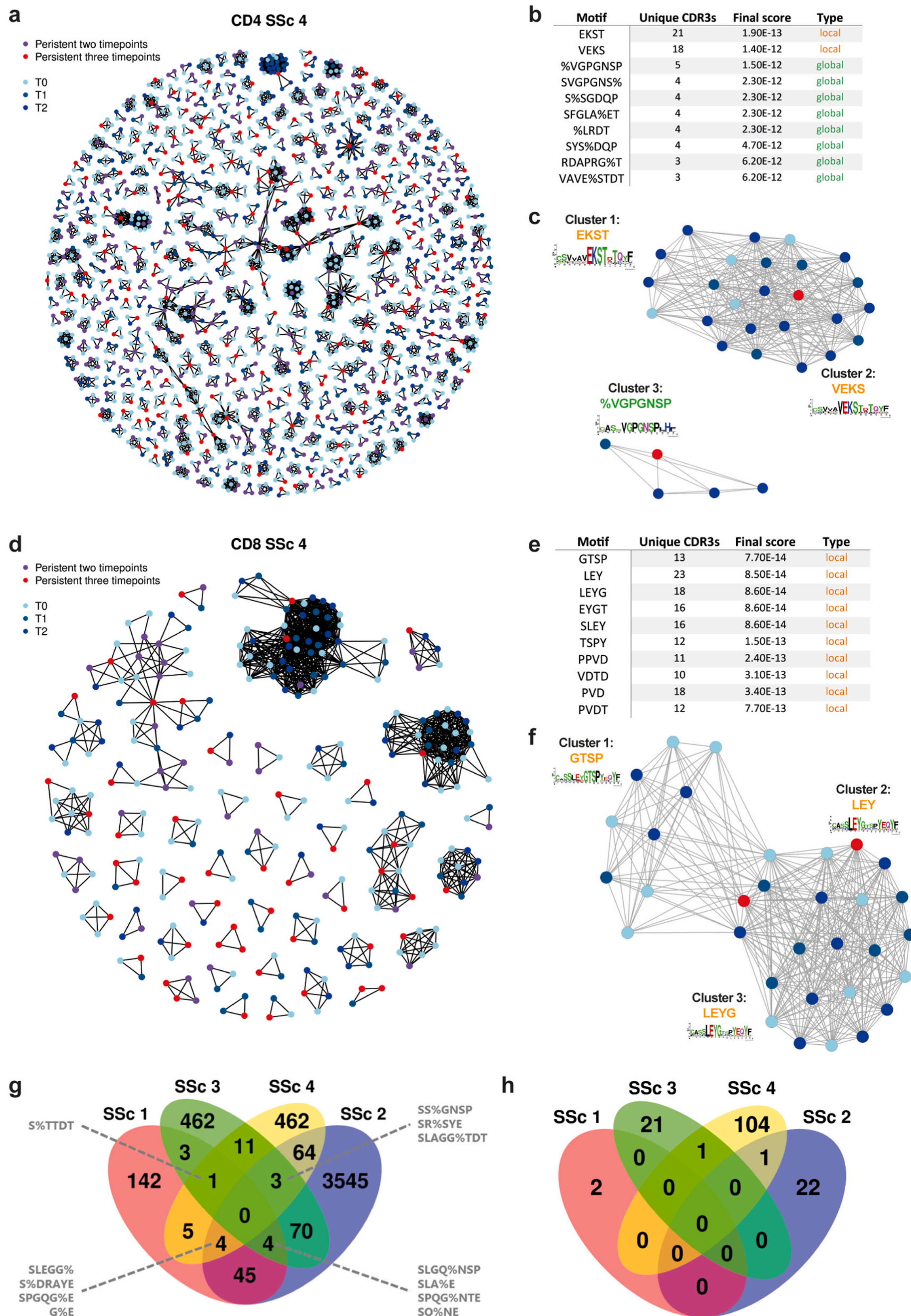

Fig. 6. Clustering analysis of similar TCRβ sequences in SSc patients over time. **a)** Network of TCRβ clusters identified by GLIPH2 in CD4^{+} T-cells from SSc patient 4. Every node represents one TCRβ sequence. Red nodes represent TCRβs persistent across three time points within a patient, purple nodes represent TCRβs persistent across two time points within a patient, and blue shaded nodes represent TCRβs present at a single time point. Nodes are connected if they share a motif or a have similar CDR3 region sequence. **b)** Top ten clusters identified by GLIPH2 in CD4^{+} T-cells from SSc patient 4. For each cluster, the motif identified by GLIPH2 is given. Motifs with global similarity (CDR3 sequence differing by maximum one amino acid) are indicated in green, while motifs with local similarity (shared motif within CDR3 amino acid region) are indicated in orange. **c)** Top three clusters in CD4^{+} T-cells of SSc 4. For each cluster a sequence logo is given based on the TCRβ sequences present in the cluster. **d)** Network of TCRβ clusters identified by GLIPH2 in CD8^{+} T-cells from SSc patient 4. **e)** Top ten clusters identified by GLIPH2 in CD8^{+} T-cells from SSc patient 4. **f)** Top three clusters in CD8^{+} T-cells of SSc 4. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4. Discussion

Identification of TCR sequences that are associated with the chronic autoimmune response in SSc will help to get more insights into the autoimmune pathogenesis of the disease, and will help to identify the antigenic triggers that underlie these responses. Our analysis reveals that the peripheral blood TCRβ repertoire of SSc patients is highly stable over time. Moreover, the TCRβ sequences that were found within a patient with a high frequency at one time point were also found with a high frequency at the other time points (even after four years), showing that frequencies of dominant TCRβs are largely consistent over time. These persistent, clonally expanded CD4$^+$ and CD8$^+$ T-cells are potentially involved in the autoimmune responses underlying SSc pathogenesis. Furthermore, we have shown that the persistent expansion of these T-cells is likely a result of antigenic selection rather than recombination bias, as TCRβ frequencies were not related to their respective generation probabilities.

When we queried the persistent TCRβs found in our SSc patients in VDJdb, we obtained several hits for TCRβ sequences that are known to be associated with viral antigens from influenza, CMV and EBV, especially in the CD8$^+$ T-cell compartment. SSc patient 4 has a CMV and EBV positive status, and TCRβ sequences associated with CMV and EBV were identified in this patient. For the other SSc patients included in this study, the CMV and EBV status are unknown. Interstingly, EBV and CMV infections have been shown to be environmental risk factors for SSc [62–65], and molecular mimicry between chronic viral antigens and human autoantigens has been proposed as a potential driver for autoimmune disease [66].

Chu *et al.* have previously shown that subsets of persistent TCRβs are also present within healthy individuals [36]. Thus, persistence of TCR sequences in itself is likely not just a characteristic of autoimmune related repertoires. However, the temporal dynamics of the TCR repertoire in healthy individuals in the aforementioned study have only been investigated over a period of one year, so it remains to be seen whether this stability is also maintained in healthy individuals over longer periods of time, as is observed in SSc in our study. Moreover, whereas the previous study looked into the total pool of memory T-cells, we show that persistent TCRβs can be identified in both the CD4$^+$ and CD8$^+$ memory T-cell compartments separately.

When further comparing the TCR repertoires of CD4$^+$ and CD8$^+$ T-cells from SSc patients to repertoires obtained from healthy memory T-cells, we found that SSc repertoires have lower diversity. Indeed, decreased diversities of TCRβ repertoires as compared to healthy have been observed in other autoimmune diseases [20,52,67], and have been proposed as a characteristic of autoimmune repertoires. Interestingly, in SSc, differences in the diversity of the TCR repertoire have also been observed between responders and non-responders after autologous hematopoietic stem-cell transplantation (AHSCT, the only therapy with long-term clinical benefit in SSc), with non-responders having a less diverse repertoire [68]. This provides further evidence that decreased TCR repertoire diversity contributes to the autoimmune pathogenesis of SSc.

Determinant spreading has been proposed as a pathogenic event in various autoimmune diseases. During determinant spreading, an antigenic epitope can induce an immune response against other, distinct epitopes on the same protein or other proteins in the same tissue. Those epitopes then become additional targets for the immune response. In systemic sclerosis, vascular damage or skin injury could induce an autoimmune response against epitopes that are normally sequestered [69]. It has been proposed that during determinant spreading, diversification of the TCR repertoire occurs as various TCRs would be expected to respond against the novel epitopes. However, this response is thought to be dynamic and although diversification of the repertoire may occur in early stages of the disease, at later time points the response might become more restricted [70]. Following this line of thought, as the SSc patients in our study have more established forms of the disease (lcSSc and dcSSc), we expect to see a more skewed TCR repertoire. Indeed when comparing the diversity of the TCR repertoire of our SSc patients to healthy donors, we observe a decrease in the repertoire diversity in SSc. In the context of determinant spreading, it would be interesting to repeat this analysis in patients at earlier stages of the disease (for example eaSSc patients) and see if the repertoire is more diverse, which could imply that determinant spreading is ongoing and contributing to disease pathogenesis.

In our generation probability analysis, we show that in both healthy memory T-cells and SSc non-naïve CD4$^+$ and CD8$^+$ T-cells, TCRβ pgen and abundance are not correlated. Therefore it is likely that in SSc patients, similar as to healthy donors, T-cell expansion is caused by antigen specific selection rather than bystander effects. Bystander activation of T-cells has been demonstrated to be driven by excessive production of cytokines including type I IFN, IL-15 and IL-18 for CD8$^+$ T-cells [71], and IL-2 for CD4$^+$ T-cells [72] during immunopathology. Interestingly, type I IFNs are implicated in SSc pathogenesis [73], and IL-15 and IL-18 have been found to be increased in circulation of SSc patients as compared to healthy individuals [74,75]. This shows that SSc patients display a skewed pro-inflammatory milieu that favors bystander activation of T-cells. Therefore, one might expect an increased number of bystander expanded T-cells within these patients as compared to healthy individuals. However, since we observe no difference between SSc and healthy memory T-cells within our generation probability analysis, we show that bystander activation does not significantly contribute to the skewed clonal expansion of T-cells that we observed in SSc patients.

Predicting T-cell reactivity towards antigens is one of the major areas currently investigated in the field of TCR research. Since prediction of TCR binding to a specific antigen is extremely challenging, current efforts are more focused on identifying groups of TCRs that contain certain motifs within their CDR3 region [58–61]. These groups of TCRs comprise clones that potentially respond to the same antigen. Apart from exact sharing of TCRβ sequences between samples, we also identified clusters of TCRβs that share sequence motifs and were persistent within patients. This indicates that antigen selection reshapes the TCRβ repertoire in SSc. Potential antigens could include self-antigens, or chronic infections with pathogens (e.g., CMV or EBV, for which we identified associated persistent TCRβs). Interestingly, we also found clusters of TCRβs from CD4$^+$ T-cells within patients that shared motifs with other TCRβ clusters between patients, even though these patients did not all share common HLA alleles. Notably, two of these motifs could not be detected in TCRβ clusters identified from memory T-cells of healthy donors. These could represent clusters of similar TCRβs that might contribute to more public autoimmune responses underlying SSc pathogenesis. The antigens that these clusters of TCRβs are potentially responding to remain to be identified. Although some epitope associated TCR CDR3 motifs have recently been identified in VDJdb [56], current information on antigen specific motifs is extremely sparse, especially for MHC II associated epitopes. Sequencing of larger (longitudinal) cohorts and MHC Class II tetramer studies including autoantigens are needed to further identify which TCR motifs can be associated with autoimmune disease, and link them to their potential pathogenic targets. For CD8$^+$ T-cells, we did not find any clusters overlapping between more than two patients. However, in general we obtained less clusters in CD8$^+$ T-cells than we did in CD4$^+$ T-cells. This could be due to the fact that we sequenced less CD8$^+$ than CD4$^+$ T-cells, and thus this difference might be explained by a difference in sequencing depth between the two cell types.

To validate our findings and further study the potential pathogenic role of antigen specific TCRβ clusters in SSc, larger patient cohorts should be studied. In this cohort we included a limited number of patients with similar clinical characteristics which makes it difficult to account for factors such as age, sex and ethnicity influencing the immune system. Thus, studying larger longitudinal cohorts are needed to further define disease specific clusters of autoimmune associated TCRβs. Lastly, it would also be interesting to perform immune sequencing of SSc

skin to see if these TCRβ clusters/motifs can be traced back in the skin (the major organ affected by the disease) of SSc patients.

5. Conclusion

Our data provide evidence for an antigen driven expansion of $CD4^{+}$/$CD8^{+}$ T-cells in SSc. We have identified clusters of T-cell clones that are highly persistent over time, and we have shown that this persistence likely is a result of antigenic selection.

Author contributions

N.H. Servaas: Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Visualization F. Zaaraoui-Boutahar: Investigation, Writing – review & editing C.G.K. Wichers: Investigation, Writing – review & editing A. Ottria: Investigation, Writing – review & editing E. Chouri: Investigation, Writing – review & editing A.J. Affandi: Investigation, Writing – review & editing S. Silva-Cardoso: Investigation, Writing – review & editing M. van der Kroef: Investigation, Writing – review & editing T. Carvalheiro: Investigation, Writing – review & editing F. van Wijk: Writing – review & editing T.R.D.J. Radstake: Writing – review & editing, Funding acquisition A.C. Andeweg: Methodology, Investigation, Resources, Writing – review & editing A. Pandit: Conceptualization, Methodology, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. All authors critically reviewed and approved the manuscript.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

We thank all the Systemic Sclerosis patients included in this study who graciously donated their time and samples. Furthermore we thank Sanne Hiddingh and the Core Flow cytometry Facility (CFF) at the UMC Utrecht for their help in sorting the T-cells samples included in this study, and Wilfred van IJcken and the Erasmus Center Biomics team (Erasmus MC Rotterdam) for Illumina sequencing. We are grateful to Anneline Hinrichs for her help with gathering of clinical data, and Rob de Boer and Peter de Greef from the Theoretical Biology & Bioinformatics department of Utrecht University for their valuable input and discussions of the data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jaut.2020.102574.

Funding

The work was supported by the Center of Translational Immunology at the UMC Utrecht, which is a non-profit research organization. A.P. was supported by Netherlands Organization for Scientific Research (NWO) grant number: 016. Veni.178.027. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] A. Gabrielli, E. V Avvedimento, T. Krieg, Scleroderma, N. Engl. J. Med. 360 (2009) 1989–2003, https://doi.org/10.1056/NEJMra0806188.
[2] A.D. Roumm, T.L. Whiteside, T.A. Medsger, G.P. Rodnan, Lymphocytes in the skin of patients with progressive systemic sclerosis, Arthritis Rheum. 27 (1984) 645–653, https://doi.org/10.1002/art.1780270607.
[3] A. Kalogerou, E. Gelou, S. Mountantonakis, L. Settas, E. Zafiriou, L. Sakkas, Early T cell activation in the skin from patients with systemic sclerosis, Ann. Rheum. Dis. 64 (2005) 1233–1235, https://doi.org/10.1136/ard.2004.027094.
[4] R. de Palma, E. D'aiuto, S. Vettori, P.P. Cuoppolo, G. Abbate, G. Valentini, Peripheral T cells from patients with early systemic sclerosis kill autologous fibroblasts in co-culture: is T-cell response aimed to play a protective role? Rheumatology 49 (2010) 1257–1266, https://doi.org/10.1093/rheumatology/keq094.
[5] T. Hügle, S. O'Reilly, R. Simpson, M.D. Kraaij, V. Bigley, M. Collin, et al., Tumor necrosis factor-costimulated T lymphocytes from patients with systemic sclerosis trigger collagen production in fibroblasts, Arthritis Rheum. 65 (2013) 481–491, https://doi.org/10.1002/art.37738.
[6] R. De Palma, F. Del Galdo, S. Lupoli, P. Altucci, G.F. Abbate, G. Valentini, Peripheral T lymphocytes from patients with early systemic sclerosis co-cultured with autologous fibroblasts undergo an oligoclonal expansion similar to that occurring in the skin, Clin. Exp. Immunol. 144 (2006) 169–176, https://doi.org/10.1111/j.1365-2249.2006.03041.x.
[7] A. Kahan, A. Kahan, F. Picard, C.J. Menkes, B. Amor, Abnormalities of T lymphocyte subsets in systemic sclerosis demonstrated with anti-CD45RA and anti-CD29 monoclonal antibodies, Ann. Rheum. Dis. 50 (1991) 354–358, https://doi.org/10.1136/ard.50.6.354.
[8] L.P. Ercole, M. Malvezzi, A.C. Boaretti, S.R. Utiyama, A. Rachid, Analysis of lymphocyte subpopulations in systemic sclerosis, J Investig. Allergol. Clin. Immunol. 13 (2003) 87–93.
[9] G. Papp, I.F. Horvath, S. Barath, E. Gyimesi, S. Sipka, P. Szodoray, et al., Altered T-cell and regulatory cell repertoire in patients with diffuse cutaneous systemic sclerosis, Scand. J. Rheumatol. 40 (2011) 205–210, https://doi.org/10.3109/03009742.2010.528021.
[10] P. Fuschiotti, A.T. Larregina, J. Ho, C. Feghali-Bostwick, T.A. Medsger, Interleukin-13-producing CD8+ T cells mediate dermal fibrosis in patients with systemic sclerosis, Arthritis Rheum. 65 (2013) 236–246, https://doi.org/10.1002/art.37706.
[11] M.E. Truchetet, N.C. Brembilla, E. Montanari, Y. Allanore, C. Chizzolini, Increased frequency of circulating Th22 in addition to Th17 and Th2 lymphocytes in systemic sclerosis: association with interstitial lung disease, Arthritis Res. Ther. 13 (2011) R166, https://doi.org/10.1186/ar3486.
[12] S. O'Reilly, T. Hügle, J.M. Van Laar, T Cells in Systemic Sclerosis: A Reappraisal, Rheumatol, United Kingdom, 2012, https://doi.org/10.1093/rheumatology/kes090.
[13] M.M. Davis, P.J. Bjorkman, T-cell antigen receptor genes and T-cell recognition, Nature 334 (1988) 395–402, https://doi.org/10.1038/334395a0.
[14] C.H. Bassing, W. Swat, F.W. Alt, The mechanism and regulation of chromosomal V(D)J recombination, Cell 109 (2002) S45–S55, https://doi.org/10.1016/S0092-8674(02)00675-X.
[15] H.S. Robins, S.K. Srivastava, P.V. Campregher, C.J. Turtle, J. Andriesen, S. R. Riddell, et al., Overlap and effective size of the human CD8+ T cell receptor repertoire, Sci. Transl. Med. 2 (2010), https://doi.org/10.1126/scitranslmed.3001442.
[16] T.P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, P. Kourilsky, A direct estimate of the human αβ T cell receptor diversity, Science 80 (1999), https://doi.org/10.1126/science.286.5441.958.
[17] Q. Qi, Y. Liu, Y. Cheng, J. Glanville, D. Zhang, J.Y. Lee, et al., Diversity and clonal selection in the human T-cell repertoire, Proc. Natl. Acad. Sci. U. S. A. 111 (2014) 13139–13144, https://doi.org/10.1073/pnas.1409155111.
[18] T. Mora, A.M. Walczak, How many different clonotypes do immune repertoires contain? Curr. Opin. Struct. Biol. 18 (2019) 104–110, https://doi.org/10.1016/j.coisb.2019.10.001.
[19] Y. Zhao, P. Nguyen, J. Ma, T. Wu, L.L. Jones, D. Pei, et al., Preferential use of public TCR during autoimmune encephalomyelitis, J. Immunol. 196 (2016), https://doi.org/10.4049/jimmunol.1501029.
[20] C.M. Chang, Y.W. Hsu, H.S.C. Wong, J.C.C. Wei, X. Liu, H.T. Liao, et al., Characterization of T-cell receptor repertoire in patients with rheumatoid arthritis receiving biologic therapies, Dis. Markers (2019), https://doi.org/10.1155/2019/2364943.
[21] X. Liu, W. Zhang, M. Zhao, L. Fu, L. Liu, J. Wu, et al., T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis, Ann. Rheum. Dis. 78 (2019) 1070–1078, https://doi.org/10.1136/annrheumdis-2019-215442.
[22] J.J. Goronzy, C.M. Weyand, T-cell senescence and contraction of T-cell repertoire diversity - catalysts of autoimmunity and chronic inflammation, Arthritis Res. Ther. 5 (2003) 225–234, https://doi.org/10.1186/ar974.
[23] L.I. Sakkas, B. Xu, C.M. Artlett, S. Lu, S.A. Jimenez, C.D. Platsoucas, Oligoclonal T cell expansion in the skin of patients with systemic sclerosis, J. Immunol. 168 (2002) 3649–3659, https://doi.org/10.4049/jimmunol.168.7.3649.
[24] V.V. Yurovsky, F.M. Wigley, R.A. Wise, B. White, Skewing of the CD8+ T-cell repertoire in the lungs of patients with systemic sclerosis, Hum. Immunol. 48 (1996) 84–97, https://doi.org/10.1016/0198-8859(96)00091-2.
[25] B. White, V. V Yurovsky, Oligoclonal expansion of V delta 1+ gamma/delta T-cells in systemic sclerosis patients, Ann. N. Y. Acad. Sci. (1995), https://doi.org/10.1111/j.1749-6632.1995.tb44542.x.
[26] M. Attaf, E. Huseby, A.K. Sewell, αβ T cell receptors as predictors of health and disease, Cell. Mol. Immunol. 12 (2015) 391–399, https://doi.org/10.1038/cmi.2014.134.
[27] Y. Pacheco, Y. Acosta-Ampudia, D.M. Monsalve, C. Chang, M.E. Gershwin, J.-M. Anaya, Bystander Activation and Autoimmunity, J Autoimmun 103 (2019) 391–399, https://doi.org/10.1016/j.jaut.2019.06.012.

[28] H.G. Lee, J.U. Lee, D.H. Kim, S. Lim, I. Kang, J.M. Choi, Pathogenic function of bystander-activated memory-like CD4 + T cells in autoimmune encephalomyelitis, Nat. Commun. 10 (2019) 1–14, https://doi.org/10.1038/s41467-019-08482-w.

[29] N.J. Burroughs, M. Ferreira, B.M.P.M. Oliveira, A.A. Pinto, Autoimmunity arising from bystander proliferation of T cells in an immune response model, Math. Comput. Model. 53 (2011) 1389–1393, https://doi.org/10.1016/j.mcm.2010.01.020.

[30] V. Venturi, K. Kedzierska, D.A. Price, P.C. Doherty, D.C. Douek, S.J. Turner, et al., Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 18691–18696, https://doi.org/10.1073/pnas.0608907103.

[31] F. Van Den Hoogen, D. Khanna, J. Fransen, S.R. Johnson, M. Baron, A. Tyndall, et al., Classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative, Ann. Rheum. Dis. 72 (2013) (2013) 1747–1755, https://doi.org/10.1136/annrheumdis-2013-204424.

[32] I.Z. Mamedov, O.V. Britanova, I.V. Zvyagin, M.A. Turchaninova, D.A. Bolotin, E. V. Putintseva, et al., Preparing unbiased T-cell receptor and antibody cDNA libraries for the deep next generation sequencing profiling, Front. Immunol. 4 (2013) 1–10, https://doi.org/10.3389/fimmu.2013.00456.

[33] J. Zhang, K. Kobert, T. Flouri, A. Stamatakis, PEAR: a fast and accurate Illumina Paired-End reAd mergeR, Bioinformatics 30 (2014) 614–620, https://doi.org/10.1093/bioinformatics/btt593.

[34] B. Gerritsen, A. Pandit, A.C. Andeweg, R.J. De Boer, RTCR: a pipeline for complete and accurate recovery of T cell repertoires from high throughput sequencing data, Bioinformatics 32 (2016) 3098–3106, https://doi.org/10.1093/bioinformatics/btw339.

[35] V.I. Nazarov, M.V. Pogorelyy, E.A. Komech, I.V. Zvyagin, D.A. Bolotin, M. Shugay, et al., tcR: an R package for T cell receptor repertoire advanced data analysis, BMC Bioinf. 16 (2015) 175, https://doi.org/10.1186/s12859-015-0613-1.

[36] N.D. Chu, H.S. Bi, R.O. Emerson, A.M. Sherwood, M.E. Birnbaum, H.S. Robins, et al., Longitudinal immunosequencing in healthy people reveals persistent T cell receptors rich in highly public receptors, BMC Immunol. 20 (2019) 19, https://doi.org/10.1186/s12865-019-0300-5.

[37] R. Emerson, A. Sherwood, C. Desmarais, S. Malhotra, D. Phippard, H. Robins, Estimating the ratio of CD4+ to CD8+ T cells using high-throughput sequence data, J. Immunol. Methods 391 (2013) 14–21, https://doi.org/10.1016/j.jim.2013.02.002.

[38] J.H. Rowe, O.M. Delmonte, S. Keles, B.D. Stadinski, A.K. Dobbs, L.A. Henderson, et al., Patients with CD3G mutations reveal a role for human CD3g in Treg diversity and suppressive function, Blood 131 (2018) 2335–2344, https://doi.org/10.1182/blood-2018-02-835561.

[39] I. Gomez-Tourino, Y. Kamra, R. Baptista, A. Lorenc, M. Peakman, T cell receptor β-chains display abnormal shortening and repertoire sharing in type 1 diabetes, Nat. Commun. 8 (2017) 1–15, https://doi.org/10.1038/s41467-017-01925-2.

[40] P. Lindau, R. Mukherjee, M.V. Gutschow, M. Vignali, E.H. Warren, S.R. Riddell, et al., Cytomegalovirus exposure in the elderly does not reduce CD8 T cell repertoire diversity, J. Immunol. 202 (2019) 476–483, https://doi.org/10.4049/jimmunol.1800217.

[41] C. Soto, R.G. Bombardi, A. Branchizio, N. Kose, P. Matta, A.M. Sevy, et al., High frequency of shared clonotypes in human B cell receptor repertoires, Nature 566 (2019) 398–402, https://doi.org/10.1038/s41586-019-0934-8.

[42] P. Savola, T. Kelkka, H.L. Rajala, A. Kuuliala, K. Kuuliala, S. Eldfors, et al., Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis, Nat. Commun. 8 (2017), https://doi.org/10.1038/ncomms15869.

[43] N. De Neuter, E. Bartholomeus, G. Elias, N. Keersmaekers, A. Suls, H. Jansens, et al., Memory CD4+ T cell receptor repertoire data mining as a tool for identifying cytomegalovirus serostatus, Gene Immun. 20 (2019) 255–260, https://doi.org/10.1038/s41435-018-0035-y.

[44] L. Wang, P. Zhang, J. Li, H. Lu, L. Peng, J. Ling, et al., High-throughput sequencing of CD4+ T cell repertoire reveals disease-specific signatures in IgG4-related disease, Arthritis Res. Ther. 21 (2019) 1–15, https://doi.org/10.1186/s13075-019-2069-6.

[45] R. R Development Core Team, A Language and Environment for Statistical computing., Vienna, Austria, 2017, ISBN 3-900051-07-0. R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org.

[46] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.

[47] Z. Sethna, Y. Elhanati, C.G. Callan, A.M. Walczak, T. Mora, OLGA: fast computation of generation probabilities of B-and T-cell receptor amino acid sequences and motifs, Bioinformatics 35 (2019) 2974–2981, https://doi.org/10.1093/bioinformatics/btz035.

[48] T.C. Hsieh, K.H. Ma, A. Chao, iNEXT: an R package for rarefaction and extrapolation of species diversity (Hill numbers), Methods Ecol. Evol. (2016), https://doi.org/10.1111/2041-210X.12613.

[49] H. Huang, C. Wang, F. Rubelt, T.J. Scriba, M.M. Davis, Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening, Nat. Biotechnol. (2020) 1–9, https://doi.org/10.1038/s41587-020-0505-4.

[50] G. Csardi, T. Nepusz, The Igraph Software Package for Complex Network Research, InterJournal Complex Syst, 2006.

[51] L. Ma, L. Yang, B. Shi, X. He, A. Peng, Y. Li, et al., Analyzing the CDR3 repertoire with respect to TCR - beta chain V-D-J and V-J rearrangements in peripheral T cells using HTS, Sci. Rep. 6 (2016) 1–10, https://doi.org/10.1038/srep29544.

[52] J.H. Cui, Y.B. Jin, K.R. Lin, P. Xiao, X.P Chen, Y.M. Pan, et al., Characterization of peripheral blood TCR repertoire in patients with ankylosing spondylitis by high-throughput sequencing, Hum. Immunol. 79 (2018) 485–490, https://doi.org/10.1016/j.humimm.2018.03.007.

[53] J.E. Park, R.A. Botting, C.D. Conde, D.M. Popescu, M. Lavaert, D.J. Kunz, et al., A cell atlas of human thymic development defines T cell repertoire formation, Science 367 (2020), https://doi.org/10.1126/science.aay3224.

[54] D. Farge, L.C.M. Arruda, F. Brigant, E. Clave, C. Douay, Z. Marjanovic, et al., Long-term immune reconstitution and T cell repertoire analysis after autologous hematopoietic stem cell transplantation in systemic sclerosis patients, J. Hematol. Oncol. 10 (2017) 21, https://doi.org/10.1186/s13045-016-0388-5.

[55] M.P. Lefranc, IMGT, the international ImMunoGeneTics database®, Nucleic Acids Res. (2003) https://doi.org/10.1093/nar/gkg085.

[56] D.V. Bagaev, R.M.A. Vroomans, J. Samir, U. Stervbo, C. Rius, G. Dolton, et al., VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium, Nucleic Acids Res. 48 (2020) D1057–D1062, https://doi.org/10.1093/nar/gkz874.

[57] X. Hou, P. Zeng, X. Zhang, J. Chen, Y. Liang, J. Yang, et al., Shorter TCR β-chains are highly enriched during thymic selection and antigen-driven selection, Front. Immunol. 10 (2019) 299, https://doi.org/10.3389/fimmu.2019.00299.

[58] M.V. Pogorelyy, A.A. Minervina, M. Shugay, D.M. Chudakov, Y.B. Lebedev, T. Mora, et al., Detecting T cell receptors involved in immune responses from single repertoire snapshots, PLoS Biol. 17 (2019), https://doi.org/10.1371/journal.pbio.3000314.

[59] J. Glanville, H. Huang, A. Nau, O. Hatton, L.E. Wagar, F. Rubelt, et al., Identifying specificity groups in the T cell receptor repertoire, Nature 547 (2017) 94–98, https://doi.org/10.1038/nature22976.

[60] P. Dash, A.J. Fiore-Gartland, T. Hertz, G.C. Wang, S. Sharma, A. Souquette, J. C. Crawford, et al., Quantifiable predictive features define epitope-specific T cell receptor repertoires, Nature 547 (2017) 89–93, https://doi.org/10.1038/nature22383.

[61] A. Madi, A. Poran, E. Shifrut, S. Reich-Zeliger, E. Greenstein, I. Zaretsky, et al., T Cell Receptor Repertoires of Mice and Humans Are Clustered in Similarity Networks Around Conserved Public CDR3 Sequences, Elife, 2017, https://doi.org/10.7554/eLife.22057.

[62] C. Lunardi, C. Bason, R. Navone, E. Millo, G. Damonte, R. Corrocher, et al., Systemic sclerosis immunoglobulin G autoantibodies bind the human cytomegalovirus late protein UL94 and induce apoptosis in human endothelial cells, Nat. Med. 6 (2000) 1183–1186, https://doi.org/10.1038/80533.

[63] Y. Arnson, H. Amital, S. Guiducci, M. Matucci-Cerinic, G. Valentini, O. Barzilai, et al., The role of infections in the immunopathogensis of systemic sclerosis-evidence from serological studies, Ann. N. Y. Acad. Sci. (2009) 627–632, https://doi.org/10.1111/j.1749-6632.2009.04808.x.

[64] L. Sternbæk, A.H. Draborg, M.T. Østerlund, L.V. Iversen, L. Troelsen, E. Theander, et al., Increased antibody levels to stage-specific Epstein–Barr virus antigens in systemic autoimmune diseases reveal a common pathology, Scand. J. Clin. Lab. Invest. 79 (2019) 7–16, https://doi.org/10.1080/00365513.2018.1550807.

[65] L.C.M. Arruda, E. Clave, C. Douay, J.R. Lima-Júnior, S.N. Slavov, K.C. R. Malmegrim, et al., CMV-specific clones may lead to reduced TCR diversity and relapse in systemic sclerosis patients treated with AHSCT, Rheumatology 59 (2020) e38–e40, https://doi.org/10.1093/rheumatology/keaa001.

[66] M.F. Cusick, J.E. Libbey, R.S. Fujinami, Molecular mimicry as a mechanism of autoimmune disease, Clin. Rev. Allergy Immunol. 42 (2012) 102–111, https://doi.org/10.1007/s12016-011-8294-7.

[67] D.R. Thapa, R. Tonikian, C. Sun, M. Liu, A. Dearth, M. Petri, et al., Longitudinal analysis of peripheral blood T cell receptor diversity in patients with systemic lupus erythematosus by next-generation sequencing, Arthritis Res. Ther. 17 (2015) 132, https://doi.org/10.1186/s13075-015-0655-9.

[68] L.C.M. Arruda, K.C.R. Malmegrim, J.R. Lima-Júnior, E. Clave, J.B.E. Dias, D. A. Moraes, et al., Immune rebound associates with a favorable clinical response to autologous HSCT in systemic sclerosis patients, Blood Adv 2 (2018) 126–141, https://doi.org/10.1182/bloodadvances.2017011072.

[69] L.S. Chan, C.J. Vanderlugt, T. Hashimoto, T. Nishikawa, J.J. Zone, M.M. Black, et al., Epitope spreading: lessons from autoimmune skin diseases, J. Invest. Dermatol. 110 (1998) 103–109, https://doi.org/10.1046/j.1523-1747.1998.00107.x.

[70] P.V. Lehmann, E.E. Sercarz, T. Forsthuber, C.M. Dayan, G. Gammon, Determinant spreading and the dynamics of the autoimmune T-cell repertoire, Immunol. Today 14 (1993) 203–208, https://doi.org/10.1016/0167-5699(93)90163-F.

[71] T.S. Kim, E.C. Shin, The activation of bystander CD8+ T cells and their roles in viral infection, Exp. Mol. Med. 51 (2019) 1–9, https://doi.org/10.1038/s12276-019-0316-1.

[72] O. Boyman, Bystander activation of CD4+ T cells, Eur. J. Immunol. 40 (2010) 936–939, https://doi.org/10.1002/eji.201040466.

[73] B. Skaug, S. Assassi, Type I Interferon Dysregulation in Systemic Sclerosis, Cytokine, 2020, https://doi.org/10.1016/j.cyto.2018.12.018.

[74] D.M. Wuttge, M. Wildt, P. Geborek, F.A. Wollheim, A. Scheja, A. Åkesson, Serum IL-15 in patients with early systemic sclerosis: a potential novel marker of lung disease, Arthritis Res. Ther. 9 (2007), https://doi.org/10.1186/ar2284.

[75] E. Lin, F.B. Vincent, J. Sahhar, G.S. Ngian, R. Kandane-Rathnayake, R. Mende, et al., Analysis of serum interleukin(IL)-1α, IL-1β and IL-18 in patients with systemic sclerosis, Clin. Transl. Immunol. 8 (2019) e1045, https://doi.org/10.1002/cti2.1045.