Clinical significance of T cell receptor repertoire in primary Sjogren's syndrome

Chenyang Lu,[a,b,1] Xuenan Pi,[c,1] Wangdong Xu,[d] Pingying Qing,[b] Honghu Tang,[b] Yanhong Li,[a,b] Yi Zhao,[b] Xiao Liu,[e,f] Huairong Tang,[g]* and Yi Liu[a,b]*

[a]Laboratory of Rheumatology and Immunology, West China Hospital, Sichuan University, Chengdu 610041, China
[b]Department of Rheumatology and Immunology, West China Hospital, Sichuan University, Chengdu 610041, China
[c]National Frontier Center of Disease Molecular Network, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, Chengdu 610041, China
[d]Department of Evidence-Based Medicine, Southwest Medical University, Luzhou, China
[e]BGI-Shenzhen, Shenzhen 518083, China
[f]Shenzhen International Graduate School, Tsinghua University, Shenzhen 518055, China
[g]Health Management Center, West China Hospital of Sichuan University, Chengdu 610041, China

Summary

Background Primary Sjogren's syndrome (SS) is a chronic inflammatory disease with unknown aetiology. Although clonal expansion of autoreactive T cells has been identified in patients with SS, the clinical correlation of T-cell receptor (TCR) variance in SS remains unclear.

Methods TCRβ repertoire sequencing was performed on 260 SS patients with 3-6 months of follow-up in a cohort study to dynamically assess the characteristics of TCR diversity and their clinical significance.

Findings We found that SS patients had lower TCR diversity, but higher frequency of public clones than healthy controls (HCs). Significant differences were identified in the usage of the variable (V) gene, joining (J) gene, and V-J pairing between SS and HCs. Eighteen SS-associated clones were identified, showing a high sensitivity and specificity for disease classification. TCR diversity was correlated with the presence of dental caries, thrombocytopenia, hepatocholangeitis, antinuclear antibody, anti-SSA/SSB, and hypergammaglobulinemia but not with disease course, number of relapses, arthritis, rheumatoid factor, hypocomplementemia or disease activity defined by SSDAI. During follow-up, the TCR abnormalities remained, represented by more altered V/J usage and higher frequencies of SS-associated clones. Among SS patients, the sensitive subgroup had increased TCR diversity after treatment. Eighty-five SS-sensitivity associated TCRs were identified and used for sensitivity classification by cross validation with high specificity and sensitivity.

Interpretation These results demonstrate that the TCR repertoire could provide insights into the disease status and prognosis in SS and other autoimmune diseases.

*Corresponding author at: No.37, Guoxue Alley, Wuhou District, Chengdu 610041, China.
E-mail addresses: 1651682099@qq.com (H. Tang), yi2006liu@163.com (Y. Liu).
[1] C. Lu and X. Pi contributed equally to this work.

Introduction

T lymphocytes (T cells) are integral components of adaptive immune responses in most vertebrates. Interacting with antigens in the context of major histocompatibility complex (MHC) molecules on antigen-presenting cells, T-cell receptors (TCRs) play an essential role in the activation of T cells. The peptide segments responsible for antigen binding are located in the

Research in context

Evidence before this study

Primary Sjogren's syndrome (SS) is a systemic autoimmune disease, characterized by the presence of autoantibodies as well as dryness of the mouth and eyes. Autoreactive T cells have been observed in the peripheral blood and organs of SS patients, where they reactivate B cells, secrete cytokines, or cause damage to tissues directly. T-cell clonal expansion and a reduction in the diversity of T-cell receptor (TCR) repertoire have been identified in a small population of SS patients. However, the clinical correlations of TCR variance and the possibility of using the TCR repertoire as a biomarker for SS remain elusive.

Added value of this study

Utilizing next-generation sequencing, we performed a dynamic analysis of the TCRβ repertoire for SS. We found that SS patients have lower TCR diversity but a higher frequency of public clones than healthy controls (HCs). SS patients use distinct variable (V) genes, joining (J) genes and V-J pairing when compared with HCs. Eighteen SS-associated clones were identified, and a random forest model was developed, which can discriminate between SS and HCs with high sensitivity and specificity. In this study, we also demonstrated that the TCR repertoire correlated with clinical features of SS, including the presence of dental caries, thrombocytopenia, hepatocholangeitis, antinuclear antibody, anti-SSA/SSB, and hypergammaglobulinemia, whereas other features, such as disease course, number of relapses, arthritis, rheumatoid factor, and disease activity (SSDAI), were not correlated with TCR diversity. Patients receiving conventional treatment showed persistent TCR abnormalities, indicating that their T-cell alterations were not reversed by these therapies in a short term. However, a subgroup of patients did show increased TCR diversity after treatment and this allowed us to build a model discriminating between treatment-sensitive and treatment-nonsensitive individuals.

Implications of all the available evidence

The data presented here contribute to the understanding of the dysregulation of the immune system in the pathogenesis of SS. Our results also suggest that assessment of the TCR repertoire could potentially serve as a complementary method for evaluating disease status and response to treatment in this disease.

third complementarity-determining regions (CDR3s) of the α and β chains in TCR.[1,2] Within the TCR β chain, the CDR3 region interacts mainly with peptide antigen and is the most variable portion of TCR, allowing for various combinations of the variable (V), diversity (D), and joining (J) regions along with palindromic and random nucleotide additions[1,2] A diverse TCR repertoire is

beneficial for immune responses to foreign pathogens; however, recognition of self-antigens by T cells in the body contributes to the development of autoimmune diseases. Therefore, TCR deciphering is crucial for understanding adaptive immunity in health and disease.[3]

Primary Sjogren's syndrome (SS) is a chronic autoimmune disease characterized by dry mouth, dry eyes, and autoantibodies to nuclear antigens.[4] Autoreactive T cells have been observed in the peripheral blood and organs of SS patients, where they reactivate B cells, secrete cytokines, or directly cause damage to tissues.[5] Accumulating evidence has shown the expansion of partial TCR clones in various autoimmune diseases, including systemic lupus erythematosus (SLE), rheumatoid arthritis (RA), celiac disease, and type 1 diabetes.[6,7] In SS, previous studies have revealed restricted TCR Vβ gene usage in T cells from peripheral blood, salivary glands and kidneys by immunostaining,[8-10] single-strand conformational polymorphism analysis,[11-13] or polymerase chain reaction (PCR) in combination with hybridization,[14-18] suggesting antigen-driven selection among T cells. Additionally, some studies have also found some conserved amino acid motifs in the TCRs of tissue-infiltrating T cells by bulk PCR tests[11,13,18] or by sequencing TCRs at the single-cell level.[19-21] However, the small number of samples in these studies limited their findings.

The TCR repertoire is sophisticated and shifts dynamically with disease status. It is difficult to comprehensively understand TCR repertoire features through single sequencing. Whether the TCR repertoire can be used as a biomarker in the diagnosis and prognosis of SS remains unknown. In this study, we demonstrated a quantitative and immunogenetic landscape of the TCR repertoire of peripheral T cells from a large population of SS patients followed longitudinally, and the classification of patients on the basis of TCR diversity could provide complementary information on immune dysregulation monitoring and the management of SS.

Methods

Patients and sample collection

Two hundred sixty SS patients who met the 2002 American-European Consensus Group (AECG) classification criteria[22] and 2016 American College of Rheumatology/European League Against Rheumatism Classification Criteria[23] were recruited from January 2017 to December 2020. Patients were ineligible to participate in this study if any of the exclusion criteria were met: 1) those with a history of infection (virus, bacteria, or fungi), cancer (e.g., lymphoma, leukaemia), or metabolic diseases (e.g., diabetes mellitus, hyperthyroidism); 2) patients diagnosed with other autoimmune disorders (e.g., RA, SLE, and myositis); 3) patients who received steroids, conventional or biological disease-modifying

antirheumatic drugs in the last 3 months; 4) those who received vaccines within the last 6 months; 5) those who had surgery or trauma in the previous 6 months; and 5) patients who were pregnant or lactating. The patient information, including age, sex, disease course, number of relapses, and symptoms, including dry eyes, dry mouth, dental caries, and arthralgia were recorded. The disease activity of SS patients was evaluated using the SS Disease Activity Index (SSDAI) scoring system.[24] Laboratory indices including platelet (PLT) counts, serum levels of alkaline phosphatase (ALP), gamma-glutamyl transpeptidase (GGT), immunoglobin (Ig) A, G, and M, presence of rheumatoid factor (RF), antinuclear antibody (ANA), anti-SSA antibody, and anti-SSB antibody, were also measured. The characteristics of all baseline samples are summarized in Supplementary Table S1. Among them, 101 had a follow-up 3 months after the start of the study, and 138 had a follow-up 6 months after the start of the study (Supplementary Figure 1). Patients who did not have follow-up information and samples were excluded. All patients received conventional treatment, including glucocorticoids, cytotoxic immunosuppressants (IS), hydroxychloroquine (HCQ) and total glucosides of paeony (TGP), as recommended by rheumatologists according to the guidelines.[25–27]

TCR sequencing

Genomic DNA was extracted using a Blood DNA kit (DP319, TIANGEN, Beijing) from the peripheral blood according to the manufacturer's instructions, and sequenced using a BGISEQ-500 platform (BGI, Shenzhen) as described previously.[28] In brief, the genomic DNA was fragmented. Multiplex polymerase chain reaction (PCR) was performed to amplify sequences of the β chain CDR3 region of TCR.[29] Then, the products were sequenced following the BGISEQ-500 protocol. Resulting raw sequence data were uploaded to the publicly available Population Health Data Archive (https://www.ncmi.cn/) under the accession number of 2016YFC0906200. The TCR sequencing data of 439 healthy controls (HCs) (28.9±6.1 years old, female 50.6%) were obtained from the pan immune repertoire database (PIRD) under the project ID P18080801 (https://db.cngb.org/pird/project/P18080801/).[6] Raw reads were processed by MiXCR[30] to align to the VJ genes and assemble the complementarity-determining region (CDR3) amino acid (CDR3aa) sequences. Downstream analyses were performed by VDJtools.[31]

TCR diversity and homeostatic space analysis

Diversity tests of inverse Simpson, d50 and normalized Shannon-Wiener were performed on resampled clonotypes to the size of the smallest sample reads. For pairwise distance evaluation, Pearson correlations of overlapping clonotype frequencies, Jensen-Shannon divergence (JSD) of V gene usage distributions (vJSD) and VJ pair usage distributions (vjJSD) were calculated. Clonotype variation was explored by principal component analysis (PCA) on V segment frequency, J segment frequency, and V-J segment pair frequency in each sample. The homeostatic space of the TCR repertoire was defined by clonotypes hyper-, large-, medium-, small- and rare-expanded, of which the proportions were above 1%, 0.1%, 0.01%, 0.001%, and 0.0001%, respectively.

Identification of disease-associated clonotypes

Disease-associated clonotypes were identified by comparing the clonotypes of SS patients with HCs as described in a previous study.[6] In brief, frequencies of clonotypes were compared between SS and HCs by Fisher's exact test. A clonotype was regarded as disease-associated if its adjusted Fisher's exact test was significant ($P<0.05$) and the relative risk was above 1. Clonotypes with the same nucleotide sequence in CDR3 were considered the same clonotype, regardless of their V/J genes. Public clonotypes were defined as the same clones present in ≥2 samples. The Levenshtein distance between CDR3aa sequences was calculated excluding the first and last three amino acids. Two CDR3aa sequences were connected if the Levenshtein distance was less than 3.

Changes in the TCR diversity during follow-up

The TCR diversity change rate was calculated as the averaged follow-up TCR diversity divided by the baseline TCR diversity. A sample was considered to be treatment sensitive if it was in the top rank (2/3 of the number of samples having diversity change rate above 1) ranked by an increase in both the inverse Simpson and normalized Shannon-Wiener change rates.

Development of the classification model using TCR clones

For model development, all CDR3aa sequence frequencies were transformed by the following formula:

$$\log_2(1000 * f + 1)$$

f means frequency

Recursive feature elimination with stratified 2-fold cross-validation was used, and random forest with max_depth=30 was built based on the selected features. Leave-one-out cross-validation was performed to assess the performance of the model, and the probability values were used to calculate the sensitivity and specificity of the model and draw ROC plots. To reduce the confounding effects of age and gender, multiple balancing approaches were applied. We were unable to get the two

groups matched in terms of both gender and age with the R package "MatchIt" or "optmatch".[32,33] However, the gender ratio and age were well balanced between SS patients and controls by a staged stratified sampling process. In detail, controls were randomly selected to make sure the female-to-male ratio equals to that of SS patients. After that, we compared ages between the two groups in all age brackets that achieved a P value >0.05 and selected the one that included the maximal number of patients and controls.

Statistics
Nonparametric groupwise comparisons were performed by the Mann-Whitney test. For V/J gene comparisons, the P value of multiple tests was corrected by step down method using Sidak adjustments. The association between discrete data was tested by linear regression, for categorical data (two categories), Fisher's exact test was used. $P<0.05$ was considered significant. Linear regression with covariates adjusted was calculated by the R package *limma*.[34] All of the other calculations were conducted with Python 3.7.8.

Ethics
This study was approved by the Ethics committees and institutional review board of West China Hospital, Sichuan University (2021-1494) and was registered on the Chinese Clinical Trial Registry (ChiCTR, http://www.chictr.org.cn) with the registration number of ChiCTR2200056100. All participants signed written informed consent forms.

Role of funders
The funders provided financial support for this study and were not involved in the study design, data collection, data interpretation, or writing of the manuscript.

Results

General characteristics of TCR in patients with SS
Overall, 29.1, 27.5, and 26.5 thousand unique TCR clones were identified in SS at baseline, the 1st follow-up, and the 2nd follow-up, respectively (Supplementary Table S2). Compared to HCs, SS patients had higher homogeneity in TCR (Pearson $P = 1.58 \times 10^{-59}$; vJSD $P < 0.0001$; vjJSD $P < 0.0001$; Mann-Whitney test)
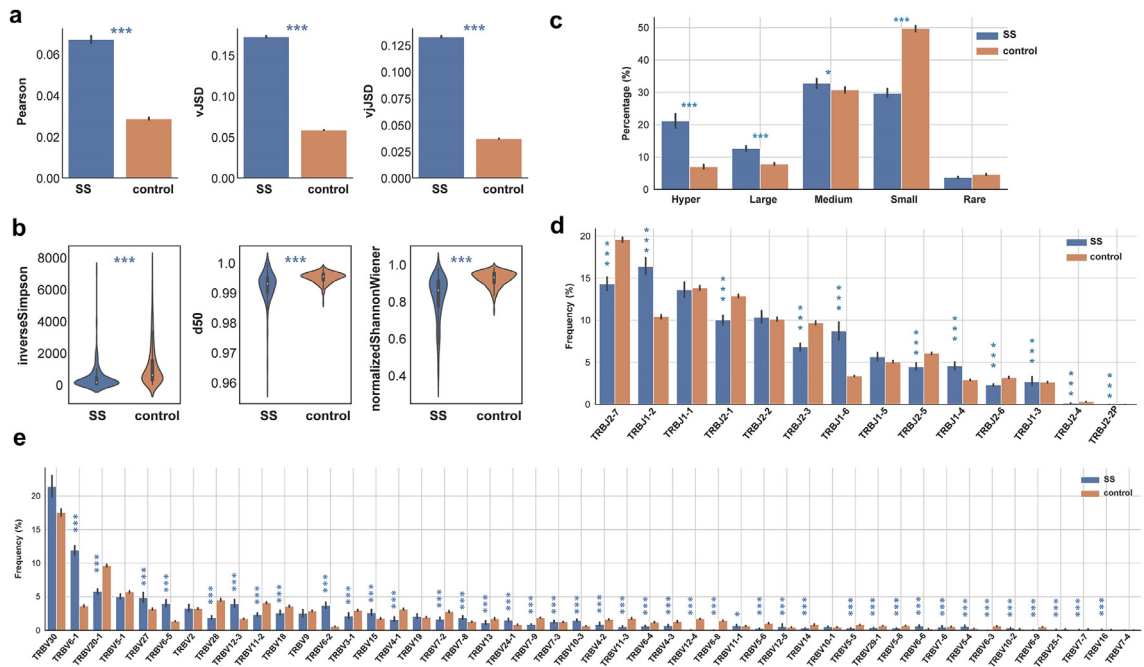

Figure 1. General characteristics of the TCRβ repertoires in SS patients at baseline and HCs. Comparison of **a**) homogeneity and **b**) diversity of TCR repertoire between SS (at baseline, n=260) and HCs (n=439), with the indicated P value adjusted by age and gender; **c**) Homeostatic space of TCR shows that the proportions of hyperexpanded (>1%), large-expanded (>0.1%) and medium-expanded (>0.01%) clonotypes in SS were significantly higher than in controls, while the small-expanded (>0.001%) and rare-expanded (>0.0001%) clonotypes were significantly lower than that in controls. Indicated p value was adjusted by age and sex. The frequencies of **d**) J gene and **e**) V gene usage in SS and HCs. pearson: Pearson correlation of overlapping clonotype frequencies; vJSD: Jensen-Shannon divergence (JSD) of V gene usage distributions; vjJSD: Jensen-Shannon divergence (JSD) of VJ pair usage distributions. *$P<0.05$, **$P<0.01$, ***$P<0.001$. a,d and e, Mann-Whitney test; b and c, linear regression.

(Figure 1a, Supplementary Figure 2) and significantly lower TCR diversity (inverse Simpson $P = 7.54 \times 10^{-5}$; d50 $P = 9.37 \times 10^{-7}$; normalized Shannon-Wiener $P = 7.69 \times 10^{-13}$; linear regression adjusted for age and sex) (Figure 1b). The proportions of hyperexpanded ($P = 6.64 \times 10^{-10}$; linear regression with age and sex-adjusted), large-expanded ($P = 1.05 \times 10^{-9}$) and medium-expanded ($P = 1.92 \times 10^{-2}$) clonotypes in SS were significantly higher than those in the control, while the small-expanded ($P = 2.81 \times 10^{-31}$) clonotypes were significantly lower than that those in the control, suggesting that SS had significant clonal expansion (Figure 1c, Supplementary Table S3). We identified 48 functional TRBV subtypes and 14 functional TRBJ subtypes by referring to the IMGT (international ImMunoGeneTics information system). In terms of V/J gene usage, SS had broadly distinct frequencies of V/J gene usage compared to the control (Figure 1d, e). The usage of TRBV6-1, TRBV6-2, TRBV6-5, TRBV27, and TRBV10-3 was more frequent, while TRBV20-1, TRBV28, TRBV11-2, TRBV18, TRBV3-1, TRBV4-1, and TRBV7-2 usage was less frequent in SS patients than HCs (corrected $P < 0.05$; Mann-Whitney test; Sidak correction) (Figure 1e, Supplementary Table S4). TRBJ1-2, TRBJ1-6, and TRBJ1-4 were more frequently used in SS, while all TRBJ2 genes were more frequently used in HCs (corrected $P<0.05$) (Figure 1d, Supplementary Table S4).

Public clones were more frequent in SS

We found that limited public clonotypes were shared between the SS and control groups among all clonotypes or public clonotypes (Figure 2a). Furthermore, public clonotypes were more frequent in SS than in the control, and the frequency of clonotypes of more than 10% publicity ($P = 0.0082$; Mann-Whitney test) was higher in SS (Figure 2b, c), indicating that SS T cells were activated by certain antigens.

Identification of disease-associated TCR clonotypes in SS patients

Eighteen SS-associated clonotypes ("SS-associated") were identified, present in 7.7% to 43.5% of patients (Figure 3a, Supplementary Table S5). Among them, 5 SS-associated clonotypes were uniquely present in SS patients. However, none of the SS-associated clones was associated with the presence of anti-SSA or anti-SSB in serum (Figure 3a). To further understand the SS-associated clones, another 18 public clonotypes were randomly chosen 10 times, representing the rest of the clones ("SS-other"). We observed that SS-associated clonotypes had a higher frequency among patients than SS-other clonotypes ($P = 3.13 \times 10^{-7}$; Mann−Whitney test), but with no significant difference in length (Figure 3b, c). No significant difference in sequence

consensus was found between SS-associated and SS-other clonotypes (Supplementary Figure 3).

Based on the 18 SS-associated clones, the random forest model generated a high classification accuracy with an area under the curve (AUC) of 94.55% with a sensitivity of 86.15% and specificity of 92.48% (Figure 3d), suggesting that SS-associated clones have good performance in the diagnosis of SS. We re-tested the model by selecting SS patients ($n = 71$) at comparable ages and gender with controls ($n = 119$, $P > 0.05$; Mann-Whitney test), and the model still achieved high accuracy (AUC = 92.64%). Moreover, we compared the 18 SS-associated clones and hyperexpanded clones (>1%) in SS patients with common TCR motifs identified in prior studies[11,13,18,19] but found no overlaps (data not shown). In comparison with disease-associated clonotypes of other autoimmune diseases including SLE and RA,[6] we identified one clonotype shared among SS, SLE, and RA and three clonotypes shared between SS and SLE (Figure 3e, Supplementary Table S6). This finding supports the existence of some common autoreactive T subpopulations in different autoimmune disorders, which could be a potential therapeutic target.

Clinical association analysis of TCR features

To understand the association of TCR diversity and clone features with clinical characteristics, we performed a series of correlation analyses. Notably, dental caries, low PLT, high ALP/GGT, positive ANA, anti-SSA/SSB, and hypergammaglobulinemia were correlated with TCR features, while disease course, number of relapses, presence of arthritis, RF, low C3/C4, and SSDAI were not (Figure 4a, Supplementary Table S7). In detail, SS patients with dental caries ($P < 0.05$; Mann-Whitney test), low PLT ($P = 0.0213$; Mann-Whitney test), high ALP/GGT ($P = 0.0243$; Mann-Whitney test), elevated IgA/M/G ($P = 0.0457$; Mann-Whitney test) and ANA ($P = 0.0173$; Mann-Whitney test) had lower TCR diversity, while SS patients with positive anti-SSA/SSB had relatively higher diversity (Figure 4b), indicating that TCR diversity could be an indicator of immune-mediated organ damage in SS.

Changes in the TCR repertoire during treatment for SS

We dynamically sequenced the TCR of SS patients receiving conventional treatment for 3-6 months. TCR homogeneity decreased during the follow-up (Pearson 3 m $P = 1.89 \times 10^{-19}$, 6 m $P = 1.20 \times 10^{-23}$; vJSD $P < 0.0001$; vjJSD $P<0.0001$; Mann-Whitney test) (Figure 5a), although no significant difference in diversity was found between the baseline and follow-ups (Figure 5b). For V gene usage, the frequency of TRBV30 decreased at the 3- and 6-month follow-ups, while some of the more frequently used V genes in SS at baseline, such as TRBV6-1, TRBV6-5, and TRBV27,
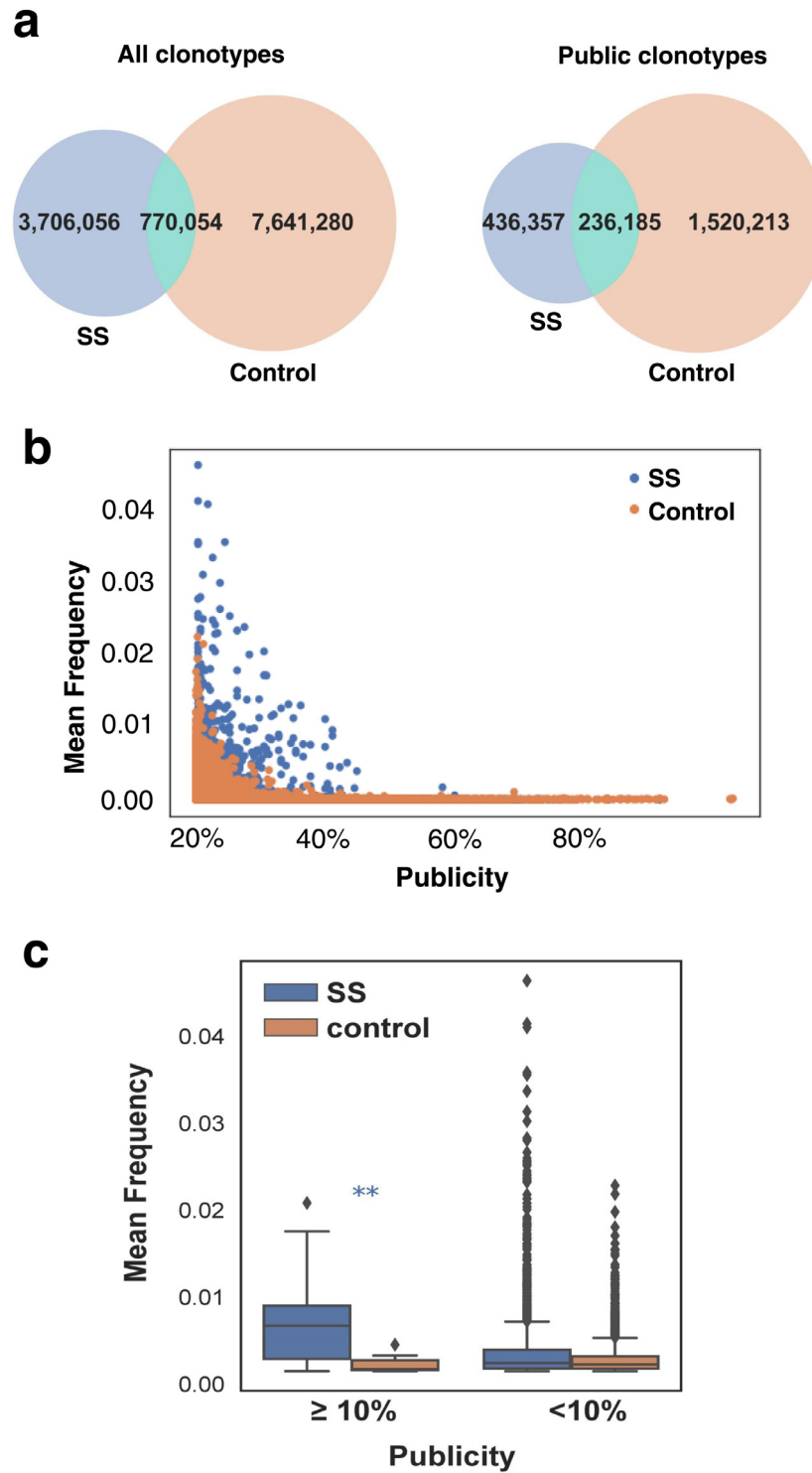
Figure 2. Public clonotypes in the TCR β repertoire of patients with SS. a) Amount of overlap of all unique clones and public clones between SS (*n* = 260) and HCs (*n* = 439). **b**) Mean frequency of clones in SS patients and HCs according to publicity. **c**) Mean frequency of clones presented in ≥10% samples and in <10% samples. **P < 0.01; Mann-Whitney test.

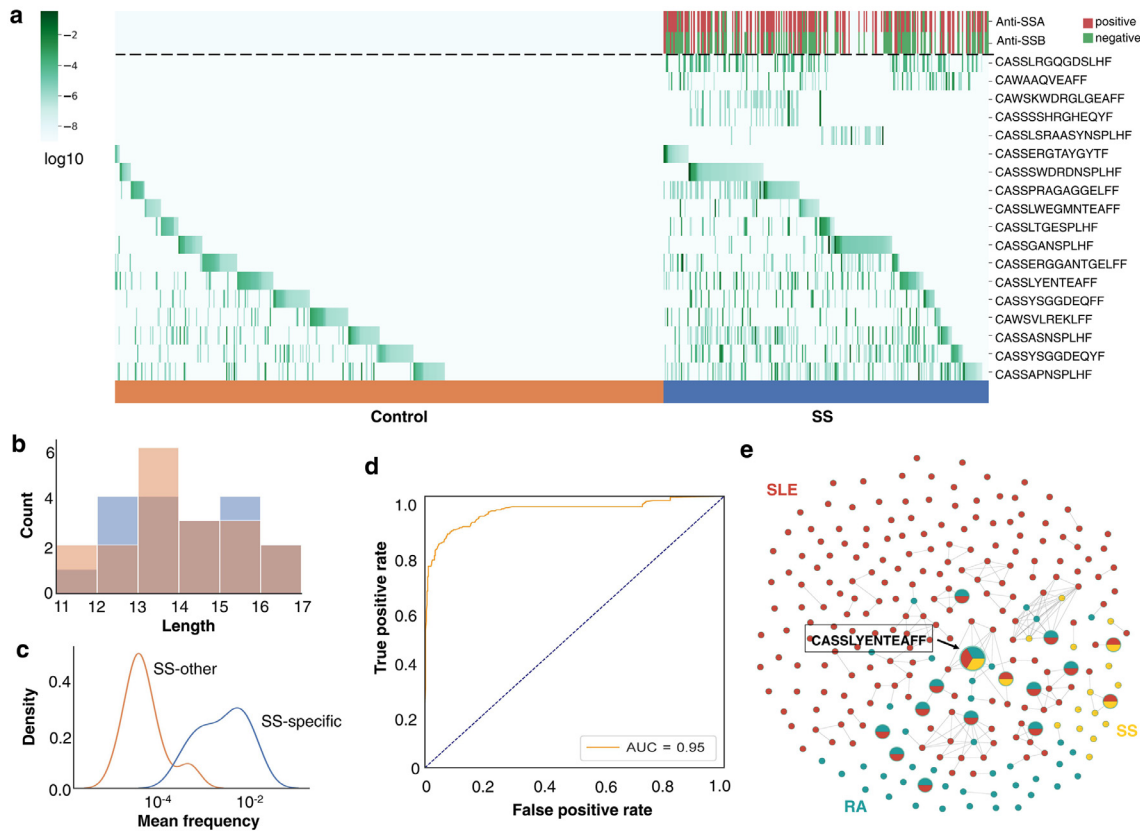**Figure 3. Identification of SS-associated TCR β clonotypes. a)** The frequency and distribution of SS-associated clones in SS patients at baseline (*n* = 260) and in controls (*n* = 439), as well as the presence of anti-SSA and anti-SSB antibodies in SS patients. The cell colour indicates the frequency of SS-associated clones, which was transformed to log10. **b)** CDR3 amino acid length distribution. **c)** Clone frequency distribution. **d)** ROC curve showing the classification performance of a classifier for two groups by leave-one-out cross-validation. **e)** Graphs of disease-associated clonotypes in SS, SLE, and RA. Vertices represent significant clonotypes identified by the algorithm, and edges connect clonotypes with at most one amino acid mismatch. Zero-degree vertices are not shown. Vertices are coloured according to the disease, and split vertices represent public sequences identified in several diseases. HC: health control; SS: Sjogren's syndrome; SLE: systematic lupus erythematosus; RA: rheumatoid arthritis; ROC: receiver operating characteristic.
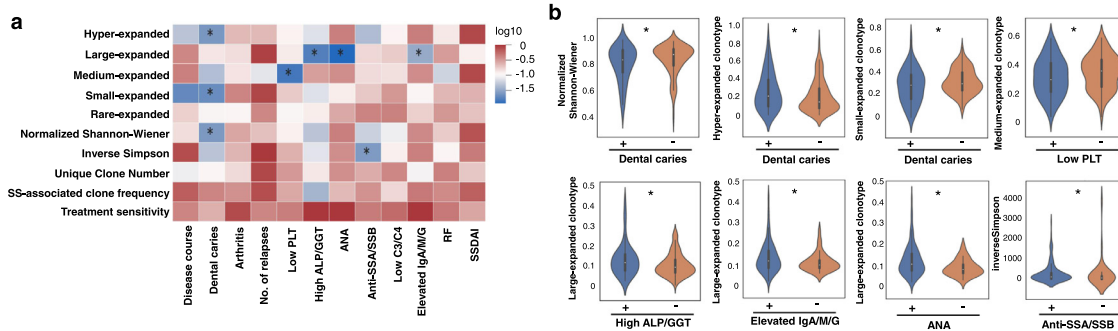

Figure 4. Clinical association analysis of TCRβ clones in SS. a) The correlation of TCR diversity indices and clinical information. **b)** Comparison of diversity indices in several clinical characteristics that have a correlation with TCR features shown in **a**). PLT: platelet; ALP: alkaline phosphatase; GGT: gamma-glutamyl transpeptidase; C3: complement 3; Ig: immunoglobin; ANA: antinuclear antibody; RF, rheumatoid factor; SSDAI: Sjogren's syndrome disease activity index; *P < 0.05, **P < 0.01. Mann-Whitney test.
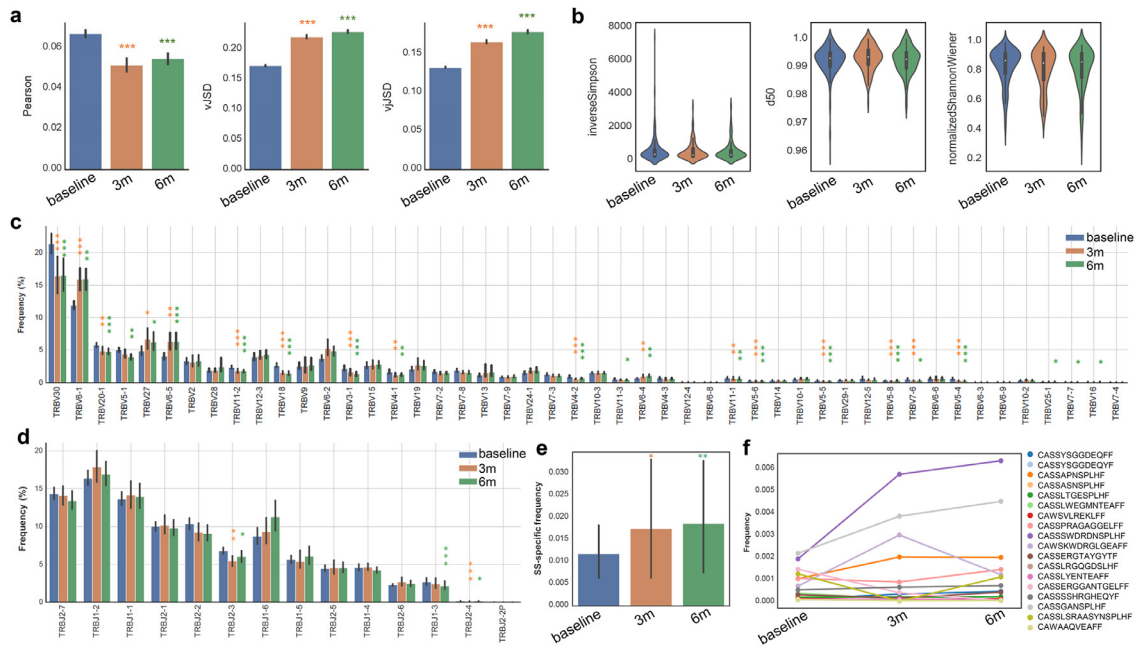
Figure 5. Changes in TCR diversity and V/J gene usage in SS with follow-ups. Comparison of **a**) homogeneity and **b**) diversity of the TCR repertoire for SS at baseline ($n = 260$), 3-month follow-up ($n=101$), and 6-month follow-up ($n = 138$). The frequencies of **c**) V gene and **d**) J gene usage in SS at baseline and at the 3-6-month follow-up. **e**) Changes in the frequencies of eighteen SS-associated clones during follow-up. **f**) Frequencies of each SS-associated clone. A dot of the same color represents the same amino acid sequence, and the connecting line indicates a varying trend. Average frequencies are shown. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. Mann-Whitney test.

increased at the follow-ups. Other less frequent V genes at baseline, such as TRBV20-1, TRBV11-2, TRBV18, TRBV3-1, and TRBV4-1, decreased at the 3- ($P = 0.0424$; Mann-Whitney test; Sidak correction) and 6-month ($P = 0.0033$) follow-ups (Figure 5c, Supplementary Table S4). Similarly, the TRBJ2-3 gene was less frequently used at follow-ups than at baseline (Figure 5d, Supplementary Table S4). Moreover, we found that the frequency of SS-associated clones increased at the 3- and 6-month follow-ups compared to baseline (Figure 5e; $P = 0.0424$ for 3 m vs baseline, $P = 0.0033$ for 6 m vs baseline, Mann-Whitney test). Among them, the frequencies of "CASSSWDRDNSPLHF", "CASSGANSPLHF", and "CASSAPNSPLHF" rose, while the frequency of "CASSERGGANTGELFF" declined at the follow-ups (Figure 5f). These findings indicate that some autoreactive T cells may continue to expand during conventional treatment.

TCR features indicate the response to treatment

High TCR diversity is an indicator of a "healthy" T-cell immune system. We compared TCR diversity changes in SS patients before and after treatment and divided them into "sensitive" and "nonsensitive" groups according to whether their TCR diversity was elevated. Under the condition of comparable medication (Figure 6c), 31

patients had increased TCR diversity after treatment and were considered "sensitive" (normalized Shannon-Wiener $P = 0.0030$; Mann-Whitney test) (Figure 6a, b, c, Supplementary Figure 4). As an example, we observed that the major clonotypes decreased in two sensitive patients (A56, A177), whereas in two nonsensitive patients (A27, A155), the major clonotypes were similar after treatment (Supplementary Figure 5). We found a decreased percentage of hyperexpanded clonotypes ($P = 0.0058$; Mann-Whitney test) and increased medium-expanded clonotypes ($P = 0.0037$) after treatment in the sensitive group but not in nonsensitive group (Figure 6d, Supplementary Table S3). No significant difference was found in V/J gene usage between them, except in the usage of TRBV10-2 (used at a low frequency) (Supplementary Figure 6, Supplementary Table S4). Furthermore, eighty-five CDR3aa sequences were selected as features for a random forest model, which achieved a high AUC of 86.91% with a sensitivity of 80.53% and specificity of 80.64% (Figure 6e, f), suggesting that CDR3 sequences are associated with the prognosis of SS patients.

Discussion

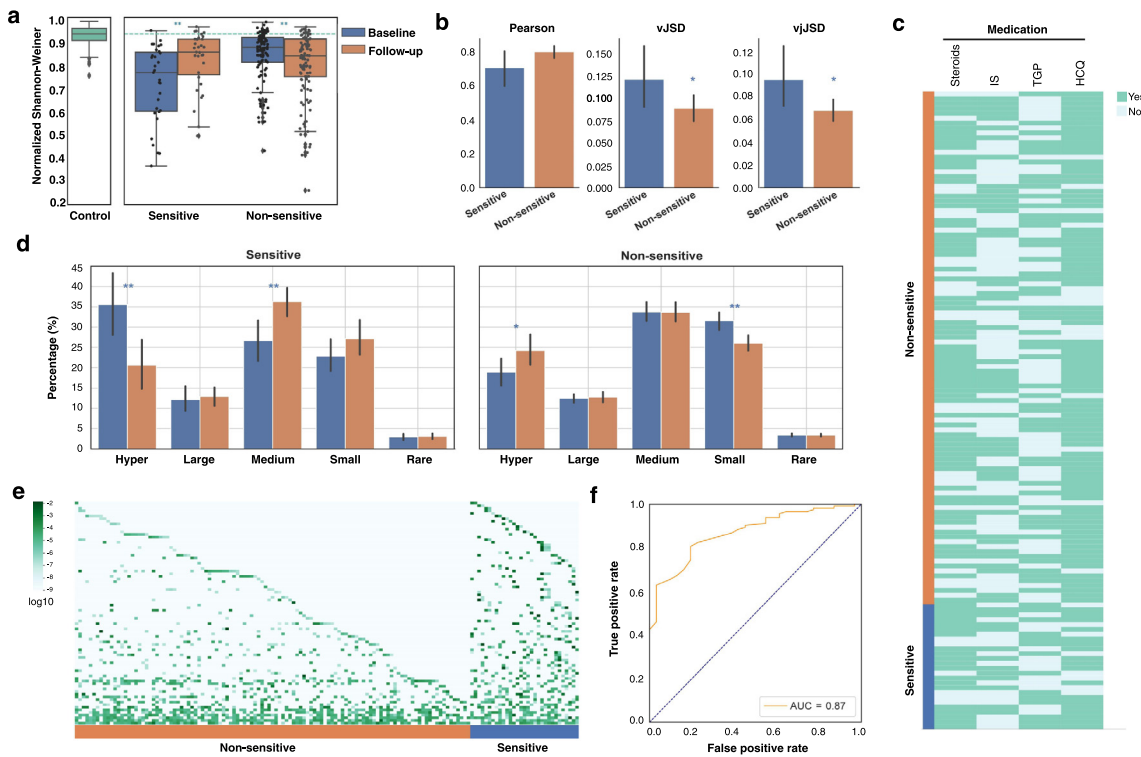In the present study, we performed a comprehensive analysis of the TCR repertoire in SS and described the

Figure 6. Treatment sensitivity-associated TCR clones in SS. a) Change in TCR diversity and **b**) homogeneity of treatment-sensitive patients ($n = 31$) and nonsensitive patients ($n = 113$). **c**) Treatment regimens of SS patients according to sensitivity. **d**) Homeostatic space of the TCR in treatment-sensitive patients and nonsensitive patients. **e**) Distribution of 85 sensitive related TCR clones in nonsensitive and sensitive patients. **f**) ROC curve showing the classification performance of a classifier for sensitive/nonsensitive groups by leave-one-out cross-validation. IS: immunosuppressants; TGP: total glucosides of paeony; HCQ: hydroxychloroquine; ROC: receiver operating characteristic. *$P < 0.05$.

role of the TCR in the evaluation of disease status and prognosis of SS. Although only Chinese patients from one center were included, we believe this study contains the largest cohort of SS patients sequenced for its TCR repertoire to date.

TCR repertoire is a potential diagnostic biomarker for immune-mediated diseases, where antigens may drive the selection of TCR clones. We found that patients with SS have lower TCR diversity, but a higher frequency of public clones than HCs, and they have broadly distinct V/J gene usage from HCs, indicating autoantigen-driven T-cell selection in SS. Similar to the study of the TCR repertoire in RA and SLE, where 198 SLE-associated and 53 RA-associated TCR clones can discriminate these autoimmune diseases from controls and between themselves,[6] we identified eighteen SS-associated clones that can be used to classify SS from HCs with high accuracy. Furthermore, TCR repertoire is also a prognostic marker, reflecting the disease status in the human body. For instance, TCR clones correlated with disease activity, organ involvement, and some autoantibodies in both SLE and RA.[6] The extent of T-cell clonal expansion in salivary glands of SS patients, indicated by TCR clones, also correlated with reduced saliva

production and increased fibrosis of salivary glands.[19,20] Consistent with this, we found that the TCR clones correlated with many clinical features of SS (Figure 4a).

T cells constitute the majority of lymphocytes that infiltrate the salivary glands in SS patients and therefore contribute to disease pathogenesis via their ability to cause tissue damage, B-cell activation and metabolic changes.[35,36] The distribution of infiltrating immune cells correlates with lesion severity and disease manifestations in SS.[37] Here, we found that some clinical characteristics of SS were associated with TCR features. Organ dysfunction represented by dental caries, thrombocytopenia, and liver dysfunction can be reflected by TCR diversity. Hypergammaglobulinemia is one of the most common serological findings in SS, reflecting the hyperactivity characteristic of the disease.[38] In the present study, patients with hypergammaglobulinemia had lower TCR diversity, indicating that enhanced antigen-driven selection of T cells and B cells occurs in these patients. Anti-SSA/SSB are hallmark antibodies of SS and are reported to be correlated with an earlier disease onset, more intense lymphocytic infiltrates in salivary glands, and more severe dysfunction of the exocrine glands.[38,39] However, we found no association between

the presence of anti-SSA or anti-SSB in serum and SS-associated clones. This observation suggests that many of the expanded T cells may not recognize the prototypic Ro and La antigens, which is in line with the findings of T cells in salivary glands.[19] Moreover, higher TCR diversity was observed in patients positive for anti-SSA/SSB. The exact mechanism remains unknown. The potential of TCR repertoire in the evaluation of disease status warrants further study.

Diverse TCRs indicate a homeostatic immune system armed with abundant naïve T cells.[40] Dynamic change in TCR repertoire is a potential parameter for monitoring treatment responses since the immune status in patients with infection and cancer may affect treatment effectiveness.[41,42] We sequenced TCR in SS during the 3-6 month follow-up and unexpectedly found that the TCR abnormalities persisted. For V/J usage and SS-associated clones, posttreatment patients had even more obvious alterations. T-cell dysregulation in SS was not reversed by conventional treatment with steroids and/or DMARDs in the short term. This may partially explain why SS is considered a chronic disease with no disease modifying or curative treatments available to date and people can accumulate morbidity over time.[25,27,43]

Among all SS patients, we found that a subgroup of SS patients had increased TCR diversity after treatment. Treatment-sensitive patients had a higher prevalence of specific sensitivity-associated TCR clones, and these clones had good performance in classification. In treatment-sensitive patients, expanded autoreactive T cells may be diminished by treatments, regardless of the regimens. TCR diversity offers a possible stratification score for responsiveness to therapy, but this requires further validation in a larger cohort.

It is believed that T cells capable of recognizing salivary gland antigens may be present in the peripheral blood of SS patients, since identical TCR clonotypes can be found in T cells from peripheral blood and from salivary glands from the same subject.[19] However, we did not observe the presence of common TCR motifs that are defined as SS-associated clones by others[11,13,18,19] in the peripheral blood of patients from our cohort. One possible reason is that different human leukocyte antigen (HLA) susceptibility haplotypes exist in different races, such as those observed in Caucasoid (DRB1*0301-DRB3*0101-DQA1*0501-DQB1*0201), Japanese (DRB1*0405-DRB4*0101-DQA1*0301-DQB1*0401), and Chinese patients (DRB1*0803-DQA1*0103-DQB1*0601).[44,45] Alternatively, multiple antigens are recognized in the same disease and/or different TCR rearrangements respond to the same antigen in the body.[46] Few clones were found to be shared among different SS patients in the limited public data.[11,13,18,19] Moreover, previous studies focused on the expanded T-cell clones in inflamed tissues of SS patients, while we provided the TCR repertoire in the peripheral blood. The relationship between expanded T-cell clones in the peripheral blood and in inflamed tissues is poorly understood, and its translation to SS pathogenesis remains unknown. In contrast, we found some overlapping clonotypes in peripheral T cells among three autoimmune diseases namely, RA, SLE, and SS. The underlying mechanism of the co-occurrence of some autoimmune diseases, including the three mentioned in patients, is not known. These shared TCR clones support the existence of common autoreactive T cells in autoimmune diseases, potentially providing insight into the pathogenesis of the co-occurrence of two or more autoimmune disorders.

There are some limitations to this study. First, all samples used in this study were from peripheral blood but not from salivary glands. Therefore, further verification should be conducted with the targeted organs of SS since different patterns are found in T cells between lesions and blood. Second, some clinical and molecular traits including disease activity and focal score were inadequate in the cohort, which may affect the significance of the correlation of TCR repertoire and some traits. The association of TCR clonotypes and inflammatory infiltrates in the salivary gland warrants further study. Third, the TCR sequencing data of HCs were generated from a different sequencing platform, and this may introduce a possible bias in the following analyses, although studies have proven that data generated by these two platforms are mostly comparable.[47,48] Last, this study only included patients from one region and unfortunately found no shared TCRs with previous studies.

In conclusion, patients with SS show lower TCR diversity and different V/J gene usage than HCs. Disease-associated clones can be used to assist in the precise diagnosis of SS. TCR repertoire correlated with several clinical parameters, and changes in the TCR repertoire to some extent reflected patients' response to treatment. Overall, TCR repertoire could be an optional tool for evaluating disease status, and predicting of treatment response in autoimmune diseases such as SS. Investigating the dynamics of TCR repertoire could provide a valuable tool to monitor T-cell driven immune responses in autoimmune diseases.

References

1   Krogsgaard M, Davis MM. How T cells 'see' antigen. *Nat Immunol*. 2005;6(3):239–245.
2   Yagüe J, White J, Coleclough C, Kappler J, Palmer E, Marrack P. The T cell receptor: the alpha and beta chains define idiotype, and antigen and MHC specificity. *Cell*. 1985;42(1):81–87.
3   Woodsworth DJ, Castellarin M, Holt RA. Sequence analysis of T-cell repertoires in health and disease. *Genome Med*. 2013;5(10):98.
4   Mariette X, Criswell LA. Primary Sjogren's syndrome. *N Engl J Med*. 2018;378(10):931–939.
5   Yao Y, Ma JF, Chang C, et al. Immunobiology of T Cells in Sjögren's syndrome. *Clin Rev Allergy Immunol*. 2021;60(1):111–131.
6   Liu X, Zhang W, Zhao M, et al. T cell receptor β repertoires as novel diagnostic markers for systemic lupus erythematosus and rheumatoid arthritis. *Ann Rheum Dis*. 2019;78(8):1070–1078.
7   Mitchell AM, Michels AW. T cell receptor sequencing in autoimmunity. *J Life Sci (Westlake Village)*. 2020;2(4):38–58.
8   Smith MD, Lamour A, Boylston A, et al. Selective expression of V beta families by T cells in the blood and salivary gland infiltrate of patients with primary Sjögren's syndrome. *J Rheumatol*. 1994;21 (10):1832–1837.
9   Kay RA, Hay EM, Dyer PA, et al. An abnormal T cell repertoire in hypergammaglobulinaemic primary Sjögren's syndrome. *Clin Exp Immunol*. 1991;85(2):262–264.
10 Mizushima N, Kohsaka H, Tsubota K, Saito I, Miyasaka N. Diverse T cell receptor beta gene usage by infiltrating T cells in the lacrimal glands of Sjögren's syndrome. *Clin Exp Immunol*. 1995;101(1):33–38.
11 Matsumoto I, Tsubota K, Satake Y, et al. Common T cell receptor clonotype in lacrimal glands and labial salivary glands from patients with Sjögren's syndrome. *J Clin Invest*. 1996;97(8):1969–1977.
12 Sasaki M, Nakamura S, Ohyama Y, et al. Accumulation of common T cell clonotypes in the salivary glands of patients with human T lymphotropic virus type I-associated and idiopathic Sjögren's syndrome. *J Immunol*. 2000;164(5):2823–2831.
13 Sumida T, Matsumoto I, Murata H, et al. TCR in Fas-sensitive T cells from labial salivary glands of patients with Sjögren's syndrome. *J Immunol*. 1997;158(2):1020–1025.
14 Ohyama Y, Nakamura S, Matsuzaki G, et al. T-cell receptor V alpha and V beta gene use by infiltrating T cells in labial glands of patients with Sjögren's syndrome. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod*. 1995;79(6):730–737.
15 Sumida T, Yonaha F, Maeda T, et al. T cell receptor repertoire of infiltrating T cells in lips of Sjögren's syndrome patients. *J Clin Invest*. 1992;89(2):681–685.
16 Ajjan RA, McIntosh RS, Waterman EA, et al. Analysis of the T-cell receptor Valpha repertoire and cytokine gene expression in Sjögren's syndrome. *Br J Rheumatol*. 1998;37(2):179–185.
17 Sumida T, Kita Y, Yonaha F, Maeda T, Iwamoto I. Yoshida S. T cell receptor V alpha repertoire of infiltrating T cells in labial salivary glands from patients with Sjögren's syndrome. *J Rheumatol*. 1994;21(9):1655–1661.
18 Murata H, Kita Y, Sakamoto A, et al. Limited TCR repertoire of infiltrating T cells in the kidneys of Sjögren's syndrome patients with interstitial nephritis. *J Immunol*. 1995;155(8):4084–4089.
19 Joachims ML, Leehan KM, Lawrence C, et al. Single-cell analysis of glandular T cell receptors in Sjögren's syndrome. *JCI Insight*. 2016;1(8):e85609.
20 Voigt A, Bohn K, Sukumaran S, Stewart CM, Bhattacharya I, Nguyen CQ. Unique glandular ex-vivo Th1 and Th17 receptor motifs in Sjogren's syndrome patients using single-cell analysis. *Clin Immunol*. 2018;192:58–67.
21 Matsumoto I, Okada S, Kuroda K, et al. Single cell analysis of T cells infiltrating labial salivary glands from patients with Sjögren's syndrome. *Int J Mol Med*. 1999;4(5):519–527.
22 Vitali C, Bombardieri S, Jonsson R, et al. Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis*. 2002;61(6):554–558.
23 Shiboski CH, Shiboski SC, Seror R, et al. 2016 American college of rheumatology/European league against rheumatism classification criteria for primary Sjögren's syndrome: a consensus and data-driven methodology involving three international patient cohorts. *Arthritis Rheumatol*. 2017;69(1):35–45.
24 Vitali C, Palombi G, Baldini C, et al. Sjögren's Syndrome Disease Damage Index and disease activity index: scoring systems for the assessment of disease damage and disease activity in Sjögren's syndrome, derived from an analysis of a cohort of Italian patients. *Arthritis Rheum*. 2007;56(7):2223–2231.
25 Vivino FB, Carsons SE, Foulks G, et al. New Treatment Guidelines for Sjögren's disease. *Rheum Dis Clin North Am*. 2016;42(3):531–551.
26 Ramos-Casals M, Brito-Zerón P, Bombardieri S, et al. EULAR recommendations for the management of Sjögren's syndrome with topical and systemic therapies. *Ann Rheum Dis*. 2020;79 (1):3–18.
27 Zhang W, Li X, Xu D, et al. Diagnosis and treamtent normas of primary Sjogren's syndrome. *Zhonghua Nei Ke Za Zhi*. 2020;59 (4):270–275.
28 Huang J, Liang X, Xuan Y, et al. A reference human genome dataset of the BGISEQ-500 sequencer. *Gigascience*. 2017;6(5):1–9.
29 Liu X, Zhang W, Zeng X, et al. Systematic comparative evaluation of methods for investigating the TCRbeta repertoire. *PLoS One*. 2016;11(3):e0152464.
30 Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods*. 2015;12 (5):380–381.
31 Shugay M, Bagaev DV, Turchaninova MA, et al. VDJtools: unifying post-analysis of T cell receptor repertoires. *PLoS Comput Biol*. 2015;11(11):e1004503.
32 HB B, Klopfer SO. Optimal full matching and related designs via network flows. *J Comput Graph Stat*. 2006;15(3):609–627.
33 Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric Preprocessing for Parametric Causal Inference. *J Stat Softw*. 2011;42(8):1–28.
34 Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.
35 Singh N, Cohen PL. The T cell in Sjogren's syndrome: force majeure, not spectateur. *J Autoimmun*. 2012;39(3):229–233.
36 Yao Y, Ma JF, Chang C, et al. Immunobiology of T Cells in Sjögren's syndrome. *Clin Rev Allergy Immunol*. 2021;60(1):111–131.

37 Christodoulou MI, Kapsogeorgou EK, Moutsopoulos HM. Characteristics of the minor salivary gland infiltrates in Sjögren's syndrome. *J Autoimmun.* 2010;34(4):400–407.

38 García-Carrasco M, Mendoza-Pinto C, Jiménez-Hernández C, Jiménez-Hernández M, Nava-Zavala A, Riebeling C. Serologic features of primary Sjögren's syndrome: clinical and prognostic correlation. *Int J Clin Rheumtol.* 2012;7(6):651–659.

39 Fayyaz A, Kurien BT, Scofield RH. Autoantibodies in Sjögren's syndrome. *Rheum Dis Clin North Am.* 2016;42(3):419–434.

40 Qi Q, Liu Y, Cheng Y, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci U S A.* 2014;111(36):13139−13144.

41 Chen YT, Hsu HC, Lee YS, et al. Longitudinal high-throughput sequencing of the T-cell receptor repertoire reveals dynamic change and prognostic significance of peripheral blood TCR diversity in metastatic colorectal cancer during chemotherapy. *Front Immunol.* 2022;12:743448.

42 Luo L, Liang W, Pang J, et al. Dynamics of TCR repertoire and T cell function in COVID-19 convalescent individuals. *Cell Discov.* 2021;7(1):89.

43 Price E, Allen A, Rauz S, et al. The management of Sjögren's syndrome: British society for rheumatology guideline scope. *Rheumatology (Oxford).* 2021;60(5):2122–2127.

44 Kang HI, Fei HM, Saito I, et al. Comparison of HLA class II genes in Caucasoid, Chinese, and Japanese patients with primary Sjögren's syndrome. *J Immunol.* 1993;150(8 Pt1):3615–3623.

45 Cruz-Tapias P, Rojas-Villarraga A, Maier-Moore S, Anaya JM. HLA and Sjogren's syndrome susceptibility. A meta-analysis of worldwide studies. *Autoimmun Rev.* 2012;11(4):281–287.

46 Klarenbeek PL, de Hair MJ, Doorenspleet ME, et al. Inflamed target tissue provides a specific niche for highly expanded T-cell clones in early human autoimmune disease. *Ann Rheum Dis.* 2012;71(6):1088–1093.

47 Mak SST, Gopalakrishnan S, Carøe C, et al. Comparative performance of the BGISEQ-500 vs Illumina HiSeq2500 sequencing platforms for palaeogenomic sequencing. *Gigascience.* 2017;6(8):1–13.

48 Zhu FY, Chen MX, Ye NH, et al. Comparative performance of the BGISEQ-500 and Illumina HiSeq4000 sequencing platforms for transcriptome analysis in plants. *Plant Methods.* 2018;14:69.