Database of synovial T cell repertoire of rheumatoid arthritis patients identifies cross-reactive potential against pathogens including unencountered SARS-CoV-2

Rheumatoid arthritis (RA) is a chronic autoimmune disease of complex aetiology in which adaptive immune responses mediated by T cells play a major role in driving synovial inflammation and joint damage.[1] Before developing effector functions, naïve T cells are conventionally understood to require an initial antigen-specific stimulus transduced through their T cell receptor (TCR). TCR-mediated recognition is also required for targeted cell lysis by cytotoxic T cells. Previous studies have sought to characterise the functional roles of various T cell populations in RA, and have demonstrated patterns of oligoclonal expansion among them.[2] It has also been shown that some T cell clones may be shared among multiple patients as a public repertoire.[3] However, the antigen specificities of synovial T cells remain poorly studied outside of a few epitopes, and open resources of TCRs and repertoires associated with RA are lacking.

To accelerate research efforts into the TCR repertoires and T cell antigen specificities in RA, we report here synovial repertoires mined using a repertoire reconstruction algorithm from two large-cohort RNA sequencing datasets totaling 196 RA patients, as well as additional controls from peripheral blood and other joint disorders (figure 1A, online supplemental methods). The first, GSE89408, had been generated to identify transcriptome signatures in the synovium of patients with RA associated with disease progression (online supplemental table S1).[4] The second, PEAC, focused on early, treatment-naïve patients and identify transcriptome signatures contributing to establishment of disease in synovial tissue and peripheral blood (online supplemental table S2).[5] By exploring these cohorts in the hitherto unnavigated immune repertoire dimension, we built the first openly accessible online database of high-frequency TCR sequences found in the synovium of patients with RA (http://repertoire.life). This reference database is an ongoing effort to collect and compile human TCR repertoires of patients with RA, and can be searched with TCR repertoires identified in future work to better identify public sequences and contextualise their antigen specificity.

Initial analysis of this database demonstrated that an increased repertoire diversity among synovial (but not peripheral blood) high-frequency clonotypes was associated with higher disease activity (Disease Activity Score-28, (DAS28)) and erythrocyte sedimentation rate (ESR) (figure 1B,C, online supplemental figure S1, online supplemental table S3), highlighting the significant correlations between synovial oligoclonal expansion and clinical indices. Through TCR clustering analyses, we observed that synovial TCR motifs from patients with RA were largely distinct from those found in other joint diseases (online supplemental figure S2), in a manner consistent with their overall transcriptome differences (online supplemental figure S3), but that these motifs could be shared in aggregate among patients with lymphoid pathotype (online supplemental figure S4). Consistently, we also observed that several of these motifs were clonally expanded and enriched in patients with highly active disease as assessed through multiple indices(figure 1D, online supplemental figures S5,S6), further indicating their potential involvement in RA progression. Through comparison with paired peripheral blood samples from the same patients, we observed that a portion of synovial clones were shared with peripheral blood independently of pathotype and disease activity levels (online supplemental figures S7,S8), indicating that clonal migration may generally occur in RA. Collectively, these results show that recovered synovial TCRs correlate with clinical measures of RA disease activity, and suggest that they may potentially originate from circulation.

Beyond observing associations between RA TCR repertoire and disease activity, we further sought to annotate the antigen specificities of these clonotypes using databases of validated antigen-specific sequences. Intriguingly, we observed that many of the clonotypes expanded in RA were highly similar to pathogen antigen-specific sequences (figure 1E, online supplemental figures S9–S11). These include commonly encountered pathogens (such as cytomegalovirus (CMV), Epstein-Barr virus (EBV), influenza), as previously described[6] (figure 1F). At the same time, some clones matched pathogens that these patients may have never encountered (such as dengue virus (DENV) and yellow fever virus (YFV)), even when absolute sequence matching criteria were used (online supplemental figure S12). More surprisingly, we observed a number of TCRs matching annotated sequences specific for multiple peptides originating from SARS-CoV-2 (figure 1G). These include three dominant epitopes shared among betacoronaviruses (online supplemental figure S13). Since these patients had been sequenced years prior to its emergence, these matches between T cell clones found in RA synovium and SARS-CoV-2 antigen-specific clones may potentially originate from cross-reactivity, possibly with uncharacterised viruses and/or autoantigens.

To confirm the presence of these SARS-CoV-2 antigen-specific matching sequences identified from recovered repertoires, we then performed additional analyses using TCRseq, which has higher sensitivity in capturing rarer clones (such as naïve cells). When we searched a pre-pandemic TCRseq dataset of peripheral blood samples of 65 patients with RA and 20 controls,[7] we observed that although annotated SARS-CoV-2 reactive clonotypes could be found in both controls and patients with RA, the frequency of these clonotypes was much higher in patients with RA, with an additional peak not seen in controls (online supplemental figure S14). At the same time, we also performed retrospective analysis on 8 TCRseq libraries of FACS-sorted CD8+T cells derived from synovial fluid of 4 patients with RA we had sequenced in 2018. These libraries included both CD57+ effector cells and an overall pool of CD57- cells (online supplemental figure S15, online supplemental table S4). Consistent with expectations, searching these samples against the reconstructed repertoires of the PEAC and GSE89408 cohorts demonstrated that a substantial number of clonotypes recovered in these patients fell into the same specificity groups as our TCRseq sequences (figure 1H,I, online supplemental figure S16). When we then searched these TCRseq libraries against antigen-specificity databases, we could once again observe that sequences annotated against SARS-CoV-2 spike antigen were found to be in the same specificity clusters (figure 1J).

Of note, TCRβ sequence sharing/similarity alone is not an absolute determinant of TCR specificity, with other factors such as TCRα structure, costimulatory molecules, and HLA-restriction/effective antigen-presentation also providing key contributions. However, these results provide a crucial necessary condition for the possibility that some patients with RA may naturally harbour TCR clones sensitive to SARS-CoV-2 antigens, particularly among CD8+T cells. These clones may then be predisposed to preferentially expand and/or adopt effector functions if the pathogen is subsequently encountered,

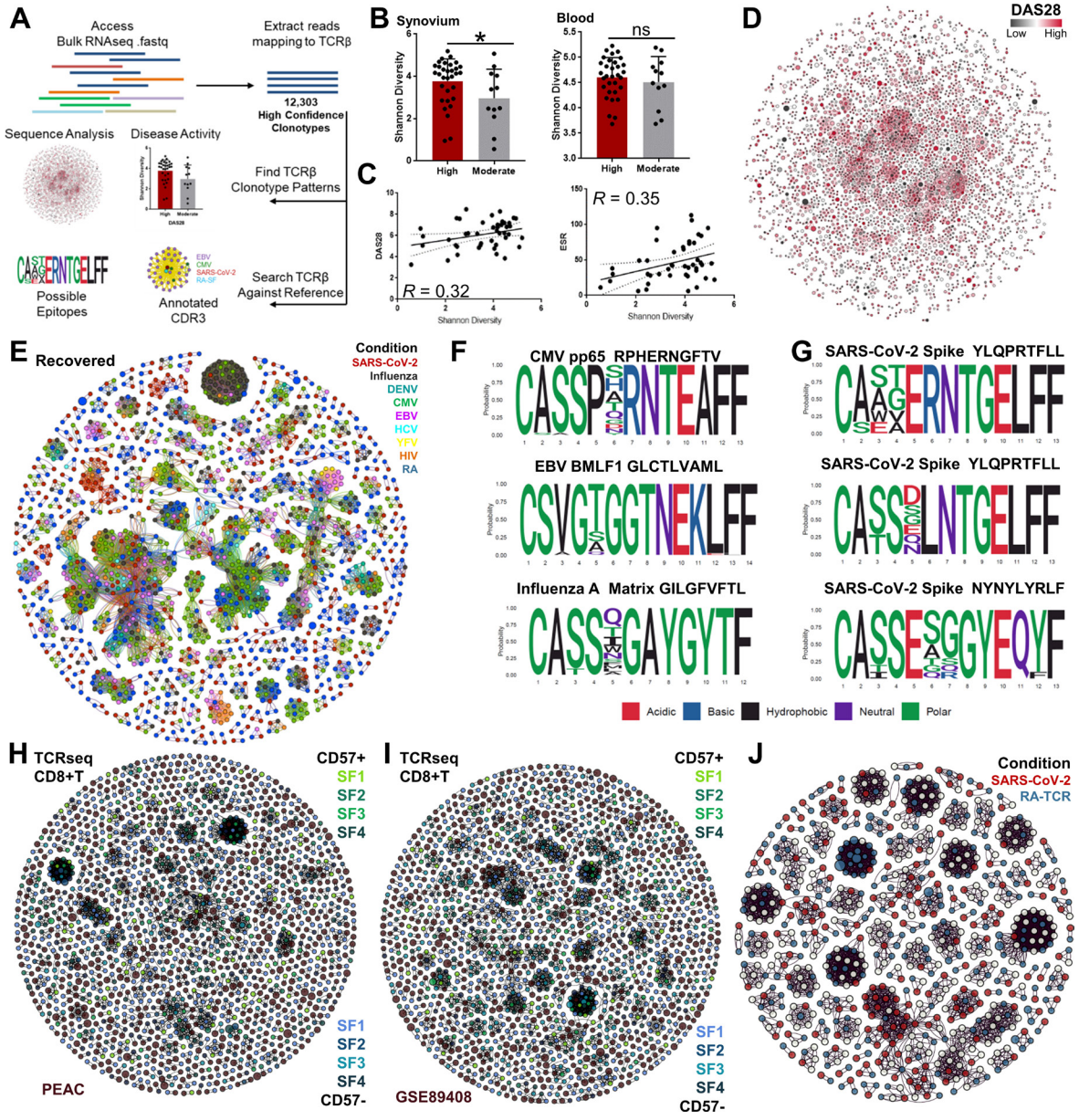


Figure 1 Extracted TCR repertoire analysis of patients with rheumatoid arthritis (RA) captures signatures of clonal expansion and pathogen antigen cross-reactivity. (A) Schematic overview of the analysis pipeline for extracting and processing RA T cell receptor (TCR) repertoire information. Raw RNA sequencing data of patients with RA from different cohorts was collected and aligned using TRUSTV.4 to extract TCR sequences. TCR sequences supported by multiple reads were retained and used to construct an RA-associated TCR repertoire database. Further analyses were performed using this database to identify associations with disease activity, as well as potentially enriched motifs. Annotation of these motifs demonstrates a potential for cross-reactivity between synovial TCR sequences and pathogen antigen-specific TCRs. (B) Shannon diversity of TCRβ clonotypes was higher in patients with high disease activity (DAS28 > 5.1) in synovial tissue, but not significantly different in peripheral blood. (C) Both overall DAS28 and ESR showed positive correlation with increasing synovial TCRβ diversity. (D) Network analyses of RA-associated TCRβ clusters according to CDR3 sequence reveals the presence of shared expanded clonotypes among patients with high disease activity, and demonstrates that many of these sequences are highly similar. Each node represents one CDR3 sequence, with edges connecting nodes assigned to the same specificity group, and larger specificity groups are densely connected together. Larger nodes correspond to higher-frequency clones. (E) Network visualisation of RA-associated TCRβ CDR3 clusters annotated against a combined database of TCRβ sequences of known antigen specificity, indicates possible associations with SARS-CoV-2 antigens as well as other pathogens. (F) Logo plots showing consensus TCRβ CDR3 motifs detected in RA synovial tissue and with known antigen specificity in three commonly encountered pathogens, labeled in order of pathogen, protein and peptide. (G) Logo plot as in (F) showing consensus TCRβ CDR3 motifs detected in RA synovial tissue but matching SARS-CoV-2 antigens, a pathogen not present at the time the sequencing studies were conducted. (H–I) Network analysis of TCR sequences recovered from focused TCRseq of 4 patients with RA (sorted into CD8+CD57+ and CD8+CD57- populations) clustered against the TCRs obtained from repertoire reconstruction of the RA synovial samples in the PEAC (H) and GSE89408 (I) cohorts. An abundance of clustered TCRs can be observed, indicating that sequences recovered from repertoire reconstruction may also be shared. (J) Network analysis of clustering primary TCRseq sequences (as in H–I) with annotated antigen-specific sequences. As in (E), associations between a number of annotated SARS-CoV-2 sequences and RA synovial sequences can be observed. *p-value <0.05. DAS28, Disease Activity Score 28; ESR, erythrocyte sedimentation rate.

and potentially contribute to the observed clinical associations between arthritis/autoimmunity and SARS-CoV-2 infection.[8] While the existence of antigen-specific memory clones in unexposed controls has been previously regarded as controversial, recent studies in the context of SARS-CoV-2 have confirmed that pre-existing antigen-specific CD4+memory clones can be found in uninfected patients,[9] and were intriguingly found to correspond with severe outcomes. Furthermore, these repertoires may be reshaped following vaccination or infection.[10] As such, we speculate that vaccination against SARS-CoV-2 may also reshape CD8+T cell clonal dynamics in the synovium and peripheral circulation of patients with RA and help to dampen the feedback between SARS-CoV-2 infection and autoimmunity. Additional exploration of possible pathogen antigen cross-reactivity in RA is needed to clarify the role of SARS-CoV-2 and other infections diseases in influencing disease progression.

Zihan Zheng ,[1,2] **Ling Chang,**[1] **Jie Mu,**[3] **Qingshan Ni,**[3] **Zhong Bing,**[4] **Qing-Hua Zou,**[4] **Ying Wan,**[3] **Yuzhang Wu,**[1] **Jingyi Li,**[4] **Liyun Zou**[1]

[1]Institute of Immunology, Army Medical University, Chongqing, Chongqing, China
[2]Department of Autoimmune Diseases, Chongqing International Institute for Immunology, Chongqing, Chongqing, China
[3]Biomedical Analysis Center, Army Medical University, Chongqing, Chongqing, China
[4]Department of Rheumatology and Immunology, First Affiliated Hospital (Southwest Hospital) of Army Medical University, Chongqing, Chongqing, China

Correspondence to Professor Liyun Zou, Institute of Immunology, Army Medical University, Chongqing, People's Republic of China; zouliyun_2012@163.comDr Jingyi Li; lijingyi_tmmu@foxmail.comProfessor Yuzhang Wu; wuyuzhang@iiicq.vipProfessor Ying Wan; wanying516@foxmail.com

Correction notice This article has been corrected since it published Online First. The funding statement has been amended.

Handling editor Josef S Smolen

Contributors ZZ, LC, JL and LZ designed and coordinated this study. ZZ, LC, JM and QN performed the primary bioinformatics analyses and built the online database. ZB and Q-HZ, performed supporting analyses. LZ, YWa and YWu supervised the analyses. ZZ, YWa, YWu, JL and LZ wrote the manuscript. All authors were involved in critical review, editing, revision and approval of the final manuscript.

Funding This work was funded by grants to LZ (National Natural Science Foundation of China No. 81971546 and No. 82171787, and the Chongqing International Institute for Immunology No. 2020YJC06), JL (National Natural Science Foundation of China No. 81971537), and WYu (Emergency Key Program of Guangzhou Laboratory No. EKPG21-30-3).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the Ethics Committee of the First Affiliated Hospital (Southwest Hospital) of Army Medical University as KY202213. Patients gave informed consent prior to participation.

Provenance and peer review Not commissioned; externally peer reviewed.

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/ard-2022-222763).

ZZ and LC contributed equally.

ZZ and LC are joint first authors.

YW, YW, JL and LZ are joint senior authors.

To cite Zheng Z, Chang L, Mu J, *et al*. *Ann Rheum Dis* 2023;**82**:438–440.

Received 6 May 2022
Accepted 8 October 2022
Published Online First 19 October 2022

Ann Rheum Dis 2023;**82**:438–440. doi:10.1136/annrheumdis-2022-222763

ORCID iD
Zihan Zheng http://orcid.org/0000-0003-1486-9586

REFERENCES

1 Smolen JS, Aletaha D, Barton A, *et al*. Rheumatoid arthritis. *Nat Rev Dis Primers* 2018;4:18001.
2 Weyand CM, Goronzy JJ. The immunology of rheumatoid arthritis. *Nat Immunol* 2021;22:10–18.
3 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;78:1070–8.
4 Lliso-Ribera G, Humby F, Lewis M, *et al*. Synovial tissue signatures enhance clinical classification and prognostic/treatment response algorithms in early inflammatory arthritis and predict requirement for subsequent biological therapy: results from the pathobiology of early arthritis cohort (PEAC). *Ann Rheum Dis* 2019;78:1642–52.
5 Guo Y, Walsh AM, Fearon U, *et al*. CD40L-dependent pathway is active at various stages of rheumatoid arthritis disease progression. *J Immunol* 2017;198:4490–501.
6 Davignon J-L, Combe B, Cantagrel A. Cytomegalovirus infection: friend or foe in rheumatoid arthritis? *Arthritis Res Ther* 2021;23:16.
7 Savola P, Kelkka T, Rajala HL, *et al*. Somatic mutations in clonally expanded cytotoxic T lymphocytes in patients with newly diagnosed rheumatoid arthritis. *Nat Commun* 2017;8:15869.
8 Redwan EM, Alghamdi MF, El-Aziz TMA, *et al*. The mechanism behind flaring/triggering of autoimmunity disorders associated with COVID-19. *Autoimmun Rev* 2021;20:102909.
9 Bacher P, Rosati E, Esser D, *et al*. Low-avidity $CD4^+$ T Cell responses to SARS-CoV-2 in unexposed individuals and humans with severe COVID-19. *Immunity* 2020;53:1258–71.
10 Minervina AA, Pogorelyy MV, Kirk AM, *et al*. SARS-CoV-2 antigen exposure history shapes phenotypes and specificity of memory $CD8^+$ T cells. *Nat Immunol* 2022;23:781–90.

Supplemental Methods

Data Access and TCR repertoire recovery

To build our database of rheumatoid arthritis-associated TCR repertoire, we searched the GEO and EBI databases for RNA-sequencing studies of rheumatoid arthritis patients. From this search, we identified two large, independent cohorts (GSE89408 and PEAC, E-MTAB-6141) with available raw sequencing data. From database records we noted that these studies were completed and publicly released prior to 2017 (GSE89408) and 2018 (PEAC). Raw .fastq files were processed through TRUSTv4[1] to recover TCR sequences. Since these TCR sequences were extracted from bulk sequencing data, where recovery of paired TCRαβ clones is not possible, we retained only TCRβ sequences to perform clonotype-level analysis. Because recovery is performed on RNAseq data, full sequences for the key complementarity-determining region 3 (CDR3) region are consequently obtained along with the relevant variable (V) and joining (J) segments used, an advantage over older profiling analyses that only identify patterns to the usage of V-J genes. However, because these TCR sequences are extracted from bulk samples containing heterogeneous cell types, the TCR data of TRUST-reconstructed repertoires is intrinsically biased towards high-frequency clonotypes that are more likely to be captured, and is likely to miss lower frequency naïve/memory clones normally identified in focused deep sequencing of the TCR repertoire in individual patients, representing a tradeoff of breadth versus depth. Clonotype diversity metrics were computed using the abdiv[2] package in R[3].

TCR Clustering

To cluster TCR sequences, we utilized the GLIPH2 algorithm[4] using the default v2.0 reference and allowing for select interchange of amino acids with positive BLOSUM-62 matrix scores. Resulting paratope-based TCR clusters were visualized as networks in Gephi[5], with nodes in each cluster connected by bi-directional edges. Smaller networks were distributed in visualizations using the Frutcherman-Reingold algorithm, while larger networks were distributed using the faster Yifan Hu proportional model. Node sizes were adjusted to reflect clonal size where indicated. Edges connect all nodes encompassing TCR sequences of the same paratope specificity group. Because some sequences were predicted to potentially be a part of more than one specificity group, some nodes are correspondingly connected to more than one group.

Online Database

A dedicated web server of the synovial TCR sequences extracted from rheumatoid arthritis patients, together with their clustered annotations, is accessible at (http://repertoire.life) (redirects to http://118.24.236.198:3838/RA). The database is an ongoing effort that allows visitors to browse and search for TCR sequences of interest, as well as motifs associated with specific pathogens of interest with annotated antigens. The website was constructed using the shiny[6] package in R, and also contains logo visualizations drawn using the ggseqlogo[7] package in R.

Human Samples and TCRseq

Synovial fluid samples from four established RA patients and high disease activity (DAS28 > 5.1) undergoing therapeutic joint fluid aspiration were collected at the First Affiliated Hospital (Southwest Hospital) of Army Medical University. Mononuclear cells were isolated from synovial fluid via gradient centrifugation over lymphoprep (StemCell, USA), washed, and subsequently stained with antibodies against CD3ε, CD19, CD4, CD8a, CD57 (all BD Biosciences, USA) and together with a viability dye (Zombie-NIR, Biolegend, USA). The CD19-CD4-CD3+CD8+ cells were sorted into CD57+ and CD57- fractions through FACS (Aria III, BD Biosciences, USA) and collected for TCR repertoire sequencing. TCR libraries were constructed following a method previously described[8,9]. Briefly, total RNA was extracted from the sorted cells using a commercially available kit (RNeasy, Qiagen, USA) and reverse transcribed using a universal human TCRB constant region primer. Multiplex PCR using optimized primers for each human variable chain gene, and sequencing was performed on the Illumina platform (Novaseq, Illumina USA). Raw sequence data was processed through MiXCR[10] to obtain processed clonotypes and their corresponding frequencies.

[1] Song L, Cohen D, Ouyang Z, et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nat Methods. 2021 Jun;18(6):627-630. doi: 10.1038/s41592-021-01142-2.
[2] Bittinger K (2020). abdiv: Alpha and Beta Diversity Measures. R package version 0.2.0.
[3] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
[4] Huang H, Wang C, Rubelt F, et al. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. Nat Biotechnol. 2020 Oct;38(10):1194-1202. doi: 10.1038/s41587-020-0505-4.
[5] Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. 2009.
[6] Chang W, Cheng J, Allaire JJ, et al (2021). shiny: Web Application Framework for R. R package version 1.7.1.
[7] Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics. 2017 Nov 15;33(22):3645-3647. doi: 10.1093/bioinformatics/btx469.
[8] Zhang B, Jia Q, Bock C, et al. Glimpse of natural selection of long-lived T-cell clones in healthy life. Proc Natl Acad Sci U S A. 2016 Aug 30;113(35):9858-63. doi: 10.1073/pnas.1601634113.
[9] Zhang J, Wang Y, Yu H, et al. Mapping the spatial distribution of T cells in repertoire dimension. Mol Immunol. 2021 Oct;138:161-171. doi: 10.1016/j.molimm.2021.08.009.
[10] Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015 May;12(5):380-1. doi: 10.1038/nmeth.3364.

FigureS1–Synovial TCRβ α diversity is associated with higher disease activity levels

A) Lowered synovial TCRβ clonotype evenness (Simpson's evenness, calculated using the reciprocal index) is observed in patients with highly active RA (as assessed by DAS28 > 5.1, from the PEAC cohort). B) No difference in clonotype evenness is observed in paired peripheral blood samples obtained from the same cohort. C) Synovial Simpson's evenness is negatively correlated with DAS28 disease activity and erythrocyte sedimentation rate (ESR). These results are consistent with Fig1B-C, calculated using Shannon diversity.

FigureS2– Clonal expansion is most prominent in RA patients with lymphoid pathotype

A) Network visualization of clustered synovial TCRβ clonotypes extracted from the GSE89408 cohort. Larger nodes correspond to larger clonotypes (higher frequency). B) Clone distribution plot in the GSE89408 cohort, wherein synovial samples from four different pathologies were analyzed. Larger clones were more frequently found in RA patients compared to osteoarthritis (OA), arthralgia (AG), and healthy controls (HC). C) Intersection plot of clonotype overlap between the four conditions demonstrates that cross-pathology clonal sharing is very rare. D) Clone distribution plot in the PEAC cohort, wherein synovial samples were classified according to pathoype. Larger clones were more frequently found in lymphoid pathotype. E) Intersection plot of clonotype overlap between the three pathotypes demonstrates that a significant number of clones can be found to overlap across pathoypes. This result indicates that synovial T cell activity of RA patients may be influenced in part by shared clonotypes independently of histological pathotype.

FigureS3: Transcriptome and repertoire differences between varying joint conditions in GSE89408 are independent of some common sources of variation

A) Visualization of the RNA transcriptome data from the GSE89408 following normalization and dimension reduction in UMAP space. Clear stratification can be seen in the majority of healthy control and osteoarthritis samples from the rheumatoid arthritis samples, while early and established samples are relatively admixed. B) Projection of the gender of the specimen donor in UMAP space. No noticeable separation can be seen between male and female donors. C) ACPA status is also not a distinguishing factor on the overall transcriptome level. D) Age of the donors is also not a distinguishing factor on the overall transcriptome level. E-G) Expression of key markers of T cells (the highly expressed, and functionally critical, molecules *CD3D, CD3E,* and *CD3G*) is most highly concentrated in a subset of the RA samples on the right half of the UMAP space. H) The overall TCR counts recovered from each sample through repertoire reconstruction mirrors the expression pattern in (E-G), indicating that repertoire reconstruction is in fact reflective of overall T cell representation in the sample.

FigureS4–Synovial TCRβ α diversity is higher in lymphoid pathotype

A) Network visualization of clustered synovial TCRβ clonotypes extracted from the PEAC cohort. Larger nodes correspond to larger clonotypes (higher frequency). One can observe that while highly interconnected nodes are dominated by the lymphoid pathotype, there is some overlap with the other two major pathotypes. B) Increased synovial TCRβ clonotype diversity (Shannon index) is observed in patients with the lymphoid pathotype. C) No significant difference in peripheral blood TCRβ clonotype diversity is observed between the three pathotypes. D) Lowered synovial TCRβ clonotype evenness (Simpson's evenness, calculated using the reciprocal index) is observed in patients with the lymphoid pathotype. E) No significant difference in peripheral blood TCRβ clonotype evenness is observed between the three pathotypes.

FigureS5– Motif analysis of RA samples stratified by disease activity assessment scores

Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort, colored according to four different measures clinical measures. Larger nodes correspond to larger clonotypes (higher frequency). A) The larger clusters correspond to TCRβ clonotypes derived from patients with higher pain levels as assessed through the visual analog scale (VAS). B) These patients also generally reported higher impact on quality of life through the healthy assessment questionnaire (HAQ). C-D) Disease duration and inflammation scores have weaker likelihoods of being overrepresented in the larger TCR clusters.

FigureS6– Motif analysis of RA samples classified by antibody levels

Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort as in FigS4, colored according to reported antibody titres. A) Larger TCR clusters contained both ACPA+ and ACPA- patients. B) While the majority of the TCR sequences originate from seropositive patients, some of these sequences are also shared in rheumatoid factor-negative patients. These results indicate that antibody status and titres may not directly influence TCR clonotype clustering.

FigureS7–Synovial TCRβ clonotypes are not wholly independent from peripheral circulation

A) Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort as in FigS4, colored according to tissue origin. Notably, the larger TCR clusters all feature sequences found in both peripheral blood and synovial tissue. B) Publicity analysis demonstrates that identical TCRβ clonotypes can be found to shared across multiple patients when both peripheral blood and synovial tissue are considered. More public clonotypes also tend to be larger. C) Pie chart summary of sharing relationships shows that clonotypes are more likely to be found in synovial tissue and peripheral blood as compared to only peripheral blood or synovial tissue alone.

FigureS8–β diversity metrics in TCR repertoire data from paired blood and synovium samples fail to delineate between pathotypes and disease activity

Calculated pairwise similarity metrics of paired blood and synovial tissue sample from the PEAC cohort fail to identify significant associations between blood-synovial tissue TCR repertoire overlap and disease pathotype (A,C), or with disease activity (B,D).

FigureS9–TCR repertoire analysis of rheumatoid arthritis patients includes sequences matching annotated pathogen-specific TCRs

A) Network visualization of RA synovial TCRβ clonotypes extracted from both the PEAC and GSE89408 cohorts, connected according to CDR3 amino acid sequence Levenshtein distance (0 perfect match, 1 a single substitution/deletion/insertion). When searched against the McPAS-TCR database of curated TCR clones, hits for both pathogens and autoimmune diseases were found. B) Network visualization of RA synovial TCRβ clonotypes extracted from both the PEAC and GSE89408 cohorts as in (A) searched against the VDJdb database of antigen-specific clones. Once again, a large number of hits were found to correspond to pathogen-specific clones.

FigureS10–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the full McPAS database

More detailed annotation of the matches to RA synovial TCR sequences and the McPAS database. A) RA synovial TCR sequences in both databases fall into a number of large clusters of highly similar sequences, and these clusters include annotated TCRs from multiple pathologies. B) CDR3 sequences originating from annotated CD4+ and CD8+ T cell populations can be found. C) The majority of these sequences originate from known antigen epitopes. D) Of note, because the vast majority of annotated TCR sequences come from pathogen-focused experiments, a vast majority of the observed hits were also found to match pathogens. However, we were unable to find matches for a large number of RA synovial TCR sequences. Further annotation efforts in the future that classify TCR sequences against autoantigens will be needed to better characterize the range of autoimmune-specific TCR sequences. We note this potential caveat here to prevent a potential misunderstanding that almost all RA synovial TCR sequences are pathogen-specific. At the same time however, the breadth of positive matches here do strongly suggest that at least a portion of RA synovial TCR sequences may respond to pathogen-derived antigens. Whether this response is due to antigen cross-reactivity arising from antigen mimics of autoantigens is an open question.

FigureS11–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the VDJdb database

More detailed annotation of the matches to RA synovial TCR sequences and the VDJdb database. A) CDR3 sequences matching antigens reported to be displayed on both class I and class II MHC complexes can be found in the synovial tissue of RA patients, although the large

majority of sequences are from class I due to representation bias in the reference. B) These annotated sequences largely reflect a number of pathogens with the nodes colored correspondingly. (CMV- cytomegalovirus, InfA- Influenza A, EBV- Epstein-Barr virus, HIV- human immunodeficiency virus, YFV- yellow fever virus, DENV- dengue fever virus).

FigureS12–Exact matches between TCR sequences recovered from RA synovium and annotated pathogen antigen-specific sequences

Beyond the prominent clusters of TCR sequences containing both SARS-CoV-2 antigen-specific sequences and RA-synovial tissue sequences shown in Fig1E, we could also observe perfect matches in CDR3 amino acid sequence between the two conditions.

FigureS13–Spike-matched epitopes recovered from RA synovium and are conserved in rare beta coronaviruses

To understand the conservation of the spike epitopes we identified in RA synovium, we performed sequence comparisons of the three spike peptides with the highest number of matches with spike proteins in other human coronaviruses (including the alphacoronaviruses HCoV-NL63, HCoV-229E, and betacoronaviruses SARS-CoV, MERS-CoV, HCoV-HKU1, HCoV-OC43) using BLAST. No matches to any alphacoronaviruses were found. However, all three peptides showed matches to SARS-CoV, and two showed some similarity to MERS-CoV. These results are consistent with our inference that the spike-specific TCRs we uncovered are unlikely to originate from antigen encounter, due to the known epidemiology of SARS-CoV and MERS-CoV. At the same time, while we cannot fully rule out patient encounter with uncharacterized circulating coronaviruses being the source, we can observe that the most common circulating seasonal coronaviruses are not the source of these antigens.

FigureS14–TCRseq of PB from RA patients and controls prior to SARS-CoV-2 emergence

A) Violin plot showing the frequencies of annotated SARS-CoV-2 antigen-specific sequences being found in the peripheral blood of healthy controls (n = 20) and RA patients (n = 65) profiled using TCRseq by Savola et al. Each sample was downsampled to 10k sequences to enable direct comparison. Strict matching criteria (identical CDR3b sequence) was used. B) Distribution of the in-sample clonal frequencies for each SARS-CoV-2 antigen specific sequence in HC and RA groups. While the majority of the clonotypes were relatively rare (~ 5E-5), a noticeable second peak in frequency can be found in RA patients at a frequency of 1E-4, and additional numbers of larger clones can also be seen.

FigureS15–TCRseq of SF from patients prior to SARS-CoV-2 emergence

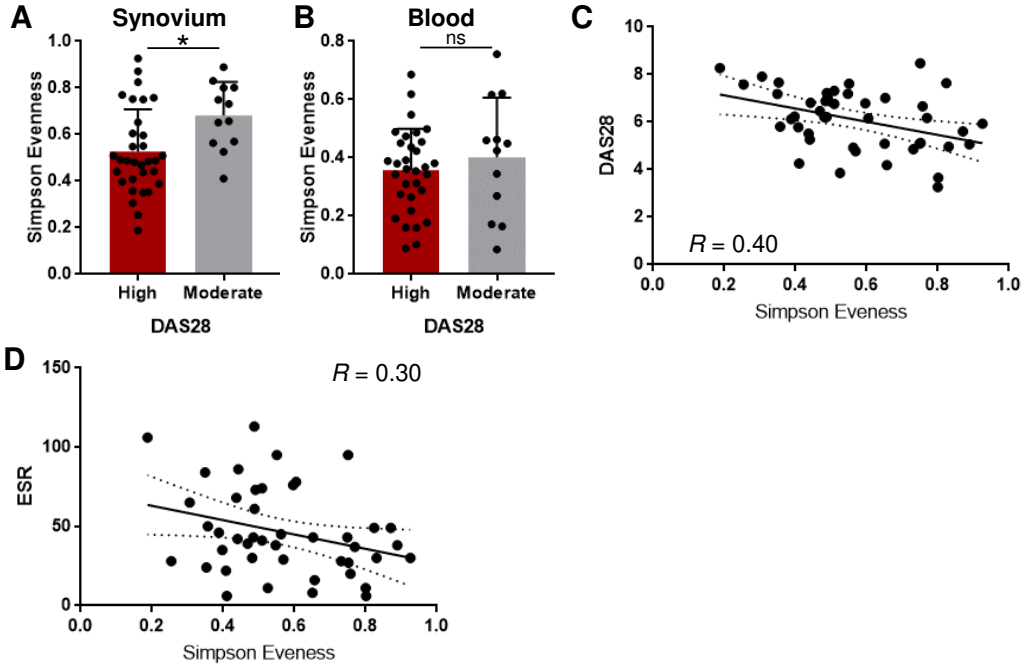
A) Shannon diversity of the CD57- samples is significantly higher than in the CD57+ samples. This is consistent with our understanding that CD57+ are dominated by effector T cells, while the CD57- pool may include more memory cells that have not undergone as dramatic of a clonal

expansion. B) Intersection analysis of the 8 TCRseq libraries reveals that a substantial number of clones shared across CD57+ and CD57- populations derived from the same individuals. Some public sharing across individuals can also be seen.

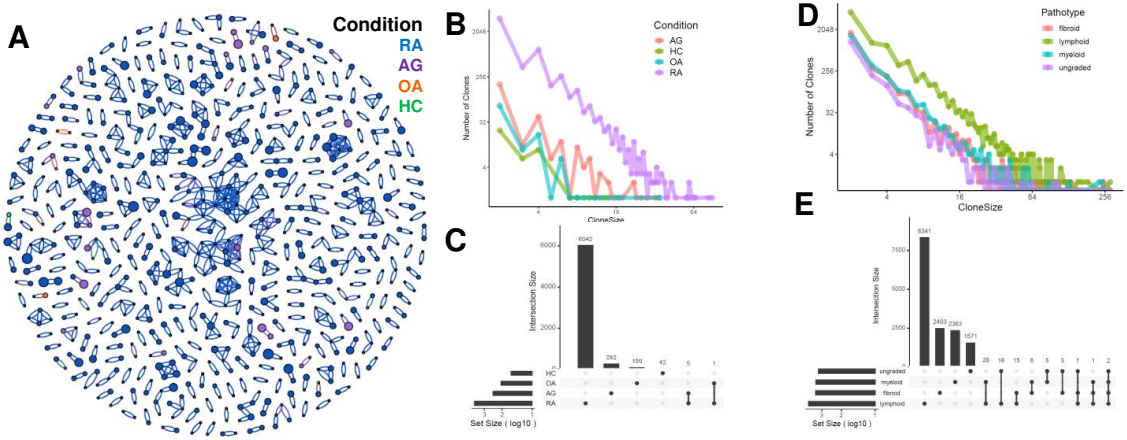
FigureS16–Spike-matched TCR sequences are found in TCRseq of SF from patients prior to SARS-CoV-2 emergence

Network visualization as in Fig1J, with nodes color split more finely into TCRseq sample origin and pathogen type. Notably, we can observe that several other antigens from SARS-CoV-2 (white nodes) can also be observed to form specificity clusters besides spike protein.

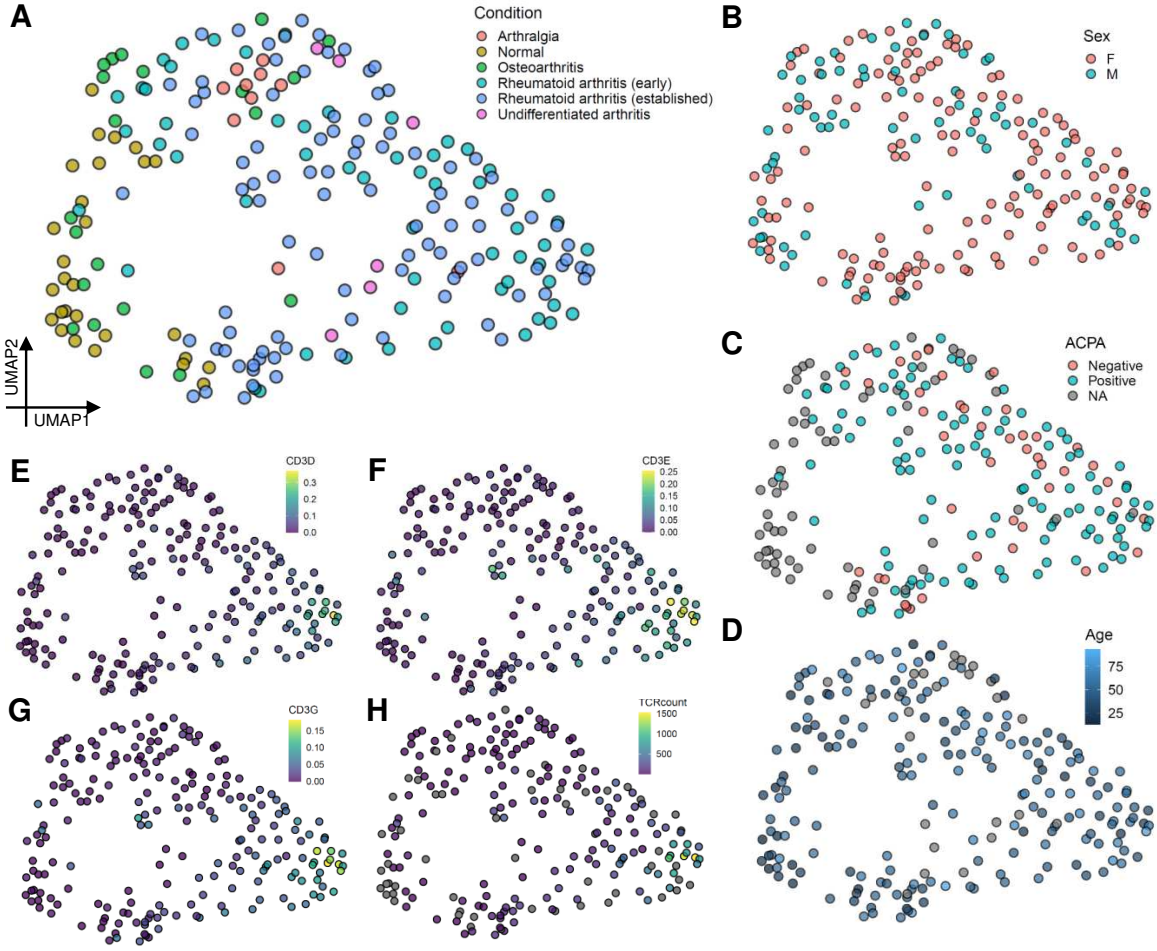
FigureS1–Synovial TCRβ α diversity is associated with higher disease activity levels



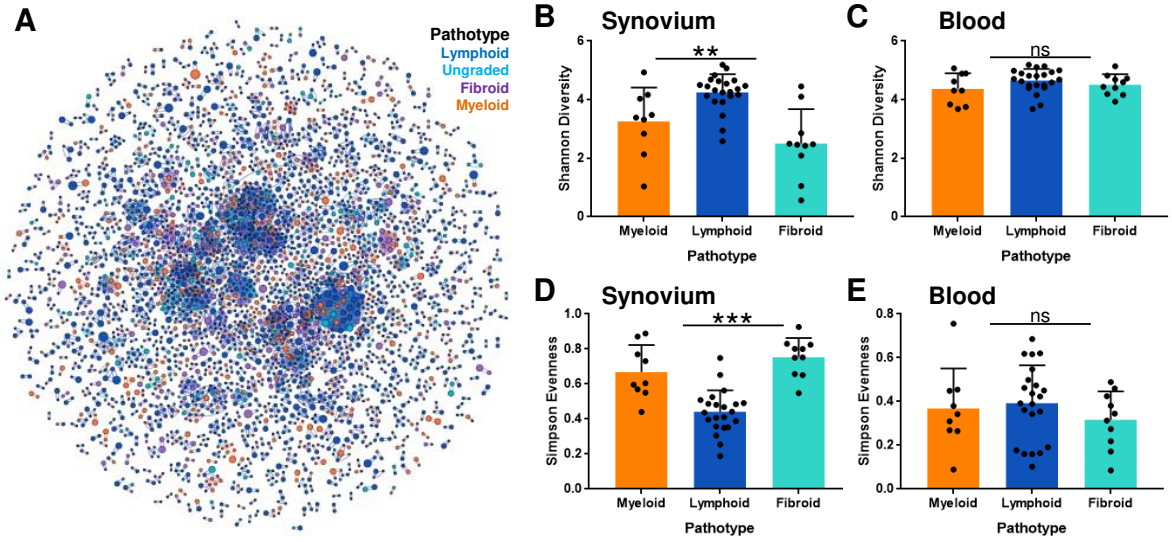
FigureS2– Clonal expansion is most prominent in RA patients with lymphoid pathotype



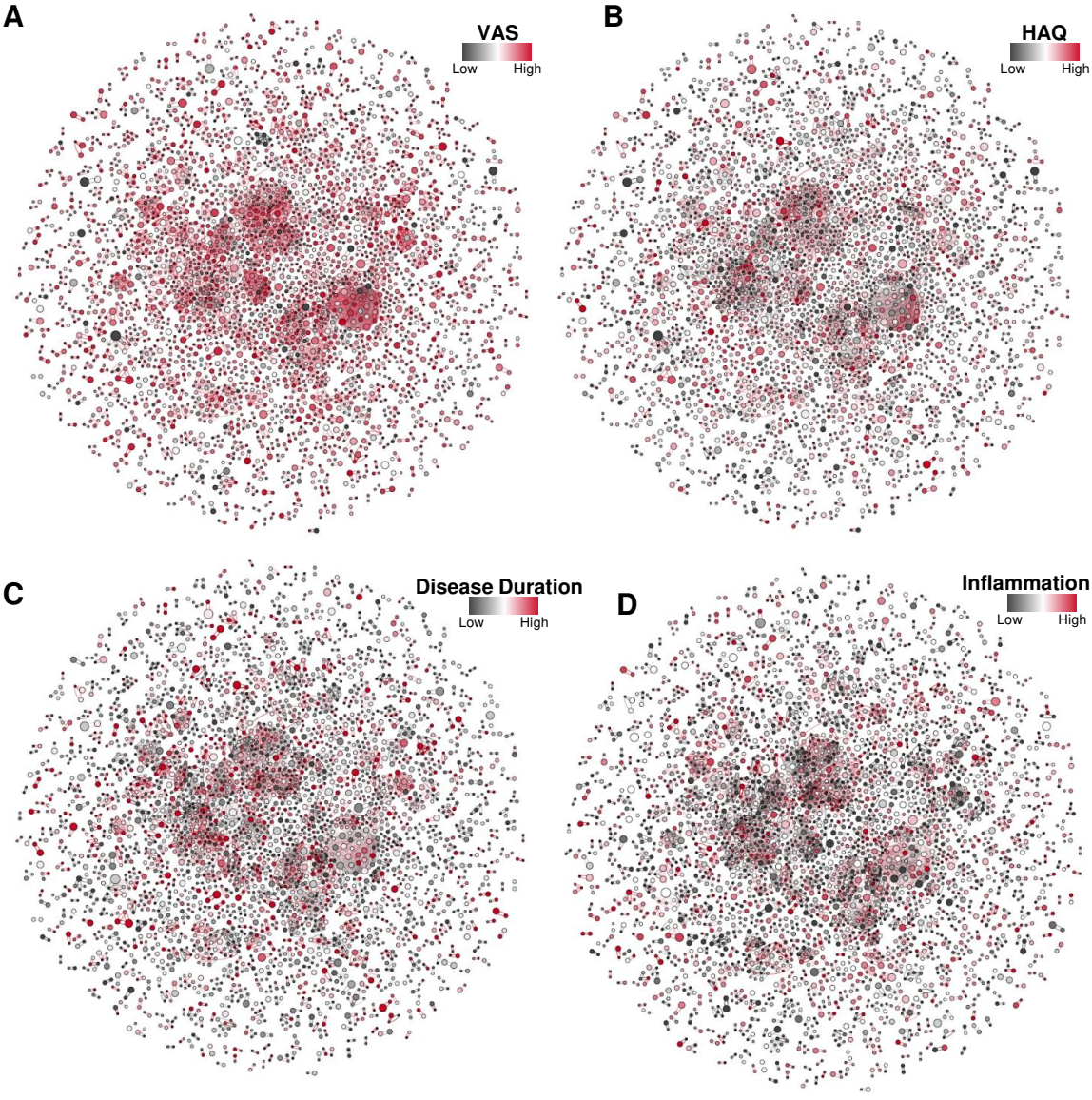
FigureS3: Transcriptome and repertoire differences between varying joint conditions in GSE89408 are independent of some common sources of variation



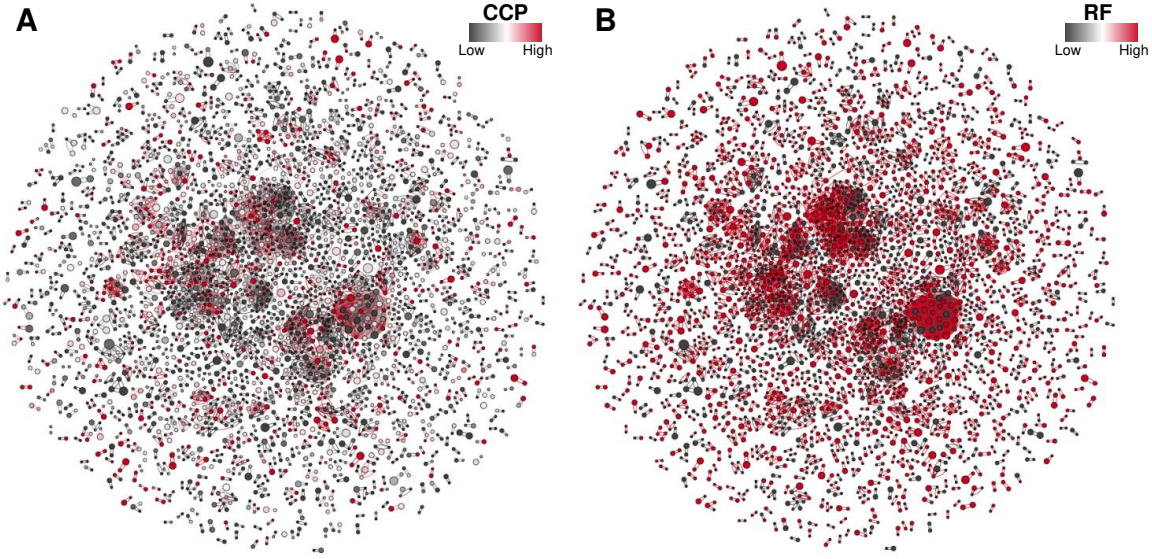
FigureS4–Synovial TCRβ α diversity is higher in lymphoid pathotype



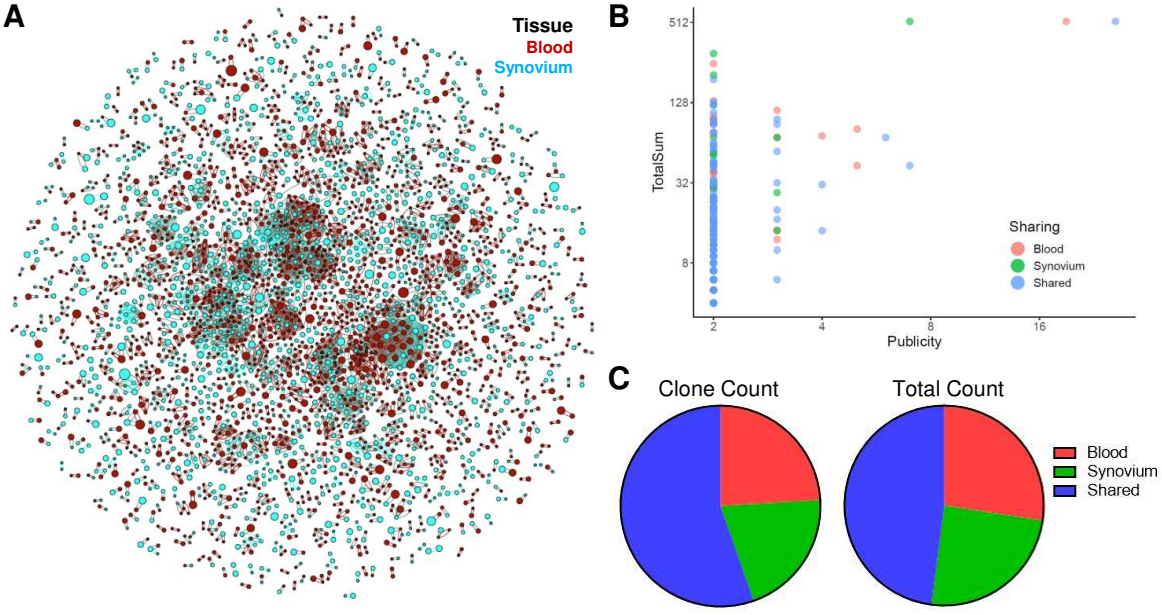
FigureS5– Motif analysis of RA samples stratified by disease activity assessment scores



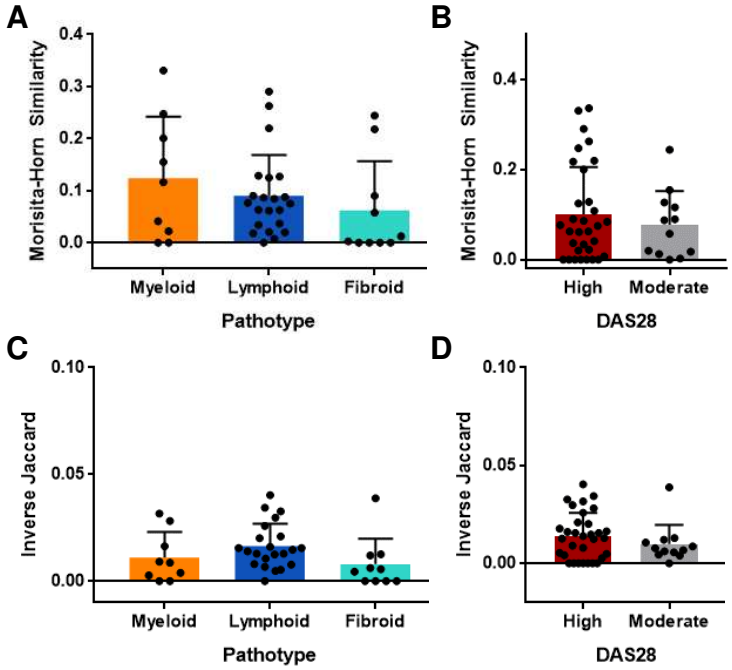
FigureS6– Motif analysis of RA samples classified by antibody levels



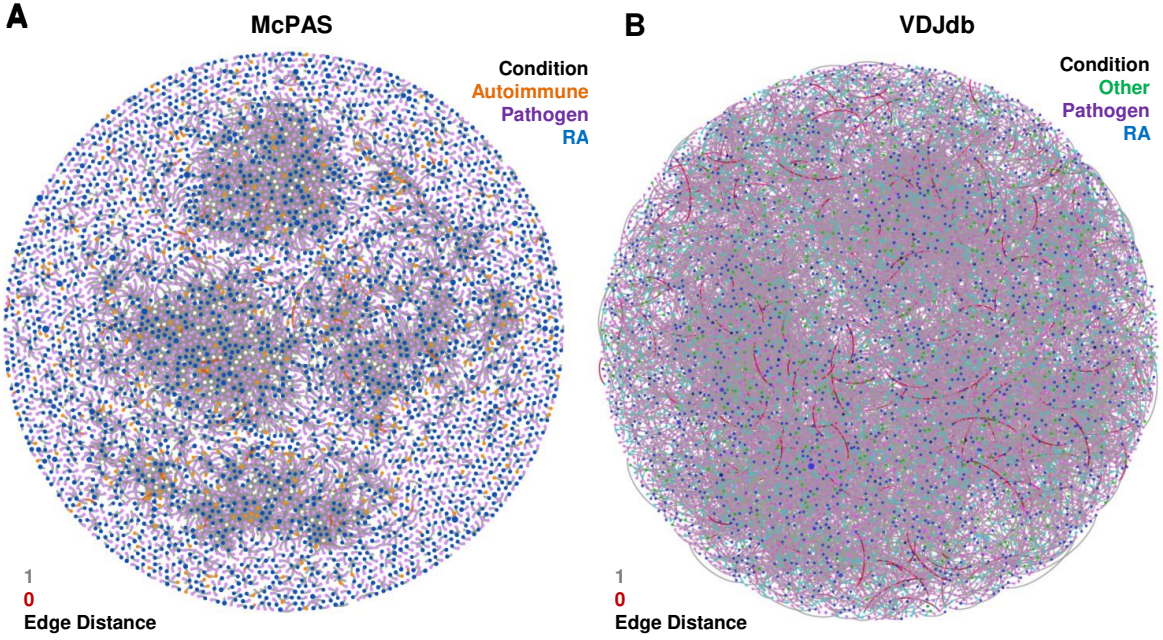
FigureS7–Synovial TCRβ clonotypes are not wholly independent from peripheral circulation



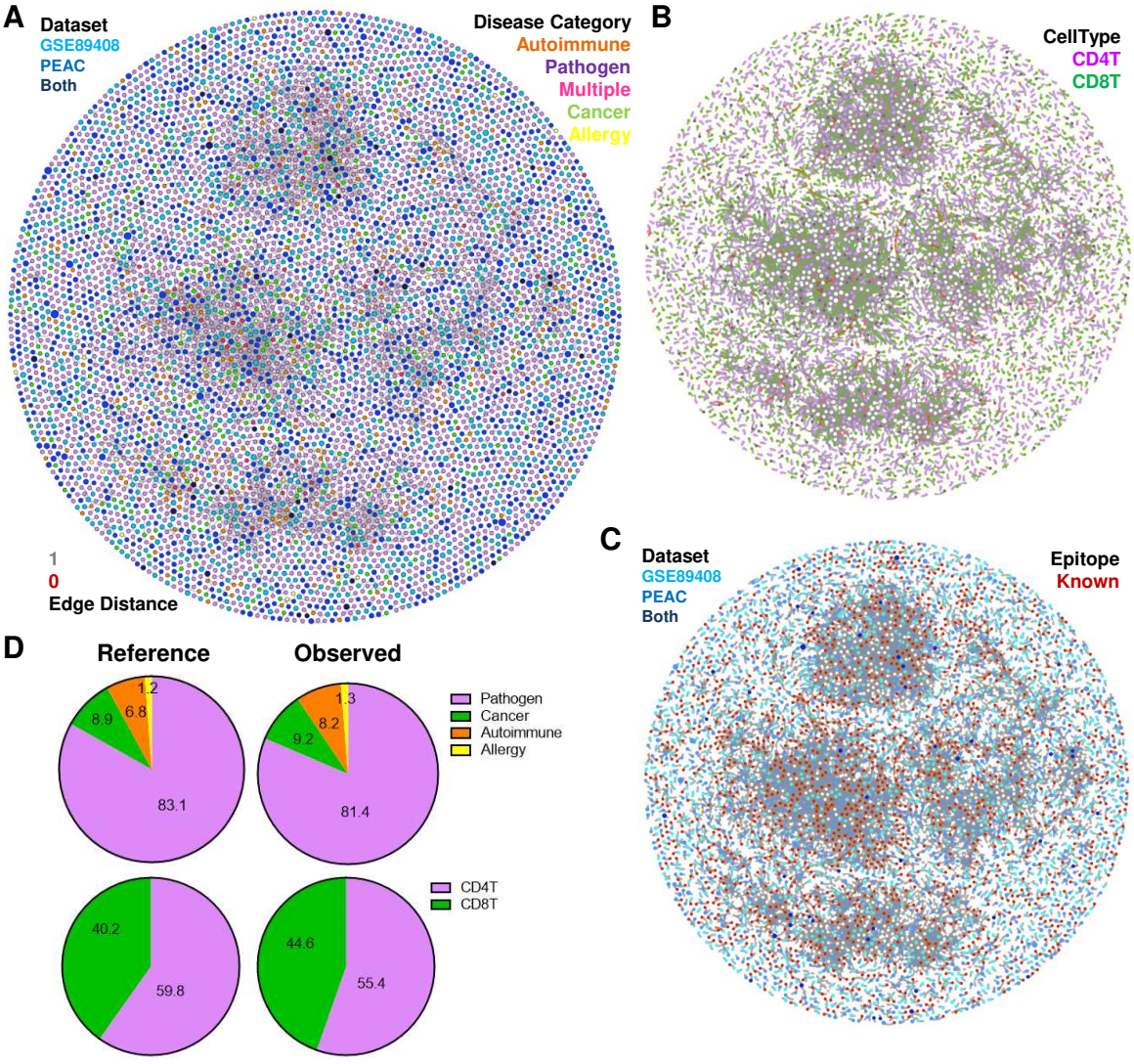
FigureS8–β diversity metrics in TCR repertoire data from paired blood and synovium samples fail to delineate between pathotypes and disease activity



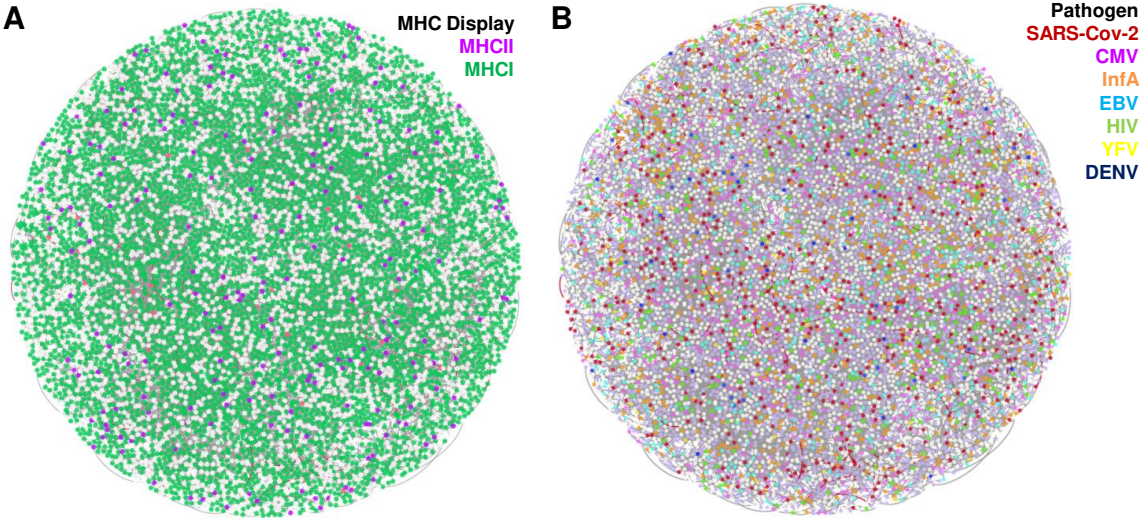
FigureS9–TCR repertoire analysis of rheumatoid arthritis patients includes sequences matching annotated pathogen-specific TCRs



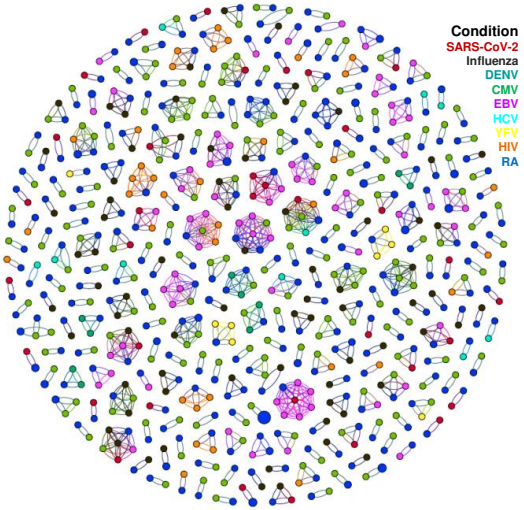
FigureS10–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the full McPAS database



FigureS11–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the VDJdb database



FigureS12–Exact matches between TCR sequences recovered from RA synovium and annotated pathogen antigen-specific sequences



FigureS13–Spike-matched epitopes recovered from RA synovium and are conserved in beta coronaviruses

A

Virus	Start	End	Sequence	Accession
SARS-CoV-2	269	277	**YLQPRTFLL**	YP_009724390.1
MERS-CoV	317	325	**KLQPLTFLL**	YP_009047204.1
SARS-CoV	256	264	**YLKPTTFML**	YP_009825051.1

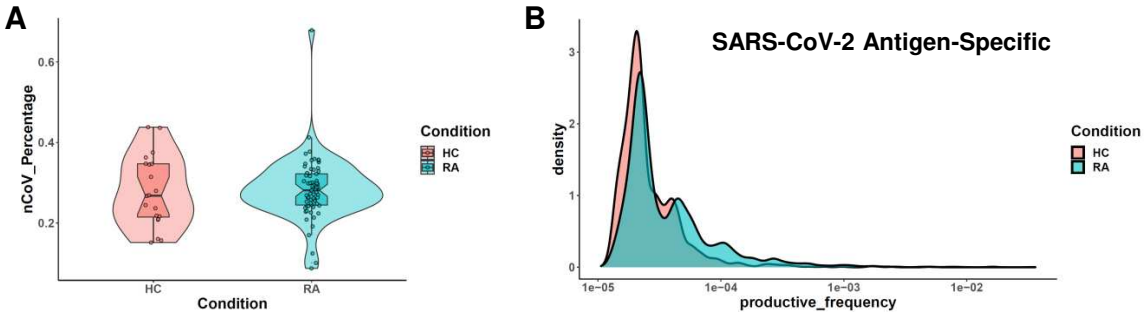
B

Virus	Start	End	Sequence	Accession
SARS-CoV-2	865	873	**LTDEMIAQY**	YP_009724390.1
MERS-CoV	1262	1270	**LTYEMLSLQ**	YP_009047204.1
SARS-CoV	847	855	**LTDDMIAAY**	YP_009825051.1

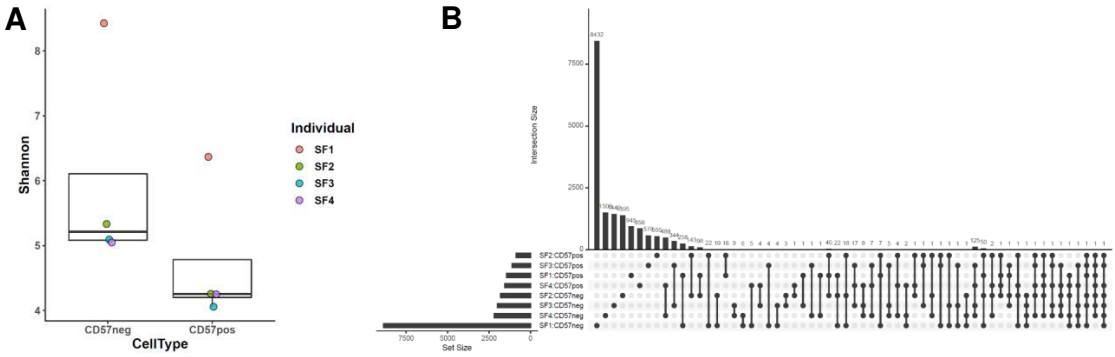
C

Virus	Start	End	Sequence	Accession
SARS-CoV-2	448	456	**NYNYLYRLF**	YP_009724390.1
SARS-CoV	435	443	**NYNYKYRYL**	YP_009825051.1

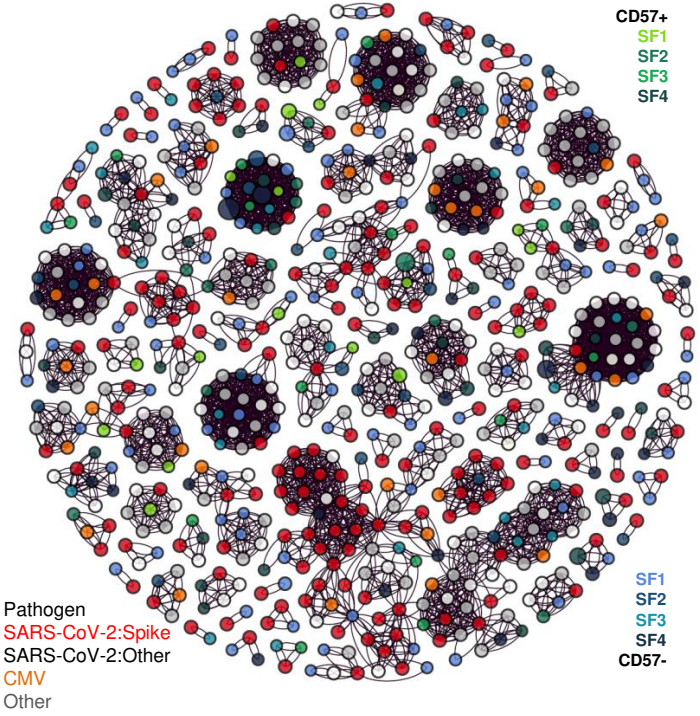
FigureS14–Spike-matched TCR sequences are found in TCRseq of PB from patients prior to SARS-CoV-2 emergence in a public dataset



FigureS15–TCRseq of SF from patients prior to SARS-CoV-2 emergence



FigureS16–Spike-matched TCR sequences are found in TCRseq of SF from patients prior to SARS-CoV-2 emergence



Supplemental Methods

Data Access and TCR repertoire recovery

To build our database of rheumatoid arthritis-associated TCR repertoire, we searched the GEO and EBI databases for RNA-sequencing studies of rheumatoid arthritis patients. From this search, we identified two large, independent cohorts (GSE89408 and PEAC, E-MTAB-6141) with available raw sequencing data. From database records we noted that these studies were completed and publicly released prior to 2017 (GSE89408) and 2018 (PEAC). Raw .fastq files were processed through TRUSTv4[1] to recover TCR sequences. Since these TCR sequences were extracted from bulk sequencing data, where recovery of paired TCRαβ clones is not possible, we retained only TCRβ sequences to perform clonotype-level analysis. Because recovery is performed on RNAseq data, full sequences for the key complementarity-determining region 3 (CDR3) region are consequently obtained along with the relevant variable (V) and joining (J) segments used, an advantage over older profiling analyses that only identify patterns to the usage of V-J genes. However, because these TCR sequences are extracted from bulk samples containing heterogeneous cell types, the TCR data of TRUST-reconstructed repertoires is intrinsically biased towards high-frequency clonotypes that are more likely to be captured, and is likely to miss lower frequency naïve/memory clones normally identified in focused deep sequencing of the TCR repertoire in individual patients, representing a tradeoff of breadth versus depth. Clonotype diversity metrics were computed using the abdiv[2] package in R[3].

TCR Clustering

To cluster TCR sequences, we utilized the GLIPH2 algorithm[4] using the default v2.0 reference and allowing for select interchange of amino acids with positive BLOSUM-62 matrix scores. Resulting paratope-based TCR clusters were visualized as networks in Gephi[5], with nodes in each cluster connected by bi-directional edges. Smaller networks were distributed in visualizations using the Frutcherman-Reingold algorithm, while larger networks were distributed using the faster Yifan Hu proportional model. Node sizes were adjusted to reflect clonal size where indicated. Edges connect all nodes encompassing TCR sequences of the same paratope specificity group. Because some sequences were predicted to potentially be a part of more than one specificity group, some nodes are correspondingly connected to more than one group.

Online Database

A dedicated web server of the synovial TCR sequences extracted from rheumatoid arthritis patients, together with their clustered annotations, is accessible at (http://repertoire.life) (redirects to http://118.24.236.198:3838/RA). The database is an ongoing effort that allows visitors to browse and search for TCR sequences of interest, as well as motifs associated with specific pathogens of interest with annotated antigens. The website was constructed using the shiny[6] package in R, and also contains logo visualizations drawn using the ggseqlogo[7] package in R.

Human Samples and TCRseq

Synovial fluid samples from four established RA patients and high disease activity (DAS28 > 5.1) undergoing therapeutic joint fluid aspiration were collected at the First Affiliated Hospital (Southwest Hospital) of Army Medical University. Mononuclear cells were isolated from synovial fluid via gradient centrifugation over lymphoprep (StemCell, USA), washed, and subsequently stained with antibodies against CD3ε, CD19, CD4, CD8a, CD57 (all BD Biosciences, USA) and together with a viability dye (Zombie-NIR, Biolegend, USA). The CD19-CD4-CD3+CD8+ cells were sorted into CD57+ and CD57- fractions through FACS (Aria III, BD Biosciences, USA) and collected for TCR repertoire sequencing. TCR libraries were constructed following a method previously described[8,9]. Briefly, total RNA was extracted from the sorted cells using a commercially available kit (RNeasy, Qiagen, USA) and reverse transcribed using a universal human TCRB constant region primer. Multiplex PCR using optimized primers for each human variable chain gene, and sequencing was performed on the Illumina platform (Novaseq, Illumina USA). Raw sequence data was processed through MiXCR[10] to obtain processed clonotypes and their corresponding frequencies.

[1] Song L, Cohen D, Ouyang Z, et al. TRUST4: immune repertoire reconstruction from bulk and single-cell RNA-seq data. Nat Methods. 2021 Jun;18(6):627-630. doi: 10.1038/s41592-021-01142-2.
[2] Bittinger K (2020). abdiv: Alpha and Beta Diversity Measures. R package version 0.2.0.
[3] R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
[4] Huang H, Wang C, Rubelt F, et al. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. Nat Biotechnol. 2020 Oct;38(10):1194-1202. doi: 10.1038/s41587-020-0505-4.
[5] Bastian M., Heymann S., Jacomy M. Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. 2009.
[6] Chang W, Cheng J, Allaire JJ, et al (2021). shiny: Web Application Framework for R. R package version 1.7.1.
[7] Wagih O. ggseqlogo: a versatile R package for drawing sequence logos. Bioinformatics. 2017 Nov 15;33(22):3645-3647. doi: 10.1093/bioinformatics/btx469.
[8] Zhang B, Jia Q, Bock C, et al. Glimpse of natural selection of long-lived T-cell clones in healthy life. Proc Natl Acad Sci U S A. 2016 Aug 30;113(35):9858-63. doi: 10.1073/pnas.1601634113.
[9] Zhang J, Wang Y, Yu H, et al. Mapping the spatial distribution of T cells in repertoire dimension. Mol Immunol. 2021 Oct;138:161-171. doi: 10.1016/j.molimm.2021.08.009.
[10] Bolotin DA, Poslavsky S, Mitrophanov I, et al. MiXCR: software for comprehensive adaptive immunity profiling. Nat Methods. 2015 May;12(5):380-1. doi: 10.1038/nmeth.3364.

FigureS1–Synovial TCRβ α diversity is associated with higher disease activity levels

A) Lowered synovial TCRβ clonotype evenness (Simpson's evenness, calculated using the reciprocal index) is observed in patients with highly active RA (as assessed by DAS28 > 5.1, from the PEAC cohort). B) No difference in clonotype evenness is observed in paired peripheral blood samples obtained from the same cohort. C) Synovial Simpson's evenness is negatively correlated with DAS28 disease activity and erythrocyte sedimentation rate (ESR). These results are consistent with Fig1B-C, calculated using Shannon diversity.

FigureS2– Clonal expansion is most prominent in RA patients with lymphoid pathotype

A) Network visualization of clustered synovial TCRβ clonotypes extracted from the GSE89408 cohort. Larger nodes correspond to larger clonotypes (higher frequency). B) Clone distribution plot in the GSE89408 cohort, wherein synovial samples from four different pathologies were analyzed. Larger clones were more frequently found in RA patients compared to osteoarthritis (OA), arthralgia (AG), and healthy controls (HC). C) Intersection plot of clonotype overlap between the four conditions demonstrates that cross-pathology clonal sharing is very rare. D) Clone distribution plot in the PEAC cohort, wherein synovial samples were classified according to pathoype. Larger clones were more frequently found in lymphoid pathotype. E) Intersection plot of clonotype overlap between the three pathotypes demonstrates that a significant number of clones can be found to overlap across pathoypes. This result indicates that synovial T cell activity of RA patients may be influenced in part by shared clonotypes independently of histological pathotype.

FigureS3: Transcriptome and repertoire differences between varying joint conditions in GSE89408 are independent of some common sources of variation

A) Visualization of the RNA transcriptome data from the GSE89408 following normalization and dimension reduction in UMAP space. Clear stratification can be seen in the majority of healthy control and osteoarthritis samples from the rheumatoid arthritis samples, while early and established samples are relatively admixed. B) Projection of the gender of the specimen donor in UMAP space. No noticeable separation can be seen between male and female donors. C) ACPA status is also not a distinguishing factor on the overall transcriptome level. D) Age of the donors is also not a distinguishing factor on the overall transcriptome level. E-G) Expression of key markers of T cells (the highly expressed, and functionally critical, molecules *CD3D, CD3E,* and *CD3G*) is most highly concentrated in a subset of the RA samples on the right half of the UMAP space. H) The overall TCR counts recovered from each sample through repertoire reconstruction mirrors the expression pattern in (E-G), indicating that repertoire reconstruction is in fact reflective of overall T cell representation in the sample.

FigureS4–Synovial TCRβ α diversity is higher in lymphoid pathotype

A) Network visualization of clustered synovial TCRβ clonotypes extracted from the PEAC cohort. Larger nodes correspond to larger clonotypes (higher frequency). One can observe that while highly interconnected nodes are dominated by the lymphoid pathotype, there is some overlap with the other two major pathotypes. B) Increased synovial TCRβ clonotype diversity (Shannon index) is observed in patients with the lymphoid pathotype. C) No significant difference in peripheral blood TCRβ clonotype diversity is observed between the three pathotypes. D) Lowered synovial TCRβ clonotype evenness (Simpson's evenness, calculated using the reciprocal index) is observed in patients with the lymphoid pathotype. E) No significant difference in peripheral blood TCRβ clonotype evenness is observed between the three pathotypes.

FigureS5– Motif analysis of RA samples stratified by disease activity assessment scores

Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort, colored according to four different measures clinical measures. Larger nodes correspond to larger clonotypes (higher frequency). A) The larger clusters correspond to TCRβ clonotypes derived from patients with higher pain levels as assessed through the visual analog scale (VAS). B) These patients also generally reported higher impact on quality of life through the healthy assessment questionnaire (HAQ). C-D) Disease duration and inflammation scores have weaker likelihoods of being overrepresented in the larger TCR clusters.

FigureS6– Motif analysis of RA samples classified by antibody levels

Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort as in FigS4, colored according to reported antibody titres. A) Larger TCR clusters contained both ACPA+ and ACPA- patients. B) While the majority of the TCR sequences originate from seropositive patients, some of these sequences are also shared in rheumatoid factor-negative patients. These results indicate that antibody status and titres may not directly influence TCR clonotype clustering.

FigureS7–Synovial TCRβ clonotypes are not wholly independent from peripheral circulation

A) Network visualization of clustered TCRβ clonotypes extracted from the PEAC cohort as in FigS4, colored according to tissue origin. Notably, the larger TCR clusters all feature sequences found in both peripheral blood and synovial tissue. B) Publicity analysis demonstrates that identical TCRβ clonotypes can be found to shared across multiple patients when both peripheral blood and synovial tissue are considered. More public clonotypes also tend to be larger. C) Pie chart summary of sharing relationships shows that clonotypes are more likely to be found in synovial tissue and peripheral blood as compared to only peripheral blood or synovial tissue alone.

FigureS8–β diversity metrics in TCR repertoire data from paired blood and synovium samples fail to delineate between pathotypes and disease activity

Calculated pairwise similarity metrics of paired blood and synovial tissue sample from the PEAC cohort fail to identify significant associations between blood-synovial tissue TCR repertoire overlap and disease pathotype (A,C), or with disease activity (B,D).

FigureS9–TCR repertoire analysis of rheumatoid arthritis patients includes sequences matching annotated pathogen-specific TCRs

A) Network visualization of RA synovial TCRβ clonotypes extracted from both the PEAC and GSE89408 cohorts, connected according to CDR3 amino acid sequence Levenshtein distance (0 perfect match, 1 a single substitution/deletion/insertion). When searched against the McPAS-TCR database of curated TCR clones, hits for both pathogens and autoimmune diseases were found. B) Network visualization of RA synovial TCRβ clonotypes extracted from both the PEAC and GSE89408 cohorts as in (A) searched against the VDJdb database of antigen-specific clones. Once again, a large number of hits were found to correspond to pathogen-specific clones.

FigureS10–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the full McPAS database

More detailed annotation of the matches to RA synovial TCR sequences and the McPAS database. A) RA synovial TCR sequences in both databases fall into a number of large clusters of highly similar sequences, and these clusters include annotated TCRs from multiple pathologies. B) CDR3 sequences originating from annotated CD4+ and CD8+ T cell populations can be found. C) The majority of these sequences originate from known antigen epitopes. D) Of note, because the vast majority of annotated TCR sequences come from pathogen-focused experiments, a vast majority of the observed hits were also found to match pathogens. However, we were unable to find matches for a large number of RA synovial TCR sequences. Further annotation efforts in the future that classify TCR sequences against autoantigens will be needed to better characterize the range of autoimmune-specific TCR sequences. We note this potential caveat here to prevent a potential misunderstanding that almost all RA synovial TCR sequences are pathogen-specific. At the same time however, the breadth of positive matches here do strongly suggest that at least a portion of RA synovial TCR sequences may respond to pathogen-derived antigens. Whether this response is due to antigen cross-reactivity arising from antigen mimics of autoantigens is an open question.

FigureS11–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the VDJdb database

More detailed annotation of the matches to RA synovial TCR sequences and the VDJdb database. A) CDR3 sequences matching antigens reported to be displayed on both class I and class II MHC complexes can be found in the synovial tissue of RA patients, although the large

majority of sequences are from class I due to representation bias in the reference. B) These annotated sequences largely reflect a number of pathogens with the nodes colored correspondingly. (CMV- cytomegalovirus, InfA- Influenza A, EBV- Epstein-Barr virus, HIV- human immunodeficiency virus, YFV- yellow fever virus, DENV- dengue fever virus).

FigureS12–Exact matches between TCR sequences recovered from RA synovium and annotated pathogen antigen-specific sequences

Beyond the prominent clusters of TCR sequences containing both SARS-CoV-2 antigen-specific sequences and RA-synovial tissue sequences shown in Fig1E, we could also observe perfect matches in CDR3 amino acid sequence between the two conditions.

FigureS13–Spike-matched epitopes recovered from RA synovium and are conserved in rare beta coronaviruses

To understand the conservation of the spike epitopes we identified in RA synovium, we performed sequence comparisons of the three spike peptides with the highest number of matches with spike proteins in other human coronaviruses (including the alphacoronaviruses HCoV-NL63, HCoV-229E, and betacoronaviruses SARS-CoV, MERS-CoV, HCoV-HKU1, HCoV-OC43) using BLAST. No matches to any alphacoronaviruses were found. However, all three peptides showed matches to SARS-CoV, and two showed some similarity to MERS-CoV. These results are consistent with our inference that the spike-specific TCRs we uncovered are unlikely to originate from antigen encounter, due to the known epidemiology of SARS-CoV and MERS-CoV. At the same time, while we cannot fully rule out patient encounter with uncharacterized circulating coronaviruses being the source, we can observe that the most common circulating seasonal coronaviruses are not the source of these antigens.

FigureS14–TCRseq of PB from RA patients and controls prior to SARS-CoV-2 emergence

A) Violin plot showing the frequencies of annotated SARS-CoV-2 antigen-specific sequences being found in the peripheral blood of healthy controls (n = 20) and RA patients (n = 65) profiled using TCRseq by Savola et al. Each sample was downsampled to 10k sequences to enable direct comparison. Strict matching criteria (identical CDR3b sequence) was used. B) Distribution of the in-sample clonal frequencies for each SARS-CoV-2 antigen specific sequence in HC and RA groups. While the majority of the clonotypes were relatively rare (~ 5E-5), a noticeable second peak in frequency can be found in RA patients at a frequency of 1E-4, and additional numbers of larger clones can also be seen.

FigureS15–TCRseq of SF from patients prior to SARS-CoV-2 emergence

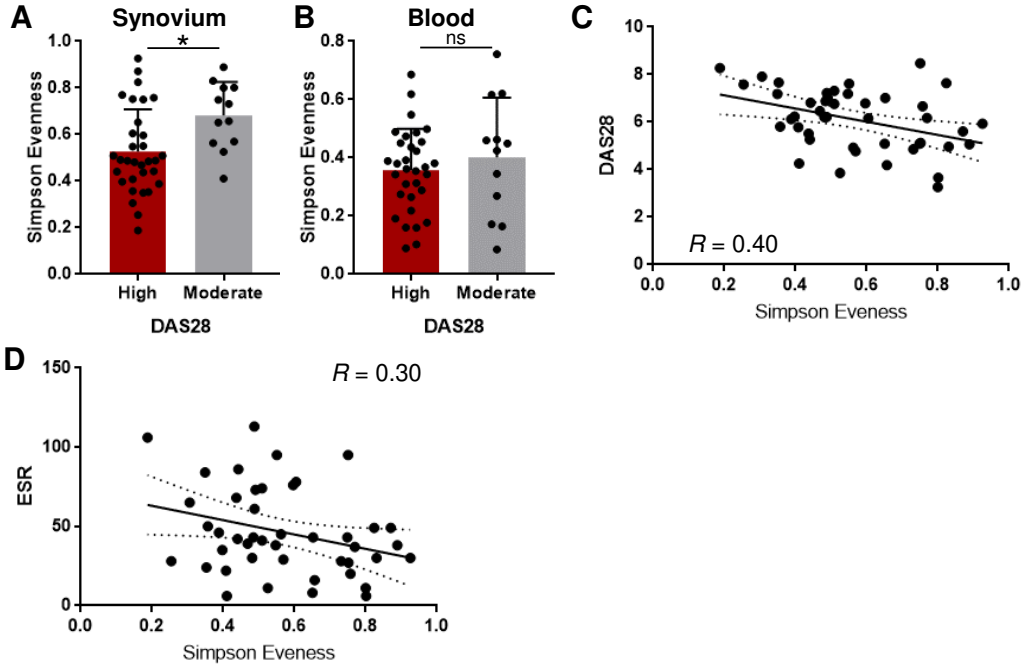
A) Shannon diversity of the CD57- samples is significantly higher than in the CD57+ samples. This is consistent with our understanding that CD57+ are dominated by effector T cells, while the CD57- pool may include more memory cells that have not undergone as dramatic of a clonal

expansion. B) Intersection analysis of the 8 TCRseq libraries reveals that a substantial number of clones shared across CD57+ and CD57- populations derived from the same individuals. Some public sharing across individuals can also be seen.

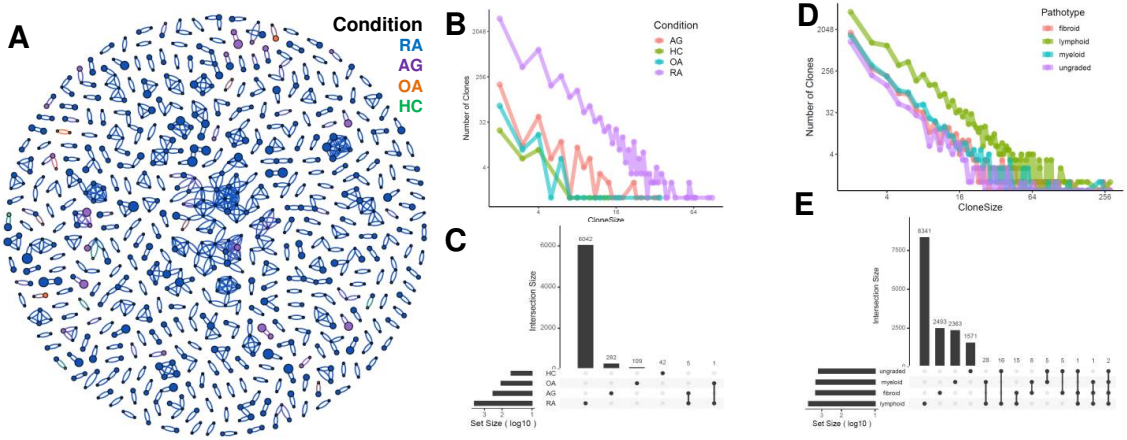
FigureS16–Spike-matched TCR sequences are found in TCRseq of SF from patients prior to SARS-CoV-2 emergence

Network visualization as in Fig1J, with nodes color split more finely into TCRseq sample origin and pathogen type. Notably, we can observe that several other antigens from SARS-CoV-2 (white nodes) can also be observed to form specificity clusters besides spike protein.

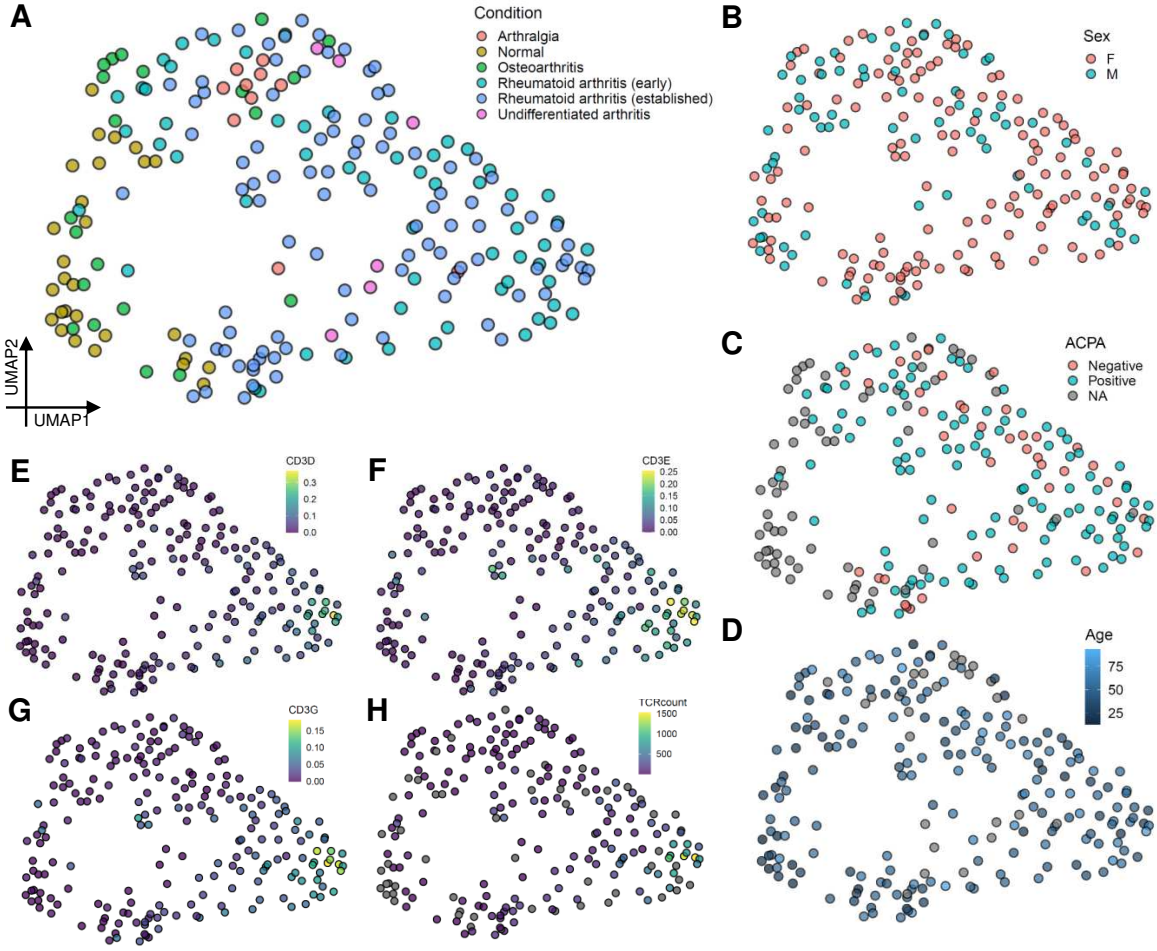
FigureS1–Synovial TCRβ α diversity is associated with higher disease activity levels



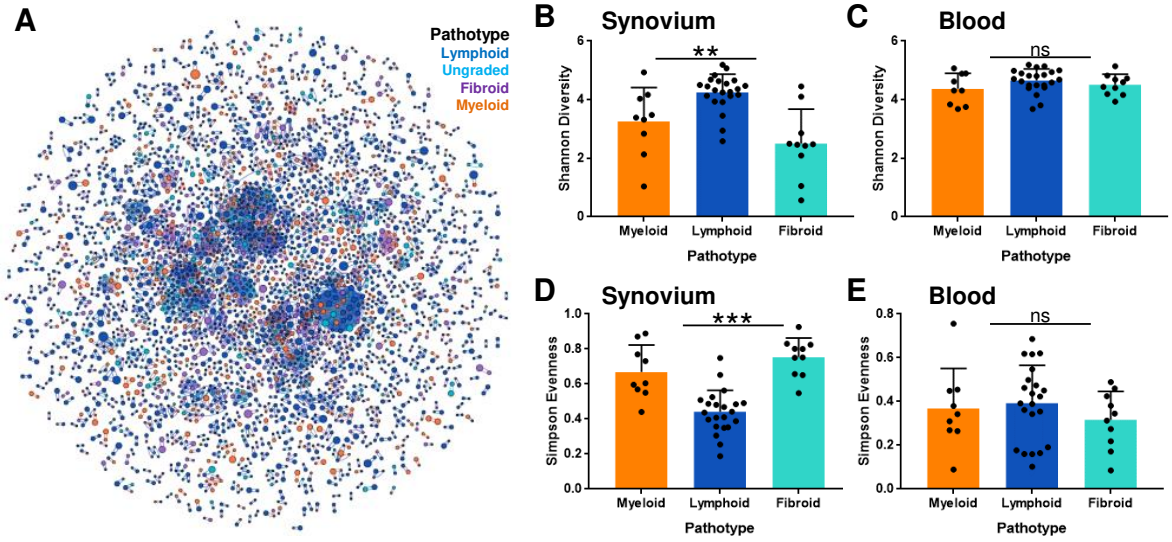
FigureS2– Clonal expansion is most prominent in RA patients with lymphoid pathotype



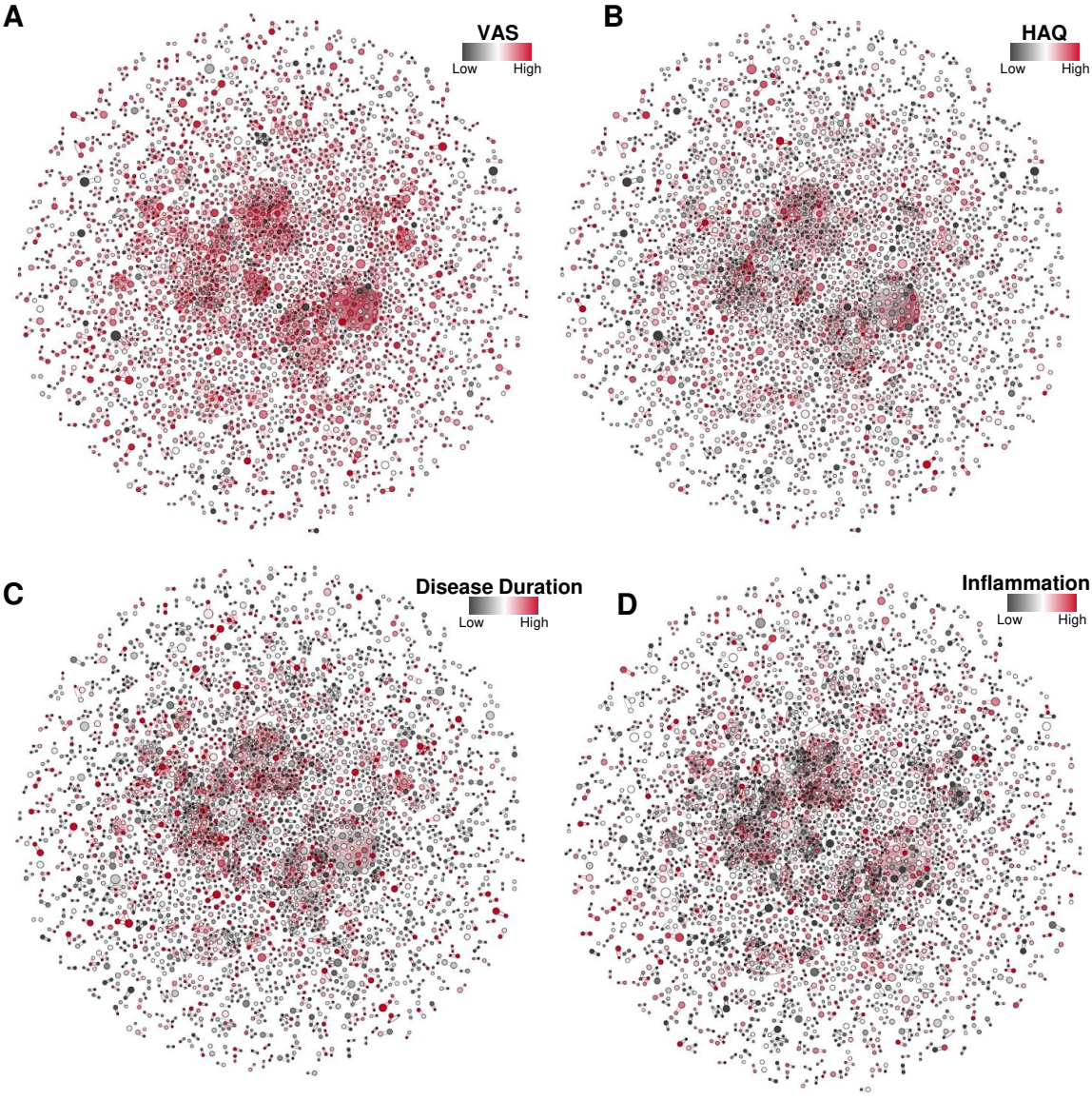
FigureS3: Transcriptome and repertoire differences between varying joint conditions in GSE89408 are independent of some common sources of variation



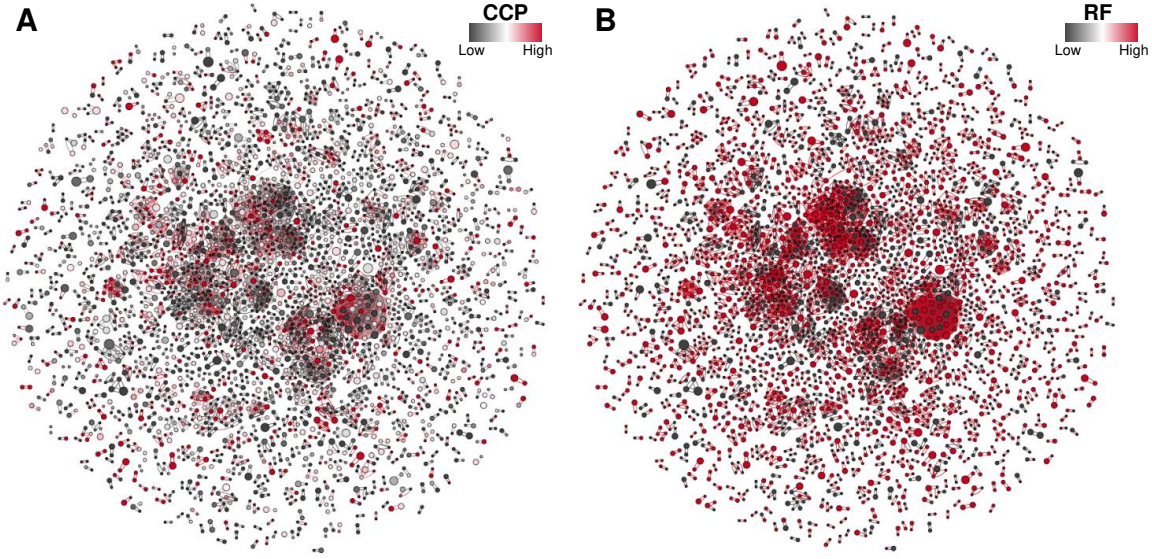
FigureS4–Synovial TCRβ α diversity is higher in lymphoid pathotype



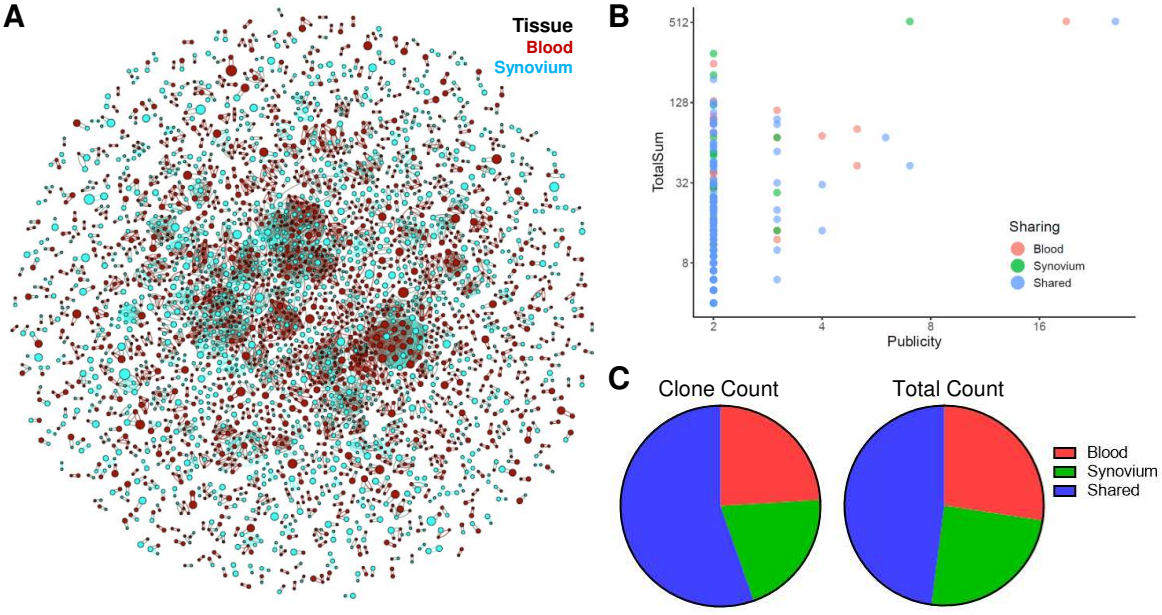
FigureS5– Motif analysis of RA samples stratified by disease activity assessment scores



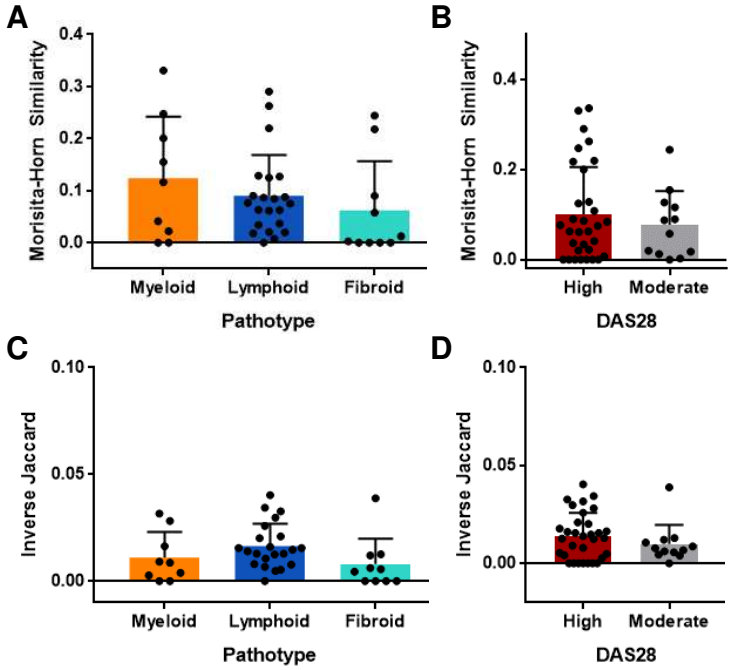
FigureS6– Motif analysis of RA samples classified by antibody levels



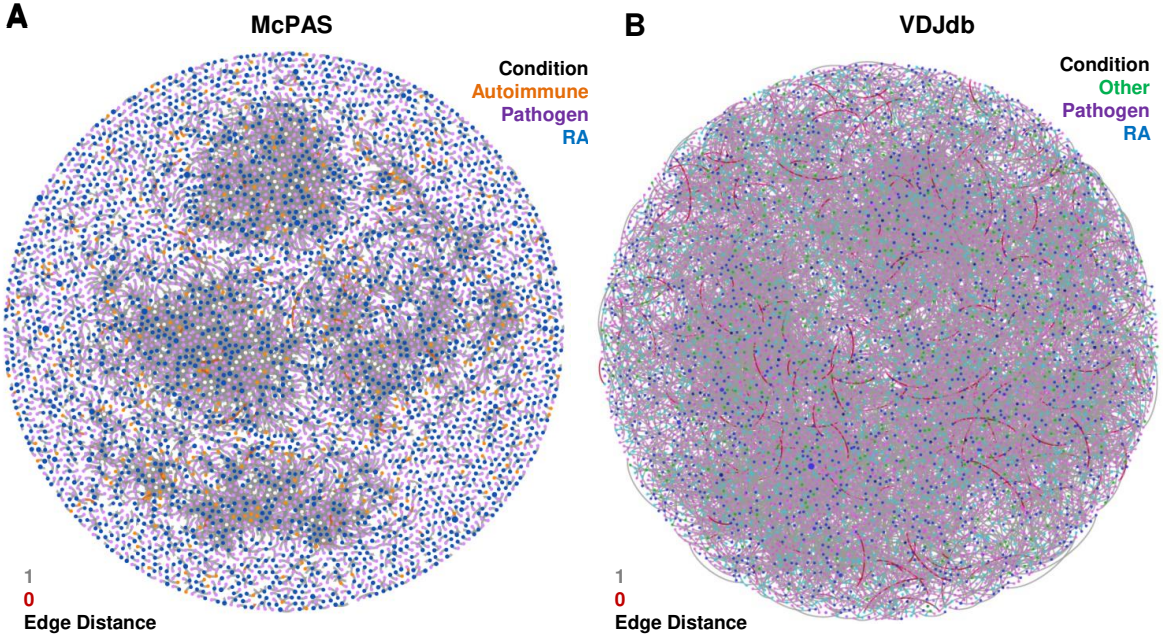
FigureS7–Synovial TCRβ clonotypes are not wholly independent from peripheral circulation



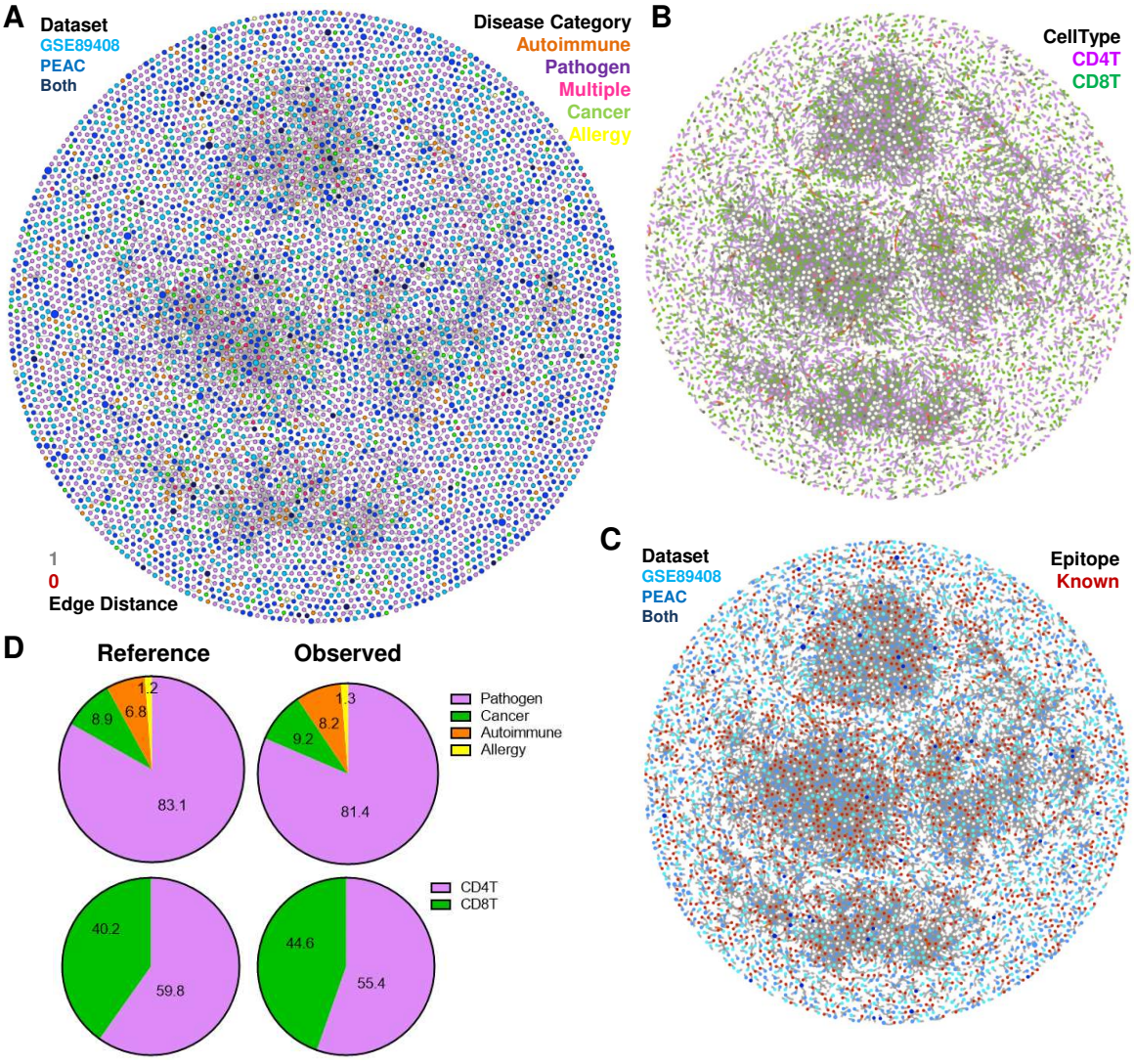
FigureS8–β diversity metrics in TCR repertoire data from paired blood and synovium samples fail to delineate between pathotypes and disease activity



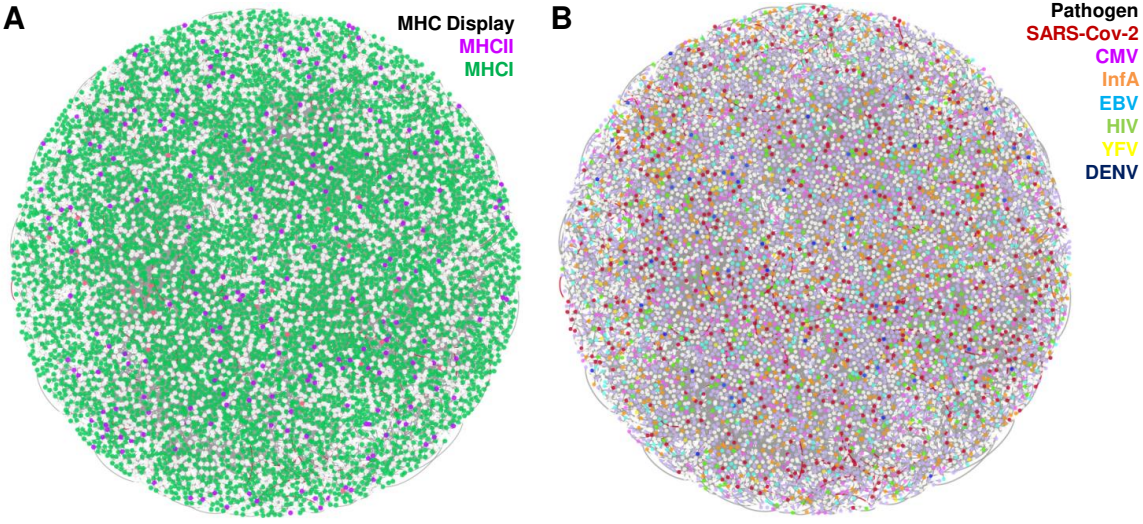
FigureS9–TCR repertoire analysis of rheumatoid arthritis patients includes sequences matching annotated pathogen-specific TCRs



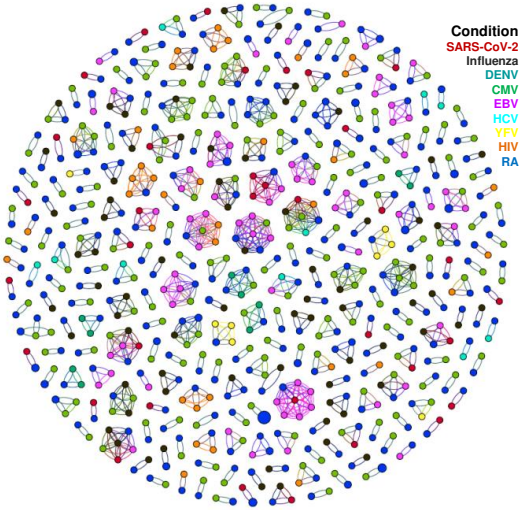
FigureS10–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the full McPAS database



FigureS11–TCR repertoire analysis of rheumatoid arthritis patients include annotated epitopes from the VDJdb database



FigureS12–Exact matches between TCR sequences recovered from RA synovium and annotated pathogen antigen-specific sequences



FigureS13–Spike-matched epitopes recovered from RA synovium and are conserved in beta coronaviruses

A

Virus	Start	End	Sequence	Accession
SARS-CoV-2	269	277	**YLQPRTFLL**	YP_009724390.1
MERS-CoV	317	325	**KLQPLTFLL**	YP_009047204.1
SARS-CoV	256	264	**YLKPTTFML**	YP_009825051.1

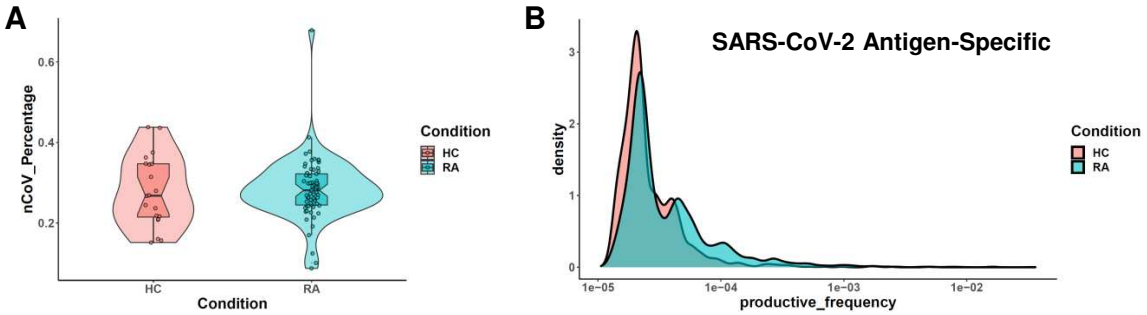
B

Virus	Start	End	Sequence	Accession
SARS-CoV-2	865	873	**LTDEMIAQY**	YP_009724390.1
MERS-CoV	1262	1270	**LTYEMLSLQ**	YP_009047204.1
SARS-CoV	847	855	**LTDDMIAAY**	YP_009825051.1

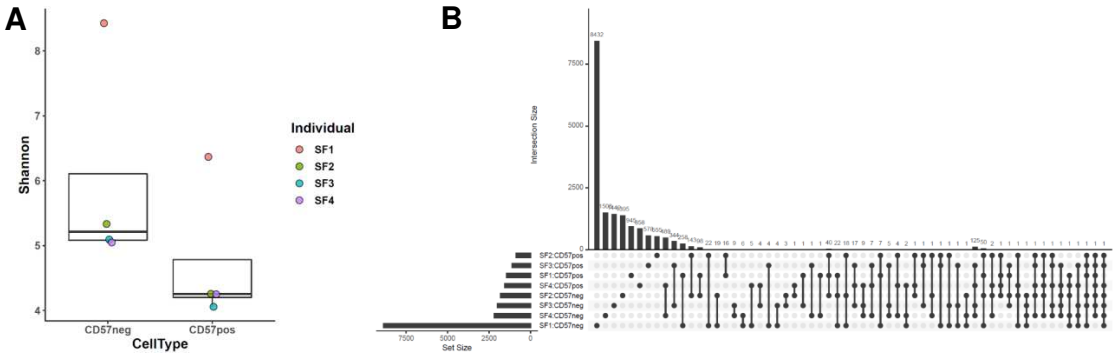
C

Virus	Start	End	Sequence	Accession
SARS-CoV-2	448	456	**NYNYLYRLF**	YP_009724390.1
SARS-CoV	435	443	**NYNYKYRYL**	YP_009825051.1

FigureS14–Spike-matched TCR sequences are found in TCRseq of PB from patients prior to SARS-CoV-2 emergence in a public dataset



FigureS15–TCRseq of SF from patients prior to SARS-CoV-2 emergence



FigureS16–Spike-matched TCR sequences are found in TCRseq of SF from patients prior to SARS-CoV-2 emergence