Title: Clonal associations of lymphocyte subsets and functional states revealed by single cell antigen receptor profiling of T and B cells in rheumatoid arthritis synovium

Authors:
Garrett Dunlap[1]†, Aaron Wagner[2]†, Nida Meednu[3]†, Fan Zhang[1,4,5,6,7,8], A. Helena Jonsson[1], Kevin Wei[1], Saori Sakaue[1,4,5,6,7], Aparna Nathan[1,4,5,6,7], Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network, Vivian P. Bykerk[9,10], Laura T. Donlin[9,10], Susan M. Goodman[9,10], Gary S. Firestein[11], David L. Boyle[11], V. Michael Holers[12], Larry W. Moreland[12,13], Darren Tabechian[3], Costantino Pitzalis[14], Andrew Filer[15], Soumya Raychaudhuri[1,4,5,6,7,16], Michael B. Brenner[1], Andrew McDavid[2]‡, Deepak A. Rao[1]‡*, Jennifer H. Anolik[3]‡*

Affiliations:
[1]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital and Harvard Medical School; Boston, MA, USA
[2]Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry; Rochester, NY, USA
[3]Division of Allergy, Immunology and Rheumatology, University of Rochester Medical Center; Rochester, NY, USA
[4]Center for Data Sciences, Brigham and Women's Hospital; Boston, MA, USA
[5]Division of Genetics, Department of Medicine, Brigham and Women's Hospital; Boston, MA, USA
[6]Department of Biomedical Informatics, Harvard Medical School; Boston, MA, USA
[7]Broad Institute of MIT and Harvard; Cambridge, MA, USA
[8]Division of Rheumatology and the Center for Health Artificial Intelligence, University of Colorado School of Medicine; Aurora, CO, USA
[9]Hospital for Special Surgery; New York, NY, USA
[10]Weill Cornell Medicine; New York, NY, USA
[11]Division of Rheumatology, Allergy, and Immunology, University of California, San Diego; La Jolla, CA, USA
[12]Division of Rheumatology, University of Colorado School of Medicine; Aurora, CO, USA
[13]Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine; Pittsburgh, PA, USA
[14]Centre for Experimental Medicine & Rheumatology, William Harvey Research Institute, Queen Mary University of London; London, UK
[15]Rheumatology Research Group, Institute for Inflammation and Ageing, University of Birmingham, NIHR Birmingham Biomedical Research Center and Clinical Research Facility, University of Birmingham, Queen Elizabeth Hospital; Birmingham, UK
[16]Versus Arthritis Centre for Genetics and Genomics, Centre for Musculoskeletal Research, Manchester Academic Health Science Centre, The University of Manchester; Manchester, UK

†These authors contributed equally to this work.
‡These authors jointly supervised this work.
Corresponding author.

Abstract

Rheumatoid arthritis (RA) is an autoimmune disease initiated by antigen-specific T cells and B cells, which promote synovial inflammation through a complex set of interactions with innate immune and stromal cells. To better understand the phenotypes and clonal relationships of synovial T and B cells, we performed single-cell RNA and repertoire sequencing on paired synovial tissue and peripheral blood samples from 12 donors with seropositive RA ranging from early to chronic disease. Paired transcriptomic-repertoire analyses highlighted 3 clonally distinct CD4 T cells populations that were enriched in RA synovium: T peripheral helper (Tph) and T follicular helper (Tfh) cells, CCL5+ T cells, and T regulatory cells (Tregs). Among these cells, Tph cells showed a unique transcriptomic signature of recent T cell receptor (TCR) activation, and clonally expanded Tph cells expressed an elevated transcriptomic effector signature compared to non-expanded Tph cells. CD8 T cells showed higher oligoclonality than CD4 T cells, and the largest CD8 T cell clones in synovium were highly enriched in *GZMK*+ cells. TCR analyses revealed CD8 T cells with likely viral-reactive TCRs distributed across transcriptomic clusters and definitively identified MAIT cells in synovium, which showed transcriptomic features of TCR activation. Among B cells, non-naive B cells including age-associated B cells (ABC), *NR4A1*+ activated B cells, and plasma cells, were enriched in synovium and had higher somatic hypermutation rates compared to blood B cells. Synovial B cells demonstrated substantial clonal expansion, with ABC, memory, and activated B cells clonally linked to synovial plasma cells. Together, these results reveal clonal relationships between functionally distinct lymphocyte populations that infiltrate RA synovium.

Introduction

Synovial inflammation in rheumatoid arthritis (RA) involves a complex set of interactions between immune and non-immune cell subsets, yet a core feature of the immune response in seropositive RA is an adaptive immune response against citrullinated proteins involving both antigen-specific B cells and T cells (Weyand and Goronzy 2021). The activation of B cells in RA has long been appreciated, given the characteristic production of disease-associated autoantibodies, including rheumatoid factor (RF) and anti-cyclic-citrullinated peptide (anti-CCP) antibodies (Roosnek and Lanzavecchia 1991; Schröder et al. 1996; Goldbach-Mansky et al. 2000; Marston, Palanichamy, and Anolik 2010). B cells may be activated locally within RA synovium, as synovial tissue studies have provided evidence of somatic hypermutation (SHM) and clonal expansion (Corsiero et al. 2016; Meednu et al. 2022). Synovial B cells may contribute antibody-independent functions as well, including antigen presentation and cytokine secretion, which involve interactions with other cell types (O'Neill et al. 2005; Wu et al. 2021; Sun et al. 2018; Meednu et al. 2016).

Populations of T cells have likewise been strongly implicated in initiating or maintaining synovial inflammation in RA (Panayi, Lanchbury, and Kingsley 1992). Genetic associations indicate a critical role for antigen presentation to CD4+ T cells via MHC class II in the development of RA (Gregersen, Silver, and Winchester 1987; Raychaudhuri et al. 2012). Cellular profiling studies of RA synovial tissue and fluid have highlighted a large population of T peripheral helper (Tph) cells, as well as T follicular helper (Tfh) cells, both of which provide help to B cells through the production of IL-21 and CD40L (Rao et al. 2017; Yoshitomi and Ueno 2021; Manzo et al. 2008; Kobayashi et al. 2013). Tph cells differ from Tfh cells in their migratory patterns, expressing chemokine receptors such as CCR2 and CCR5 to home to sites of peripheral inflammation such as the rheumatoid joint (Rao 2018; Fortea-Gordo et al. 2019). Large populations of CD8+ T cells also accumulate within RA synovium, including a prominent

granzyme K-expressing population, which may contribute to synovial inflammation through inflammatory cytokine production rather than cytotoxicity (Jonsson et al. 2022).

Complementing studies of specific subsets of immune cells, a holistic picture of both immune and non-immune populations in RA is emerging through single-cell RNA-sequencing (scRNA-seq) atlases of synovial tissue samples (Stephenson et al. 2018; Zhang et al. 2019; Alivernini et al. 2020; Zhang et al. 2022). These studies have highlighted the diversity of cell states present in the inflamed tissue of these patients, as well as how the presence and effects of these states may differ among patient subpopulations (Alivernini et al. 2020; Zhang et al. 2022). For lymphocytes, analyses of the T cell receptors (TCR) or B cell receptors (BCR) can provide unique insights into the expansion and developmental relationships of lymphocyte subsets, leveraging the feature that each new lymphocyte generates a unique TCR/BCR that is shared with all progeny. Studies tracking TCRs across tissues or longitudinally in RA patients have identified shared T cell clones in different joints (Musters et al. 2018; Klarenbeek et al. 2012), clonal expansion of specific cell subsets (Argyriou et al. 2022), persistence of expanded clones over time (Ishigaki et al. 2015), and overrepresented gene rearrangements that may suggest shared antigenic targets (Klarenbeek et al. 2012; Chang et al. 2019; X. Liu et al. 2019); however, reactivity of expanded TCRs from RA synovial CD4 T cells to citrullinated peptides has been difficult to demonstrate (Turcinov et al. 2022). Studies of BCR repertoires of RA patients have indicated somatic hypermutation in RA synovial B cells and identified potential specificities across B cells collected from synovial tissue or fluid (Scheel et al. 2011; Corsiero et al. 2016; Titcombe et al. 2018; Hardt et al. 2022). A comprehensive examination of both the T cell receptor (TCR) and B cell receptor (BCR) repertoires of synovial tissue lymphocyte populations and across tissue and blood at the single-cell level has not been described. Such studies have the potential to directly link clonal features to the functional roles, developmental relationships, and cell-cell interactions of specific lymphocyte phenotypes, as has been

achieved in studies of cancer immunotherapy and infectious disease (Luoma et al. 2020; Yost et al. 2019; Collora et al. 2022; Mathew et al. 2021).

Here, we implemented 5' droplet-based scRNA-seq on T and B cells of synovial tissue and matched peripheral blood samples from 12 RA patients to simultaneously study their transcriptomes and antigen receptor repertoires. Our study provides a high-resolution landscape of the clonal relationships within and between cell states, and further between inflamed synovial tissue and peripheral blood.

Results

Single-cell profiling of synovial tissue and peripheral blood lymphocytes

We collected synovial tissue (n = 12) and matched peripheral blood (n = 10) from individuals with RA that comprised a subset of a larger cohort analyzed as part of the Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network, prioritizing samples with high synovial cell yields and evident lymphocyte populations by flow cytometry (Zhang et al. 2022). Donors had a mean age of 63.6 years (range 28-80) and were predominantly female (n = 11). Patients in the cohort had a Clinical Disease Activity Index (CDAI) classification of moderate or high with a mean score of 26.4 (range 10.5-64.0), and were mainly lymphoid in pathotype, but otherwise had a range of disease duration, treatment, and cell type abundance phenotype (CTAP) **(Figure S1A, Table S1)** (Canhão et al. 2018; Humby et al. 2019; Zhang et al. 2022).

Tissue samples were previously disaggregated into single-cell suspensions for an unbiased analysis of the cell states present in RA synovium using RNA and cell surface protein profiling (Zhang et al. 2022). Cryopreserved synovial cells remaining after the initial analysis were thawed and sorted to isolate CD45+CD3+ and CD45+CD19+ populations, which were subsequently encapsulated into droplets and used to generate gene expression, cell-surface protein, and TCR/BCR single-cell sequencing libraries **(Figure 1A, S1B)**. Cryopreserved

PBMCs were thawed and sorted in parallel. Following a unified single-cell analysis pipeline of all samples, we recovered a total of 84,750 cells. After applying QC criteria, 83,159 cells remained, which we used to perform an initial round of unsupervised clustering at low resolution **(Figure 1B, S2)**. On average, we obtained 2,888 cells per synovial tissue sample (range 804-5,188) and 4,851 cells per blood sample (range 3,442-6,601) **(Figure 1C, S1A)**.

Through parallel examination of gene expression and protein detection, we identified several populations of T cells expressing markers such as *CD3E* and *IL7R*, as well as a cluster containing B and plasma cells expressing *CD79A* and CD20 **(Figure 1D, S3A, B)**. Cells from both lineages could be found in all samples **(Figure 1E,F)**. We further identified a cluster containing proliferating T and B cells, characterized by the expression of markers including *MKI67* and *TYMS*. Lastly, we identified two rare populations of contaminating cells expressing markers of fibroblasts (*PRG4* and *FN1*) and monocytes (*S100A8* and *LYZ*), which were excluded from subsequent analyses **(Figure 1D, S3A,B)**.

Of the 81,708 lymphocytes captured across all samples, we obtained paired TCR or BCR information for 73,185 cells. In the synovial tissue, an average of 84.7% of lymphocytes per sample had an associated TCR or BCR (range 72.9-96.1%), while in the blood, 93.1% of lymphocytes per sample had this information on average (range 89.3-96.8%) **(Figure 1G, S3C)**.

Effector CD4+ populations are significantly enriched in synovial tissue

Sub-clustering of the 35,301 CD4+ T cells obtained from synovial tissue and blood samples identified 14 CD4+ T cell subsets **(Figure 2A,B, S4A, Table S2)**. We identified two clusters expressing the B cell chemoattractant *CXCL13,* one of which was composed solely of Tph cells, while the other included some cells with detectable *CXCR5*, suggesting the cluster contained a mixture of Tfh and Tph cells **(Figure S4B)** (Rao et al. 2017; Zhang et al. 2019, 2022). Both clusters expressed genes associated with B cell help, such as *MAF,* and both scored highly for a Tfh cell gene signature **(Figure 2C,D, S4C, Table S3)**. While the Tfh/Tph cluster displayed higher expression of markers such as *IL7R* and *CD69* compared to the Tph

cluster, the Tph cluster had significantly elevated expression of *PDCD1, CTLA4, LAG3,* and others. Further, the pure Tph subset produced higher gene scores of TCR signaling and T cell activation compared to other CD4+ clusters, suggesting that the cells in this cluster are activated **(Figure 2D,E, Table S3)**.

Several other clusters of memory cells with distinctive expression of chemokines, chemokine receptors, and granzymes were present, including *CCR7*+ (which likely represents central memory cells), *IL7R*+*CCL5*+, *GZMA*+*CCL5*+, and *GZMK*+ populations. In addition, we identified a population of cytotoxic CD4+ cells, marked by the expression of *GNLY, PRF1,* and *GZMB,* as well as a cluster characterized by a strong interferon-response signature. A small population likely containing a mix of naive and memory cells was distinguished primarily by increased expression of members of the GIMAP family (e.g. *GIMAP4*, *GIMAP5*), which has been associated with survival and quiescence in lymphocytes **(Table S2)** (Barnes et al. 2010).

Sub-clustering also revealed two populations of regulatory T cells (Tregs). Both expressed *FOXP3, CTLA4,* and *TIGIT,* though one population was marked by stronger expression of *IL2RA* (encoding CD25) and IL32, while the other displayed higher *CCR7* and *TCF7,* suggesting the presence of populations of effector and central memory-like Tregs, respectively (Shevyrev and Tereshchenko 2019). A large cluster containing naive T cells expressing *SELL, TCF7,* and *CCR7* displayed the strongest expression signature for a previously-identified naive T cell gene set (Abbas et al. 2009) **(Figure 2B-D, S4C)**. Lastly, an actively proliferating cluster of cells, as well as a subset with elevated mitochondrially encoded RNAs could be detected **(Figure 2C,D, S4C)**. In total, T cell populations identified by this present clustering analysis were well aligned with those identified in analyses of a larger set of synovial tissue samples (of which the 12 synovial samples studied here are a subset), where clustering was performed through a combination of RNA and cell surface protein profiling, **(Figure S4D)** (Zhang et al. 2022).

We next identified T cell subsets that had differential representation in either synovial tissue or peripheral blood. Several memory/effector cell populations were enriched in synovium compared to blood, including the Tph, Tfh/Tph, CD25-high Treg, IL7R+CCL5+ memory, GZMA+CCL5+ memory, GZMK+ memory, and proliferating clusters **(Figure 2F, S4E)**. In contrast, blood samples contained increased abundances of the naive, GIMAP+, and interferon-stimulated clusters. On average, 45% of the CD4+ population for each blood sample was composed of cells from the naive cluster (range 18-66%), while only 5% of synovial tissue CD4 cells had a naive phenotype (range 2-15%). The increased proportions of many memory/effector populations in synovium support the importance of these cells in propagating inflammation within the synovium.

Clonally expanded Tph cells display increased effector and cytotoxic profiles

We then sought to examine the TCR repertoire of the CD4+ populations across synovial tissue and blood. Among the CD4+ populations, the GNLY+ cytotoxic cluster displayed the highest clonal expansion and comprised nearly all clones larger than 20 cells. Clonal expansion, defined as two or more cells with an identical TCR, was also identified among the Tph and GZMA+CCL5+ memory clusters **(Figure 2G, S4F,G)**. An analysis of clonal sharing among the synovial tissue clusters revealed a high degree of clonal overlap between the Tph and Tfh/Tph clusters, suggesting that cells in these clusters are developmentally related despite their transcriptomic differences **(Figure 2H,I)**. In addition, the Tph cluster shared the highest clonal overlap with the cycling cluster compared to all other subsets, highlighting the active proliferation of the Tph population within the synovial tissue. Conversely, among the blood CD4+ cells, we noted clonal sharing across the memory populations and also found the GZMA+CCL5+ subset to have elevated clonal overlap with the proliferating blood CD4+ T cells **(Figure 2H,I)**.

After identifying the Tph population as highly represented in the synovial tissue, clonally expanded, and actively proliferating, we sought to identify features that differentiate clonally expanded and unexpanded Tph cells in the tissue. Across donors, an average of 12.5% of Tph cells belonged to an expanded TCR clone (range 4-32.5%) **(Figure 2J)**. By examining transcriptional differences between these cells and Tph cells belonging to an unexpanded clone, clonally expanded Tph cells showed significantly increased expression of signatures of effector function and cytotoxicity, including elevated expression of *IFNG, PRF1, CD40LG,* and *CCL5* **(Figure 2K,L)**. In contrast, no difference in the Tfh gene signature score was present between expanded and unexpanded cells, suggesting that clonally expanded Tph cells do not lose their B cell-helping functions. Interestingly, a similar comparison using cells from the Tfh/Tph cluster yielded no significant differences between expanded and unexpanded cells, further suggesting a unique set of features among cells in the Tph clusters **(Figure 2M)**.

Expanded CD8+ clones across tissues largely differ by patterns of GZMK and GZMB expression

Next, we isolated and subclustered the CD8+ T cells present in the synovial tissue and blood samples, which revealed 9 CD8+ T cell populations that could be found across samples **(Figure 3A,B, S5A, Table S2)**. Among these, 3 populations were distinguished by expression patterns of *GZMK* and *GZMB,* all of which had strong corresponding matches to clusters identified in a recent study of synovial tissue **(Figure 3D, S5C)** (Zhang et al. 2022). One of these clusters solely expressed *GZMK*, while another expressed GZMK and low levels of *GZMB*. Both of these populations also had elevated *GZMA* and *CCL5*, while the GZMK/B+ cluster was further differentiated through increased expression of markers, including *CCL4* and *HLA-DRA*, and decreased *IL7R*. The third population, characterized by expression of *GZMB* only, appeared highly cytotoxic through gene module analysis and elevated expression of *GNLY* and *PRF1* **(Figure 3C, D, S5B)**. We also isolated a population of likely resident memory CD8+ T (Trm) cells, characterized by increased *ZNF683* and *XCL1*, which was confirmed more broadly

through examination of a previously-published Trm gene list **(Table S3)** (Zhao et al. 2020). We identified one naive CD8+ population with high expression of *SELL* and *LEF1* and a naive gene module signature **(Figure 3C,D, S5B)**. Another population of likely naive cells was also identified, characterized by increased expression of *IGTB1* and *LMNA.* Lastly, we found separate clusters of CD8+ T cells with elevated interferon-response, proliferation, or mitochondrial gene modules **(Figure 3C,D, S5B)**.

An analysis of cluster representation between synovial tissue and peripheral blood samples found the GZMK/B+ cluster to be highly increased in synovial tissue samples **(Figure 3E, S5D)**, consistent with a recent report (Jonsson et al. 2022). Clusters solely expressing *GZMK* or *GZMB* were not significantly different in abundance between tissue compartments, though the GZMK+ memory population trended higher in synovial tissue, while the GZMB T_{EMRA} cluster appeared slightly elevated in the blood **(Figure S5D)**. Additionally, the Trm, proliferating, and mitochondrial-high clusters had elevated representation in synovial tissue samples. In contrast, the naive and ITGB1-elevated populations could both be found at higher frequencies in the blood **(Figure 3E, S5D)**.

After characterizing the cell states present among the CD8+ population, we sought to connect these clusters' clonal attributes. Broadly, we found a much larger degree of clonal expansion in the CD8+ compartment compared to the CD4+ subsets. We found the strongest clonal expansion among the GZMK/B+ memory and GZMB+ T_{EMRA} clusters, and also noted cells in the ISG-high and proliferating clusters to belong to expanded clones **(Figure 3F, S5E)**. Comparing the top 50 largest clones across the synovial tissue and blood revealed striking differences in cluster composition. Cells from the top 50 clones in the synovial tissue belonged overwhelmingly to the GZMK/B+ memory population. In contrast, the top 50 clones in the blood belonged to the GZMB+ T_{EMRA} cluster, with minor representation of other populations **(Figure 3G,H)**. This analysis of the largest clones, and a broader examination of the clonal overlap between subclusters in either tissue, showed sharing between the GZMK/B+ memory cluster

and proliferating cells of the synovial tissue, while the GZMB+ T_{EMRA} was found to overlap with the proliferating component of the blood **(Figure S5F,G)**. To further analyze this, we subsetted the proliferating cluster and mapped these cells onto the other CD8+ cells of this dataset. Compared to the general cluster proportions, the tissue and blood proliferating cells mapped heavily to the GZMK/B+ (201/261 tissue proliferating cells) and GZMB+ T_{EMRA} (24/38 blood proliferating cells) clusters, respectively, suggesting an active, though likely differing, role of these two populations in their distinct tissues **(Figure 3I)**.

A further examination of clonal overlap across tissues revealed a small degree of sharing between the tissue and blood components of either the GZMK/B+ memory or GZMB+ T_{EMRA} clusters separately, but clonal overlap between synovial tissue GZMK/B+ cells and blood GZMB+ T_{EMRA} cells was essentially absent **(Figure S5H)**. Of the top 50 clones in both tissue compartments, only 6 could be found in both, nearly all of which were larger in the blood than the synovial tissue. Notably, 4 of these clones retained a GZMB+ skew even in the synovial tissue **(Figure 3G,H)**. Together, these results suggest functionally distinct roles for the expanded GZMK/B+ cells of the synovial tissue and the GZMB+ T_{EMRA} cells of the blood. The findings corroborate recent work suggesting that the GZMK/B+ cells likely do not arrive in the synovial tissue as *GZMB*+ cytotoxic T cells that subsequently alter their phenotype, and instead may enter the synovium as *GZMK*+ cells or expand locally in the tissue as they adopt a phenotype that includes *GZMK* expression (Jonsson et al. 2022).

Predicted virus cross-reactive CD8+ T cells do not display altered expansion or phenotypic characteristics

Viral infection has long been connected with the potential for the development and maintenance of autoimmune diseases (Fujinami et al. 2006). One possibility that may instigate this phenomenon is the presence of cross-reactive epitopes between a virus and endogenous proteins, which may drive the activation and expansion of a set of T cells. We sought to identify potentially viral-specific T cells in the synovial tissue to evaluate the extent of potential viral

cross-reactivity among RA synovial CD8 T cells. To accomplish this, we gathered databases of previously identified CMV, EBV, and influenza A-specific T cells (Tickotsky et al. 2017; Shugay et al. 2018). We then identified exact matches with the beta chain CDR3 sequence and HLA between the RA synovial CD8+ T cells analyzed here and those in the database **(Figure S6A)**. Although three of the RA patients (RA04, RA05, and RA07) expressed less common HLA alleles, precluding the ability to find matches with previously-discovered specific clones, we were able to identify matching clones in the majority of patients **(Figure 3J, S6B)**. Within and between patients, we further found these matches to be directed against multiple different epitopes for each of the viruses tested **(Figure S6C)**. When examining which cluster these matching T cells belong to, we found a spread across clusters, with no cluster being significantly overrepresented compared to non-matching cells **(Figure 3K)**. Further, we noted that few matching clones were expanded (16/56), with the largest matching clone comprising only 7 cells.

Requiring an exact match of clone sequence and HLA can provide strong evidence of the capacity for reactivity against a virus; however, identifying shared motifs between viral-specific T cells and those in our dataset may allow the identification of a larger set of T cells with the potential to detect viral epitopes. Thus, we employed GLIPH2 to identify motifs within the beta chain CDR3 sequence enriched in viral-specific T cells and our CD8+ T cells (Huang et al. 2020). After running the GLIPH2 algorithm individually with each virus (CMV, EBV, and influenza) for RA synovial T cells, we filtered motifs that contained both virus and RA clones, and only those with an HLA match **(Figure S6D)**. We obtained a larger number of clones belonging to a viral-associated GLIPH motif (369 unique clones) compared to our exact matches (56 unique clones), which was variable across donors **(Figure 3L, S6E,F)**. To strengthen the association of the hits within our data, we sought to identify a relationship between patient age and the percent of CD8+ T cells associated with potential viral reactivity. Though not significantly correlated, we found trends between donor age and the percent of

potential CMV and EBV reactive clones, consistent with the dynamics of the anti-viral repertoire with age (Khan et al. 2002; Looney et al. 1999) **(Figure S6G)**. Similar to results with exact matches, a breakdown of the cluster makeup of motif-matching cells and non-matching cells again revealed no significant differences **(Figure 3M)**. Further, we identified no differences in the clone size distributions between these two groups, and also found no clones belonging to viral GLIPH motifs among the 50 largest synovial tissue clones **(Figure 3N)**.

Identification of activated innate T cell populations in RA synovium

Populations of innate T cells have previously been associated with RA, including natural killer (NK), γδ, and mucosal-associated invariant T (MAIT) cells, though clear roles for many of these subsets in RA remain elusive (M. F. Liu et al. 1999; Andreu et al. 1991; Yamin et al. 2019; Koppejan et al. 2019). We identified and subclustered innate T cells in the synovial tissue and blood samples, resulting in 7 subsets representing multiple innate lineages **(Figure 4A, S7A, Table S2)**. Two populations of γδ T cells, including a Vδ1 subset characterized by expression of *TRDV1* and a Vδ2 subset expressing *TRDV2* and *TRGV9* (the TCR γ variable gene commonly paired with TRDV2), were retrieved (Lawand, Déchanet-Merville, and Dieu-Nosjean 2017). The Vδ1 population had a elevated expression of *GZMB* and *TIGIT,* while Vδ2 cells had higher levels of *TYROBP*. MAIT cells were also identified among the innate cells, expressing markers including *SLC4A10, AQP3,* and *ZBTB16* **(Figure 4B,C)**. Two populations of NK cells were detected, including CD56-dim and CD56-bright subsets, that aligned with corresponding NK populations in a scRNA-seq reference **(Figure S7C)** (Zhang et al. 2022). These CD3- subsets were not intended to be included in the sorting scheme used in this study; thus, their frequency may not reflect the true representation in these samples. In addition to these clusters, we also found a population characterized by the expression of *ZNF683* (encoding Hobit), which appears to contain both γδ and NK cells and a population of innate cells with elevated mitochondrial gene expression **(Figure 4B)**. When comparing the frequency of these populations between synovial tissue and peripheral blood, only the Vδ1 subset had significantly increased

representation in the synovium, comprising a mean of 2.1% of T cells in the tissue (range 0.1-4.5%) and 0.5% (range 0.1-1.4%) in the blood **(Figure 4D, S7D)**.

Leveraging the dataset's repertoire information, we confirmed the presence of MAIT cells through an examination of the TCR alpha chain rearrangements. The MAIT population was characterized by its use of TRAV1-2, often accompanied by TRAJ33, TRAJ20, or TRAJ12 rearrangements (Mori, Lepore, and De Libero 2016). The pairing of the TRAV1-2 and TRAJ33/20/12 gene rearrangements could be detected in over half of the cells with an associated TCR in the cluster, while it was largely absent across other T cells in the dataset **(Figure 4E, S7E,F)**. We then sought to clonally track these MAIT cells between tissue and blood and found expanded clones uniquely represented in either tissue, as well as a subset of clones that were present in both synovium and blood **(Figure 4F, S7G)**.

To characterize transcriptional differences between the clonally expanded and non-expanded MAIT across these tissues, we examined signatures of activation. Synovial MAIT cells had elevated activation scores compared to MAIT cells in the blood **(Figure 4G)**. A similar pattern was observed with Vδ1 and Vδ2 cells **(Figure S7H)**. Because MAIT cells are known to become activated through both TCR-dependent and -independent mechanisms (Hinks and Zhang 2020), understanding how this relates to differences between tissues is relevant to better decipher the potential role of these cells in RA synovium. Gene signatures of TCR-dependent and -independent MAIT activation from multiple independent sources showed significantly enriched scores for both mechanisms in synovium compared to blood (Leng et al. 2019; Lamichhane et al. 2021). Both clonally expanded and non-expanded MAIT cells from synovial tissue had higher signatures for both mechanisms compared to their blood counterparts, but no significant difference could be detected between the expanded and non-expanded subsets of either tissue, suggesting subsets of MAIT cells in synovial tissue may become activated by either TCR-dependent and independent mechanisms **(Fig 4H,I, S7I,J, Table S3)**. Together,

these transcriptomic and repertoire data provide strong evidence for the presence of a defined MAIT cell population within RA synovium that appears activated.

B cells exhibit tissue-specific enrichment

Analogous to the T cell analysis, we characterized 27,869 B cells through subclustering to obtain 8 B cell and 2 plasma cell populations **(Figure 5A-C, Figure S8A-C, Table S2)**. We annotated 4 of these populations as naive subsets based on relatively higher expression of naive markers such as IgD and *TCL1A*, strong mapping to naive B cells from blood and tonsil in the published literature **(Figure 5E)** and recently identified in RA synovial tissue using RNA and surface protein expression **(Figure S8B)** (Zhang et al. 2022). Two of the naive B clusters were distinguished by relative levels of IgD expression (Naive-IgD-low, Naive-IgD-high). The Naive-IgD-low is also distinguished by higher *FCER2* (CD23). The other two naive clusters had elevated expression of *HSPA1B* (Naive-*HSPA1B*+) or higher mapped mitochondrial reads (Naive-MT-high) and had weaker mapping to a naive B cell state **(Figure 5E, Figure S8B,D)**. These latter two clusters were the dominant naive population in the synovium. Naive-*HSPA1B*+ has an activated phenotype with up-regulation of *NR4A1* and *DNAJB1* in addition to *HSPA1B*. Given the upregulation of *ZEB2* and *ITGAX* (CD11c) **(Figure S8D,E)**, this subset resembles an activated naive B cell described as expanded in the blood of patients with lupus (Jenks et al. 2018). The other six clusters are non-naive and dominate the synovial B cells. On average, 64.5% of the B cells for each blood sample were composed of cells from the naive clusters (IQR 59.5 - 86.7%), while only 9.1% of synovial tissue B cells had a naive phenotype (IQR 5.5 - 12.0%). We identified a cluster of memory B cells expressing memory markers, including *CD27*, *TNFRSF13B, S100A10,* and *S100A4* **(Figure 5B, S8C-E)**. This memory cluster maps to a peripheral blood memory B cell signature **(Figure 5E)** and the switched memory population described by Zhang et al. **(Figure S8B)**. Finally, many of these B cells are class-switched based on V(D)J sequencing **(Figure 5H)**.

Additional B cell populations included age-associated B cells (ABCs), activated-, and *LILRA4+ B* cells. Age-associated B cells (ABCs) expressed the canonical markers CD11c *and FCRL5* as previously described **(Figure 5B, S8C)** (Zhang et al. 2019, 2022; Jenks et al. 2018; Wang et al. 2018), as well as other ABC makers such as *TBX21* (T-bet*)* and *ZEB2* **(Figure 5B and S8D)**. Interestingly, one of the top differentially expressed genes in the ABC cluster is *IFI30*, highlighting activation of interferon signaling as a potential driver of ABCs in situ **(Figure S8C)**. The B-cell activated cluster was annotated based on high expression of *NR4A1,* related early response and germinal center light zone (GC LZ) genes **(Figure 5B,E, Table S3)**. Notably, both ABCs and activated B cells showed evidence of class-switch recombination **(Figure 5H,I)**, consistent with previous reports (Meednu et al. 2022; Jenks et al. 2018). Another small B cell population was named clonal-*LILRA4+* based on high BCR clonality and high *LILRA4* and G-protein signaling molecules. This small population only appeared in three PBL samples, with 87 out of 89 total cells belonging to one sample, and was omitted from downstream analyses. We used *XBP1* and CD27 expression to identify plasmablast (*XPB1+CD27int*) and plasma cells (*XBP1+CD27hiIgA+IgG+)*. Plasmablasts also mapped to a GC dark zone (GC DZ) signature **(Figure 5E)**, likely reflective of a high proliferation state (see Figure S8A for specific genes). Both populations exhibited high levels of class-switch recombination and SHM **(Figure 5G-I, Table S3)**.

Next, we tested each population for enrichment in the blood versus synovium. All populations were found in both the blood and synovium and across multiple donors **(Figure S8A)**. Naive-*HSPA1B+*, Activated B cells and plasma cells were significantly more abundant in the synovium. In contrast, Naive-IgD-low, Naive-IgD-high, B-memory, ABCs, and Naive-MT-high clusters were significantly enriched in the blood **(Figure 5D, S10D)**. The increased proportions of activated B cell populations in synovium support the active participation of these cells in synovial immune responses. We next sought to identify features that differentiate synovial from blood B cells by performing GSEA **(Figure S10B)**. Many of the B cell populations in the

synovium showed enrichment for GO pathways associated with cell activation and cytokine-mediated signaling **(Figure S10B)**. The former is consistent with the overall activated state of B cells in the synovium. It is interesting that the most significant enrichment for cytokine-mediated signaling in any synovial B cell state is within the plasmablasts, suggesting that cytokine signaling is a critical mediator of plasmablast generation in the tissue **(Figure S10B)**.

Accumulation of somatic hypermutation and class switch in synovial B cells

Utilizing the paired repertoire BCR data, we examined the degree of SHM in each of the previously described cell populations. We calculated the percent of SHM in the V/J region of the CDR3 for both heavy and light chains, then compared the rate between blood and synovial cells for each population **(Figure 5F)**. Mutation rates differed between cell states. Activated B cells, ABCs, memory, plasmablasts, plasma cell, and MT-high populations all had significantly higher levels of SHM compared to the Naive-IgD-high population as a reference **(Figure 5G)**.

Interestingly, 3 naive-like B cell populations – Naive-IgD-low, Naive-*HSPA1B*+, and Naive-MT-high – had significantly higher mutation rates in the synovium compared to blood, suggesting a spectrum of *in situ* naive B cell activation as we recently described (Meednu et al. 2022). Notably, ABCs and activated B cell populations also exhibited significantly higher mutation in the synovium, suggesting *in situ* activation and selection. The only population that had significantly higher mutation rates in the blood were plasma cells. Naive-IgD-high, memory, and plasmablasts did not show significance for tissue-specific mutation differences **(Figure 5F)**.

Additionally, we examined the amount of class-switching occurring within each cell population and between blood and synovial cells. Synovial B cell states exhibited significantly more class-switched BCRs (IgG or IgA+, IgD-) compared to their blood counterparts across all populations, except for naive IgD-high B cells **(Figure 5H)**. All cell populations showed a statistically significant increase in the amount of class-switched BCRs compared to the least class-switched population: naive IgD-high B cells **(Figure 5I)**. Of note, over 50% of the synovial ABCs showed evidence of class-switching.

Evidence of *in situ* antigen exposure and clonal expansion in synovial B cells

Using the BCRs recovered in our dataset, we identified groups of clonally-related B cells by quantifying the similarity between their IgH CDR3 sequences. Cell populations that were more highly mutated produced larger B cell clones, as has been reported in analyses of blood B cell repertoires from healthy individuals **(Figure 6A,B)** (DeWitt et al. 2016). For example, plasmablasts and plasma cells had a higher proportion of large clones (20-100 cells) than other populations, consistent with antigen-driven clonal expansion **(Figure 6A,B)**. For plasmablasts/plasma cells, clonal expansion was more prominent in the blood than in the synovium **(Figure 6B)**. This suggests that plasma cells in the synovium are experiencing different selection pressures than those circulating in the periphery. In the synovium, B cell populations other than plasma cells had mostly singletons or, within the activated B cells, ABCs, and MT-high, smaller clones (2-5 cells). Overall, the presence of both clonal expansion and higher SHM is consistent with stimulation and provision of T cell help within these more activated B cell populations. Further, consistent with an antigen-experienced B cell repertoire, activated B cells and ABCs had shorter CDR3 length and higher charge overall in both the synovium and blood **(Figure S10A)** (Meffre et al. 2001). In the blood, clones were also mostly singletons but with a proportion of small, medium, and even large clones detected in the activated, memory, MT-high, ABCs, and Naive-*HSPA1B*$^{+}$ B cells **(Figure 6B)**.

To assess clonal relationships between cell types, co-occurrence of expanded clone members between cell types was reported for each clone that contained a member within two different cell types. Though the vast majority of clones were contained within a specific population, we did identify clones that were shared between populations. **Figure 6C** depicts the clonal sharing across populations within each compartment. Within the synovium, plasma cells and plasmablasts share a large number of clones (24 clones, **Figure 6D**), strongly supporting a developmental relationship between newly generated plasmablasts and more mature plasma cells. Notably, we also observed shared clones between ABCs and activated B cells, as well as

between both these B cell states and the plasma cells. A small number of clones are also shared between the MT-high B cell state and both activated B cells and ABCs (**Figure 6C,D**). Overall, this data suggests a developmental relationship from Naive-MT-high (a naive B cell population already showing some signs of antigenic stimulation based on higher SHM rates compared to resting naive - **Figure 5G)** alternatively down an ABC vs. activated B cell pathway, as well as between these cell states and then downstream to plasma cells. It is also of interest that memory B cells share clones with activated B cells as well as differentiate to plasma cells, consistent with memory B cells participating in synovial immune reactions (**Figure 6C,D**). Within the peripheral blood, there is substantially more clonal sharing between multiple cell states **(Figure 6C)**. This is also evident when examining the cell population composition of the 50 most dominant clones in the blood compared to the synovium **(Figure 6F,G)**. As an example, the 5th most expanded blood clone was observed in the ABC, plasmablast, Naive-IgD-low, MT-high and memory B cells **(Figure 6F)**. In contrast, the 50 most dominant synovial clones are heavily weighted toward plasma cells **(Figure 6G)**. Though clones were shared between the synovium and blood, these clones were not expanded **(Figure 6H)**. Tissue trafficking clones were found within plasma cells, ABCs, activated and memory populations consistent with some trafficking of these B cells between the two compartments **(Figure 6I)**. Interestingly, clones shared between blood and tissue had distinct phenotypes **(Figure 6I)**. As an example, there are multiple cases of memory B cells in the blood becoming activated B cells or ABCs or plasma cells in the synovium, strongly suggesting that memory B cells participate in synovial immune reactions.

In order to identify features that differentiate clonally expanded from unexpanded B cells in the synovium we performed GSEA. We focused this analysis on plasmablasts and plasma cells given the limited numbers of clones in other cell states. Clonally expanded plasmablasts showed increased expression of signatures of BCR signaling, cytokine mediated signaling and response to cytokines as compared to non-clonally expanded cells **(Figure 6E)**.

Identification of altered T cell-B cell communication patterns in synovium

Given the association of BCR signaling, B cell activation, and response to cytokines with the enriched and clonally expanded B cells in the synovial analysis, we next sought to systematically investigate potential T cell-B cell interactions. We constructed cell–cell communications networks using CellChat (Jin et al. 2021), with an initial focus on communication differences between the CD4+ T cell subpopulations and total B cells. The proliferating, Tph, and Tfh/Tph subsets had the largest numbers of significant interactions identified with B cells **(Figure 7A)**. Detection of significant CXCL13-CXCR5 interactions between Tph-B cell and Tfh/Tph-B cell pairs is consistent with prior reports and supports the performance of the analysis method (Rao et al. 2017). A significant IFNG-IFNGR interaction was further detected between Tph-B but not Tfh/Tph-B pairs, consistent with the increased expression of *IFNG* in the Tph cluster compared to the Tfh/Tph cluster **(Figure S11A,B)**. We then generated an inferred communication network using CD4+ T subpopulations and B cell subpopulations. Examination of the cumulative incoming and outgoing interactions for each population identified the Tph population as having elevated signals for both directions, while the ABC population had the strongest outgoing signal of any cluster **(Figure S11B)**. A pair-wise analysis of significant interactions between these CD4+ T and B cell populations found the proliferating cluster to have the largest number of predicted interactions with all B cell subsets, though this may reflect a heterogeneous nature of proliferating cells. Aside from the proliferating cluster, elevated interactions between Tph cells and the ABC, Plasmablast, Memory, and Clonal B cell populations were identified. The number of significant interactions for each of these pairs was higher than with any other CD4+ population, suggesting an elevated signaling potential of the Tph cluster **(Figure 7B)**.

Next, we leveraged the cross-tissue nature of our dataset to compare T cell-B cell signaling differences between synovial tissue and blood. We generated a separate communication network for each, both of which returned roughly similar numbers of significant

interactions detected with a slight elevation in the synovium **(Figure S11C)**. Within the synovial tissue, the TNF, CXCL, IL-2, and IFN type II signaling pathway families were elevated, together highlighting characteristics of an inflammatory and immune-activated state of the tissue. Among the signaling pathways underrepresented in the tissue compared to signaling in the blood were CCL, SELPLG, ICAM, and ITGB, which are often expressed on cells migrating in the blood **(Figure 7C)**. Finally, we sought to identify cell-cell interactions elevated in the synovial tissue between Tph or Tfh/Tph cells and either ABC, Memory, or Activated B cells. In all analyzed pairs, BTLA-TNFRSF14 (HVEM) and CXCL13-CXCR5 interactions were elevated in the synovium compared to the blood. In contrast, ITGB2-ICAM2 and LGALS9-CD44/45 interactions were higher in most blood pairs. Comparing different interactions between Tph-B subsets and Tfh/Tph-B subsets, LTA-TNFRSF14 (HVEM) was identified in all synovial Tph-B cell pairings but absent in Tfh/Tph-B cell pairings. In the synovium, IFNG-IFNGR was a significant interaction in both Tph/ABC and Tph/Activated B cell pairings, and LTA-TNFRSF1B (TNFR2) was specific to the Tph/ABC pairing only **(Figure 7D)**.

Discussion

By leveraging paired single-cell RNA and TCR/BCR sequencing, this work provides a detailed assessment of the relationship between the immune repertoire and cell state composition, gene expression, and cell-cell interactions in RA synovium. These data provide new insights into the developmental relationships between specific synovial lymphocyte populations, for example demonstrating clonal links connecting ABC and activated B cells with plasma cells, while separating clonally distinct Tph/Tfh vs *CCL5*+ CD4 T cells and *GZMK*+ vs *GZMB*+ CD8 T cells.

Within the CD4+ T cell compartment, the Tph cells express among the highest effector and activation signatures compared to other subsets and are significantly enriched in synovium compared to the blood, consistent with prior observations (Rao et al. 2017; Aldridge et al. 2020).

Here, we leverage the paired repertoire information of this population, identifying it as one of the most clonally expanded CD4 T cell populations and clonally related to both proliferating cells and Tfh cells. These observations of Tph cells from synovial tissue are consistent with recent work tracking clonal relationships of Tph cells in synovial fluid, which also showed clonal overlap with proliferating cells and a population of Tph cells with lower expression of *CXCL13* (Argyriou et al. 2022). Further, our work suggests additional roles for Tph cells, where upon T cell activation and clonal expansion, they upregulate factors related to effector function and cytotoxicity, such as *GZMB*, *IFNG*, and *CCL4.* This result may suggest a direct contribution by this population in promoting tissue inflammation or injury through cytotoxic activity in addition to its well-characterized B cell helper function.

While analyses of CD8 T cells have largely focused on expression of cytotoxic features such as granzyme B, recent observations have highlighted a prominent CD8 T cell population in RA synovium with distinct expression of granzyme K (Jonsson et al. 2022). Our analyses further underscore a key distinction between *GZMK*+ and *GZMB*+ CD8+ T cells in RA synovium and blood. Whereas in the blood *GZMB*-expressing cytotoxic cells formulate the largest clones, CD8+ cells that express *GZMK* comprise the majority of the most expanded cells in the tissue. A striking finding through the current and prior analyses was the near-absence of clonal overlap between these cohorts across tissues, suggesting that these GZMK+ cells do not arrive at the synovium as *GZMB*-expressing cytotoxic cells, and may instead receive antigenic stimulation locally that drives clonal expansion and functions such as cytokine production (Jonsson et al. 2022).

The presence of T cell cross-reactivity between virus and self has previously been associated with the capacity for driving autoimmunity, including in RA, yet the potential for virus-reactive T cells to contribute to synovial inflammation has remained uncertain (Wucherpfennig and Strominger 1995; Pender et al. 2017; Ashton et al. 2016; Lanz et al. 2022). Our analyses identify multiple synovial T cell clones that match those previously identified to be viral-reactive,

a result consistent with the recent demonstration of viral reactivity among TCRs from RA synovial CD4 T cells (Turcinov et al. 2022). Our paired RNA-seq/TCR analysis enabled interrogation of the phenotypes of potentially viral-reactive CD8 T cells, yet no broad-scale difference in the cluster composition or clonal characteristics of these cells was apparent. Because our methodology relied on utilizing viral-specific clones that had previously been identified on defined HLA alleles, there are likely a number of “false-negative” clones in the dataset that could be viral-reactive yet did not have a match in the examined databases. Future work to better define the viral-reactive capacity of T cells within the joint may rely on isolating viral-specific cells using viral peptides bound to tetramers. Still, this work shows the presence of likely cross-reactive T cells within the synovium, yet with no enrichment for specific activated or effector phenotypes.

Populations of innate T cells are thought to contribute to RA (Bank 2020; Fathollahi et al. 2021). Here, we define the subsets of innate T cells present in synovium, including populations of γδ T, NK, and MAIT cells, with selective enrichment of Vdelta1 γδ T cells in the synovium compared to blood, consistent their enrichment in other tissues including gut and skin (Bos et al. 1990; Deusch et al. 1991). Leveraging paired TCR information, this is the first single-cell RNA sequencing study in RA to confirm the presence of MAIT cells by analyzing VDJ gene rearrangements. Our detection of shared MAIT cell clones in synovium and blood suggests that these cells may traffic into and out of synovium. Synovial MAIT cells showed an increased activation signatures compared to those in blood, as did γδ T cells, suggesting an active role for these cells in inflammatory arthritis.

One of the striking findings of our study is the enrichment of activated B cell populations in the synovium with evidence of clonal expansion and clonal sharing between different B cell states. We achieved unprecedented resolution of discrete B cell states, and in contrast to a previous study (Hardt et al. 2022), our data demonstrated clonal sharing between multiple B cell states extending beyond memory B cell and PC pools. We further extend the observation from

previous studies (Scheel et al. 2011; Hardt et al. 2022) that synovial plasma cells are generated from locally activated B cells including activated B cells, ABCs and memory B cells. Further, our work highlights the likely important signals promoting synovial B cell activation and selection, including antigen (reflected in upregulation of BCR signaling), cytokines (most notably IFNG), and direct cell-cell interactions, mainly involving Tph/Tfh cells. Citrullinated peptides may be the main antigenic drivers in the synovium as has been suggested by other studies (Corsiero et al. 2016; Kristyanto et al. 2020).

Our study highlighted a spectrum of B cell activation unique to the synovium. We identified multiple naïve-like B cell states, with surprising evidence of antigen encounter/activation based on higher mutation rates. The majority of the synovial B cells in our study are non-naive, dominated by a B-activated cluster expressing *NR4A1 (Meednu et al. 2022)*, ABCs, and plasmablasts/plasma cells. The presence of higher SHM, clonal expansion, and clonal sharing in these B cell subsets is consistent with *in situ* synovial activation and selection. We have hypothesized that activated B cells are involved in initiating ectopic lymphoid structures (ELS), as evidenced by their higher expression of ELS-inducing cytokines like LT and IL6 (Meednu et al. 2022). This may promote the generation of memory B cells and plasma cells *in situ*. Our data also suggest that *NR4A1*+ activated B cells may go down an extrafollicular pathway as they share clones with ABCs. Further, Naive-MT-high B cells, although present in small numbers in the synovium, share clones with both ABCs and activated B cells. It is thus tempting to speculate that this naive-like state may represent the precursor of ABCs and activated B cells in the synovium.

The availability of paired analysis of T and B cells from the same synovial samples allowed us to directly link clonally expanded T cell populations with synovial tissue expanded B cells for the first time. The most striking predicted interactions were identified between proliferating T cells and Tph with activated B cells, ABCs, and plasmablasts. Notably, Tph and ABCs had among the largest numbers of incoming and outgoing interactions, suggesting an

elevated signaling potential for these populations. Signaling pathways enriched in the synovium and identified in the Tph-ABC interaction included cytokines (e.g. IFNG- produced by Tph and IFNGR on ABCs and LTA-TNFRSF14 (HVEM) interactions) and chemokines. Notably, GSEA also identified cytokine-mediated signaling as prominent in clonally expanded synovial B cells, including within the ABC population. This subset of B cells has been previously reported to be expanded in autoimmune disease (Jenks et al. 2018) with accumulation in inflamed tissue (Zhang et al. 2019, 2022), but the signals that promote ABCs in the synovium remain unclear. Our data strongly point to Tph cells as a key driver. This is in accord with a recent report in juvenile idiopathic arthritis demonstrating that clonally expanded *IL21* and *IFNG* coexpressing Tph promote CD11c+ double negative B cell differentiation (Fischer et al. 2022).

Together, these findings across T and B cells highlight the altered cell state composition and clonal characteristics that may work together to maintain inflammation in RA. Our study utilized a cross-sectional cohort unified in high disease activity, but otherwise heterogeneous across treatment history, disease duration, and cell-type abundance phenotype (CTAP). Further studies with a larger cohort of patients may assist in connecting clonal characteristics such as those identified here with patient stratifications, which may serve to increase our knowledge of the inherent cellular and molecular heterogeneity of the disease. As the understanding of pathogenic roles of B and T cell subsets in RA continues to evolve, this dataset will be a useful resource to generate or test insights related to the antigen receptor repertoires of synovial lymphocytes.

Methods

Sample processing

For this study, patients were recruited and consented through the Accelerating Medicines Partnership (AMP) Network for RA and SLE (Zhang et al. 2022). Synovial tissue samples and matched peripheral blood mononuclear cells were cryopreserved after collection as described

(Donlin et al. 2018). Stored synovial tissue samples were then thawed and disaggregated into single-cell suspensions by mincing and digesting with 100 µg/mL LiberaseTL (Roche) and 100 µg/mL DNaseI (Roche) in RPMI (Life Technologies) for 15 min, with occasional inversion during disaggregation. Disaggregated cells were passed through a 70 µm cell strainer and washed prior to antibody staining with anti-CD235a antibodies (clone 11E4B-7-6 (KC16), Beckman Coulter) and Fixable Viability Dye eFluor 780 (eBioscience/ThermoFisher). Live non-erythrocyte cells (viability dye- CD235-) were collected by fluorescence-activated cell sorting (BD FACSAria Fusion) and were initially cryopreserved in Cryostor CS10 (Sigma-Aldrich). The disaggregated synovial tissue cells and matched cryopreserved peripheral blood mononuclear cells were then thawed in batches, and both T and B cells were collected by fluorescence-activated cell sorting (BD FACSAria II) (Figure S1B, Table S4).

Single-cell library preparation and sequencing

Cells collected from cell sorting were encapsulated into oil droplets using a Chromium NextGEM Chip G (10X Genomics). Following reverse transcription and cDNA amplification, 5' gene expression, immune repertoire, and feature barcode libraries were constructed following manufacturer protocols (v1.1). The libraries were finally pooled for sequencing on an Illumina Novaseq 6000 using an S4 flow cell. Gene expression libraries were sequenced to obtain a read depth of 100,000 reads per cell, feature barcode libraries were sequenced at 5,000 reads per cell, and immune repertoire libraries were sequenced at 5,000 reads per cell. FASTQ file demultiplexing for gene expression libraries was performed using the *mkfastq* function in CellRanger (10X Genomics, v4.0). Following this, alignment to a reference genome (GRCh38) and counting was completed using the *count* function to generate expression matrices for each sample. Immune repertoire FASTQ files were separately demultiplexed, and the *vdj* function was used to perform sequence assembly and clonotype calling for TCR and BCR sequences in each sample.

Initial quality control

Gene count matrices were imported in R for downstream analysis. Quality control was first performed jointly on all cells collected and sequenced in this experiment. Several metrics were explored to assess the quality of each cell. First, low-quality cells were distinguished from high-quality cells in each tissue compartment. Cells were kept for downstream analysis if they had at least 1000 mapped reads in either the blood or the synovium. After initial filtering, we generated several metrics to identify doublets using software packages scDblFinder (Germain et al. 2022) and scds (Bais and Kostka 2020), as well as marking cells that coexpressed at least 1 TCR and 1 BCR. At this point, a single synovial sample (AMP ID# 300_0415) was discarded due to only 12 cells passing these initial QC thresholds. Log-normalization was then applied to the gene expression counts for remaining cells. Final QC thresholding was performed on the log-normalized counts, keeping cells with greater than 500 genes detected and less than 20 percent of detected reads coming from mitochondrial genes. Additional quality control was performed for T and B cells separately in downstream analysis.

Broad cell type identification

After initial QC, unsupervised clustering was performed on the remaining cells to identify major cell types present in the data. Principal components were first generated in order to reduce the dimensionality of the feature space before clustering. Using Seurat's clustering functionalities on the first 30 principal components, a 20 nearest-neighbors network graph was computed. Then, we performed Louvain clustering with a resolution parameter of 0.3, and visualized the cells in 2D space using Uniform Manifold Approximation and Projection (UMAP). Differentially-expressed genes between clusters were identified (Student's T-test) by including only genes exhibiting a greater-than 0.25 log-fold difference between clusters. In order to annotate each cluster with a biologically meaningful name, genes with the highest log-fold changes were considered, as well as marker genes that are cell-type specific.

T cell subclustering

Cells broadly labeled as “T cells” and “Proliferating” from the combined object were subset, and Harmony (Korsunsky et al. 2019) was then used to perform batch correction at the level of the patient and tissue using theta = 2 and max.iter.cluster = 20. Using the top 50 Harmony embeddings, Louvain clustering was performed, which was then visualized in UMAP space. Nearest neighbors were identified using the Harmony embeddings, and clustering was iteratively performed with a resolution of 1.4 finally selected. Broad T cell markers (eg. CD4, CD8, TRDC) were used along with differential gene expression (Wilcoxon Rank Sum test) to identify the T cell lineages for each of these initial clusters. At this point, a small number of remaining contaminating cells (such as B cells from the Proliferating cluster), were removed. Based on gene expression, clusters of CD4, CD8, and innate T cells were separated and individually clustered using a similar strategy with slightly different parameters for each subset (CD4: 40 Harmony embeddings, 0.5 cluster resolution; CD8: 40 Harmony embeddings, 0.4 cluster resolution; innate T: 10 Harmony embeddings, 0.4 cluster resolution).

B cell subclustering

Cells labeled as “B cells” from the broad clustering, we further characterized B cell subpopulations. Before reclustering the B cells, we discard cells marked as doublets according to scDblFinder and B cells which simultaneously coexpressed at least 1 BCR and at least 1 TCR. On the remaining cells, Seurat’s default normalization and scaling was performed and principal components were generated. Harmony was used to perform batch correction at the patient level using theta = 2 and max.iter.cluster = 20. From here, the 20-nearest-neighbors network graph was generated using the first 30 harmonized principal components. We then applied Louvain clustering using a resolution parameter of 0.5 to identify clusters of similar cells to visualize in the UMAP space. Differential gene expression was then performed (Student’s t-test) to provide markers for cluster annotation. Before selecting the final set of input parameters, results were explored at multiple resolutions, variable number of included principal components, and using both harmonized and non-harmonized principal components.

Gene signature analysis

Gene signatures used in this study were obtained from the sources listed in Table S3. The *AddModuleScore* function in Seurat was used to assign a value for each cell for each signature, corresponding to the average expression of the signature subtracted by the aggregate expression of randomly-selected control genes.

Dataset reference mapping

A reference object was built using Symphony (Kang et al. 2021) with cells from Zhang et al, 2022 that correspond to the populations being assayed (CD4 T, CD8 T, innate T, or B cells), integrating at the sample level and using the first 20 PCs. Data from the current study was then projected onto this reference using the mapQuery and knnPredict functions, to generate confidence scores for each reference cluster's mapping, which we further visualized using pheatmap. The most-likely identities for each cluster were then cross-referenced with DEG lists from our dataset to aid in generating final cluster identities.

Single-cell TCR receptor profiling

For each sample, the filtered_contig_annotations.csv file output from cellranger was used to identify TCR sequences obtained for each cell barcode using scRepertoire (Borcherding, Bormann, and Kraus 2020). Using the filterMulti argument, only the top 2 expressed chains were retained when a cell barcode was associated with more than 2 chains (eg. ααβ or αββ). These TCRs and their cell barcodes were then matched with corresponding cell barcodes obtained from the sample's RNA library. Combined, 45,096 cells in the initial T cell subset had available TCR information, including 22,634 cells obtained from synovial tissue and 22,462 cells obtained from blood. Clones were further characterized into discrete groups for their extent of clonal expansion for downstream analysis.

Single-cell BCR receptor profiling

Analogous to the TCR profiling, output from CellRanger was used to identify BCR sequences for each cell barcode. For downstream analysis, we included BCR that were considered high-confidence, full-length. We also excluded BCR that were associated with more than 2 heavy chains or more than 2 light chains. BCR passing QC were then matched with cell barcodes from the sample's single-cell RNA library. Combined, 30,779 cells in the B cell subset had associated BCR information, of which, 7,763 cells were from the synovial tissue and 23,016 were from blood. We assigned each BCR sequence to its closest sequence in the IMGT database using the Change-O tool (Gupta et al. 2015). The degree of somatic hypermutation in each B cell was quantified by determining the number of V and J substitutions in each B cell's IgH CDR3 sequence when compared to its closest IMGT sequence. Clones were characterized in B cells according to the similarity of their CDR3 regions. Using a similarity threshold of 96.5%, CD-HIT (Fu et al. 2012) was used to define discrete clonal groups for downstream analysis.

Mixed-effect modeling

A number of mixed effect models were fit in our analysis, which all generally took a similar form. To adjust for patient-specific effects in our data, mixed-effect models were fit using sample_ID as a scalar random effect and fixed effects for other covariates of interest. Lme4 (Bates et al. 2015) was used to obtain point estimates for all mixed-effect models, with 95% confidence intervals.

Gene-set enrichment analysis

Gene set enrichment analysis was performed using Fast Gene Set Enrichment Analysis (FGSEA) (Korotkevich et al. 2021). FGSEA calculates an enrichment score for each gene set, given a ranked vector of gene-level statistics. A null distribution of the enrichment score is estimated through random sampling of gene sets. P-values are estimated as the number of random gene sets with more extreme enrichment scores than the gene set of interest, divided by the number of random gene sets generated. Multiple testing correction is then performed to get adjusted P-values. We first performed differential expression for our comparisons of interest

within each population (SYN vs PBL and Clonal vs Non-clonal) to obtain a ranked vector of gene-level statistics. We then sampled 1000 gene random sets to estimate the null distribution of each gene set's enrichment score and calculate adjusted P-values.

Cell-Cell interactions

Cell-cell interaction inference was performed using CellChat (Jin et al. 2021), which uses a manually curated database of ligand-receptor interactions gathered from KEGG signaling pathways and published literature. Interaction networks are constructed by identifying differentially expressed genes related to these interactions, computing the average expression of each ligand-receptor pair across cell cluster pairings, and finally calculating a communication probability value based on permutation testing. For finding communication probabilities, we used the tool's Tukey triMean method, which performs a weighted average of the median and upper and lower quartiles, and further used 100 bootstraps (nboot = 100) to calculate p values.

HLA imputation

To obtain estimated two-field HLA alleles in each donor, we performed HLA imputation from SNP genotype data. We genotyped donors from this study by using Illumina Multi-Ethnic Genotyping Array. We performed quality control of genotype by sample call rate > 0.99, variant call rate > 0.99, minor allele frequency > 0.01, and $P_{HWE} > 1.0\times10^{-6}$. We extracted the extended MHC region (28–34Mb on chromosome 6) and performed haplotype phasing with SHAPEIT2 software (Delaneau, Marchini, and Zagury 2011). We then performed HLA imputation by using a multi-ancestry HLA reference panel version 2 (Luo et al. 2021) and minimac3 software (Das et al. 2016). From the imputed dosage of two-field HLA alleles of each HLA gene in each donor, we defined the most likely set of two two-field alleles.

Data availability

Data generated by this study are deposited at Synapse (doi.org/10.7303/syn47217489.1).

Code availability

Source code to reproduce analyses used in this study are available at

https://github.com/dunlapg/amp2repertoire.

Acknowledgments

We thank the participants who provided synovial tissue and blood samples for this study. We also acknowledge the Flow Cytometry Core and Genomics Research Center at the University of Rochester for performing the flow sorting, scRNA library preparation, sequencing, and primary data analysis for this study. This work was supported by the Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network. The AMP Program is a public-private partnership that includes AbbVie, the Arthritis Foundation, Bristol-Myers Squibb Company, the Foundation for the National Institutes of Health, GlaxoSmithKline, Janssen Research and Development, the Lupus Foundation of America, the Lupus Research Alliance, Merck, the National Institute of Allergy and Infectious Diseases, the National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer, the Rheumatology Research Foundation, Sanofi, and Takeda Pharmaceuticals. Funding for AMP RA/SLE work was provided through grants from the National Institutes of Health (UH2-AR067676, UH2-AR067677, UH2-AR067679, UH2-AR067681, UH2-AR067685, UH2-AR067688, UH2-AR067689, UH2-AR067690, UH2-AR067691, UH2-AR067694, and UM2-AR067678). F.Z. is supported by the PhRMA Foundation Faculty Starter Grant for Translational Medicine. K.W. is supported by a Burroughs Wellcome Fund Career Awards for Medical Scientists, a Doris Duke Charitable Foundation Clinical Scientist Development Award, and a Rheumatology Research Foundation Innovative Research Award. D.A.R. is supported by a Burroughs Wellcome Fund Career Awards for Medical Scientists.

Author Contributions

G.S.F., D.L.B., D.T., and C.P. recruited patients and obtained synovial tissues. L.W.M., S.M.G., V.M.H., A.F., V.P.B., and J.H.A contributed to the procurement of samples and overall design of the AMP study. N.M., A.H.J, K.W., A.N., L.T.D., S.R., and M.B.B. contributed to the design and implementation of the tissue disaggregation, cell sorting, and single-cell sequencing pipeline.

G.S.D., A.W., F.Z., A.H.J., S.S., and A.M. conducted the computational analysis. G.S.D, A.W., N.M., D.A.R., and J.H.A. wrote and edited the initial manuscript draft. G.S.D, A.W., N.M., F.Z., A.H.J., K.W., V.M.H, S.R., D.A.R., and J.H.A. revised the manuscript. A.M., D.A.R., and J.H.A. supervised the study.

Competing Interests

K.W. received a sponsored-research agreement from Gilead Sciences, and served as a consultant for Gilead Sciences and Horizon Therapeutics.

Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network includes:

Jennifer Albrecht[3], William Apruzzese[17], Jennifer L. Barnas[3], Joan M. Bathon[18], Ami Ben-Artzi[19], Brendan F. Boyce[20], S. Louis Bridges Jr.[9,10], Debbie Campbell[3], Hayley L. Carr[15], Arnold Ceponis[11], Adam Chicoine[1], Andrew Cordle[21], Michelle Curtis[1,4,5,6,7], Kevin D. Deane[12], Edward DiCarlo[22], Patrick Dunn[23,24], Lindsy Forbess[19], Laura Geraldino-Pardilla[18], Ellen M. Gravallese[1], Peter K. Gregersen[25], Joel M. Guthridge[26], Diane Horowitz[25], Laura B. Hughes[27], Kazuyoshi Ishigaki[1,4,5,6,7,28], Lionel B. Ivashkiv[9,10], Judith A. James[26], Joyce B. Kang[1,4,5,6,7], Gregory Keras[1], Ilya Korsunsky[1,4,5,6,7], Amit Lakhanpal[9,10], James A. Lederer[29], Yuhong Li[1], Zhihan J. Li[1], Katherine P. Liao[1,6], Holden Maecker[30], Arthur M. Mandelin II[31], Ian Mantel[9,10], Mark Maybury[15], Mandy J. McGeachy[13], Joseph Mears[1,4,5,6,7], Alessandra Nerviani[14], Dana E. Orange[9,32], Harris Perlman[31], Javier Rangel-Moreno[3], Karim Raza[15], Yakir Reshef[1,4,5,6,7], Christopher Ritchlin[3], Felice Rivellese[14], William H. Robinson[33], Laurie Rumker[1,4,5,6,7], Ilfita Sahbudin[15], Karen Salomon-Escoto[34], Dagmar Scheel-Toellner[15], Jennifer A. Seifert[12], Lorien Shakib[9,10], Anvita Singaraju[9,10], Melanie H. Smith[9], Paul J. Utz[33], Kathryn Weinand[1,4,5,6,7], Dana Weisenfeld[1], Michael H. Weisman[19], Qian Xiao[1,4,5,6,7], Zhu Zhu[1]

[17]Accelerating Medicines Partnership Program: Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network
[18]Division of Rheumatology, Columbia University College of Physicians and Surgeons; New York, NY, USA
[19]Division of Rheumatology, Cedars-Sinai Medical Center; Los Angeles, CA, USA
[20]Department of Pathology and Laboratory Medicine, University of Rochester Medical Center; Rochester, NY, USA
[21]Department of Radiology, University of Pittsburgh Medical Center; Pittsburgh, PA, USA
[22]Department of Pathology and Laboratory Medicine, Hospital for Special Surgery; New York, NY, USA

[23]Division of Allergy, Immunology, and Transplantation, National Institute of Allergy and Infectious Diseases, National Institutes of Health; Bethesda, MD, USA
[24]Northrop Grumman Health Solutions; Rockville, MD, USA
[25]Feinstein Institute for Medical Research, Northwell Health, Manhasset; New York, NY, USA
[26]Department of Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation; Oklahoma City, OK, USA
[27]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA
[28]Laboratory for Human Immunogenetics, RIKEN Center for Integrative Medical Sciences; Yokohama, Japan
[29]Department of Surgery, Brigham and Women's Hospital and Harvard Medical School; Boston, MA, USA
[30]Human Immune Monitoring Center, Stanford University; Stanford, CA, USA
[31]Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School of Medicine; Chicago, IL, USA
[32]Laboratory of Molecular Neuro-Oncology, The Rockefeller University; New York, NY, USA
[33]Division of Immunology and Rheumatology, Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine; Stanford, CA, USA
[34]Division of Rheumatology, Department of Medicine, University of Massachusetts Medical School; Worcester, MA, USA

References

Abbas, Alexander R., Kristen Wolslegel, Dhaya Seshasayee, Zora Modrusan, and Hilary F. Clark. 2009. “Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus.” *PloS One* 4 (7): e6098.

Aldridge, Ekwall, Mark, and Bergström. 2020. “T Helper Cells in Synovial Fluid of Patients with Rheumatoid Arthritis Primarily Have a Th1 and a CXCR3+ Th2 Phenotype.” *Arthritis Research & Therapy*. https://doi.org/10.1186/s13075-020-02349-y;

Alivernini, Stefano, Lucy MacDonald, Aziza Elmesmari, Samuel Finlay, Barbara Tolusso, Maria Rita Gigante, Luca Petricca, et al. 2020. “Distinct Synovial Tissue Macrophage Subsets Regulate Inflammation and Remission in Rheumatoid Arthritis.” *Nature Medicine* 26 (8): 1295–1306.

Andreu, J. L., A. Trujillo, J. M. Alonso, J. Mulero, and C. Martínez. 1991. “Selective Expansion of T Cells Bearing the Gamma/delta Receptor and Expressing an Unusual Repertoire in the Synovial Membrane of Patients with Rheumatoid Arthritis.” *Arthritis and Rheumatism* 34 (7): 808–14.

Argyriou, Alexandra, Marc H. Wadsworth 2nd, Adrian Lendvai, Stephen M. Christensen, Aase H. Hensvold, Christina Gerstner, Annika van Vollenhoven, et al. 2022. “Single Cell Sequencing Identifies Clonally Expanded Synovial CD4+ TPH Cells Expressing GPR56 in Rheumatoid Arthritis.” *Nature Communications* 13 (1): 4046.

Ashton, Michelle P., Anne Eugster, Denise Walther, Natalie Daehling, Stephanie Riethausen, Denise Kuehn, Karin Klingel, et al. 2016. “Incomplete Immune Response to Coxsackie B Viruses Associates with Early Autoimmunity against Insulin.” *Scientific Reports* 6 (September): 32899.

Bais, Abha S., and Dennis Kostka. 2020. “Scds: Computational Annotation of Doublets in Single-Cell RNA Sequencing Data.” *Bioinformatics* 36 (4): 1150–58.

Bank, Ilan. 2020. “The Role of Gamma Delta T Cells in Autoimmune Rheumatic Diseases.” *Cells* 9 (2). https://doi.org/10.3390/cells9020462.

Barnes, Michael J., Halil Aksoylar, Philippe Krebs, Tristan Bourdeau, Carrie N. Arnold, Yu Xia, Kevin Khovananth, et al. 2010. “Loss of T Cell and B Cell Quiescence Precedes the Onset of Microbial Flora-Dependent Wasting Disease and Intestinal Inflammation in Gimap5-Deficient Mice.” *Journal of Immunology* 184 (7): 3743–54.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. “Fitting Linear Mixed-Effects Models Using lme4.” *Journal of Statistical Software*. https://doi.org/10.18637/jss.v067.i01.

Borcherding, Nicholas, Nicholas L. Bormann, and Gloria Kraus. 2020. “scRepertoire: An R-Based Toolkit for Single-Cell Immune Receptor Analysis.” *F1000Research* 9 (January): 47.

Bos, J. D., M. B. Teunissen, I. Cairo, S. R. Krieg, M. L. Kapsenberg, P. K. Das, and J. Borst. 1990. “T-Cell Receptor Gamma Delta Bearing Cells in Normal Human Skin.” *The Journal of Investigative Dermatology* 94 (1): 37–42.

Canhão, Helena, Ana Maria Rodrigues, Maria João Gregório, Sara S. Dias, José António Melo Gomes, Maria José Santos, Augusto Faustino, et al. 2018. “Common Evaluations of Disease Activity in Rheumatoid Arthritis Reach Discordant Classifications across Different Populations.” *Frontiers of Medicine* 5 (March): 40.

Chang, Che-Mai, Yu-Wen Hsu, Henry Sung-Ching Wong, James Cheng-Chung Wei, Xiao Liu, Hsien-Tzung Liao, and Wei-Chiao Chang. 2019. “Characterization of T-Cell Receptor Repertoire in Patients with Rheumatoid Arthritis Receiving Biologic Therapies.” *Disease Markers* 2019 (July): 2364943.

Collora, Jack A., Runxia Liu, Delia Pinto-Santini, Neal Ravindra, Carmela Ganoza, Javier R. Lama, Ricardo Alfaro, et al. 2022. “Single-Cell Multiomics Reveals Persistence of HIV-1 in Expanded Cytotoxic T Cell Clones.” *Immunity* 55 (6): 1013–31.e7.

Corsiero, Elisa, Michele Bombardieri, Emanuela Carlotti, Federico Pratesi, William Robinson, Paola Migliorini, and Costantino Pitzalis. 2016. “Single Cell Cloning and Recombinant Monoclonal Antibodies Generation from RA Synovial B Cells Reveal Frequent Targeting of Citrullinated Histones of NETs.” *Annals of the Rheumatic Diseases* 75 (10): 1866–75.

Das, Sayantan, Lukas Forer, Sebastian Schönherr, Carlo Sidore, Adam E. Locke, Alan Kwong, Scott I. Vrieze, et al. 2016. “Next-Generation Genotype Imputation Service and Methods.” *Nature Genetics* 48 (10): 1284–87.

Delaneau, Olivier, Jonathan Marchini, and Jean-François Zagury. 2011. “A Linear Complexity Phasing Method for Thousands of Genomes.” *Nature Methods* 9 (2): 179–81.

Deusch, K., F. Lüling, K. Reich, M. Classen, H. Wagner, and K. Pfeffer. 1991. “A Major Fraction of Human Intraepithelial Lymphocytes Simultaneously Expresses the Gamma/delta T Cell Receptor, the CD8 Accessory Molecule and Preferentially Uses the V Delta 1 Gene Segment.” *European Journal of Immunology* 21 (4): 1053–59.

DeWitt, William S., Paul Lindau, Thomas M. Snyder, Anna M. Sherwood, Marissa Vignali, Christopher S. Carlson, Philip D. Greenberg, Natalie Duerkopp, Ryan O. Emerson, and Harlan S. Robins. 2016. “A Public Database of Memory and Naive B-Cell Receptor Sequences.” *PloS One* 11 (8): e0160853.

Donlin, Laura T., Deepak A. Rao, Kevin Wei, Kamil Slowikowski, Mandy J. McGeachy, Jason D. Turner, Nida Meednu, et al. 2018. “Methods for High-Dimensional Analysis of Cells Dissociated from Cryopreserved Synovial Tissue.” *Arthritis Research & Therapy* 20 (1): 139.

Fathollahi, Anwar, Leila Nejatbakhsh Samimi, Maassoumeh Akhlaghi, Ahmadreza Jamshidi, Mahdi Mahmoudi, and Elham Farhadi. 2021. “The Role of NK Cells in Rheumatoid Arthritis.” *Inflammation Research: Official Journal of the European Histamine Research Society ... [et Al.]* 70 (10-12): 1063–73.

Fischer, Jonas, Johannes Dirks, Julia Klaussner, Gabriele Haase, Annette Holl-Wieden, Christine Hofmann, Stephan Hackenberg, Hermann Girschick, and Henner Morbach. 2022. “Effect of Clonally Expanded PD-1high CXCR5-CD4+ Peripheral T Helper Cells on B Cell Differentiation in the Joints of Patients with Antinuclear Antibody-Positive Juvenile Idiopathic Arthritis.” *Arthritis & Rheumatology* 74 (1): 150–62.

Fortea-Gordo, Paula, Laura Nuño, Alejandro Villalba, Diana Peiteado, Irene Monjo, Paloma Sánchez-Mateos, Amaya Puig-Kröger, Alejandro Balsa, and María-Eugenia Miranda-Carús. 2019. “Two Populations of Circulating PD-1hiCD4 T Cells with Distinct B Cell Helping Capacity Are Elevated in Early Rheumatoid Arthritis.” *Rheumatology* 58 (9): 1662–73.

Fujinami, Robert S., Matthias G. von Herrath, Urs Christen, and J. Lindsay Whitton. 2006. “Molecular Mimicry, Bystander Activation, or Viral Persistence: Infections and Autoimmune Disease.” *Clinical Microbiology Reviews* 19 (1): 80–94.

Fu, Limin, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. 2012. “CD-HIT: Accelerated for Clustering the next-Generation Sequencing Data.” *Bioinformatics* 28 (23): 3150–52.

Germain, Pierre-Luc, Aaron Lun, Will Macnair, and Mark D. Robinson. 2022. “Doublet Identification in Single-Cell Sequencing Data Using scDblFinder.” *F1000Research* 10: 979.

Goldbach-Mansky, R., J. Lee, A. McCoy, J. Hoxworth, C. Yarboro, J. S. Smolen, G. Steiner, et al. 2000. “Rheumatoid Arthritis Associated Autoantibodies in Patients with Synovitis of Recent Onset.” *Arthritis Research* 2 (3): 236–43.

Gregersen, P. K., J. Silver, and R. J. Winchester. 1987. “The Shared Epitope Hypothesis. An Approach to Understanding the Molecular Genetics of Susceptibility to Rheumatoid Arthritis.” *Arthritis and Rheumatism* 30 (11): 1205–13.

Gupta, Namita T., Jason A. Vander Heiden, Mohamed Uduman, Daniel Gadala-Maria, Gur Yaari, and Steven H. Kleinstein. 2015. “Change-O: A Toolkit for Analyzing Large-Scale B Cell Immunoglobulin Repertoire Sequencing Data.” *Bioinformatics* 31 (20): 3356–58.

Hardt, Uta, Konstantin Carlberg, Erik Af Klint, Peter Sahlström, Ludvig Larsson, Annika van Vollenhoven, Susana Hernandez Machado, et al. 2022. “Integrated Single Cell and Spatial Transcriptomics Reveal Autoreactive Differentiated B Cells in Joints of Early Rheumatoid Arthritis.” *Scientific Reports* 12 (1): 11876.

Hinks, Timothy S. C., and Xia-Wei Zhang. 2020. “MAIT Cell Activation and Functions.” *Frontiers in Immunology* 11 (May): 1014.

Huang, Huang, Chunlin Wang, Florian Rubelt, Thomas J. Scriba, and Mark M. Davis. 2020. “Analyzing the Mycobacterium Tuberculosis Immune Response by T-Cell Receptor Clustering with GLIPH2 and Genome-Wide Antigen Screening.” *Nature Biotechnology* 38 (10): 1194–1202.

Humby, Frances, Myles Lewis, Nandhini Ramamoorthi, Jason A. Hackney, Michael R. Barnes, Michele Bombardieri, A. Francesca Setiadi, et al. 2019. “Synovial Cellular and Molecular Signatures Stratify Clinical Response to csDMARD Therapy and Predict Radiographic Progression in Early Rheumatoid Arthritis Patients.” *Annals of the Rheumatic Diseases*. https://doi.org/10.1136/annrheumdis-2018-214539.

Ishigaki, Kazuyoshi, Hirofumi Shoda, Yuta Kochi, Tetsuro Yasui, Yuho Kadono, Sakae Tanaka, Keishi Fujio, and Kazuhiko Yamamoto. 2015. “Quantitative and Qualitative Characterization of Expanded CD4+ T Cell Clones in Rheumatoid Arthritis Patients.” *Scientific Reports* 5 (August): 12937.

Jenks, Scott A., Kevin S. Cashman, Esther Zumaquero, Urko M. Marigorta, Aakash V. Patel, Xiaoqian Wang, Deepak Tomar, et al. 2018. “Distinct Effector B Cells Induced by Unregulated Toll-like Receptor 7 Contribute to Pathogenic Responses in Systemic Lupus Erythematosus.” *Immunity* 49 (4): 725–39.e6.

Jin, Suoqin, Christian F. Guerrero-Juarez, Lihua Zhang, Ivan Chang, Raul Ramos, Chen-Hsiang Kuan, Peggy Myung, Maksim V. Plikus, and Qing Nie. 2021. “Inference and Analysis of Cell-Cell Communication Using CellChat.” *Nature Communications* 12 (1): 1088.

Jonsson, A. Helena, Fan Zhang, Garrett Dunlap, Emma Gomez-Rivas, Gerald F. M. Watts, Heather J. Faust, Karishma Vijay Rupani, et al. 2022. “Granzyme K^+ CD8 T Cells Form a Core Population in Inflamed Human Tissue.” *Science Translational Medicine* 14 (649): eabo0686.

Kang, Joyce B., Aparna Nathan, Kathryn Weinand, Fan Zhang, Nghia Millard, Laurie Rumker, D. Branch Moody, Ilya Korsunsky, and Soumya Raychaudhuri. 2021. “Efficient and Precise Single-Cell Reference Atlas Mapping with Symphony.” *Nature Communications* 12 (1): 5890.

Khan, Naeem, Naseer Shariff, Mark Cobbold, Rachel Bruton, Jenni A. Ainsworth, Alan J. Sinclair, Laxman Nayak, and Paul A. H. Moss. 2002. “Cytomegalovirus Seropositivity Drives the CD8 T Cell Repertoire toward Greater Clonality in Healthy Elderly Individuals.” *Journal of Immunology* 169 (4): 1984–92.

Klarenbeek, P. L., M. J. H. de Hair, M. E. Doorenspleet, B. D. C. van Schaik, R. E. E. Esveldt, M. G. H. van de Sande, T. Cantaert, et al. 2012. “Inflamed Target Tissue Provides a Specific Niche for Highly Expanded T-Cell Clones in Early Human Autoimmune Disease.” *Annals of the Rheumatic Diseases* 71 (6): 1088–93.

Kobayashi, Shio, Koichi Murata, Hideyuki Shibuya, Mami Morita, Masahiro Ishikawa, Moritoshi Furu, Hiromu Ito, et al. 2013. “A Distinct Human CD4+ T Cell Subset That Secretes CXCL13 in Rheumatoid Synovium.” *Arthritis and Rheumatism* 65 (12): 3063–72.

Koppejan, Hester, Diahann T. S. L. Jansen, Marjolijn Hameetman, Ranjeny Thomas, Rene E. M. Toes, and Floris A. van Gaalen. 2019. “Altered Composition and Phenotype of Mucosal-Associated Invariant T Cells in Early Untreated Rheumatoid Arthritis.” *Arthritis Research & Therapy* 21 (1): 3.

Korotkevich, Gennady, Vladimir Sukhov, Nikolay Budin, Boris Shpak, Maxim N. Artyomov, and Alexey Sergushichev. 2021. “Fast Gene Set Enrichment Analysis.” *bioRxiv*.

https://doi.org/10.1101/060012.

Korsunsky, Ilya, Nghia Millard, Jean Fan, Kamil Slowikowski, Fan Zhang, Kevin Wei, Yuriy Baglaenko, Michael Brenner, Po-Ru Loh, and Soumya Raychaudhuri. 2019. “Fast, Sensitive and Accurate Integration of Single-Cell Data with Harmony.” *Nature Methods* 16 (12): 1289–96.

Kristyanto, Hendy, Nienke J. Blomberg, Linda M. Slot, Ellen I. H. van der Voort, Priscilla F. Kerkman, Aleida Bakker, Leonie E. Burgers, et al. 2020. “Persistently Activated, Proliferative Memory Autoreactive B Cells Promote Inflammation in Rheumatoid Arthritis.” *Science Translational Medicine* 12 (570). https://doi.org/10.1126/scitranslmed.aaz5327.

Lamichhane, Rajesh, Fran Munro, Thomas W. R. Harrop, Sara M. de la Harpe, Peter K. Dearden, Andrea J. Vernall, John L. McCall, and James E. Ussher. 2021. “Human Liver-Derived MAIT Cells Differ from Blood MAIT Cells in Their Metabolism and Response to TCR-Independent Activation.” *European Journal of Immunology* 51 (4): 879–92.

Lanz, Tobias V., R. Camille Brewer, Peggy P. Ho, Jae-Seung Moon, Kevin M. Jude, Daniel Fernandez, Ricardo A. Fernandes, et al. 2022. “Clonally Expanded B Cells in Multiple Sclerosis Bind EBV EBNA1 and GlialCAM.” *Nature* 603 (7900): 321–27.

Lawand, Myriam, Julie Déchanet-Merville, and Marie-Caroline Dieu-Nosjean. 2017. “Key Features of Gamma-Delta T-Cell Subsets in Human Diseases and Their Immunotherapeutic Implications.” *Frontiers in Immunology* 8 (June): 761.

Leng, Tianqi, Hossain Delowar Akther, Carl-Philipp Hackstein, Kate Powell, Thomas King, Matthias Friedrich, Zoe Christoforidou, et al. 2019. “TCR and Inflammatory Signals Tune Human MAIT Cells to Exert Specific Tissue Repair and Effector Functions.” *Cell Reports* 28 (12): 3077–91.e5.

Liu, M. F., C. Y. Yang, S. C. Chao, J. S. Li, T. H. Weng, and H. Y. Lei. 1999. “Distribution of Double-Negative (CD4- CD8-, DN) T Subsets in Blood and Synovial Fluid from Patients with Rheumatoid Arthritis.” *Clinical Rheumatology* 18 (3): 227–31.

Liu, Xiao, Wei Zhang, Ming Zhao, Longfei Fu, Limin Liu, Jinghua Wu, Shuangyan Luo, et al. 2019. “T Cell Receptor β Repertoires as Novel Diagnostic Markers for Systemic Lupus Erythematosus and Rheumatoid Arthritis.” *Annals of the Rheumatic Diseases* 78 (8): 1070–78.

Looney, R. J., A. Falsey, D. Campbell, A. Torres, J. Kolassa, C. Brower, R. McCann, et al. 1999. “Role of Cytomegalovirus in the T Cell Changes Seen in Elderly Individuals.” *Clinical Immunology* 90 (2): 213–19.

Luoma, Adrienne M., Shengbao Suo, Hannah L. Williams, Tatyana Sharova, Keri Sullivan, Michael Manos, Peter Bowling, et al. 2020. “Molecular Pathways of Colon Inflammation Induced by Cancer Immunotherapy.” *Cell* 182 (3): 655–71.e22.

Luo, Yang, Masahiro Kanai, Wanson Choi, Xinyi Li, Saori Sakaue, Kenichi Yamamoto, Kotaro Ogawa, et al. 2021. “A High-Resolution HLA Reference Panel Capturing Global Population Diversity Enables Multi-Ancestry Fine-Mapping in HIV Host Response.” *Nature Genetics* 53 (10): 1504–16.

Manzo, Antonio, Barbara Vitolo, Frances Humby, Roberto Caporali, David Jarrossay, Francesco Dell’accio, Laura Ciardelli, Mariagrazia Uguccioni, Carlomaurizio Montecucco, and Costantino Pitzalis. 2008. “Mature Antigen-Experienced T Helper Cells Synthesize and Secrete the B Cell Chemoattractant CXCL13 in the Inflammatory Environment of the Rheumatoid Joint.” *Arthritis and Rheumatism* 58 (11): 3377–87.

Marston, Bethany, Arumugam Palanichamy, and Jennifer H. Anolik. 2010. “B Cells in the Pathogenesis and Treatment of Rheumatoid Arthritis.” *Current Opinion in Rheumatology* 22 (3): 307–15.

Mathew, Nimitha R., Jayalal K. Jayanthan, Ilya V. Smirnov, Jonathan L. Robinson, Hannes Axelsson, Sravya S. Nakka, Aikaterini Emmanouilidi, et al. 2021. “Single-Cell BCR and Transcriptome Analysis after Influenza Infection Reveals Spatiotemporal Dynamics of

Antigen-Specific B Cells." *Cell Reports* 35 (12): 109286.

Meednu, Nida, Javier Rangel-Moreno, Fan Zhang, Katherine Escalera-Rivera, Elisa Corsiero, Edoardo Prediletto, Edward DiCarlo, et al. 2022. "Dynamic Spectrum of Ectopic Lymphoid B Cell Activation and Hypermutation in the RA Synovium Characterized by NR4A Nuclear Receptor Expression." *Cell Reports*. https://doi.org/10.1016/j.celrep.2022.110766.

Meednu, Nida, Hengwei Zhang, Teresa Owen, Wen Sun, Victor Wang, Christopher Cistrone, Javier Rangel-Moreno, Lianping Xing, and Jennifer H. Anolik. 2016. "Production of RANKL by Memory B Cells: A Link Between B Cells and Bone Erosion in Rheumatoid Arthritis." *Arthritis & Rheumatology (Hoboken, N.J.)* 68 (4): 805–16.

Meffre, E., M. Milili, C. Blanco-Betancourt, H. Antunes, M. C. Nussenzweig, and C. Schiff. 2001. "Immunoglobulin Heavy Chain Expression Shapes the B Cell Receptor Repertoire in Human B Cell Development." *The Journal of Clinical Investigation* 108 (6): 879–86.

Mori, Lucia, Marco Lepore, and Gennaro De Libero. 2016. "The Immunology of CD1- and MR1-Restricted T Cells." *Annual Review of Immunology* 34 (1): 479–510.

Musters, Anne, Paul L. Klarenbeek, Marieke E. Doorenspleet, Giulia Balzaretti, Rebecca E. E. Esveldt, Barbera D. C. van Schaik, Aldo Jongejan, et al. 2018. "In Rheumatoid Arthritis, Synovitis at Different Inflammatory Sites Is Dominated by Shared but Patient-Specific T Cell Clones." *Journal of Immunology* 201 (2): 417–22.

O'Neill, Shannon K., Mark J. Shlomchik, Tibor T. Glant, Yanxia Cao, Paul D. Doodes, and Alison Finnegan. 2005. "Antigen-Specific B Cells Are Required as APCs and Autoantibody-Producing Cells for Induction of Severe Autoimmune Arthritis." *Journal of Immunology* 174 (6): 3781–88.

Panayi, G. S., J. S. Lanchbury, and G. H. Kingsley. 1992. "The Importance of the T Cell in Initiating and Maintaining the Chronic Synovitis of Rheumatoid Arthritis." *Arthritis and Rheumatism* 35 (7): 729–35.

Pender, Michael P., Peter A. Csurhes, Jacqueline M. Burrows, and Scott R. Burrows. 2017. "Defective T-Cell Control of Epstein-Barr Virus Infection in Multiple Sclerosis." *Clinical & Translational Immunology* 6 (1): e126.

Rao, Deepak A. 2018. "T Cells That Help B Cells in Chronically Inflamed Tissues." *Frontiers in Immunology* 9 (August): 1924.

Rao, Deepak A., Michael F. Gurish, Jennifer L. Marshall, Kamil Slowikowski, Chamith Y. Fonseka, Yanyan Liu, Laura T. Donlin, et al. 2017. "Pathologically Expanded Peripheral T Helper Cell Subset Drives B Cells in Rheumatoid Arthritis." *Nature* 542 (7639): 110–14.

Raychaudhuri, Soumya, Cynthia Sandor, Eli A. Stahl, Jan Freudenberg, Hye-Soon Lee, Xiaoming Jia, Lars Alfredsson, et al. 2012. "Five Amino Acids in Three HLA Proteins Explain Most of the Association between MHC and Seropositive Rheumatoid Arthritis." *Nature Genetics* 44 (3): 291–96.

Roosnek, E., and A. Lanzavecchia. 1991. "Efficient and Selective Presentation of Antigen-Antibody Complexes by Rheumatoid Factor B Cells." *The Journal of Experimental Medicine* 173 (2): 487–89.

Scheel, Tobias, Angelika Gursche, Josef Zacher, Thomas Häupl, and Claudia Berek. 2011. "V-Region Gene Analysis of Locally Defined Synovial B and Plasma Cells Reveals Selected B Cell Expansion and Accumulation of Plasma Cell Clones in Rheumatoid Arthritis." *Arthritis and Rheumatism* 63 (1): 63–72.

Schröder, A. E., A. Greiner, C. Seyfert, and C. Berek. 1996. "Differentiation of B Cells in the Nonlymphoid Tissue of the Synovial Membrane of Patients with Rheumatoid Arthritis." *Proceedings of the National Academy of Sciences of the United States of America* 93 (1): 221–25.

Shevyrev, Daniil, and Valeriy Tereshchenko. 2019. "Treg Heterogeneity, Function, and Homeostasis." *Frontiers in Immunology* 10: 3100.

Shugay, Mikhail, Dmitriy V. Bagaev, Ivan V. Zvyagin, Renske M. Vroomans, Jeremy Chase

Crawford, Garry Dolton, Ekaterina A. Komech, et al. 2018. “VDJdb: A Curated Database of T-Cell Receptor Sequences with Known Antigen Specificity.” *Nucleic Acids Research* 46 (D1): D419–27.

Stephenson, William, Laura T. Donlin, Andrew Butler, Cristina Rozo, Bernadette Bracken, Ali Rashidfarrokhi, Susan M. Goodman, et al. 2018. “Single-Cell RNA-Seq of Rheumatoid Arthritis Synovial Tissue Using Low-Cost Microfluidic Instrumentation.” *Nature Communications*. https://doi.org/10.1038/s41467-017-02659-x.

Sun, Wen, Nida Meednu, Alexander Rosenberg, Javier Rangel-Moreno, Victor Wang, Jason Glanzman, Teresa Owen, et al. 2018. “B Cells Inhibit Bone Formation in Rheumatoid Arthritis by Suppressing Osteoblast Differentiation.” *Nature Communications* 9 (1): 5127.

Tickotsky, Nili, Tal Sagiv, Jaime Prilusky, Eric Shifrut, and Nir Friedman. 2017. “McPAS-TCR: A Manually Curated Catalogue of Pathology-Associated T Cell Receptor Sequences.” *Bioinformatics* 33 (18): 2924–29.

Titcombe, Philip J., Gustaf Wigerblad, Natalie Sippl, Na Zhang, Anna K. Shmagel, Peter Sahlström, Yue Zhang, et al. 2018. “Pathogenic Citrulline-multispecific B Cell Receptor Clades in Rheumatoid Arthritis.” *Arthritis & Rheumatology* 70 (12): 1933–45.

Turcinov, Sara, Erik af Klint, Bertrand Van Schoubroeck, Arlette Kouwenhoven, Sohel Mia, Karine Chemin, Hans Wils, et al. 2022. “The T Cell Receptor Repertoire and Antigen Specificities in Small Joints of Early Rheumatoid Arthritis - Diversity and Clonality.” *Arthritis & Rheumatology*. https://doi.org/10.1002/art.42407.

Wang, Shu, Jingya Wang, Varsha Kumar, Jodi L. Karnell, Brian Naiman, Phillip S. Gross, Saifur Rahman, et al. 2018. “IL-21 Drives Expansion and Plasma Cell Differentiation of Autoreactive CD11chiT-Bet+ B Cells in SLE.” *Nature Communications* 9 (1): 1758.

Weyand, Cornelia M., and Jörg J. Goronzy. 2021. “The Immunology of Rheumatoid Arthritis.” *Nature Immunology* 22 (1): 10–18.

Wucherpfennig, K. W., and J. L. Strominger. 1995. “Molecular Mimicry in T Cell-Mediated Autoimmunity: Viral Peptides Activate Human T Cell Clones Specific for Myelin Basic Protein.” *Cell* 80 (5): 695–705.

Wu, Fengping, Jinfang Gao, Jie Kang, Xuexue Wang, Qing Niu, Jiaxi Liu, and Liyun Zhang. 2021. “B Cells in Rheumatoid Arthritis：Pathogenic Mechanisms and Treatment Prospects.” *Frontiers in Immunology* 12 (September): 750753.

Yamin, Rachel, Orit Berhani, Hagit Peleg, Suhail Aamar, Natan Stein, Moriya Gamliel, Issam Hindi, Anat Scheiman-Elazary, and Chamutal Gur. 2019. “High Percentages and Activity of Synovial Fluid NK Cells Present in Patients with Advanced Stage Active Rheumatoid Arthritis.” *Scientific Reports* 9 (1): 1351.

Yoshitomi, Hiroyuki, and Hideki Ueno. 2021. “Shared and Distinct Roles of T Peripheral Helper and T Follicular Helper Cells in Human Diseases.” *Cellular & Molecular Immunology* 18 (3): 523–27.

Yost, Kathryn E., Ansuman T. Satpathy, Daniel K. Wells, Yanyan Qi, Chunlin Wang, Robin Kageyama, Katherine L. McNamara, et al. 2019. “Clonal Replacement of Tumor-Specific T Cells Following PD-1 Blockade.” *Nature Medicine* 25 (8): 1251–59.

Zhang, Fan, Anna Helena Jonsson, Aparna Nathan, Kevin Wei, Nghia Millard, Qian Xiao, Maria Gutierrez-Arcelus, et al. 2022. “Cellular Deconstruction of Inflamed Synovium Defines Diverse Inflammatory Phenotypes in Rheumatoid Arthritis.” *bioRxiv*. https://doi.org/10.1101/2022.02.25.481990.

Zhang, Fan, Kevin Wei, Kamil Slowikowski, Chamith Y. Fonseka, Deepak A. Rao, Stephen Kelly, Susan M. Goodman, et al. 2019. “Defining Inflammatory Cell States in Rheumatoid Arthritis Joint Synovial Tissues by Integrating Single-Cell Transcriptomics and Mass Cytometry.” *Nature Immunology* 20 (7): 928–42.

Zhao, Juanjuan, Shuye Zhang, Yang Liu, Xiaomeng He, Mengmeng Qu, Gang Xu, Hongbo Wang, et al. 2020. “Single-Cell RNA Sequencing Reveals the Heterogeneity of Liver-

Resident Immune Cells in Human." *Cell Discovery* 6 (1): 22.

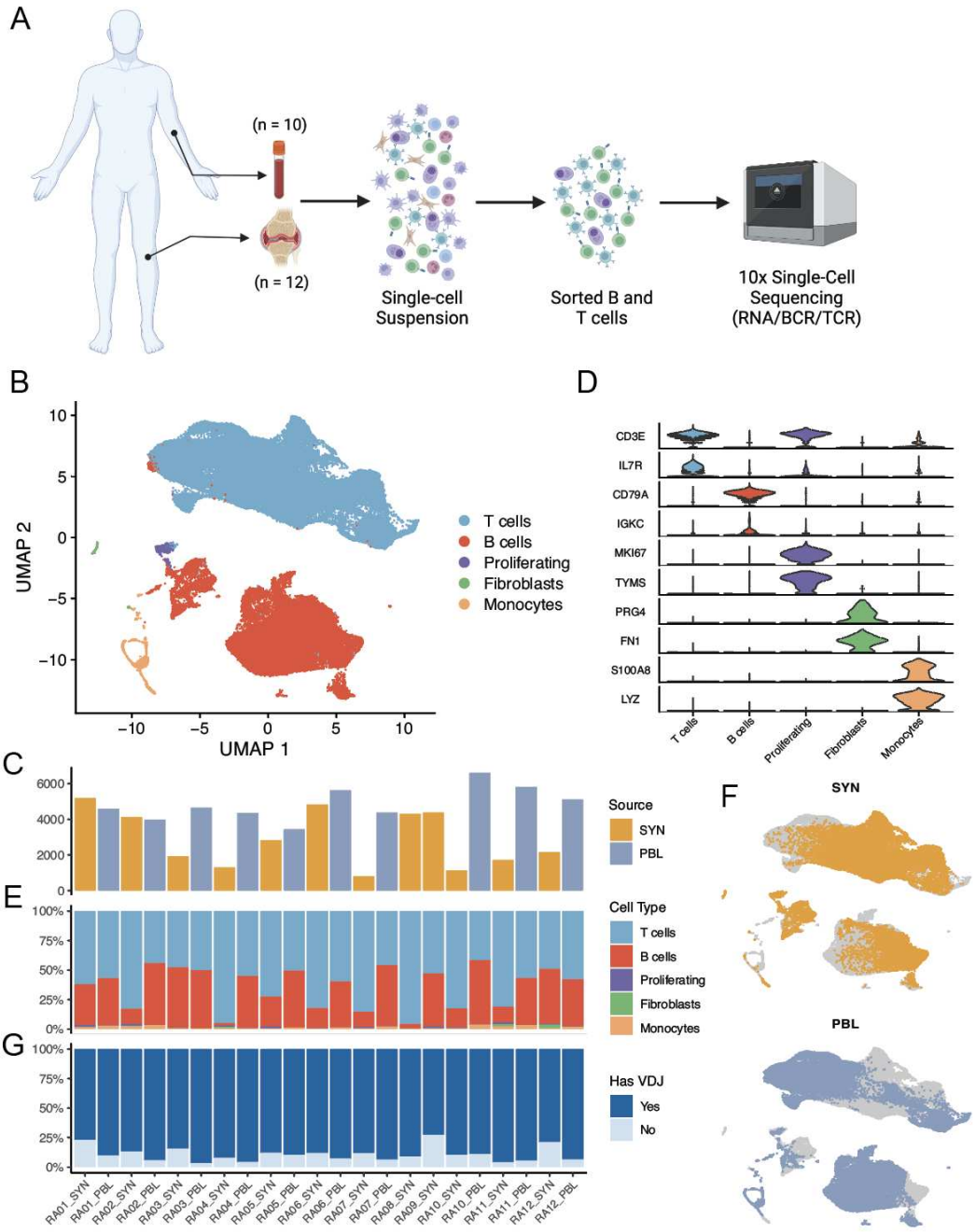

Figure 1. Sorting and single-cell analysis of matched blood and synovial tissue T cells and B cells. A. Schematic showing the overall study design. T and B cells were isolated from synovial tissue biopsies (n=12) and matched peripheral blood (n = 10). Single-cell libraries for 5' gene expression and receptor repertoires were generated using the 10X Genomics platform. **B.** Unsupervised clustering and UMAP projection of 83,159 cells that passed QC. **C.** Bar plot of the number of cells recovered for each sample, color denotes tissue origin. **D.** Violin plots showing the expression distribution of select markers for each identified cell population. **E.** Bar plot of the cluster composition for each sample. **F.** UMAPs of the combined clustering, separated by peripheral blood (top) and synovial tissue (bottom). **G.** Bar plot highlighting the recovery of VDJ

information for each sample, excluding fibroblast and monocyte populations. PBL, peripheral blood lymphocytes. SYN, synovial tissue. Figure (A) created using BioRender.

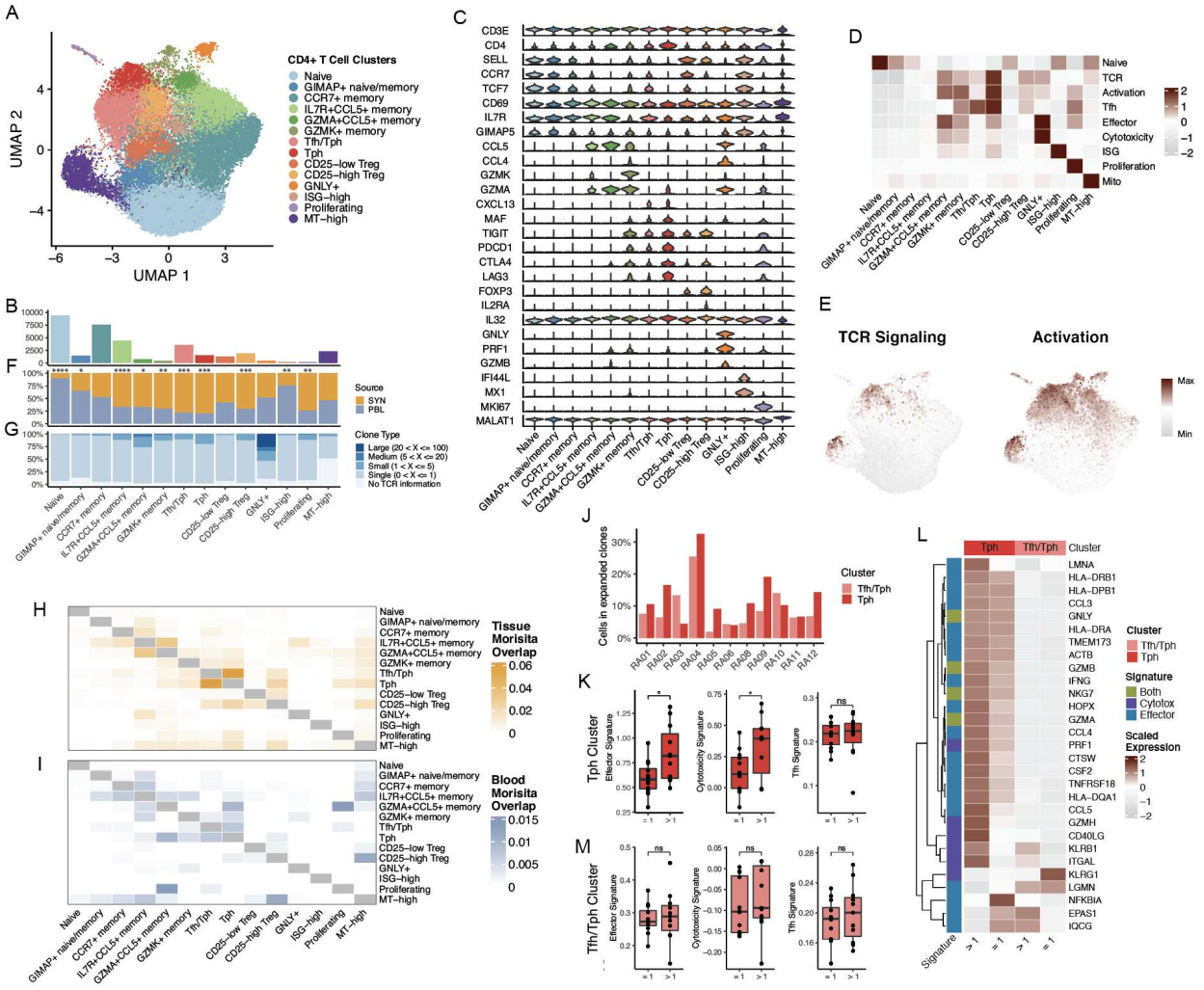

Figure 2. Enriched representation of effector CD4+ populations in synovial tissue. A. UMAP projection of CD4+ T cell reclustering. **B.** Bar plot of the number of cells included in each cluster. **C.** Violin plots of the expression distribution of select, differentially-expressed markers for each cluster. **D.** Heat map showing the scaled module score of select gene signatures. **E.** UMAPs of TCR signaling (left) and activation (right) signatures. **F.** Bar plot of the tissue distribution for cells in each cluster, with significance determined by paired T-tests in Figure S4E. **F.** Bar plot of the clone size distribution for each cluster. **H and I**. Heat map of pairwise clonal overlap values calculated using Morisita's index for synovial tissue **(H)** and blood **(I)**. **J.** Bar plots of the percent of synovial tissue Tph and Tfh/Tph cells found to be expanded, per donor. **K.** Box plots of the effector signature (left), cytotoxicity signature (middle), and Tfh signature (right) distribution for expanded and non-expanded Tph cells. Each dot represents a donor, and P values were determined by paired T-tests. **L.** Heat map of the average expression of effector and cytotoxicity signature genes, comparing expanded and non-expanded Tph and Tfh/Tph cells. **M.** Box plots of the effector signature (left), cytotoxicity signature (middle), and Tfh signature (right) distribution for expanded and non-expanded Tfh/Tph cells. Each dot represents

a donor, and P values were determined by paired T-tests. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$; ns, not significant.

Figure 3. GZMK and GZMB+ T cells are not clonally related. A. UMAP projection of CD8+ T cell reclustering. **B.** Bar plot of the number of cells included in each cluster. **C.** Violin plots of the expression distribution of select, differentially-expressed markers for each cluster. **D.** Heat map showing the scaled module score of select gene signatures. **E.** Bar plot of the tissue distribution for cells in each cluster, with significance determined by paired T-tests in Figure S5D. **F.** Bar plot of the clone size distribution for each cluster. **G and H**. Bar plots of the clone sizes and cluster breakdowns of the top 50 most expanded clones for synovial tissue **(G)** and blood **(H)** separately. Letters above bars denote clones shared between synovial tissue and blood. **I.** Donut plots of the prolfieration cluster mapping proportions compared to the general cluster proportions. **J.** Bar plot of the percent of CD8+ cells that are exact viral-reactive matching, per patient, split by virus. Box size denotes size of clone. **K.** Bar plot of the cluster distribution for exact viral matching and non-matching cells. **L.** Bar plot of the percent of CD8+ cells that have a GLIPH2 motif matching viral-reactive clones, per patient, split by virus. **M.** Bar plot of the cluster

distribution for GLIPH2 motif viral-reactive matching and non-matching cells. **N.** Bar plot of the clone size distribution for GLIPH2 motif viral-reactive matching and non-matching cells. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; **** $p \leq 0.0001$.

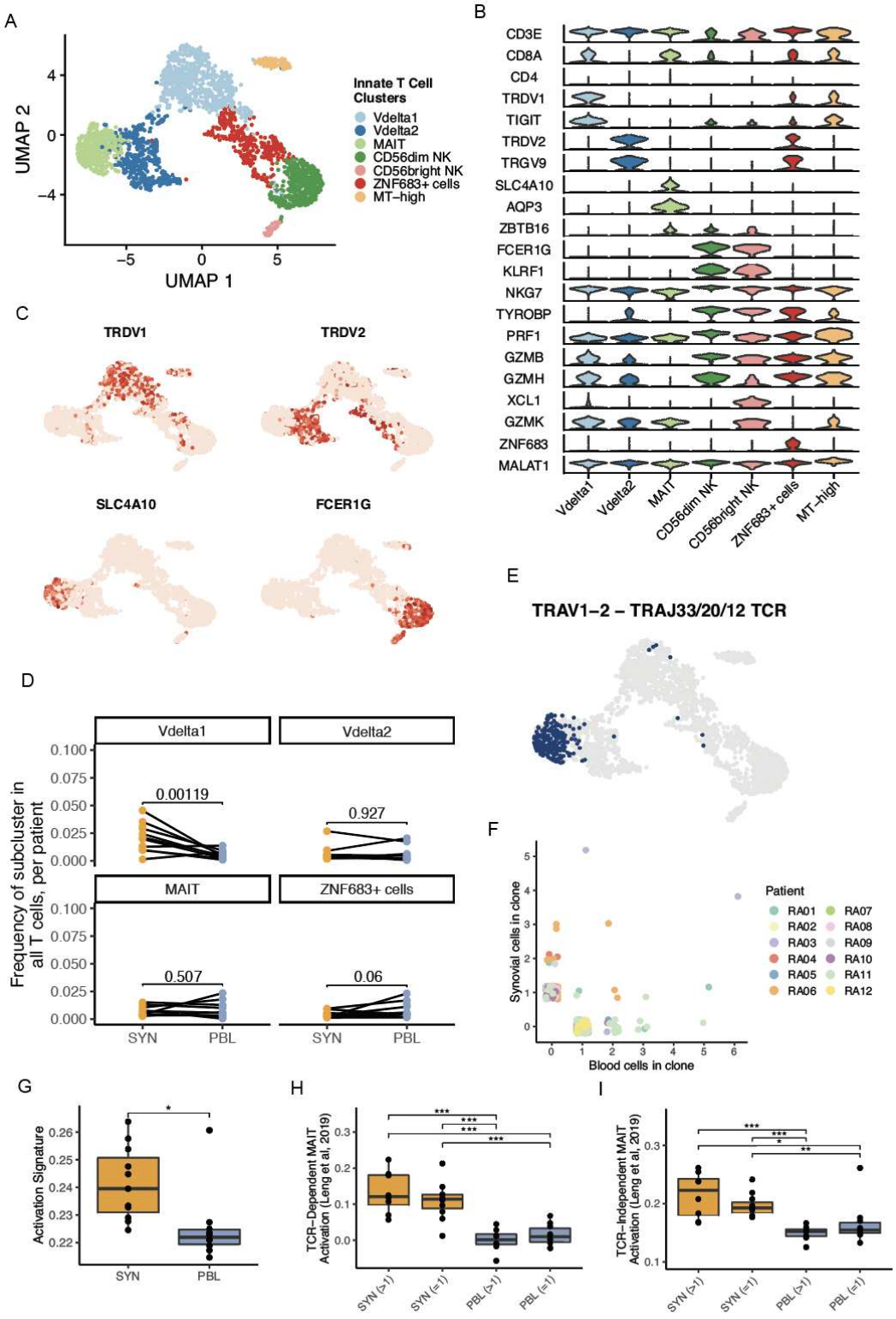

Figure 4. Innate T cell populations have increased activation signatures in the synovium. **A.** UMAP projection of innate T cell reclustering. **B.** Violin plots of the expression distribution of select, differentially-expressed markers for each cluster. **C.** UMAPs of the gene expression levels of select innate T cell lineage markers. **D.** For each selected cluster, frequency of cell

representation as a proportion of all T cells in a sample. Each dot represents a single sample, and lines denote paired blood and tissue for a donor. Significance is determined by paired T-tests, with multiple testing correction. **E.** UMAP highlighting innate T cells with a TRAV1-2 and either TRAJ33, 20, or 12 gene rearrangement. **F.** Scatter plot of the size of each MAIT cell clone across tissue sources, colored by patient. **G.** Box plot of an activation signature in the MAIT cell cluster. **H and I.** TCR-dependent **(H)** and independent **(I)** MAIT cell activation signatures from Leng et al, 2019, split by tissue and clonal expansion. Significance calculated using paired Wilcox testing with Holm correction. * $p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$; ns, not significant.

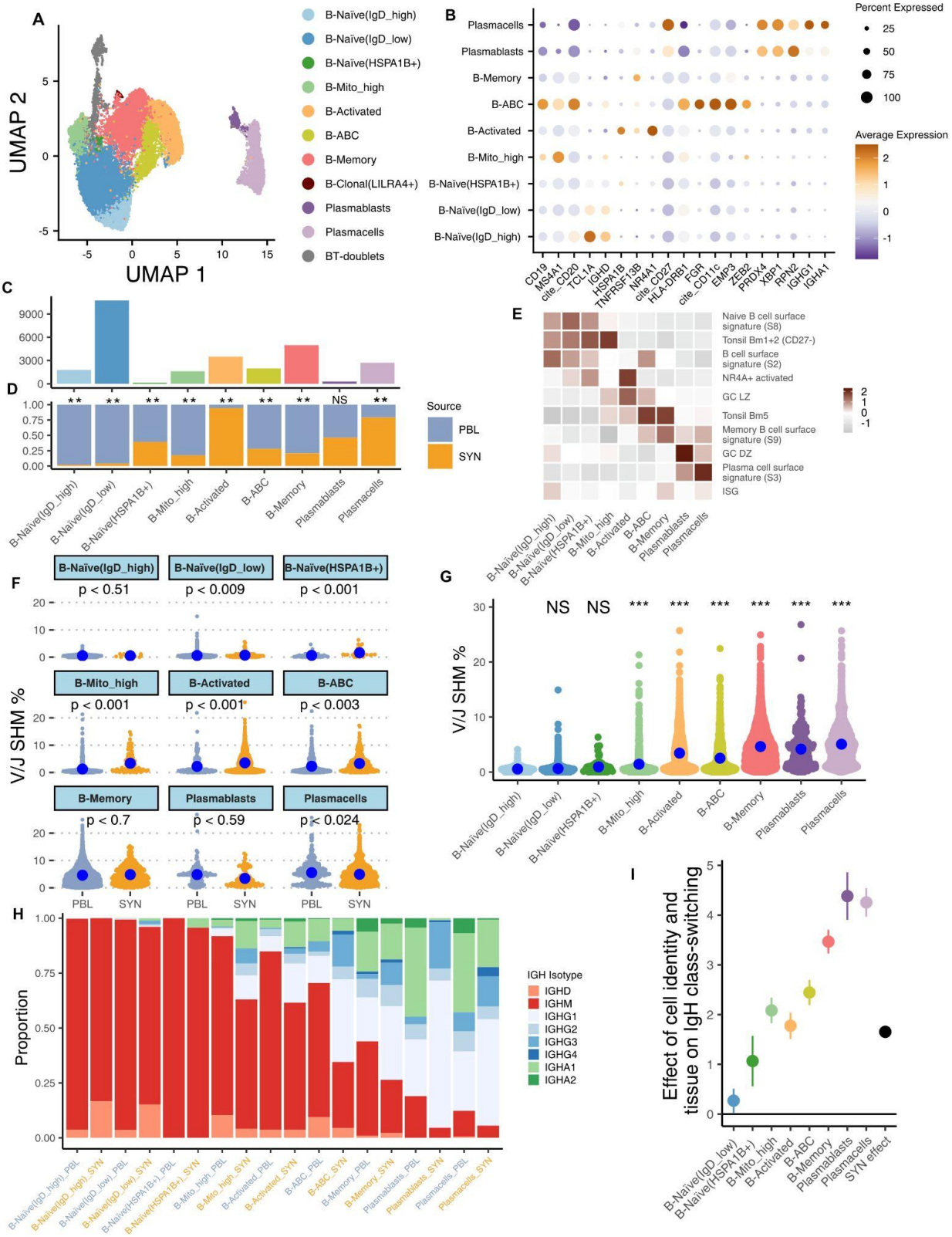

Figure 5. Accumulation of activated, somatically mutated B cell in the RA synovium. A. UMAP projection of B cell clustering. Unlabeled gray cells ("NA") were called B-T doublets as determined by high expression of *CD3E* and high doublet scores; these cells were omitted from downstream analysis. The population labeled B-Clonal(*LILRA4*+) was also omitted from downstream analysis due to having a small number of cells which were only found in 2 blood samples. **B.** Dot plots of salient markers used in annotating clusters. **C.** Bar plot of the number of cells included in each cluster. **D.** Bar plot of the tissue distribution within each cluster. Significance determined by mixed effect model using MASC, shown in Figure S8D. **E.** Heat map showing the scaled module score of select gene signatures. Heatmaps of selected genes for pathways available in Figure S10A. **F.** Plot quantifying differences in SHM rate between tissue within each population. P-values assessed through a linear mixed-effect model with random effect for donor. **G.** Plot quantifying differences in SHM rate between clusters. P-values assessed through a linear mixed-effect model with random effect for donor. Reference population set to B-Naive(IgD-high). **H.** Bar chart of productive immunoglobulin heavy chain (IgH) isotype usage for each cluster split by tissue after QC. **I.** Plot of effect sizes and 95% confidence intervals for a mixed-effect logistic regression model which regresses a cell's class-switch state (IgG or IgA -> "class-switched", IgM or IgD -> "not class-switched") on its phenotype and tissue source. Random effects included for donor, B-Naive(IgD-high) set as the reference population and PBL set as the reference tissue source. *** $p \leq 0.01$; NS, not significant.

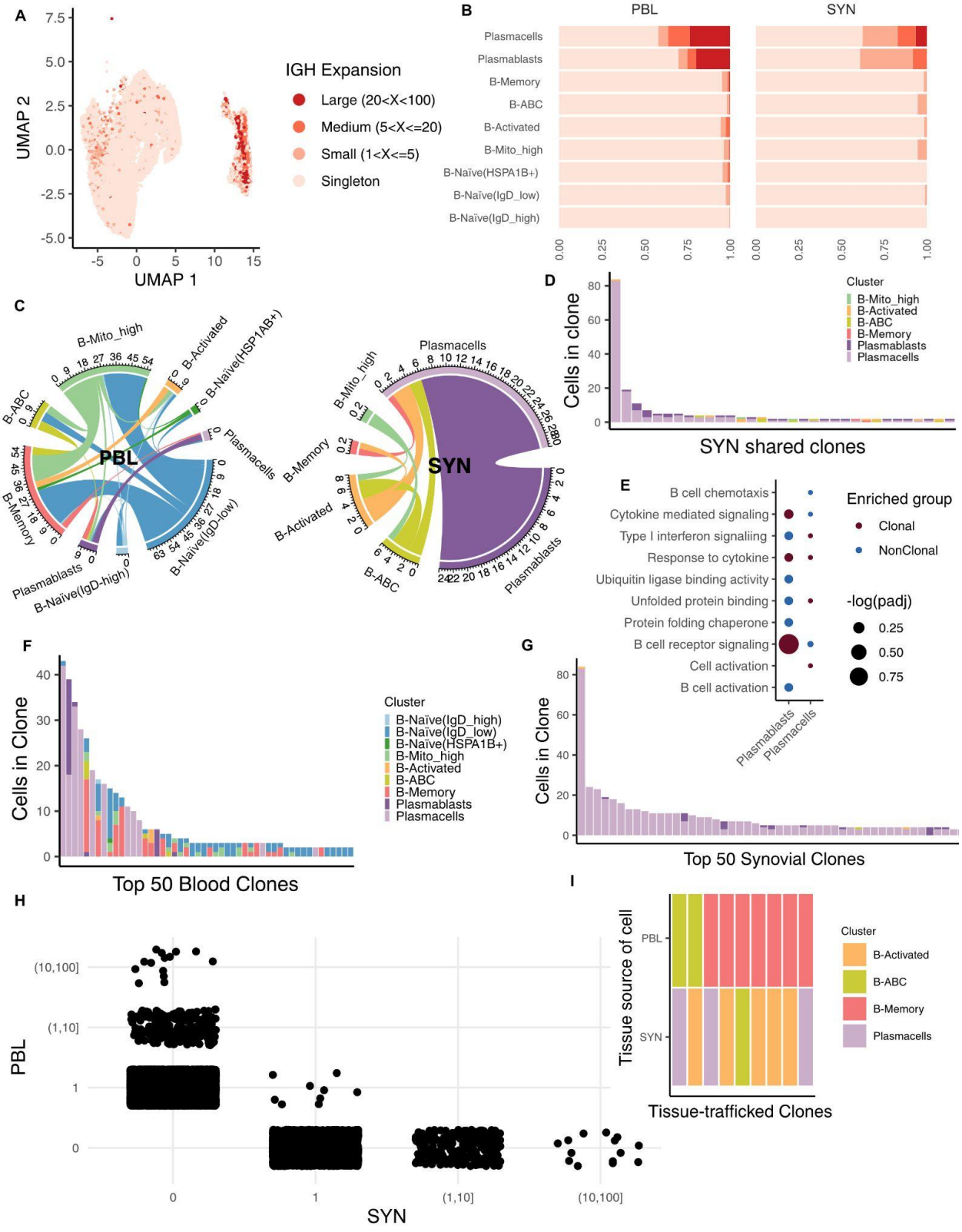

Figure 6. Clonal expansion and clonal sharing between B cell states in the synovium. A. UMAP projection of IgH-defined clonal expansion in B cells. Cells without BCR are not included in this plot or other plots in this figure. **B.** Bar chart showing proportion of clone sizes for each cluster split by tissue. **C.** Circos plot showing clonal sharing between cell populations split by tissue compartment. The color of the edge corresponds to the proposed originating population. **D.** Bar plot showing the cell identity and clone size for each synovial trafficked clone. **E.** Dotplot of selected pathways checked for gene set enrichment between clonal cells and non-clonal cells in the synovium. Cell states were omitted that had a limited number of clones. **F.** Bar plot displaying the 50 largest clones in the blood and their cell population composition. **G.** Bar plot displaying the 50 largest clones in the synovium and their cell population composition. **H.** Plot displaying the distribution of clones across each tissue and clones shared between tissues. **I.** Plot of cell population composition of the tissue-trafficked clones.

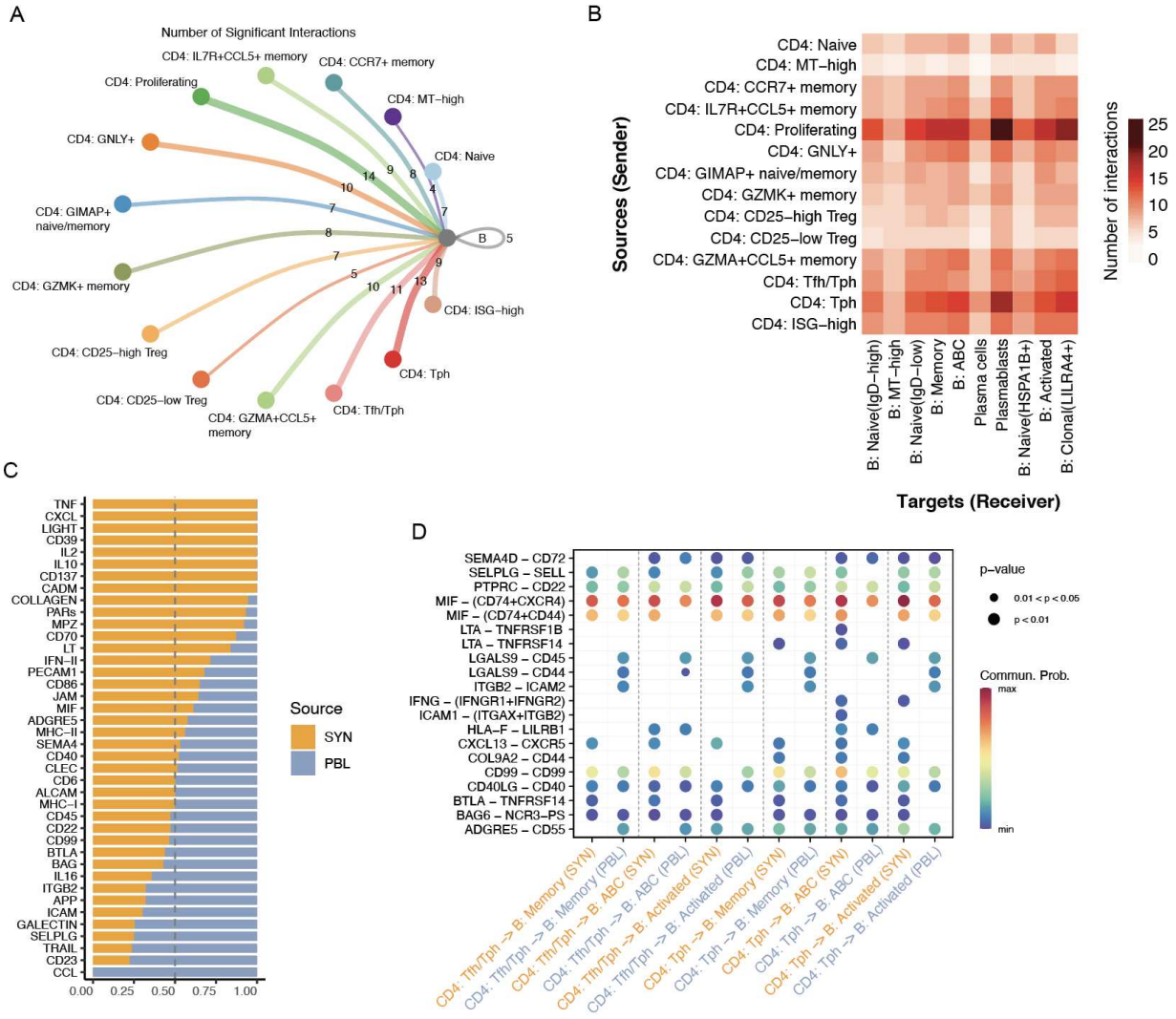

Figure 7. Identification of altered T-B cell communication pathways. A. Circle plot of the number of significant cell-cell interactions identified between CD4+ T cell subsets and B cells. **B.** Heatmap of the number of significant cell-cell interactions identified between CD4+ T cell subsets as senders and B cell subsets as receivers. **C.** Bar plot of the proportion of a signaling pathway's detection in synovial tissue or blood. **D.** Dot plot of significant interactions detected between Tph or Tfh/Tph cells and ABC, Memory, or Activated B cells, split by tissue.