Single cell analysis of RA synovial B cells reveals a dynamic spectrum of ectopic lymphoid B cell activation and hypermutation characterized by NR4A nuclear receptor expression

Nida Meednu[1], Javier Rangel-Moreno[1], Fan Zhang[2,3,4,5,6], Katherine Escalera-Rivera[1], Elisa Corsiero[7], Edoardo Prediletto[7], Edward DiCarlo[8], Susan Goodman[8], Laura T Donlin[8], Soumya Raychaudhuri[2,3,4,5,6,9], Michele Bombardieri[7], Costantino Pitzalis[7], Dana E Orange[8,10], Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Network[*], Andrew McDavid[1,13], and Jennifer H Anolik[1,11,12]

[1]Division of Allergy, Immunology and Rheumatology, Department of Medicine, University of Rochester Medical Center, Rochester, NY, USA.
[2]Center for Data Sciences, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
[3]Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA.
[4]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.
[5]Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
[6]Division of Rheumatology, Inflammation, and Immunity, Brigham and Women's Hospital, Boston, MA, USA.
[7]Centre for Experimental Medicine & Rheumatology, William Harvey Research Institute, Queen Mary University of London, EC1M 6BQ, London, UK
[8]Division of Rheumatology, Hospital for Special Surgery, New York, NY 10021, USA
[9]Faculty of Medical and Human Sciences, University of Manchester, Manchester, UK.
[10]Rockefeller University, New York, NY, USA.
[11]Department of Pathology, University of Rochester Medical Center, Rochester, NY, USA.
[12]Center for Musculoskeletal Research, University of Rochester Medical Center, Rochester, NY, USA.
[13]Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA.
[*]A list of members and affiliations appears at the end of the paper

Corresponding author
E-mail: Jennifer_anolik@urmc.rochester.edu

ABSTRACT

Ectopic lymphoid structures (ELS) are present in rheumatoid arthritis (RA) synovial tissue, but the precise pathways of B cell activation and the role of *in situ* synovial B cell differentiation and selection in disease are not well understood. Here, we identified a B cell population in the synovium characterized by expression of NR4A1-3, a family of orphan nuclear receptors, that is highly enriched at both early and late stages of RA. NR4A B cells are rare in healthy peripheral blood, RA blood, and SLE kidney, but share markers with blood transcriptomic signatures that peak during RA disease flare. Using combined single cell transcriptomics and B cell receptor (BCR) sequencing, we demonstrate that NR4A synovial B cells have an activated transcriptomic profile that significantly overlaps with germinal center (GC) light zone (LZ) B cells and an accrual of somatic hypermutation that correlates with loss of naïve B cell status. NR4A B cells uniquely co-express lymphotoxin β and IL6, supporting important functions in ELS promotion and pro-inflammatory cytokine production. The presence of shared clones in this activated B cell state, NR4A expressing synovial plasma cells (PC), and $CCR6^{+}$ memory B cell (MBC) precursors further points to *in situ* differentiation. NR4A1 was expressed at the protein level in RA synovial B cells and PC, was high in tonsil GC B cells with a light zone (LZ)-dark zone (DZ) intermediate phenotype, and was rapidly induced at both the RNA and protein level upon activation through the BCR. Taken together, we identified a dynamic progression of B cell activation in RA synovial ELS, with NR4A as a read-out of likely antigen activation and local adaptive immune responses.

INTRODUCTION

Rheumatoid arthritis is a chronic autoimmune disease characterized by inflammation of the synovial tissue leading to joint damage and disability. Although dramatic advances in treatment options for RA have been achieved over the last two decades, a significant number of RA patients do not achieve remission or low disease activity (*1, 2*) , highlighting the need for new therapies and biomarkers of response. B cells play a key role in RA disease pathogenesis both through autoantibody mediated and antibody independent functions and are key treatment targets in the disease (*3-6*). Though B cell clonal expansion is detected in RA blood (*7*) and synovium (*8*) and has even been found in pre-RA with anti-citrullinated protein autoantibodies (ACPA) autoimmunity (*9*), the mechanisms and location of antigen specific B cell priming remains unclear.

B cell aggregates are present in the RA synovium (*10, 11*) at both early and late stages of disease and have been linked to important clinical outcomes including treatment response, bone erosion, and radiographic progression (*3, 11*). Ectopic lymphoid structures (ELS) are found in at least 40% of RA patient synovia with 10-25% of those structures displaying features of a functional germinal center (GC) (*12-15*). ELS organization can recapitulate the architecture of secondary lymphoid organs (SLO), with areas that resemble dark zones (DZ) (proliferating and AID expressing B cells) and areas that resemble light zones (LZ) (rich in follicular dendritic cells and T cells in addition to B cells) (*16*). There are, however, important differences including that ELS are exposed directly to antigens and regulatory mediators produced within the inflamed tissue. In RA, the chronic exposure to autoantigens in synovial tissue may favor ELS formation (*15*). ELS in RA synovial tissue has been shown to support antigen-driven selection and differentiation of autoreactive B cells, facilitating diversification, somatic hypermutation (SHM) (*15*) and class switching (*17*) for in situ production of autoantibodies (*18*), suggesting that these structures are

functional. Many critical factors driving ectopic lymphoid neogenesis, including lymphoid chemokines, are shared with factors driving SLO development, but the cellular components that produce these key factors may differ and are the subject of ongoing investigation. In mouse colitis lymphotoxin (LT) expressing B cells can support ELS formation and lead to severe inflammatory disease (*19*). The role of B cells during ELS in RA synovial tissue, however, is unclear and which B cell subsets may be crucial for this process is still unknown.

As part of the Accelerating Medicines Partnership (AMP) network, we recently identified four B cell subsets including naïve, memory B cells, age-associated B cells, and PCs in RA synovial tissue using scRNA-seq (*20*). To improve our understanding of the cellular, transcriptional, and antibody repertoire dynamics during human B cell activation in ELS, we performed unbiased single-cell transcriptomic and repertoire profiling of RA synovial B cells. These single-cell antibody repertoires paired with single-cell transcriptomics allowed us to define novel transcriptional B cell states, resolving functional B cell heterogeneity and gene expression dynamics. We identified an abundant population of synovial B cells characterized by up-regulation of NR4A1-3 along with other immediate early response genes (EGR1/3, FOS, JUN) and activation markers (CD69, CD83). These cells show evidence of SHM, class-switching, and clonal relationship with a subset of NR4A expressing PCs. Gene expression profiling of synovial NR4A B cells demonstrated enrichment for genes associated with GC LZ B cells and expression of elevated levels of chemokine receptors and chemotactic factors involved in ectopic lymphoid neogenesis, including LT-β and IL6. This subset of B cells is enriched in synovial tissue and found at very low levels in peripheral blood and another autoimmune target tissue, the kidney in lupus nephritis. Notably, RA synovial B cells spontaneously express NR4A1 protein by flow cytometric and histologic analysis, and NR4A cluster genes are enriched in synovial samples with a lymphoid

histologic pathotype (*10*). In an RA flare cohort (*21*), an NR4A B cell transcriptomic signature peaked in the blood during flare. Finally, stimulation of normal B cells through the BCR *in vitro* up-regulated NR4A1-3. Together our data support *in situ* activation of B cells with accrual of SHM in synovium ELS and NR4A as a read-out of antigen stimulation, local adaptive immunity, and pathological B cell responses.

RESULTS

Unbiased scRNA-seq analysis of RA B cells reveals multiple distinct subsets including a unique NR4A B cell population in the synovium

In order to define the diversity of B cells in the synovium, we isolated B cells from lymphoid-rich RA synovial tissues and paired blood and performed single cell sequencing on the 10X Chromium platform (Fig. 1 and table S1). In addition, we obtained paired heavy and light V(D)J sequences for the BCR of individual B cells from these samples. B cells were sorted by flow cytometry as $CD19^+$. The mean frequency of B cells by flow cytometry was 12.65% (2.7%-21.6%, n=4) of $CD45^+$ synovial cells. The majority of synovial B cells were IgD^-, whereas 80% of blood B cells were naïve (IgD^+CD27^-). We also observed synovial enrichment of $CD24^-CD27^{hi}$ plasmablasts/plasma cells (1.9%-7.6%) (fig. S1B).

We sequenced 4,115 B cells (2,589 synovial and 1,526 peripheral blood) with an average median genes per cell of 1,427 (755-1,868). After exclusion of cells with high mitochondrial RNA content, dimensionality reduction using t-distributed stochastic neighbor embedding (t-SNE) (*22*) and unsupervised clustering using Seurat package (*23*) on the combined data set of 3,786 cells, 8 conserved B cell subsets or clusters were defined (see *Single cell RNA and repertoire sequencing analysis in Methods).* Differentially expressed genes (DEGs) were identified by calculating the

difference between the average expression by cells in the subset versus cells not in the subset (Fig. 2). Two subpopulations of plasma cells (PC(i) and PC (ii)) were characterized by high expression of transcription factors required for PC differentiation and maintenance (*XPB1, IRF4* and *PRDM1)* and immunoglobulin genes (Fig. 2, A-B and D). Expression of *IGHD* and *TCL1A* identified three naïve B cell subpopulations (Naïve(i), Naïve (ii) and Naïve (iii)) (Fig. 2, A and B). Two B cell subpopulations expressed high levels of *CD27, HOPX*, *S100A10 and S100A4*, markers associated with memory B cells (*24-27*). The $LMNA^+$ designated subpopulation also uniquely expressed the *LMNA* gene, not previously reported in memory B cells (Fig. 2, A-B and D). The final subpopulation was labeled $NR4A^+$ as it expressed high levels of the nuclear family receptor 4 (*NR4A1, NR4A2, NR4A3*). This subpopulation also showed high expression of markers associated with activation and GC formation like *CD69*, *CD83* and *GPR183*, the latter a molecule mediating B cell migration in lymphoid follicles (*28, 29*) (Fig. 2, A-B and D).

The $NR4A^+$ and $LMNA^+$ subpopulations were restricted to synovial samples. Naïve B cells were the main population in the blood, whereas the memory and PC subpopulations were present in both blood and synovial tissue (Fig. 2C, and fig. S2C).

Unique enrichment of NR4A B cells in RA synovium compared to blood and other tissues

Using a supervised classification method SingleR (*30*), we next examined the presence of NR4A B cells in an independent RA synovial tissue cohort and other tissues where single cell data was available. In SLE kidney (*31*), peripheral blood B cells from our RA134 subject, and blood from a healthy donor (*32*), NR4A B cells were very rare at 0.7% -1.5% abundance. In contrast NR4A B cells were highly abundant (>40% abundance) in synovial tissue from AMP phase I (*20*), similar to samples in our current study (**Fig. 2E**). The AMP Phase I study also performed bulk RNA

sequencing from sorted B cells from a larger number of synovial samples, including 16 RA biopsies, 13 RA arthroplasties, and 10 osteoarthritis (OA) arthroplasties (*20*). Several markers including *CD83* and *NR4A1,* which defined the NR4A$^+$ cluster were upregulated in B cells from RA biopsy samples compared to B cells from RA and OA arthroplasty samples (Fig. 2F). When samples were separated into inflamed leucocyte-rich vs leucocyte-poor (*20*), this NR4A cluster enrichment appeared to be driven by the inflammatory state of the tissue (1.7-fold and 2.1-fold increase in *CD83* and *NR4A1* gene expression leucocyte-rich RA vs OA, $p<0.05$; NS for leucocyte-poor RA vs OA).

Evidence of somatic hypermutation in synovial B cells

To define the Ig repertoire in single B cells, we performed V(D)J sequencing on the same sorted CD19$^+$ cells used to generate the 5'-based gene expression library, allowing a joint analysis of transcriptional profile and immune repertoire in the same cell. Of the total cellular barcodes recovered from sequencing, 64% (range: 29%-90%) are cells with productive V-J spanning pairs, with 47% containing a paired Ig heavy and light chain and 16% with only a single Ig heavy or light chain (fig. S2, A and B). There was a difference in the frequency of BCR recovery between B cell clusters, with the LMNA$^+$ cluster having the lowest frequency (fig. S2C).

Somatic hypermutation (SHM) is a critical component of the affinity maturation process to generate high-affinity antibodies. Ig SHM classically occurs in the GC dark zone (DZ), and high SHM rates are characteristics of post-GC B cells including memory and plasma cells (*33*). We calculated SHM rates in each B cell as the number of substitutions from germline in heavy and light chains vs the total length recovered chains. Similar to previously published data all three naïve clusters exhibited very low SHM frequencies, while memory and plasma cell clusters

showed the highest mutation rates (*34-36*) (Fig. 3A and fig. S2F). The LMNA$^+$ cluster has a mutation rate comparable to that of the memory B cells, supporting the classification of LMNA$^+$ synovial B cells as memory cells that may have developed in the synovium. Though mutated cells are apparent in the NR4A$^+$ cluster, the average SHM rate was not significantly higher than naive B cells (1.5%, 95% CI -1%--4%) while memory and plasma cells had clear differences compared to naive, suggesting that these cells may be in an earlier stage of the affinity selection and maturation process (Fig. 3A). Furthermore, we detected *IGHA* expression in a small number of NR4A B cells, providing evidence of class-switch recombination (fig. S2E). Next, we looked at the relationship between SHM rate and the expression levels of *CD27*, *NR4A* genes and *IGHD*. As expected, all subpopulations showed increased SHM with increased expression of CD27, while NR4A and memory subpopulations showed inverse associations of SHM with *IGHD*. However, particular to only the NR4A$^+$ cluster, an accumulation of SHM positively associated with the expression levels of *NR4A1* and *NR4A2* (Fig. 3B), suggesting a potential role for *NR4A* genes in the RA synovial immune reaction. There was no significant association found between the expression of either the *NR4A* genes nor *IGHD* in naïve and LMNA$^+$ clusters. Importantly, we detected expansion of 43 families of putative clones in the synovium and five in the peripheral blood, most prominently featured in the PC and NR4A clusters. Each clonal family was confined to a single transcriptomic subpopulation (naïve, memory, LMNA, NR4A, or PC), and this would not be expected to occur by random chance ($p<0.001$ by permutation test). The only exception was the clonal family CARHWRGKKPFDSW which was detected in both the NR4A and PC cluster, consistent with a developmental relationship between these subsets (Fig. 3, C and D and fig. S2D).

Enrichment of an activation and GC formation transcriptomic profile in the NR4A subset

To further clarify the 8 synovial B cell clusters, we next performed gene set enrichment analysis (GSEA). All three naïve clusters were significantly enriched with naïve B cells from blood and tonsil (Bm1+Bm2) (Fig. 4A). In contrast, memory and $LMNA^+$ clusters were enriched for gene sets expressed by tonsil Bm5 (a memory subset) and to a lesser extent memory B cells from blood, supporting their identity as memory B cells (Fig. 4A). The analysis also confirmed the identity of PC clusters (*37*) (Fig. 4A). The NR4A cluster was strongly enriched for genes expressed by GC LZ B cells (Fig. 4, A and C). These genes include *EGR1/2/3, CD83, BCL2A1* and *GPR183* (*38*) (Fig. 4B).

Recently, single cell RNA sequencing of GC B cells from conventional SLO has revealed a dynamic spectrum of DZ, LZ and intermediate zone B cells (*39, 40*). In particular, between the DZ and LZ there was a gradient of down-regulation of classic DZ genes like *CXCR4* and up-regulation of classic LZ genes like *CD83*. Along this gradient, there was an initial up-regulation of BCR signaling followed by NFkB signaling. The later stages of the intermediate zone GC B cells were characterized by up-regulation of CD40 and MYC signaling, possibly indicative of their potential to interact with T cells (*40*). In our data, the NR4A B cells in RA synovial tissue displayed characteristics compatible with these intermediate cell stages. First, in contrast to classic LZ, NR4A cluster cells expressed *CXCR4*, *BCL6* and *FOXP1*, genes associated with GC DZ (*38, 40, 41*) (Fig. 4, B and C and fig. S3). Second, NR4A B cells showed evidence of BCR stimulation (fig. S3) and high levels of immediate early genes associated with recent BCR stimulation including *NR4A1, NR4A3, EGR1* and *c-Fos* (*42-44*) and NFkB signaling (fig. S3). In contrast to classic LZ cells, the NR4A cluster does not show consistent up-regulation of CD40 and MYC signaling with the exception of *BCL2A1* (CD40 signaling) and *GPR183* (MYC signaling) (fig. S3). Our NR4A B

cell cluster shares remarkable transcriptomic similarity with a recently described activated B cell state in the human tonsil that appears to be on a trajectory to GC formation (*36*).

Highly expressed chemokines and cytokines in NR4A B cells

Chemokines and their interaction with corresponding chemokine receptors are vital for the organization of lymphoid organs and B cell functions (*16*). Thus, LTβ/LTβ-R, CXCL13/CXCR5 and CCL19/CCL21/CCR7 are critical for the recruitment and segregation of B/T cells in distinct areas in both SLO and ELS. A number of these ELS promoting factors are up-regulated in lymphoid-like RA synovial tissue (*45, 46*). To investigate the relevant chemokine and cytokine pathways in synovial tissue, we next examined differential gene expression of chemokines, cytokines and their receptors in each B cell cluster. We found that multiple synovial B cell subsets express relevant chemokine receptors. Thus, both naïve and NR4A B cells showed significant enrichment for *CXCR5*, *CCR7* and *CXCR4* (Fig. 5, A and B). These chemokine receptors are instrumental to orchestrate homing to B cell follicles (*CXCR5*), migration across high endothelial venules (HEVs) (*CCR7*), and dynamic mobilization inside GC (*CXCR4*) (*47, 48*), suggesting that these B cells can be recruited to the synovium and localize in response to specific chemokine signals. Receptors for B cell survival factors are also expressed by multiple B cell subsets. Naïve and NR4A subsets showed significant up-regulation of *TNFRSF13C* (BAFFR), while $LMNA^+$ and memory shared up-regulation of *TNFRSF13B* (TACI), a receptor that can bind to both B cell survival factors BAFF and APRIL. Consistent with previous reports, PC uniquely showed up-regulation of *TNFRSF17* (BCMA), a BAFF receptor important for PC survival (Fig. 5A). Furthermore, $NRA4^+$, $LMNA^+$ and memory showed significant enrichment of *TNFRSF1* (TNFR2), a receptor for TNF (Fig. 5A). We also observed that $LMNA^+$ and memory subsets

uniquely up-regulate *SIGIRR*, a negative regulator of IL-1 receptor signaling pathway implying response to IL1 family cytokines (Fig. 5A).

B cells can play a direct role in the organization of SLO by producing lymphotoxin β (*49*). More recently, B cell-derived IL6 has been implicated in spontaneous GC formation in systemic lupus (*50*). We found that all non-PC B cell subsets express lymphotoxin β (Fig. 5, C and D). However, the NR4A^{+} subset uniquely expressed *IL6*, and co-expressed *LTB* (Fig. 5C). As previously noted, *GPR183* (EBI2) is highly expressed in the NR4A subset. GPR183 has pleiotropic roles in B cell positioning depending on the balance of other chemokines. It is important for distributing B cells along the T-B boundary in the early stage of antibody responses (*29*) as well as B cell exit from the GC and re-entry to the periphery (*29*). *GPR183* together with *CXCR4*, also expressed at high levels by the NR4A subset, are critical factors for the segregation of GC B cells in the DZ and LZ (Fig. 5A). NR4A B cells also have elevated expression of a number of co-stimulatory and signaling molecules important for putative cell-cell interactions with APCs and T cells. In addition to *IL6* (interaction with IL6R on APCs and Tfh), this includes high levels of *TNFSF9* (4-1BBL) (Fig. 5, C and D), a ligand that binds to 4-1BB on activated CD4 and CD8 T cells to provide co-stimulatory signals (*51, 52*). Overall, these data suggest NR4A B cells may be a critical initiator of ELS formation in synovial tissue. Independent analysis of the RA cohort from AMP also supports these findings (fig. S5). Of note, we observed the presence of histologic ELS in our samples by immunofluorescent staining with classic GC markers (fig. S4).

One of the differentially expressed chemokine receptors in the NR4A subset is CCR6 (Fig. 5, A and B). Single cell analysis of GC B cells from human tonsil has identified a distinct cell cluster of memory B cell precursors (PreM), originating within the LZ that shares a gene signature with previously reported murine CCR6^{+} PreM B cells (*40, 53*). The range of *CCR6* expression

within the NR4A$^+$ cluster (Fig. 5E) suggests that memory precursor cells are originating here and is consistent with memory precursor commitment to post-GC differentiation in the early stage of the LZ (*40*). Consistent with *CCR6* expression, we found that NR4A$^+$ and LMNA$^+$ cells are significantly enriched in a PreM signature (Fig. 5E). In contrast to naïve subsets, CCR6$^+$ cells in the NR4A$^+$ subset showed higher expression of most of the genes described as a signature of PreM including *CELF2*, *BANK1* and *CD44* compared to CCR6$^-$ counterparts, while LMNA$^+$ CCR6$^+$ displayed a smaller subset of PreM signature genes (Fig. 5E). In addition, the NR4A subset contained high fractions of cells co-expressing *CCR6, BANK1* and *CELF2* (Fig. 5E).

A gene expression activation continuum from naïve to NR4A$^+$ B cell

The variability in *NR4A1, NR4A2, IGHD* and *CD27* expression in the NR4A subset (Fig. 2, B and D) led us to hypothesize that B cells in synovium might exist along a continuum from a naïve state. Projecting the gene expression data of the B cells onto two dimensions using principal component analysis, we sought to further define the relationship between B cell subsets. Plasma cells were excluded from this analysis. Dimension 1 (4% variance) separated NR4A cells from naïve. In contrast, dimension 2 explained the difference between the NR4A and memory subsets (2% variance) (fig. S6A). LMNA$^+$ cells were positive in both dimension 1 and 2. In dimension 1, we found that genes with large gene loadings are the ones involved in lymphocyte activation and GC molecular events including *CDKN1A*, *NFKBID*, *GPR183*, *CD83*, *LY9*, *KLF6* and *IGHG1*, and are expressed in increasing levels from naïve to NR4A$^+$ and LMNA$^+$ clusters. Consistently, the NR4A cluster is significantly enriched for gene sets associated with cell activation (fig. S6D). Additionally, a number of DNA-binding transcription factors (*NR4A2, NFKBID, NR4A3, JUND, DNAJA1, CREM, FOS*) gradually increase from naïve to NR4A cluster, while genes characteristic

of naïve B cells (*TCL1A, IGHM, FCER2*) are down-regulated (fig. S6B). We observed along dimension 1 cells that are in transition states between naïve and NR4A$^+$. Thus, expression of NR4A$^+$ markers including *NR4A1* and *DUSP1* are gradually increased, while concurrently markers associated with naïve status such as *TXNIP* and *CD79B* showed a gradual decrease in expression (fig. S6B). In dimension 2, genes with increased expression levels are those associated with memory B cells including *HOPX*, *IGHA1*, *S100A4* and *CD99*. We observed evidence of transitioning between NR4A$^+$ and memory state with increased expression of memory genes (*S100A6*, *HOPX*) and decreased expression of NR4A$^+$ genes (*JUN*, *DUSP1*). Our data are consistent with a continuum of states from naïve to NR4A$^+$ with both loss of naïve and acquisition of activation status. Whether LMNA$^+$ cells differentiate from memory, or NR4A$^+$ or both remains to be defined.

NR4A1 protein expression in SLO and synovial ELS

Endogenous NR4A1 (Nur77) expression has been shown to be an indicator of antigen receptor signaling in both mouse and human B cells (*54-56*) . In mice, the Nur77-GFP transgene is expressed by a small subset of LZ centrocytes revealing active BCR signaling in the GC (*55*), and supporting the role of NR4A1 in the GC reaction and adaptive immune response. To examine for endogenous NR4A1 expression in classic human SLO, we first examined human tonsil by immunofluorescent staining. We detected NR4A1$^+$ B cells in GCs in close proximity to CD21$^+$ LZ follicular dendritic cell networks (FDC) inside GC (Fig. 6A). Interestingly, we also observed NR4A1 expression in CD138$^+$ plasma cells surrounding GCs (Fig. 6B). Additional analysis by flow cytometry of tonsil B cells shows that multiple subsets (Bm1+Bm2, Bm5, GC and PC) of tonsil B cells express NR4A1, but with significantly more NR4A1$^+$ cells in the GC subset (Fig.

6C). Paralleling immunofluorescent staining, we identified NR4A1$^+$ PCs by flow cytometry (Fig. 6C). This is also consistent with our transcriptomic profiling data demonstrating *NR4A* gene expression in a population of PCs. Furthermore, when we subset GC B cells based on CD83 and CXCR4, significantly more NR4A1$^+$ cells were present in CD83$^+$CXCR4$^+$ intermediate GC B cells than in CD83$^-$CXCR4$^+$ (DZ), CD83$^+$CXCR4$^-$ (LZ) or CD83$^-$CXCR4$^-$ GC B cells, also consistent with our RNA sequencing data that synovial NR4A$^+$ B cells express high levels of both *CXCR4* and *CD83* (Fig. 6D).

To investigate endogenous NR4A1 protein expression in synovial B cells, we performed immunofluorescent staining on synovial tissue sections. Interestingly, the most intense NR4A1 expression in the synovium is in PCs (Fig. 7A), consistent with the tonsil staining and NR4A1 expression in a PC subpopulation in the single cell RNA sequencing data (Fig. 2B). We also observed NR4A1$^+$ B cells in aggregates (Fig. 7A), lacking Ki67 expression, in agreement with a GC intermediate zone. Flow cytometry of B cells from disaggregated RA tissue and synovial fluid also revealed high NR4A1 expression. Based on surface markers, 8.5 % of NR4A1$^+$ cells in synovium are IgD$^-$CD27^{++}, 42% are IgD$^-$CD27$^+$, 36% are IgD$^-$CD27$^-$, 6.3% are IgD$^+$CD27$^+$ and 7.2% are IgD$^+$CD27$^-$ (naïve) B cells (n=7, Fig. 7, B and C). In marked contrast, only a small percentage of peripheral blood B cells (from healthy controls or RA) spontaneously express NR4A1 (Fig. 7C).

When we stimulated peripheral blood and synovial B cells with a cocktail of anti-IgG, anti-IgA and anti-IgM to engage BCR of multiple isotypes, NR4A1 expression was significantly upregulated (Fig. 7D). The NR4A1 mean fluorescent intensity (MFI) was significantly increased in stimulated synovial B cells compared to untreated and stimulated peripheral blood healthy and

RA B cells (Fig. 7D). In addition to NR4A1 up-regulation, we also found that BCR stimulation up-regulated NR4A2 and NR4A3 expression at the mRNA level (fig. S7).

Lymphoid specific NR4A subset genes correlate with RA synovial tissue pathotype and increase in the blood during RA flare

Next, we sought to relate NR4A subset signature genes to synovial histology and clinical characteristics in an independent cohort of synovial biopsies from treatment-naive patients: The Pathobiology of Early Arthritis Cohort (PEAC) (*10*). Synovial biopsies were analyzed by immunohistochemistry and scored semi-quantitatively (0–4) for the presence of B cell aggregates ($CD20^+$), plasma cells ($CD138^+$), T cells ($CD3^+$), and monocytes or macrophages ($CD68^+$) in the synovial lining (CD68L) or sub-lining (CD68SL) layers. Based on histology scores, synovial samples were classified as lympho-myeloid (CD20 B cell aggregate rich), diffuse-myeloid (CD68 rich in the lining or sub-lining layer but poor in B cells), or fibroid (paucity of immune-inflammatory cell infiltrate), as previously described (*10*). We focused on the expression of *CD83* and *GPR183* as lymphoid specific genes representative of the NR4A cluster as RNA sequencing of total tissue is available in this study (Fig. 8, A and B). Both genes were strongly enriched in the lymphoid pathotype tissues and also showed significant correlation with synovial histology scores for CD20, CD3, CD138, and sub-lining CD68. These data suggest that infiltration of multiple immune cell types associated with ectopic lymphoid responses in the synovial tissue may be linked to the B cell NR4A subset. Additionally, *CD83* expression showed a relationship with ultrasonographic joint synovial thickness confirming that ELS gene expression strongly matches imaging signs of active joint inflammation in the particular joint undergoing biopsy. Surprisingly, *NR4A1* and *NR4A2* whole synovial gene expression was lower in the lymphoid vs fibroid

pathotype (p=0.00025 for *NR4A1* and p=0.22 for *NR4A2*) and inversely correlated with histologic immune cell infiltration (*NR4A1*: p=0.00051(CD3), p=1.7x10^{-5}(CD20), p=2.3x10^{-5}(CD138); *NR4A2*: p=7.5x10^{-5}(CD3), p=2.6x10^{-5}(CD20), p=5.2x10^{-5}(CD138)) consistent with B cell independent roles for the NR4A family in the synovium in other cell populations that may be anti-inflammatory (*57*).

Although our initial data suggested that NR4A B cells are rare in RA blood (**Fig. 7C**), we wondered whether circulating numbers of this activated B cell population may vary over time depending on disease activity. To begin to understand this dynamic, we compared our synovial B cell cluster signatures with blood transcriptional profiles recently defined relative to flare (*21*). In this data set two transcriptomic clusters appeared in the blood preceding flare: AC2 was increased 2 weeks before flare and AC3 was increased 1 week before flare. AC2 has characteristics of naïve B cells, while AC3 is a unique CD45−CD31−PDPN+ mesenchymal population, termed PRIME (preinflammatory mesenchymal) cells. An additional cluster 1 was increased at the onset of flare. When we examined the overlap between these blood clusters and our single cell B cell clusters (Fig. 8C, and table S2) using hypergeometric tests for enrichment, we found that AC2 was significantly enriched with transcripts characteristic of peripheral blood dominant naïve B cell clusters, though the synovial NR4A and LMNA clusters were represented (Fig. 8C). In contrast, the flare cluster 1 was significantly enriched with naïve (iii), NR4A, and LMNA synovial populations (7%, 6.6%, and 5.6% overlapping markers with the NR4A, LMNA and Naive (iii) clusters, respectively). This is consistent with a trajectory of B cell activation whereby resting blood B cells are recruited from the blood to the synovium pre-flare and then activated within the synovial microenvironment to express NR4A during flare.

DISCUSSION

Our integrated single-cell analysis of RA synovial B cells has revealed a surprising level of B cell activation, with a large fraction of synovial B cells demonstrating evidence of recent antigen stimulation as revealed by upregulation of the NR4A gene family. NR4A B cells appear primed for cell-cell interactions with APCs and T cells and, further, have evidence of molecular events associated with active GC reactions, including class-switch recombination, somatic hypermutation, and clonal expansion with selection into the PC compartment. Our findings suggest a spectrum of activation of B cells within the synovial microenvironment with a pluripotent fate that includes differentiation to memory or PC lineage *in situ*.

Prior studies have supported local synovial activation of RA B cells with evidence of SHM and clonal expansion (*8, 18*) using a number of approaches including enzymatic digestion or focal microdissection with bulk or single cell BCR sequencing. An important advance in the current study is the high-throughput combination of single cell RNA sequencing and Ig gene repertoire. This allowed the unprecedented identification of a spectrum of B cell activation in synovial B cells downstream of BCR signaling as revealed by *NR4A1* expression. A key finding from our study was the relationship between the degree of NR4A family gene expression and SHM. Notably, this dynamic process was unique to the synovium and not observed in B cells from RA peripheral blood or lupus kidney. NR4A1-3 encode a small family of orphan nuclear hormone receptors (NUR77, NURR1, and NOR1 respectively) which are rapidly induced by acute and chronic antigen stimulation in B cells and T cells (*56, 58*) . Our results raise key questions regarding the precise synovial microenvironmental signals that promote NR4A family expression and 2nd signals that may direct subsequent differentiation to memory B or plasma cell. Another study did combine transcriptomic and repertoire analysis of peripheral blood RA B cells and interestingly suggested

that ACPA enriched B cell responses are imprinted with a T cell dependent transcriptional network (*7*). Given the strong association of NR4A1-NUR77 expression in B cells and BCR engagement (*54*), we speculate that the enrichment of B cell NR4A1 expression in the RA synovium represents chronic auto-antigen stimulation and an ACPA response. This would be in keeping with other data in the literature indicating that B cells differentiating within ELS frequently target citrullinated peptides (*18*). However, the antigen specificity of the NR4A B cells in our study remains to be determined.

In terms of B cell fate, it is notable that NR4A family members may trigger B cell apoptosis after encounter with self-antigen in the absence of a 2nd signal provided by T cell (or other) co-stimulation and act as a negative regulator of B cell activation (*59*). Thus, in murine B cells, the NR4A family, particularly NR4A1/NUR77, actively participates in B cell tolerance. NUR77 is up-regulated in self-reactive B cells in response to chronic antigen stimulation and selectively restricts survival by limiting B cell survival factors such as BAFF (*60*). In addition, a small subset of GC LZ B cells exhibiting a high SHM rate, up-regulate NR4A1/NUR77, suggesting active BCR signaling and a role for NR4A1 in selection and regulation of GC differentiation pathways into the PC compartment (*55*). Recently, NR4A1 and NR4A3 have been identified as partially redundant mediators of an antigen-dependent negative feedback loop in B cells to reinforce tolerance by increasing B cell dependence on T cell help and regulating clonal competition, when T cell help is limited (*59*). These data point to an expanding role for the NR4A family in adaptive immune response occurring in SLO. However, the role of NR4A in ELS and autoimmunity had not been studied previously. Whether NR4A attempts to restrain B cell activation within the synovium is not clear. Of interest, an anti-inflammatory synovial tissue resident macrophage population was recently reported as expanded in the tissue in disease remission and characterized by the expression

of NR4A1 and NR4A2 (*57*). This is also in accord with the negative correlation of NR4A2 expression in RA synovium and disease activity in the independent PEAC cohort (*10*).

Our finding of shared clonality between the NR4A B cell and PC clusters suggests that some NR4A B cells do indeed receive the signals required to drive their differentiation to memory and PC *in situ*. We identified several genes within the NR4A cluster that may facilitate T-B cell interactions including CD69 (which prevents activated lymphocyte egress from SLO), CD86 (a ligand for CD28 on T cells), IL6 (which can promote both APC and T cell interactions), and ICAM1 (which can support B cell interaction with T follicular helper cells) (*61*). This is also in accord with a recent report showing that tetramer identified ACPA positive B cells in RA blood and synovial fluid express T cell-stimulating ligands and a persistently activated phenotype suggestive of continuous antigenic triggering (*62*). A key role for B cells in RA flare was also suggested by the recent discovery of a circulating expanded B cell subset (AC2) prior to emergence of a circulating CD45-/CD31-/PDPN+, PRe-Inflammatory MEsenchymal ("PRIME") cell in RA patient blood (*21*). We find that this pre-PRIME circulating B cell cluster (AC2) was enriched with resting naïve (naïve (i), naïve (ii)) and memory B cells that were dominant in the blood sample we included in our single cell analysis. In contrast, the blood transcriptomic cluster appearing during flare (cluster 1) was enriched in the NR4A, LMNA, and naïve (iii) clusters. This is consistent with a trajectory of B cell activation whereby blood B cells are recruited to the synovium pre-flare, activated within the synovial microenvironment to express NR4A, and then possibly re-circulate during flare. Re-circulation of B cells between the synovium and the periphery is also supported by a recent publication demonstrating shared clonal families between these compartments in RA (*63*). The origin of the B cells increasing in the blood pre-flare remains to be determined.

An interesting feature of the NR4A B cell cluster identified here is its unique chemokine receptor expression pattern (CXCR4, CXCR5, CCR7, and CCR6), which we speculate promotes both tissue localization and ELS formation. Of note, the NR4A cluster was the only one to express CCR6 at a significant level, with evidence of a range of expression. Recent studies have found evidence for distinct precursors of memory B cells and plasma cells arising within the human tonsil GC (*40*) with CCR6 a marker of memory B cell precursors arising within the GC LZ in SLO (*53*). Though GC B cells are classically categorized into two subsets by location and phenotype with SHM occurring in the DZ and selection in the LZ facilitated by contacts with FDCs and T follicular helper cells, recent single cell transcriptomic studies of secondary lymphoid tissue have revealed that B cells display a spectrum of activation states (including activated and pre-GC) (*36*) and also dynamically transit through multiple states in the GC (*40*). Interestingly, this includes intermediate states with both DZ and LZ marker expression that may dominate GC B cells (*39*), similar to the NR4A synovial B cell cluster, identified in our study, which expresses both classic GC DZ (*BCL6*, *CXCR4*) and LZ (*CD83*, *GP183*) genes. To the best of our knowledge, this is the first demonstration of multiple activation states and memory B cell precursors arising in ELS. The signals that control differentiation of activated B cells into memory or plasma cells in ELS require further study, but similar to SLO may depend on BCR affinity and the strength and stability of their interaction with T cells (*36, 53, 64*).

B cells may contribute to RA pathogenesis in multiple ways, not only as the precursors of PCs and production of autoantibodies, but also by antibody independent functions including antigen presentation and cytokine production. B cells can produce cytokines and chemokines that promote ELS. Thus, lymphotoxin β expression by B cells has been shown to support ELS formation in a mouse colitis model (*19*). Lymphotoxin β and α are elevated in RA synovium and

relate to the level of inflammation in the tissue (*65*), and LT-β was expressed by synovial B cells but also other cell populations (*12*). The precise role of B cells in ELS formation in RA synovial tissue and which sub-populations may be critical is unclear. We found a significant relationship between NR4A B cell expression and the degree of B cell infiltration and lymphoid organization in tissue implicating NR4A B cells as a potential ELS initiating and promoting B cell subset. In accord with this, NR4A B cells expressed LT-β and were the dominant source of IL6, the latter important for Tfh interactions and demonstrated to promote spontaneous GC formation in SLE (*50*) .

Overall, our results highlight the unique enrichment of NR4A expressing B cells in the RA synovium and their identity as an antigen experienced B cell in the synovial ELS with the capability of differentiating to plasma cell or memory B cell in situ depending on additional signals. NR4A may serve as a biomarker of chronic autoantigen stimulation in the synovial microenvironment and a novel therapeutic target.

MATERIALS AND METHODS

Patient samples

Arthroplasty samples were acquired from RA patients after surgical removal as part of standard of care at Hospital of Special Surgery, NY (n=3) or collected as part of the AMP network (ImmPort SDY998) (n=1). Peripheral blood was collected prior to the procedure. Written informed consent for research use of patient samples was obtained prior to study inclusion at the time of sample collection. The study received institutional review board approval at each site. Synovial tissue fragments were cryopreserved in CryoStor CS10 (BioLife Solution) for subsequent disaggregation at University of Rochester (*66*). PBMC were isolated from whole blood by Ficoll-hypaque density gradient centrifugation and cryopreserved in CryoStor CS10 (BioLife Solution). Samples from the AMP network were collected as part of multi-center cross-sectional study of patients undergoing elective surgical procedures and a prospective observational study of synovial biopsy specimen from patients with RA aged ≥ 18 years, with at least one inflamed joint (*20*). This cohort included synovial tissues from 16 RA biopsies, 13 RA arthroplasties and 10 OA arthroplasties. 14 RA biopsies have B cell single cell RNA sequencing data available for analysis. The Pathobiology of Early Arthritis cohort (PEAC) contains data from 90 RA patients fulfilling 2010 ACR/EULAR RA Classification Criteria that were enrolled as part of the Medical Research Council funded multi-center PEAC project (*10*) (https://peac.hpc.qmul.ac.uk/). Normal control and RA peripheral blood were obtained from patients at University of Rochester under institutional review board approved study protocols. PBMCs were isolated by Ficoll-hypaque density gradient centrifugation and cryopreserved in freezing medium (90%FBS/10%DMSO) until used. Tonsil tissue samples were obtained as discarded tissue from tonsillectomy procedures. Portions of tissue was fixed in formalin and paraffin-embedded for histologic staining. The remaining tissue was

mechanically disaggregated to obtain cell suspension. The cell suspension was further purified by Ficoll-hypaque density gradient centrifugation. Isolated mononuclear tonsil cells were cryopreserved in freezing medium (90%FBS/10%DMSO) until used.

Preparation of synovial tissue and PBMC

Cryopreserved synovial fragments were thawed per SOP (*66*). Synovial tissue fragments were then disaggregated by combination of an enzymatic digestion and mechanical disruption as described. PBMCs were thawed by rapidly warming the cryovial in a 37 °C water bath. Thawed cells were washed in 10% FBS/RPMI, resuspended in fresh 10%FBS/RPMI and kept on ice for downstream analysis.

Flow cytometry cell sorting

To stain cells for sorting, cells were first incubated with Fc-Block (Biolegend) for 5 minutes at room temperature. Cells were then incubated with a mixture of fluorochrome-conjugated antibodies (Pdpn, CD90, CD45, CD14, IgD, CD19, CD27, CD3, CD4, CD11c, and CD24) (table S3) for 30 min on ice. Cells were washed with PBS and then resuspended in PBS solution containing DAPI (Biolegend). Cells were then sorted on a FACSAria II (BD biosciences) into RPMI media supplemented with 5% FBS. Intact cells were gated based on FSC-A and SSC-A prior to doublet exclusion by serial FSC-H/FSC-W and SSC-H/SS-W gates. Nonviable cells were excluded based on DAPI uptake. B cells were gated as $CD45^{+}CD3^{-}CD14^{-}CD19^{+}$ and sorted through a 100-µm nozzle (fig. S1A).

Single cell RNA and BCR repertoire sequencing

Cellular suspensions were loaded on a Chromium Single-Cell Instrument (10x Genomics, Pleasanton, CA, USA) to generate single-cell Gel Bead-in-Emulsions (GEMs). Single-cell RNA-Seq libraries were prepared using Chromium Single-Cell 5′ Library & Gel Bead Kit (PN-1000006, PN-1000014, 10x Genomics). The beads were dissolved and cells were lysed per manufacturer's recommendations. GEM reverse transcription (GEM-RT) was performed to produce a barcoded, full-length cDNA from poly-adenylated mRNA. After incubation, GEMs were broken and the pooled post-GEM-RT reaction mixtures were recovered and cDNA was purified with silane magnetic beads (DynaBeads MyOne Silane Beads, PN37002D, ThermoFisher Scientific). The entire purified post GEM-RT product was amplified by PCR. This amplification reaction generated sufficient material to construct a B cell-enriched library and a 5' gene expression library. The barcoded, full-length V(D)J segments were enriched by PCR with primers specific to the BCR constant regions using the Chromium Single Cell V(D)J Enrichment Kit, Human B cell (PN-1000016, 10x Genomics). For both B-cell enriched library and 5' gene expression library, enzymatic fragmentation and size selection was used to optimize the cDNA amplicon size and indexed sequencing libraries were constructed by End Repair, A-tailing, Adaptor Ligation, and PCR. Final libraries contain the P5 and P7 priming sites used in Illumina bridge amplification. Paired end reads of 150nt were generated for each B-cell enriched library using the MiSeq sequencer (Illumina). The 5' gene expression libraries were sequenced following 10X Genomics read length guidelines on the NovaSeq6000 sequencer (Illumina).

Single cell RNA and repertoire sequencing analysis

Raw reads generated from the Illumina basecalls were demultiplexed using Cell Ranger mkfastq version 2.1.1 utilizing bcl2fastq version 2.19.1.403. Cells were aligned and counted against the

Cell Ranger reference refdata-cellranger-GRCh38-1.2.0 using Cell Ranger count version 2.1.1, utilizing only the library index, project, and sample ID for a given sample. We performed quality control to remove outlier events and doublets based on the cell's total UMI size versus the number of distinct genes expressed in SeqGeq™ v1.4 (FlowJo, LLC) by gating on bivariate scatter plots. We also filtered putatively degraded cells based on the proportion of UMIs in the cell derived from the mitochondrial genome. These steps removed 252 cells. After filtering, the data were exported for further analysis in scran 1.8.4 and Seurat 2.3.4 (*23*). The top 1000 variable genes, excluding IGHV, IGKV and IGLV genes were used to derive 6 principal components. These were used to define clusters using the Louvain method and resolution parameter 0.5. Based on expression of non-B cell markers (*COL1A2*$^{+}$, *FN1*$^{+}$, *MS4A1*$^{-}$), we identified 77 putative fibroblasts, and accordingly excluded these cells from further analyses. BCR clonotypes were assigned by requiring 97% identity of the DNA sequence of the CDR3 region of the heavy or light chain using CellaRepertorium version 0.8.1. (http://bioconductor.org/packages/release/bioc/html/CellaRepertorium.html). For somatic hypermutation (SHM) rate calculation, the consensus FASTQ sequences were aligned to the human immunoglobulin reference from the international ImMunoGeneTics information system (IMGT) using HighVQuest to calculate the percent identity in V and J in heavy and light chains which then averaged, weighting by the length of each segment.

Projection of clustering using SingleR

Gene-cell count matrices were downloaded for the AMP Phase I RA and SLE cohorts (ImmPort accessions SDY998 and SDY997). Filtered gene-cell count matrices for CD19^{+} B cells from Zheng et al 2017 (*32*) were downloaded from https://support.10xgenomics.com/single-cell-

gene-expression/datasets/1.1.0/b_cells. 12848 gene symbols expressed in common in these datasets and our data were retained and the data merged. QC was evaluated using the quickPerCellQC function in scran and 874 were cells filtered. A SingleR model was trained on our data and its 8 cluster labels, and the inferred cluster identities in the other datasets were queried. Binomial mixed models were used to estimate the frequency of the $NR4A^+$ cluster among these other data sets. For bulk RNAseq from AMP RA Cohort, data was downloaded (ImmPort accession SDY1299). DESeq2 version 1.20.0 was used to evaluate differential expression of $NR4A^+$ marker genes in sorted B cells from this study.

Gene set enrichment analysis

For GSEA analysis, we used gene set from Blood transcription modules from Li et al (*37*), bulk RNA sequencing of sorted tonsil B cells (Bm1+Bm2 and Bm5) and DZ and LZ GC gene lists from Victora et al. (*38*). We used clusterProfiler 3.8.1 to conduct GSEA analysis. Heatmaps displaying the signed -log10(FDR q value) of the result was created using GraphPad Prism (GraphPad Software, LLC).

In vitro stimulation

Freshly isolated or thawed PBMC were washed with complete medium and counted. For qPCR, B cells were positively selected using anti-CD19 magnetic beads (Miltenyi Biotech). 0.5-1.0x10^6 cells per well were seeded in round-bottom 96-well plates in 200µl 10%FBS/RPMI 1640. Cells were stimulated with anti- human IgM F(ab′2) or anti- human Ig(A+G+M) F(ab′2) (Jackson Immuno-Research Laboratories) for indicated times. After stimulation, cells were collected for flow cytometry or qPCR.

Quantitative real-time polymerase chain reaction (qPCR).

B cells were collected in RLTplus lysis buffer (Qiagen) supplemented with 1% 2-mercaptoethanol, and RNA were extracted using *Quick*-RNA kit (Zymo Research) and concentration was measured by NanoDrop (Thermo Scientific). Complementary DNA (cDNA) synthesis was done using a qScript cDNA synthesis kit (Quantabio). TaqMan expression assay system for NR4A1 (Hs00374226_m1), NR4A2 (Hs01117527_g1), NR4A3 (Hs00545009_g1) from Applied Biosystems were used to detect NR4A1, NR4A2 and NR4A3 mRNA expression, respectively. PPIA (Hs04194521_s1, Applied Biosystems) was used as a control housekeeping gene. The fold change in mRNA expression relative to control was calculated using the comparative threshold cycle ($\Delta\Delta C_t$) method.

Intracellular staining and Flow cytometry

To stain for flow cytometry, PBMC or tonsil cells were first incubated with Fc-Block (Biolegend) for 5 minutes at room temperature. Then 100ul of surface antibody cocktail (CD19, CD3, CD27, IgD, CD38, IgG, CD83, CD77, GPR183, CD69 and CXCR4, table S4) was added to each tube and incubated on ice for 30 minutes. Next, live/dead cell assay (Invitrogen) was applied for 15 minutes on ice. Intracellular staining for NR4A1, BCL-6, Ki67 and IgG was performed using Foxp3/Transcription factor staining buffer kit (eBioscience) following manufacture protocol. After intracellular staining, cells were washed and fixed in 100µl PBS/1% paraformaldehyde for 20 minutes at room temperature and then 200µl of PBS/5%BSA was added to dilute the fixative. Samples were analyzed on an 18-color LSR II flow cytometer (eBioscience) either on the same day or after kept at 4°C overnight. Analyses were performed using FlowJo v.10 (Tree Star) and doublet exclusion was performed on all samples. Gating strategy is depicted in fig. S8.

Immunofluorescent staining

List of primary and secondary antibodies used for immunofluorescent staining can be found in table S5. 5 μm formalin-fixed paraffin sections were incubated at 60°C overnight (ON). Tissue sections were quickly transferred to xylenes and gradually hydrated by transferring slides to absolute, 96%, 70% alcohol, and water. Slides were immersed in an antigen retrieval solution, boiled for 30 minutes, and cooled down for 10 minutes at room temperature (RT). Slides were rinsed several times in water and transferred to PBS. Non-specific binding was blocked with 5% normal donkey serum in PBS containing 0.1% Tween 20, 0.1% Triton-X-100 for 30 minutes, at RT in a humid chamber. Primary antibodies were added to slides and incubated in a humid chamber at RT, ON. Slides were quickly washed in PBS, and fluorescently labeled, secondary antibodies were incubated for 2 hours at RT in a humid chamber. Finally, slides were rinsed for 1 hour in PBS and mounted with Vectashield antifade mounting media with DAPI (H-1200, Vector Laboratories). 200x and 200x 3x3 mosaic pictures were taken with a Zeiss Axioplan 2 microscope and recorded with a Hamamatsu camera.

Statistical analysis

The statistical tests used are as described in each figure legend. P-values ≤ 0.05 are considered significant.

AUTHOR CONTRIBUTIONS

JHA, AM and NM conceived and designed the work. ED, DEO, SG and LTD were responsible for tissue sample acquisition and sample characterization. NM performed the sample processing, sorting, single-cell capture and flow cytometry experiments. JRM performed immunofluorescent

staining of tissues. KER performed in vitro stimulation of B cells. JHA, AM and NM established the analytical strategies and analyzed the data. AM performed bioinformatic analyses. FZ, SR and the AMP network provided bioinformatic assistance. EC, EP, MB and CP provided tissue samples from the PEAC cohort. DEO provided analytic input and assistance with analysis of the RA flare cohort. JHA supervised the study. JHA, AM and NM wrote the manuscript. All authors contributed to reviewing and editing of the manuscript.

ACKNOWLEDGEMENTS

Funding: This work was funded by R21 AR071670 to JHA, RUCCTS Grant #UL1 TR001866, the Robertson Foundation, and the Bernard and Irene Schwartz Foundation to DEO. JHA is also supported by the Bertha and Louis Weinstein research fund and the Accelerating Medicines Partnership (AMP) in RA and SLE Network. AMP is a public-private partnership (AbbVie Inc., Arthritis Foundation, Bristol-Myers Squibb Company, GlaxoSmithKline LLC, Janssen Research & Development LLC, Lupus Foundation of America, Lupus Research Alliance, Merck Sharp & Dohme Corp., National Institute of Allergy and Infectious Diseases, National Institute of Arthritis and Musculoskeletal and Skin Diseases, Pfizer Inc., Rheumatology Research Foundation, and Sanofi and Takeda Pharmaceuticals International, Inc.) created to develop new ways of identifying and validating promising biological targets for diagnostics and drug development. Funding 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). See Supplemental Acknowledgements for network details. We acknowledge the expertise and support of the University of Rochester Center for Musculoskeletal Research Histology Biochemistry and

Molecular Imaging Core NIAMS AR069655 and the University of Rochester Resource Cores in Flow Cytometry and the Genomic Research Center.

REFERENCE

1. E. Kaltsonoudis, E. Pelechas, P. V. Voulgari, A. A. Drosos, Unmet needs in the treatment of rheumatoid arthritis. An observational study and a real-life experience from a single university center. *Semin Arthritis Rheum* **48**, 597-602 (2019).
2. A. A. Drosos, E. Pelechas, P. V. Voulgari, Rheumatoid Arthritis Treatment. A Back to the Drawing Board Project or High Expectations for Low Unmet Needs? *J Clin Med* **8**, 1237-1242 (2019).
3. N. Meednu, H. Zhang, T. Owen, W. Sun, V. Wang, C. Cistrone, J. Rangel-Moreno, L. Xing, J. H. Anolik, Production of RANKL by Memory B Cells: A Link Between B Cells and Bone Erosion in Rheumatoid Arthritis. *Arthritis Rheumatol* **68**, 805-816 (2016).
4. W. Sun, N. Meednu, A. Rosenberg, J. Rangel-Moreno, V. Wang, J. Glanzman, T. Owen, X. Zhou, H. Zhang, B. F. Boyce, J. H. Anolik, L. Xing, B cells inhibit bone formation in rheumatoid arthritis by suppressing osteoblast differentiation. *Nat Commun* **9**, 5127-5140 (2018).
5. U. Harre, D. Georgess, H. Bang, A. Bozec, R. Axmann, E. Ossipova, P. J. Jakobsson, W. Baum, F. Nimmerjahn, E. Szarka, G. Sarmay, G. Krumbholz, E. Neumann, R. Toes, H. U. Scherer, A. I. Catrina, L. Klareskog, P. Jurdic, G. Schett, Induction of osteoclastogenesis and bone loss by human autoantibodies against citrullinated vimentin. *J Clin Invest* **122**, 1791-1802 (2012).
6. V. Derksen, T. W. J. Huizinga, D. van der Woude, The role of autoantibodies in the pathophysiology of rheumatoid arthritis. *Semin Immunopathol* **39**, 437-446 (2017).
7. D. R. Lu, A. N. McDavid, S. Kongpachith, N. Lingampalli, J. Glanville, C. H. Ju, R. Gottardo, W. H. Robinson, T Cell-Dependent Affinity Maturation and Innate Immune Pathways Differentially Drive Autoreactive B Cell Responses in Rheumatoid Arthritis. *Arthritis Rheumatol* **70**, 1732-1744 (2018).
8. T. Scheel, A. Gursche, J. Zacher, T. Haupl, C. Berek, 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 Rheum* **63**, 63-72 (2011).
9. P. P. Tak, M. E. Doorenspleet, M. J. H. de Hair, P. L. Klarenbeek, M. H. van Beers-Tas, A. H. C. van Kampen, D. van Schaardenburg, D. M. Gerlag, F. Baas, N. de Vries, Dominant B cell receptor clones in peripheral blood predict onset of arthritis in individuals at risk for rheumatoid arthritis. *Ann Rheum Dis* **76**, 1924-1930 (2017).
10. M. J. Lewis, M. R. Barnes, K. Blighe, K. Goldmann, S. Rana, J. A. Hackney, N. Ramamoorthi, C. R. John, D. S. Watson, S. K. Kummerfeld, R. Hands, S. Riahi, V. Rocher-Ros, F. Rivellese, F. Humby, S. Kelly, M. Bombardieri, N. Ng, M. DiCicco, D. van der Heijde, R. Landewe, A. van der Helm-van Mil, A. Cauli, I. B. McInnes, C. D. Buckley, E. Choy, P. C. Taylor, M. J. Townsend, C. Pitzalis, Molecular Portraits of Early Rheumatoid Arthritis Identify Clinical and Treatment Response Phenotypes. *Cell Rep* **28**, 2455-2470 e2455 (2019).
11. F. Humby, M. Lewis, N. Ramamoorthi, J. A. Hackney, M. R. Barnes, M. Bombardieri, A. F. Setiadi, S. Kelly, F. Bene, M. DiCicco, S. Riahi, V. Rocher, N. Ng, I. Lazarou, R. Hands, D. van der Heijde, R. B. M. Landewe, A. van der Helm-van Mil, A. Cauli, I. McInnes, C. D. Buckley, E. H. Choy, P. C. Taylor, M. J. Townsend, C. Pitzalis, Synovial cellular and

molecular signatures stratify clinical response to csDMARD therapy and predict radiographic progression in early rheumatoid arthritis patients. *Ann Rheum Dis* **78**, 761-772 (2019).

12. S. Takemura, A. Braun, C. Crowson, P. J. Kurtin, R. H. Cofield, W. M. O'Fallon, J. J. Goronzy, C. M. Weyand, Lymphoid neogenesis in rheumatoid synovitis. *J Immunol* **167**, 1072-1080 (2001).
13. G. Yanni, A. Whelan, C. Feighery, W. Quinlan, J. Symons, G. Duff, B. Bresnihan, Contrasting levels of in vitro cytokine production by rheumatoid synovial tissues demonstrating different patterns of mononuclear cell infiltration. *Clin Exp Immunol* **93**, 387-395 (1993).
14. M. Rooney, D. Condell, W. Quinlan, L. Daly, A. Whelan, C. Feighery, B. Bresnihan, Analysis of the histologic variation of synovitis in rheumatoid arthritis. *Arthritis Rheum* **31**, 956-963 (1988).
15. A. Manzo, M. Bombardieri, F. Humby, C. Pitzalis, Secondary and ectopic lymphoid tissue responses in rheumatoid arthritis: from inflammation to autoimmunity and tissue damage/remodeling. *Immunological reviews* **233**, 267-285 (2010).
16. A. Nerviani, C. Pitzalis, Role of chemokines in ectopic lymphoid structures formation in autoimmunity and cancer. *J Leukoc Biol* **104**, 333-341 (2018).
17. F. Humby, M. Bombardieri, A. Manzo, S. Kelly, M. C. Blades, B. Kirkham, J. Spencer, C. Pitzalis, Ectopic lymphoid structures support ongoing production of class-switched autoantibodies in rheumatoid synovium. *PLoS Med* **6**, e1 (2009).
18. E. Corsiero, M. Bombardieri, E. Carlotti, F. Pratesi, W. Robinson, P. Migliorini, C. Pitzalis, Single cell cloning and recombinant monoclonal antibodies generation from RA synovial B cells reveal frequent targeting of citrullinated histones of NETs. *Ann Rheum Dis* **75**, 1866-1875 (2016).
19. M. Lochner, C. Ohnmacht, L. Presley, P. Bruhns, M. Si-Tahar, S. Sawa, G. Eberl, Microbiota-induced tertiary lymphoid tissues aggravate inflammatory disease in the absence of RORgamma t and LTi cells. *J Exp Med* **208**, 125-134 (2011).
20. F. Zhang, K. Wei, K. Slowikowski, C. Y. Fonseka, D. A. Rao, S. Kelly, S. M. Goodman, D. Tabechian, L. B. Hughes, K. Salomon-Escoto, G. F. M. Watts, A. H. Jonsson, J. Rangel-Moreno, N. Meednu, C. Rozo, W. Apruzzese, T. M. Eisenhaure, D. J. Lieb, D. L. Boyle, A. M. Mandelin, 2nd, A. Accelerating Medicines Partnership Rheumatoid, C. Systemic Lupus Erythematosus, B. F. Boyce, E. DiCarlo, E. M. Gravallese, P. K. Gregersen, L. Moreland, G. S. Firestein, N. Hacohen, C. Nusbaum, J. A. Lederer, H. Perlman, C. Pitzalis, A. Filer, V. M. Holers, V. P. Bykerk, L. T. Donlin, J. H. Anolik, M. B. Brenner, S. Raychaudhuri, Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat Immunol* **20**, 928-942 (2019).
21. D. E. Orange, V. Yao, K. Sawicka, J. Fak, M. O. Frank, S. Parveen, N. E. Blachere, C. Hale, F. Zhang, S. Raychaudhuri, O. G. Troyanskaya, R. B. Darnell, RNA Identification of PRIME Cells Predicting Rheumatoid Arthritis Flares. *N Engl J Med* **383**, 218-228 (2020).
22. L. J. P. H. van der Maaten, G.E., Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

23. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* **36**, 411-420 (2018).
24. U. Klein, K. Rajewsky, R. Kuppers, Human immunoglobulin (Ig)M+IgD+ peripheral blood B cells expressing the CD27 cell surface antigen carry somatically mutated variable region genes: CD27 as a general marker for somatically mutated (memory) B cells. *J Exp Med* **188**, 1679-1689 (1998).
25. R. Riedel, R. Addo, M. Ferreira-Gomes, G. A. Heinz, F. Heinrich, J. Kummer, V. Greiff, D. Schulz, C. Klaeden, R. Cornelis, U. Menzel, S. Kroger, U. Stervbo, R. Kohler, C. Haftmann, S. Kuhnel, K. Lehmann, P. Maschmeyer, M. McGrath, S. Naundorf, S. Hahne, O. Sercan-Alp, F. Siracusa, J. Stefanowski, M. Weber, K. Westendorf, J. Zimmermann, A. E. Hauser, S. T. Reddy, P. Durek, H. D. Chang, M. F. Mashreghi, A. Radbruch, Discrete populations of isotype-switched memory B lymphocytes are maintained in murine spleen and bone marrow. *Nat Commun* **11**, 2570-2583 (2020).
26. D. Bautista, C. Vasquez, P. Ayala-Ramirez, J. Tellez-Sosa, E. Godoy-Lozano, J. Martinez-Barnetche, M. Franco, J. Angel, Differential Expression of IgM and IgD Discriminates Two Subpopulations of Human Circulating IgM(+)IgD(+)CD27(+) B Cells That Differ Phenotypically, Functionally, and Genetically. *Front Immunol* **11**, 736-754 (2020).
27. M. Descatoire, S. Weller, S. Irtan, S. Sarnacki, J. Feuillard, S. Storck, A. Guiochon-Mantel, J. Bouligand, A. Morali, J. Cohen, E. Jacquemin, M. Iascone, C. Bole-Feysot, N. Cagnard, J. C. Weill, C. A. Reynaud, Identification of a human splenic marginal zone B cell precursor with NOTCH2-dependent differentiation properties. *J Exp Med* **211**, 987-1000 (2014).
28. D. Gatto, K. Wood, R. Brink, EBI2 operates independently of but in cooperation with CXCR5 and CCR7 to direct B cell migration and organization in follicles and the germinal center. *J Immunol* **187**, 4621-4628 (2011).
29. J. P. Pereira, L. M. Kelly, Y. Xu, J. G. Cyster, EBI2 mediates B cell segregation between the outer and centre follicle. *Nature* **460**, 1122-1126 (2009).
30. D. Aran, A. P. Looney, L. Liu, E. Wu, V. Fong, A. Hsu, S. Chak, R. P. Naikawadi, P. J. Wolters, A. R. Abate, A. J. Butte, M. Bhattacharya, Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172 (2019).
31. A. Arazi, D. A. Rao, C. C. Berthier, A. Davidson, Y. Liu, P. J. Hoover, A. Chicoine, T. M. Eisenhaure, A. H. Jonsson, S. Li, D. J. Lieb, F. Zhang, K. Slowikowski, E. P. Browne, A. Noma, D. Sutherby, S. Steelman, D. E. Smilek, P. Tosta, W. Apruzzese, E. Massarotti, M. Dall'Era, M. Park, D. L. Kamen, R. A. Furie, F. Payan-Schober, W. F. Pendergraft, 3rd, E. A. McInnis, J. P. Buyon, M. A. Petri, C. Putterman, K. C. Kalunian, E. S. Woodle, J. A. Lederer, D. A. Hildeman, C. Nusbaum, S. Raychaudhuri, M. Kretzler, J. H. Anolik, M. B. Brenner, D. Wofsy, N. Hacohen, B. Diamond, S. L. E. n. Accelerating Medicines Partnership in, The immune cell landscape in kidneys of patients with lupus nephritis. *Nat Immunol* **20**, 902-914 (2019).
32. G. X. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R.

Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049-14060 (2017).

33. K. L. Good-Jacobson, M. J. Shlomchik, Plasticity and heterogeneity in the generation of memory B cells and long-lived plasma cells: the influence of germinal center interactions and dynamics. *J Immunol* **185**, 3117-3125 (2010).
34. U. Klein, R. Kuppers, K. Rajewsky, Evidence for a large compartment of IgM-expressing memory B cells in humans. *Blood* **89**, 1288-1298 (1997).
35. S. G. Tangye, K. L. Good, Human IgM+CD27+ B cells: memory B cells or "memory" B cells? *J Immunol* **179**, 13-19 (2007).
36. H. W. King, N. Orban, J. C. Riches, A. J. Clear, G. Warnes, S. A. Teichmann, L. K. James, Single-cell analysis of human B cell maturation predicts how antibody class switching shapes selection dynamics. *Sci Immunol* **6**, eabe6291 (2021).
37. S. Li, N. Rouphael, S. Duraisingham, S. Romero-Steiner, S. Presnell, C. Davis, D. S. Schmidt, S. E. Johnson, A. Milton, G. Rajam, S. Kasturi, G. M. Carlone, C. Quinn, D. Chaussabel, A. K. Palucka, M. J. Mulligan, R. Ahmed, D. S. Stephens, H. I. Nakaya, B. Pulendran, Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. *Nat Immunol* **15**, 195-204 (2014).
38. G. D. Victora, D. Dominguez-Sola, A. B. Holmes, S. Deroubaix, R. Dalla-Favera, M. C. Nussenzweig, Identification of human germinal center light and dark zone cells and their relationship to human B-cell lymphomas. *Blood* **120**, 2240-2248 (2012).
39. P. Milpied, I. Cervera-Marzal, M. L. Mollichella, B. Tesson, G. Brisou, A. Traverse-Glehen, G. Salles, L. Spinelli, B. Nadel, Human germinal center transcriptional programs are de-synchronized in B cell lymphoma. *Nat Immunol* **19**, 1013-1024 (2018).
40. A. B. Holmes, C. Corinaldesi, Q. Shen, R. Kumar, N. Compagno, Z. Wang, M. Nitzan, E. Grunstein, L. Pasqualucci, R. Dalla-Favera, K. Basso, Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J Exp Med* **217**, e20200483 (2020).
41. G. Caron, S. Le Gallou, T. Lamy, K. Tarte, T. Fest, CXCR4 expression functionally discriminates centroblasts versus centrocytes within human germinal center B cells. *J Immunol* **182**, 7595-7602 (2009).
42. J. Ye, L. Gradoville, G. Miller, Cellular immediate-early gene expression occurs kinetically upstream of Epstein-Barr virus bzlf1 and brlf1 following cross-linking of the B cell antigen receptor in the Akata Burkitt lymphoma cell line. *J Virol* **84**, 12405-12418 (2010).
43. J. Ollila, M. Vihinen, Stimulation-induced gene expression in Ramos B-cells. *Genes Immun* **4**, 343-350 (2003).
44. J. G. Monroe, Up-regulation of c-fos expression is a component of the mIg signal transduction mechanism but is not indicative of competence for proliferation. *J Immunol* **140**, 1454-1460 (1988).
45. A. Manzo, B. Vitolo, F. Humby, R. Caporali, D. Jarrossay, F. Dell'accio, L. Ciardelli, M. Uguccioni, C. Montecucco, C. Pitzalis, Mature antigen-experienced T helper cells synthesize and secrete the B cell chemoattractant CXCL13 in the inflammatory environment of the rheumatoid joint. *Arthritis Rheum* **58**, 3377-3387 (2008).

46. A. Manzo, S. Bugatti, R. Caporali, R. Prevo, D. G. Jackson, M. Uguccioni, C. D. Buckley, C. Montecucco, C. Pitzalis, CCL21 expression pattern of human secondary lymphoid organ stroma is conserved in inflammatory lesions with lymphoid neogenesis. *Am J Pathol* **171**, 1549-1562 (2007).
47. T. Okada, V. N. Ngo, E. H. Ekland, R. Forster, M. Lipp, D. R. Littman, J. G. Cyster, Chemokine requirements for B cell entry to lymph nodes and Peyer's patches. *J Exp Med* **196**, 65-75 (2002).
48. M. Miyasaka, T. Tanaka, Lymphocyte trafficking across high endothelial venules: dogmas and enigmas. *Nat Rev Immunol* **4**, 360-370 (2004).
49. A. Tumanov, D. Kuprash, M. Lagarkova, S. Grivennikov, K. Abe, A. Shakhov, L. Drutskaya, C. Stewart, A. Chervonsky, S. Nedospasov, Distinct role of surface lymphotoxin expressed by B cells in the organization of secondary lymphoid tissues. *Immunity* **17**, 239-250 (2002).
50. T. Arkatkar, S. W. Du, H. M. Jacobs, E. M. Dam, B. Hou, J. H. Buckner, D. J. Rawlings, S. W. Jackson, B cell-derived IL-6 initiates spontaneous germinal center formation during systemic autoimmunity. *J Exp Med* **214**, 3207-3217 (2017).
51. J. L. Cannons, P. Lau, B. Ghumman, M. A. DeBenedette, H. Yagita, K. Okumura, T. H. Watts, 4-1BB ligand induces cell division, sustains survival, and enhances effector function of CD4 and CD8 T cells with similar efficacy. *J Immunol* **167**, 1313-1324 (2001).
52. D. S. Vinay, B. S. Kwon, 4-1BB signaling beyond T cells. *Cell Mol Immunol* **8**, 281-284 (2011).
53. D. Suan, N. J. Krautler, J. L. V. Maag, D. Butt, K. Bourne, J. R. Hermes, D. T. Avery, C. Young, A. Statham, M. Elliott, M. E. Dinger, A. Basten, S. G. Tangye, R. Brink, CCR6 Defines Memory B Cell Precursors in Mouse and Human Germinal Centers, Revealing Light-Zone Location and Predominant Low Antigen Affinity. *Immunity* **47**, 1142-1153 e1144 (2017).
54. J. Zikherman, R. Parameswaran, A. Weiss, Endogenous antigen tunes the responsiveness of naive B cells but not T cells. *Nature* **489**, 160-164 (2012).
55. J. Mueller, M. Matloubian, J. Zikherman, Cutting edge: An in vivo reporter reveals active B cell receptor signaling in the germinal center. *J Immunol* **194**, 2993-2997 (2015).
56. J. F. Ashouri, A. Weiss, Endogenous Nur77 Is a Specific Indicator of Antigen Receptor Signaling in Human T and B Cells. *J Immunol* **198**, 657-668 (2017).
57. S. Alivernini, L. MacDonald, A. Elmesmari, S. Finlay, B. Tolusso, M. R. Gigante, L. Petricca, C. Di Mario, L. Bui, S. Perniola, M. Attar, M. Gessi, A. L. Fedele, S. Chilaka, D. Somma, S. N. Sansom, A. Filer, C. McSharry, N. L. Millar, K. Kirschner, A. Nerviani, M. J. Lewis, C. Pitzalis, A. R. Clark, G. Ferraccioli, I. Udalova, C. D. Buckley, E. Gremese, I. B. McInnes, T. D. Otto, M. Kurowska-Stolarska, Distinct synovial tissue macrophage subsets regulate inflammation and remission in rheumatoid arthritis. *Nat Med* **26**, 1295-1306 (2020).
58. P. R. Mittelstadt, A. L. DeFranco, Induction of early response genes by cross-linking membrane Ig on B lymphocytes. *J Immunol* **150**, 4822-4832 (1993).
59. C. Tan, R. Hiwa, J. L. Mueller, V. Vykunta, K. Hibiya, M. Noviski, J. Huizar, J. F. Brooks, J. Garcia, C. Heyn, Z. Li, A. Marson, J. Zikherman, NR4A nuclear receptors restrain B cell responses to antigen when second signals are absent or limiting. *Nat Immunol* **21**, 1267-1279 (2020).

60. C. Tan, J. L. Mueller, M. Noviski, J. Huizar, D. Lau, A. Dubinin, A. Molofsky, P. C. Wilson, J. Zikherman, Nur77 Links Chronic Antigen Stimulation to B Cell Tolerance by Restricting the Survival of Self-Reactive B Cells in the Periphery. *J Immunol* **202**, 2907-2923 (2019).
61. I. Zaretsky, O. Atrakchi, R. D. Mazor, L. Stoler-Barak, A. Biram, S. W. Feigelson, A. D. Gitlin, B. Engelhardt, Z. Shulman, ICAMs support B cell interactions with T follicular helper cells and promote clonal selection. *J Exp Med* **214**, 3435-3448 (2017).
62. H. Kristyanto, N. J. Blomberg, L. M. Slot, E. I. H. van der Voort, P. F. Kerkman, A. Bakker, L. E. Burgers, R. M. Ten Brinck, A. H. M. van der Helm-van Mil, H. Spits, D. L. Baeten, T. W. J. Huizinga, R. E. M. Toes, H. U. Scherer, Persistently activated, proliferative memory autoreactive B cells promote inflammation in rheumatoid arthritis. *Sci Transl Med* **12**, eaaz5327 (2020).
63. S. E. Elliott, S. Kongpachith, N. Lingampalli, J. Z. Adamska, B. J. Cannon, L. K. Blum, M. S. Bloom, M. Henkel, M. J. McGeachy, L. W. Moreland, W. H. Robinson, B cells in rheumatoid arthritis synovial tissues encode focused antibody repertoires that include antibodies that stimulate macrophage TNF-alpha production. *Clin Immunol* **212**, 108360 (2020).
64. W. Ise, K. Fujii, K. Shiroguchi, A. Ito, K. Kometani, K. Takeda, E. Kawakami, K. Yamashita, K. Suzuki, T. Okada, T. Kurosaki, T Follicular Helper Cell-Germinal Center B Cell Interaction Strength Regulates Entry into Plasma Cell or Recycling Germinal Center Cell Fate. *Immunity* **48**, 702-715 e704 (2018).
65. K. P. O'Rourke, G. O'Donoghue, C. Adams, H. Mulcahy, C. Molloy, C. Silke, M. Molloy, F. Shanahan, F. O'Gara, High levels of Lymphotoxin-Beta (LT-Beta) gene expression in rheumatoid arthritis synovium: clinical and cytokine correlations. *Rheumatol Int* **28**, 979-986 (2008).
66. L. T. Donlin, D. A. Rao, K. Wei, K. Slowikowski, M. J. McGeachy, J. D. Turner, N. Meednu, F. Mizoguchi, M. Gutierrez-Arcelus, D. J. Lieb, J. Keegan, K. Muskat, J. Hillman, C. Rozo, E. Ricker, T. M. Eisenhaure, S. Li, E. P. Browne, A. Chicoine, D. Sutherby, A. Noma, R. A. S. L. E. N. Accelerating Medicines Partnership, C. Nusbaum, S. Kelly, A. B. Pernis, L. B. Ivashkiv, S. M. Goodman, W. H. Robinson, P. J. Utz, J. A. Lederer, E. M. Gravallese, B. F. Boyce, N. Hacohen, C. Pitzalis, P. K. Gregersen, G. S. Firestein, S. Raychaudhuri, L. W. Moreland, V. M. Holers, V. P. Bykerk, A. Filer, D. L. Boyle, M. B. Brenner, J. H. Anolik, Methods for high-dimensonal analysis of cells dissociated from cyropreserved synovial tissue. *Arthritis Res Ther* **20**, 139 (2018).

Supplemental Acknowledgements

Accelerating Medicines Partnership Rheumatoid Arthritis & Systemic Lupus Erythematosus (AMP RA/SLE) Network member list

William Apruzzese[6], Arnon Arazi[29,23], Jennifer Barnard[1], Jennifer Barnas[1], Joan M. Bathon[14], H. Michael Belmont[30], Ami Ben-Artzi[15], Celine Berthier[31], Brendan F. Boyce[1], David L. Boyle[16], Michael B. Brenner[6], S. Louis Bridges Jr[17], Jill P. Buyon[30], Vivian P. Bykerk[8], Phillip Carlucci[30], Arnold Ceponis[16], Adam Chicoine[6], Robert Clancy[30], Sean Connery[32], Carla M. Cuda[18], Maria Dall'Era[33], Anne Davidson[23], Kevin Deane[19], Wade DeJager[20], Betty Diamond[23], Kristina Deonaraine[30], Salina Dominguez[18], Patrick J. Dunn[21], Thomas Eisenhaure[29], Andrea Fava[34], Andrew Filer[22], Gary S. Firestein[16], Lindsy Forbess[15], Jennifer Goff[1], Beatrice Goilav[35], Ellen M. Gravallese[6], Peter K. Gregersen[24], Jennifer Grossman[37], Joel M. Guthridge[21], Nir Hacohen[29], David Hildeman[37,38], Jeffrey Hodgin[31], V. Michael Holers[19], Paul Hoover[29], Diane Horowitz[23], Raymond Hsu[33], Laura B. Hughes[17], Kazuyoshi Ishigaki[2,3,5], Mariko Ishimori[15], Peter Izmirly[30], Judith A. James[20], Tony Jones[29], A. Helena Jonsson[6], Kenneth C. Kalunian[16], Diane L. Kamen[40], Joshua Keegan[24], Gregory Keras[6], Ilya Korsunsky[2,3,4,5,6], Matthias Kretzler[31], Manjunath Kustagi[41], Amit Lakhanpal[8], James A. Lederer[6], Myles Lewis[7], Jessica Li[34], Yuhong Li[6], Zhihan Jian Li[6], David Lieb[29], Susan Macwana[20], Holden Maecker[25,26], Arthur M. Mandelin II[18], Rong Mao[25,26], Mandy J. McGeachy[27], Maureen A. McMahon[36], Joseph R. Mears[2,3,4,5], Raji Menon[31], Nghia Millard[2,3,4,5], Larry Moreland[19], Pavel Morozov[41], Aparna Nathan[2,3,4,5], Alessandra

Nerviani[7], Fernanda Payan-Schober[32], Harris Perlman[18], Michael Peters[29], Michelle A. Petri[34], Chaim Putterman[35], Deepak A. Rao[6], Karim Raza[22], Christopher Ritchlin[1], Felice Rivellese[7], William H. Robinson[25,26], Ilfita Sahbudin[22], Saori Sakaue[2,3,4,5], Karen Salomon-Escoto[28], Jennifer Seifert[19], Lorien Shakib[8], Heather Sherman[8], Daimon Simmons[6], Anvita Singaraju[8], Melanie Smith[8], Hemant Suryawanshi[41], Darren Tabechian[1], Anjali Thakrar[18], Thomas Tuschl[41], Paul J. Utz[25,26], Gerald Watts[6], Kevin Wei[6], Kathryn Weinand[2,3,4,5], Michael Weisman[15], David Wofsy[33], E. Steve Woodle[39], Qian Xiao[2,3,4,5], Zhu Zhu[6]

[14]Division of Rheumatology, Department of Medicine, NewYork-Presbyterian/Columbia University Irving Medical Center, New York, NY, USA.

[15]Division of Rheumatology, Cedars Sinai Medical Center, Los Angeles, CA, USA.

[16]Department of Medicine, Division of Rheumatology, Allergy and Immunology, University of California, San Diego, La Jolla, CA, USA.

[17]Division of Clinical Immunology and Rheumatology, Department of Medicine, University of Alabama at Birmingham, Birmingham, AL, USA.

[18]Division of Rheumatology, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

[19]Division of Rheumatology, University of Colorado School of Medicine, Aurora, CO, USA.

[20]Department of Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA.

[21] ImmPort Curation Team, NG Health Solutions, 2101 Gaither Road Rockville, MD 20850, USA.

[22]Rheumatology Research Group, Institute for Inflammation and Ageing, NIHR Birmingham Biomedical Research Center and Clinical Research Facility, University of Birmingham, Queen Elizabeth Hospital, Birmingham, UK.

[23]Institute of Molecular Medicine, Feinstein Institute for Medical Research, North Shore-LIJ Health System, Manhasset, NY, USA.

[24]Department of Surgery, Brigham and Women's Hospital and Harvard Medical School,

Boston, MA, USA.

[25]Division of Immunology and Rheumatology, Department of Medicine, Stanford University School of Medicine, Palo Alto, CA, USA.

[26]The Institute for Immunity, Transplantation, and Infection, Stanford University School of Medicine, Stanford, CA, USA.

[27]Division of Rheumatology and Clinical Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA.

[28]Division of Rheumatology, Department of Medicine, University of Massachusetts Medical School, Worcester, MA, USA.

[29]Broad Institute of MIT and Harvard, Cambridge, MA, USA.

[30]Department of Medicine, Division of Rheumatology, New York University School of Medicine, New York, NY, USA.

[31]Internal Medicine, Department of Nephrology, University of Michigan, Ann Arbor, MI, USA.

[32]Department of Medicine, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, El Paso, TX, USA.

[33]Rheumatology Division, University of California San Francisco, San Francisco, CA, USA.

[34]Division of Rheumatology, Johns Hopkins University, Baltimore, MD, USA.

[35]Division of Rheumatology and Department of Microbiology and Immunology, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA.

[36]Department of Medicine, University of California Los Angeles, Los Angeles, CA, USA.

[37]Department of Pediatrics, University of Cincinnati, Cincinnati, OH, USA.

[38]Division of Immunobiology, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio, USA.

[39]Division of Transplantation, Department of Surgery, University of Cincinnati College of Medicine, Cincinnati, OH, USA.

[40]Division of Rheumatology and Immunology, Medical University of South Carolina, Charleston, SC, USA.

[41]Laboratory of RNA Molecular Biology, Rockefeller University, New York, NY, USA.

FIGURE LEGENDS

Fig 1. Overview of work flow. Synovial tissues were disaggregated using a validated protocol (*66*) and B cells were sorted by flow cytometry cell sorting. We then performed scRNA-seq on sorted tissue B cells using the 10x genomic platform with poly-A selected, 5' initiated expression and V(D)J libraries generated from each single cell.

Fig 2. **Single-cell RNA sequencing identifies an enrichment of B cells expressing nuclear orphan family receptors (NR4As) in RA synovial tissue.** (**A**) tSNE visualization of 8 B cell clusters from 3786 B cells from 4 RA synovial tissues (RA134_SYN, RA172_SYN, RA195_SYN and RA221_SYN) and one paired blood sample (RA134_BLD). (**B**) Markers identified 2 clusters of plasma cells (PC (i), (ii)), 3 clusters of naïve (Naïve (i), (ii), (iii), a memory cluster, LMNA$^+$ cluster and NR4A$^+$ cluster. (**C**) The same tSNE map as in (A) with cells labeled by sample ID. SYN and BLD denoted synovial tissue sample and blood sample, respectively. (**D**) Color heatmap displays top differentially expressed genes in each cell clusters. Top 10 differentially expressed genes of NR4A$^+$ cluster are magnified on the right of heatmap. (**E**) Box plots display frequency of B cells from published data sets that labeled with NR4A$^+$ cluster genes calculated using a supervised classification method SingleR. Single cell RNA sequencing of Synovium (AMP) (n=14) from Zhang et al, 2019 (*20*) (ImmPort SDY998), SLE (AMP)(n=13) from Arazi et al, 2019 (*31*) (ImmPort SDY997), peripheral B cells from Zheng et al, 2017 (*32*) and the blood sample from this study, and data from our current study (Synovium (10x). Horizontal line inside the box plots is the mean and the blue box represent 90% confidence interval. (**F**) Bar graphs plotting log2 fold change of genes associated with NR4A$^+$ cluster in leukocyte poor RA vs. OA, leukocyte rich RA vs. OA or RA biopsy vs. OA. Bulk RNA sequencing data were used (ImmPort SDY1299). Leukocyte rich (n=19), leukocyte poor (n=17) RA and OA (n=15) samples were from Zhang et al, 2019 (*20*). Leukocyte rich and leukocyte poor RA were separated on the basis of flow cytometry of lymphocyte and monocyte infiltration using OA samples as a reference as described in Zhang et al, 2019 (*20*). * indicates significant at 5% and ** indicates significant at 1%

Fig 3. **B cells in NR4A+ cluster display somatic hypermutation (SHM), clonal expansion, and shared clonality with plasma cells.** (**A**) Violin plots display SHM rate, averaged across detected chains, in each B cell cluster obtained from single cell BCR sequencing. Each cluster's SHM rate was tested against Naive(i) using linear mixed models, with $p<0.05$ displayed. (**B**) Scatter plots display the associations between % SHM and expression level of *CD27*, *NR4A1*, *NR4A2* and *IGHD* in NR4A$^+$, memory and plasma cell (combined PC(i) and PC(ii) clusters) clusters. Blue lines display a linear mixed model fit and grey area depict 95% confidence intervals. p-values are from a linear mixed effect model that adjusts for sample and number of genes detected. Only p-values < 0.05 are displayed. (**C**) Number of cells with >97% DNA identity of the heavy-chain CDR3. The amino acid sequence of each putative clone is shown on the left. (**D**) Heavy-chain BCR data were superimposed on cluster t-SNE plot. None (gray) indicates cells that BCR was not recovered, Recovered (black) marks cells that BCR was recovered, Expanded (blue) show cells with clonality in $\geq$2 cells, CARHWRGKKPFDSW (red) marks cells with a shared clone recovered from both NR4A+ and Plasma cells.

Fig 4. **NR4A+ cluster is enriched in genes expressed by germinal center light zone (GC LZ) B cells.** (**A**) Heatmap displays signed-log10 (FDR q-value) of gene set enrichment analysis between our B cell clusters and published blood and tonsil B cell subsets. (**B**) Log2 fold change of germinal center dark zone (GC DZ) and light zone (GC LZ) genes from a published study (*38*) in each B cell cluster. (**C**) Violin plots display log2 fold change of all LZ genes combined (top) or all DZ genes combined (bottom) from (B) in each cluster. Solid lines represent median and the dotted lines depict quartiles.

Fig 5. **NR4A+ cluster expresses chemokine receptors and cytokines important for ELS formation** (**A**) Heatmap displays log2 fold change of cytokine and chemokine receptors in each B cell subset (n=1250 naive, 746 NR4A$^+$, 633 LMNA$^+$, 330 memory and 779 PC). Naive subset is combined Naïve (i), (ii) and (iii). Plasma cell (PC) is combined PC(i) and (ii). Shown are receptors that are expressed in at least 20% of the cells of at least one subset. Volcano plot display log2 fold change and –log10 p value of genes in the heatmap in NR4A$^+$ subset. p-values are adjusted using the Benjamini-Hochberg correction for multiple tests. (**B**) Bar graphs display percentage of cells of each subset that express *CXCR5, CCR7, CXCR4* or *CCR6*. (**C**) Heatmap displays log2 fold change of cytokines and chemokines in each B cell subset. Shown are cytokines or chemokines that are expressed in at least 20% of the cells of at least one subset. Volcano plot display log2 fold change and –log10 p value of genes in the heatmap in NR4A$^+$ and LMNA$^+$ subset. p-values are adjusted using the Benjamini-Hochberg correction for multiple tests. (**D**) Bar graph represents percentage of cells of each subset that express *IL6, TNFSF9, CD70, TGFB1, CCL5* or *LTB*. (**E**) GSEA analysis revealed significant enrichment of PreM signature (*CCR6, CD44, CD69, BANK1, RASGRP2, STAT1, IFNGR1, CELF2, IFITM1, GPR183, TNFRSF13B, SELL, FXYD5, CXCR4* and *MYC)* in NR4A$^+$ and LMNA$^+$ subsets. The bars display –log10 (p-value from GSEA analysis to determine the enrichment of PreM signature in each subset, *p < 0.05. Middle heatmaps show log2 fold change of PreM genes (*40*) in CCR6$^+$ and CCR6$^-$ cells within NR4A$^+$, LMNA$^+$ or naïve subset. Violin plots showed log2 expression of CCR6 in each cluster. Right bar graph shows percentage of cells of each subset that co-express *CCR6, CELF2* and *BANK1*.

Fig 6. **NR4A1 is expressed by B cells and plasma cells in tonsil SLO.** (**A**) Immunofluorescent staining of NR4A1 (green) in combination with CD20 (red) and CD21(white) in 5 μm formalin-fixed tonsil serial sections. Representative 3x3 200X mosaic (left image) and 200X magnification pictures show the composite and individual channels. Box area represent area that was magnified. Dashed yellow lines outline the tonsillar germinal centers. Scale bars represents 100 μm. White arrows show NR4A1$^+$ B cells in close proximity to CD21$^+$ FDC in the GC LZ. (**B**) Immunofluorescent staining of 5 μm formalin-fixed tonsil serial sections. First section was stained with CD20 (red), CD138 (white) and CD3 (green). The adjacent section was stained with NR4A1 (green) in combination with CD20 (red) and Ki67 (white) in tonsil serial sections. Area in the white box was magnified. White arrows point out CD138$^+$ plasma cells in the first section and the corresponding NR4A1$^+$ plasma cells in the adjacent section. (**C**) Flow cytometry for NR4A1 expression in tonsil B cells. Cell suspension was prepared from de-identified tonsil (n=6) and intracellular staining was performed to stain for NR4A1. Additional B cell markers were included to facilitate characterization of B cell subsets in tonsil. CD38 and IgD are used to characterize Bm1+Bm2, Bm5, GC and PC. Dot plot representative of phenotype of NR4A1+ tonsil B cells with respect to CD38 and IgD expression.

Bar graph display the percentage of NR4A1+ tonsil B cells that are Bm1+Bm2, Bm5, GC or PC. Line graph shows geometric mean fluorescent intensity (gMFI) of NR4A1+ Bm1+Bm2, Bm5, GC or PC. Dash line connect data from the same sample. Error bars are SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.005$, **** $p < 0.0001$ by non-parametric one-way ANOVA and Tukey's multiple comparisons test. (**D**) GC B cells were defined as IgD-CD38+. GC B cells were further characterized based on CXCR4 and CD83 expression. CD83-CXCR4+ are dark zone (DZ) cells, CD83+CXCR4+ are intermediate zone (IZ) cells whereas CD83+CXCR4- are light zone (LZ) cells. Percentage of $NR4A1^+$ cells were determined in each CD83-CXCR4 population and plotted per sample. The lines connect data from the same sample. One-way ANOVA and Tukey's multiple comparisons were used to test for statically significance, *p<0.05, and **p<0.005.

Fig 7. **NR4A1 is expressed by B cells and plasma cells in synovium ELS.** (**A**) Immunofluorescent staining of NR4A1(green) in combination with CD20 (red) and Ki67 (white) in 5 μm formalin-fixed synovial serial sections. Representative 3x3 200X mosaic show the composite and individual channels. White (*) and orange (#) boxes are magnified in the images on the right. White arrow points out an aggregate of $Ki67^+$ B cells in an ectopic GC DZ. Gray arrows show $NR4A1^+$ B cells outside of GC. Yellow arrows on the right most image show $NR4A1^+$ plasma cells. (**B**) Example dot plot showing spontaneous NR4A1 expression by B cells and plasma cells from RA synovial tissue. (**C**) Histogram represents intracellular NR4A1 expression immediately ex vivo from RA and normal control (NC) peripheral blood B cells (PBL) and RA synovial fluid B cells (RA SynF). Bar graph represents mean of % $NR4A1^+$ B cells in immediate ex vivo sample from NC (n=5), RA (n=5) and synovial fluid (n=6). Dash line is the median FMI of NR4A1 in NC PBL. Error bars are SEM and * $p < 0.05$ by non-parametric one-way ANOVA and Tukey's multiple comparisons test. (**D**) NR4A is up-regulated with BCR engagement. Histogram represents intracellular NR4A1 expression in CD19+ B cells from PBMC treated with anti-Ig(A+M+G) for 4 hours. Peripheral blood B cells (PBL) from normal control (NC), RA and synovial fluid B cells (RA SynF). Scatter plots represent % $NR4A1^+$ B cells after treatment with anti-Ig(A+M+G) for 4 hours (S, open symbols) compared to untreated control (U, closed symbols). Dash line is isotype control for NR4A1.** $p < 0.01$ by paired t-test.

Fig. 8. Lymphoid specific NR4A subset genes correlate with RA synovial tissue pathotype and increase in the blood during RA flare. Boxplots and linear regression of histology and clinical parameters by tertile demonstrating correlation with (**A**) CD83 (P_{adj} lymphoid v myeloid=3.0e-0.2; P_{adj} myeloid v fibroid=3.4e-0.2; P_{adj} lymphoid v fibroid=3.3e-0.7) and (**B**) GPR183 (P_{adj} lymphoid v myeloid=5.4e-0.3; P_{adj} myeloid v fibroid=4.6e-0.2; P_{adj} lymphoid v fibroid=4.1e-0.8). Ultrasound biopsy joint parameters: (ST, synovial thickness; PD, power doppler) at the biopsy joint (Ultrasound ST/PD BJ). p values were calculated by linear regression models. The plots and p-values were obtained from https://peac.hpc.qmul.ac.uk/ (*10*). (**C**) The plot shows significance of enrichment scores of marker genes from various synovial B cells in clusters of genes differentially expressed over time to flare (*21*). The dash vertical line represents the threshold for significance ($-\log_{10}$ FDR > 1.3). The numbers in parentheses display the gene ratio (# overlapping genes/total genes in the set) of $LMNA^+$, $NR4A^+$ and Naïve (iii) with Cluster 1.

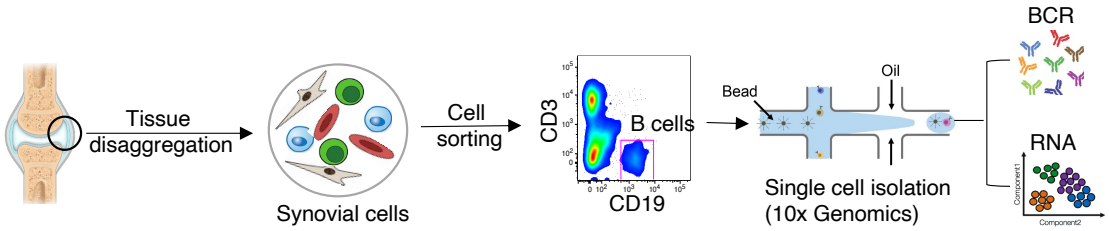


Figure 1

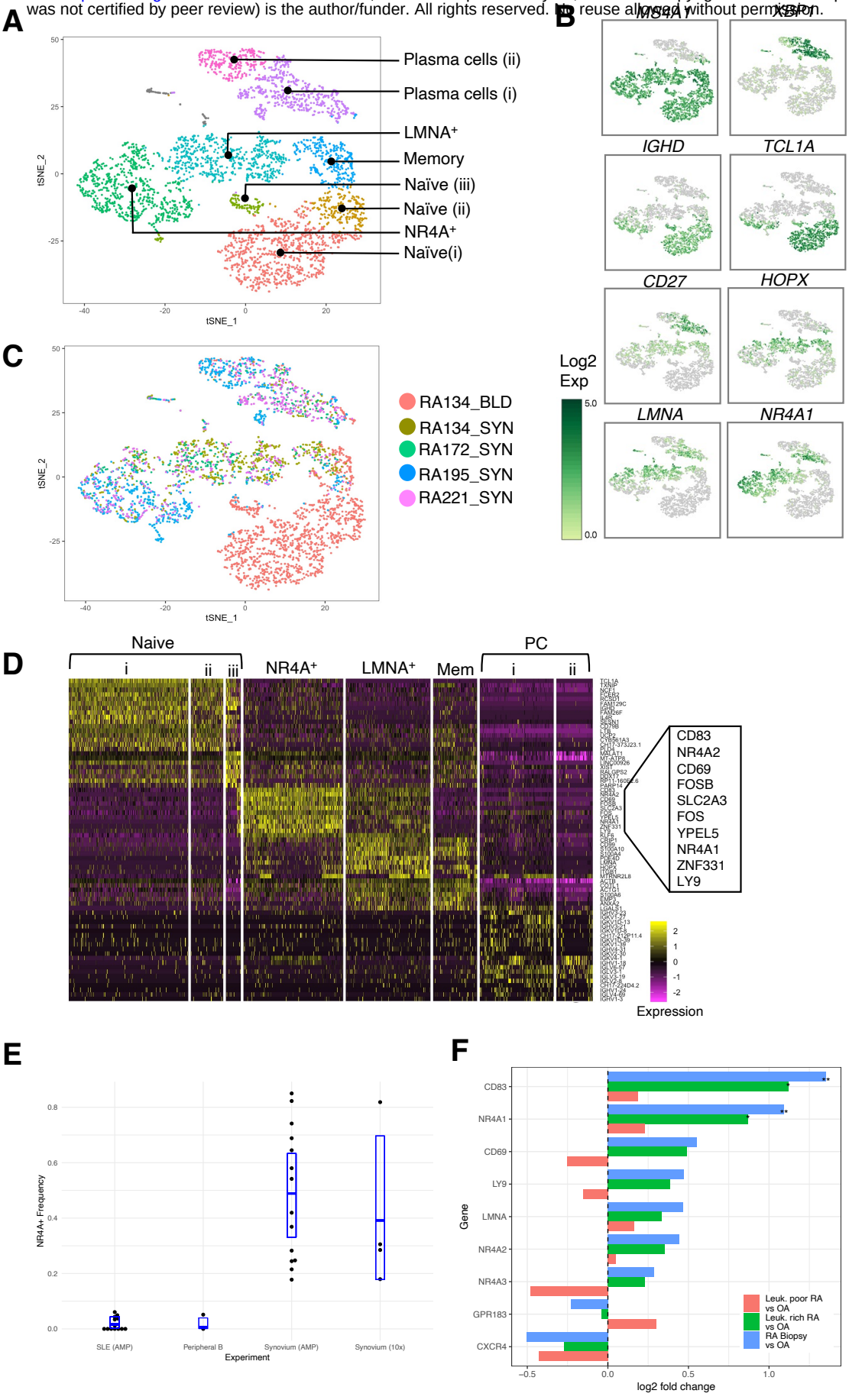


Figure 2

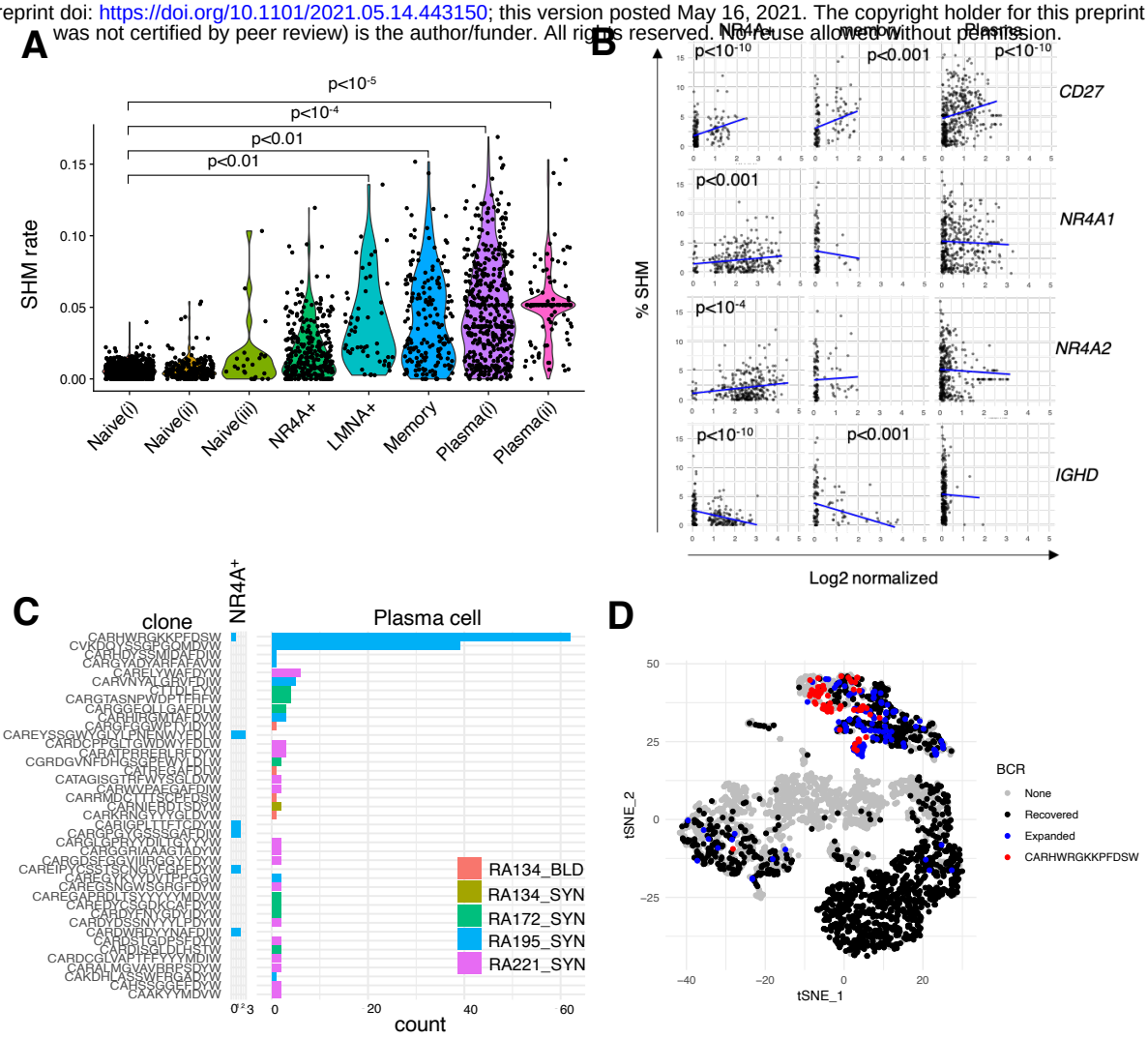


Figure 3

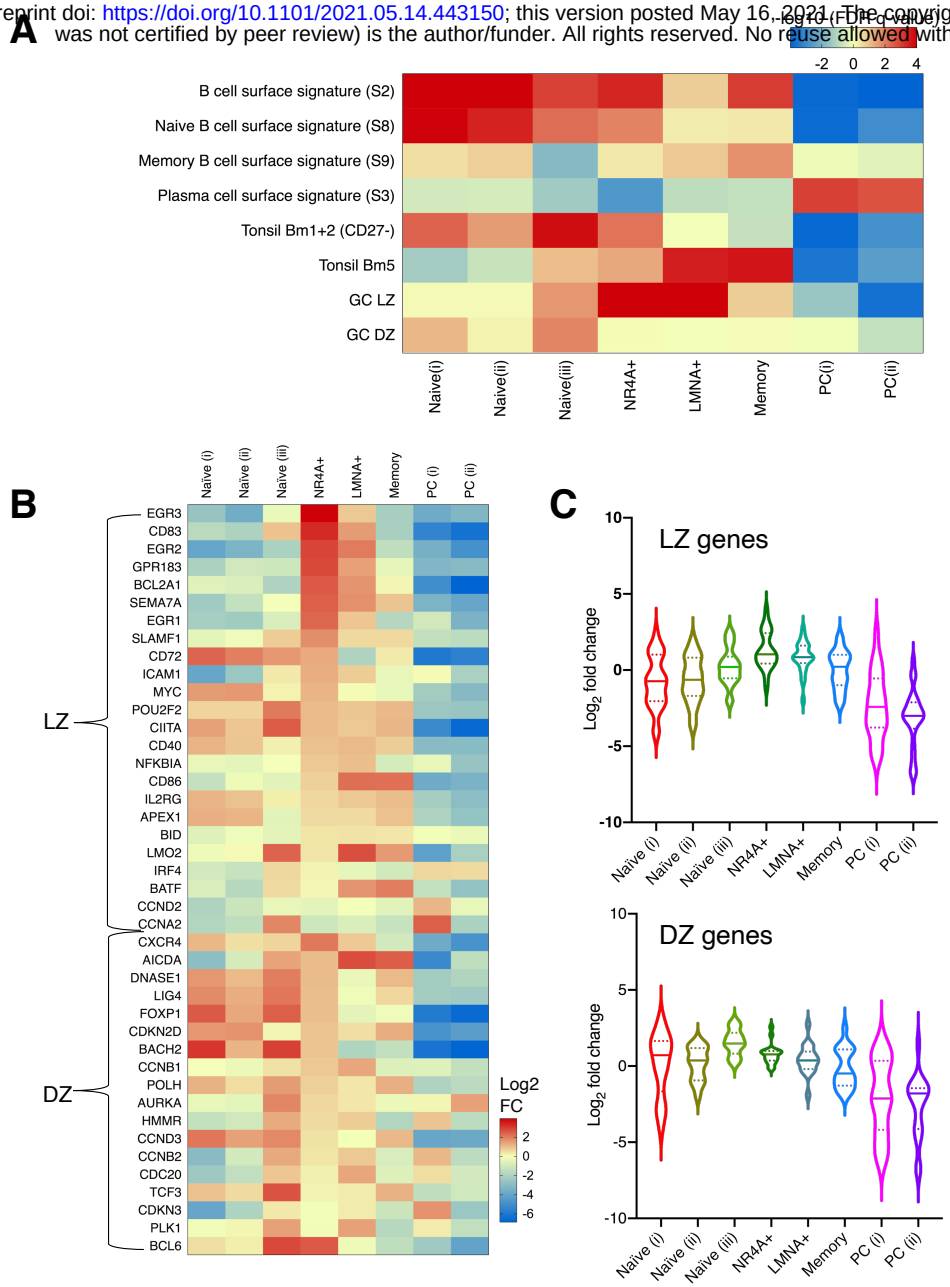


Figure 4

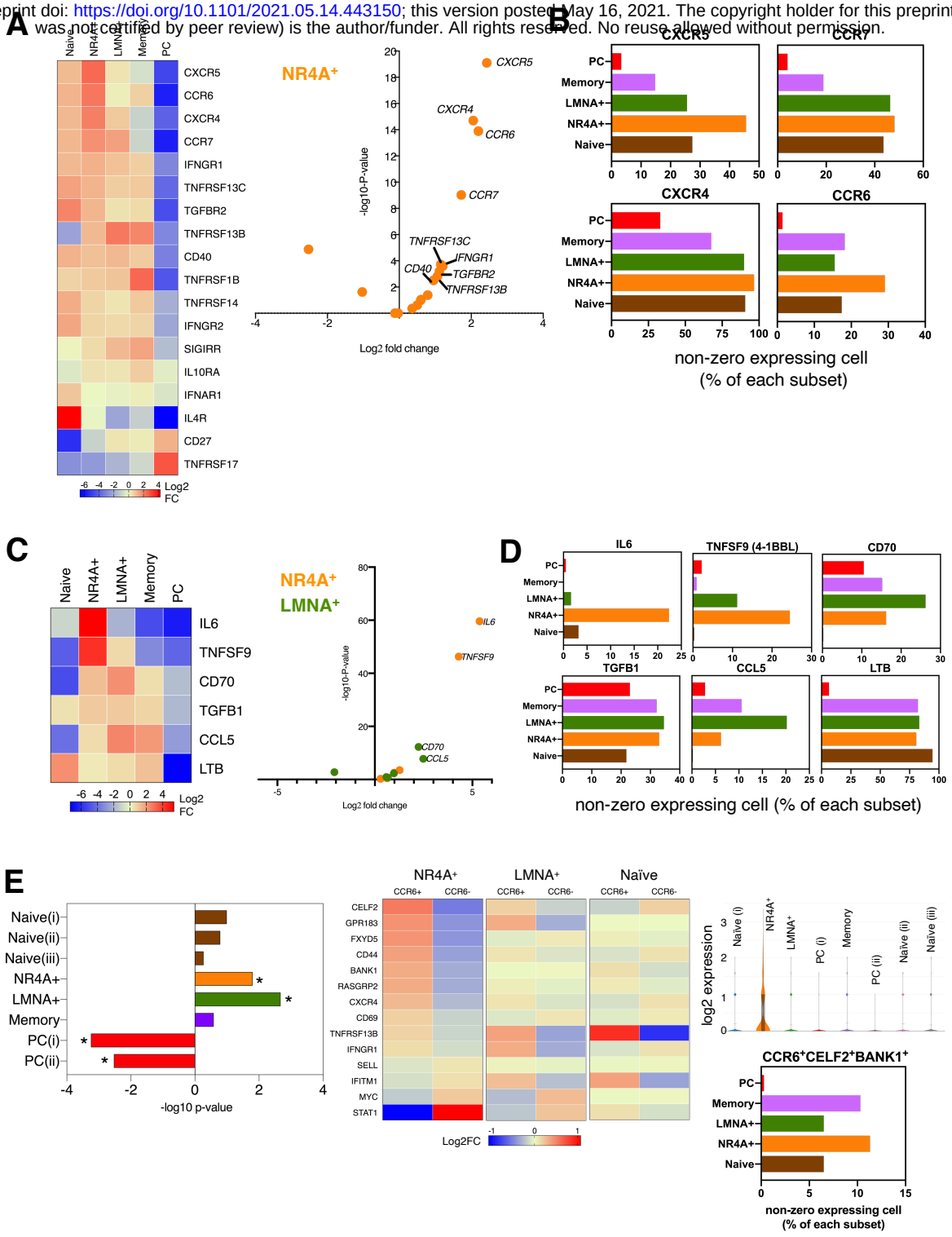


Figure 5

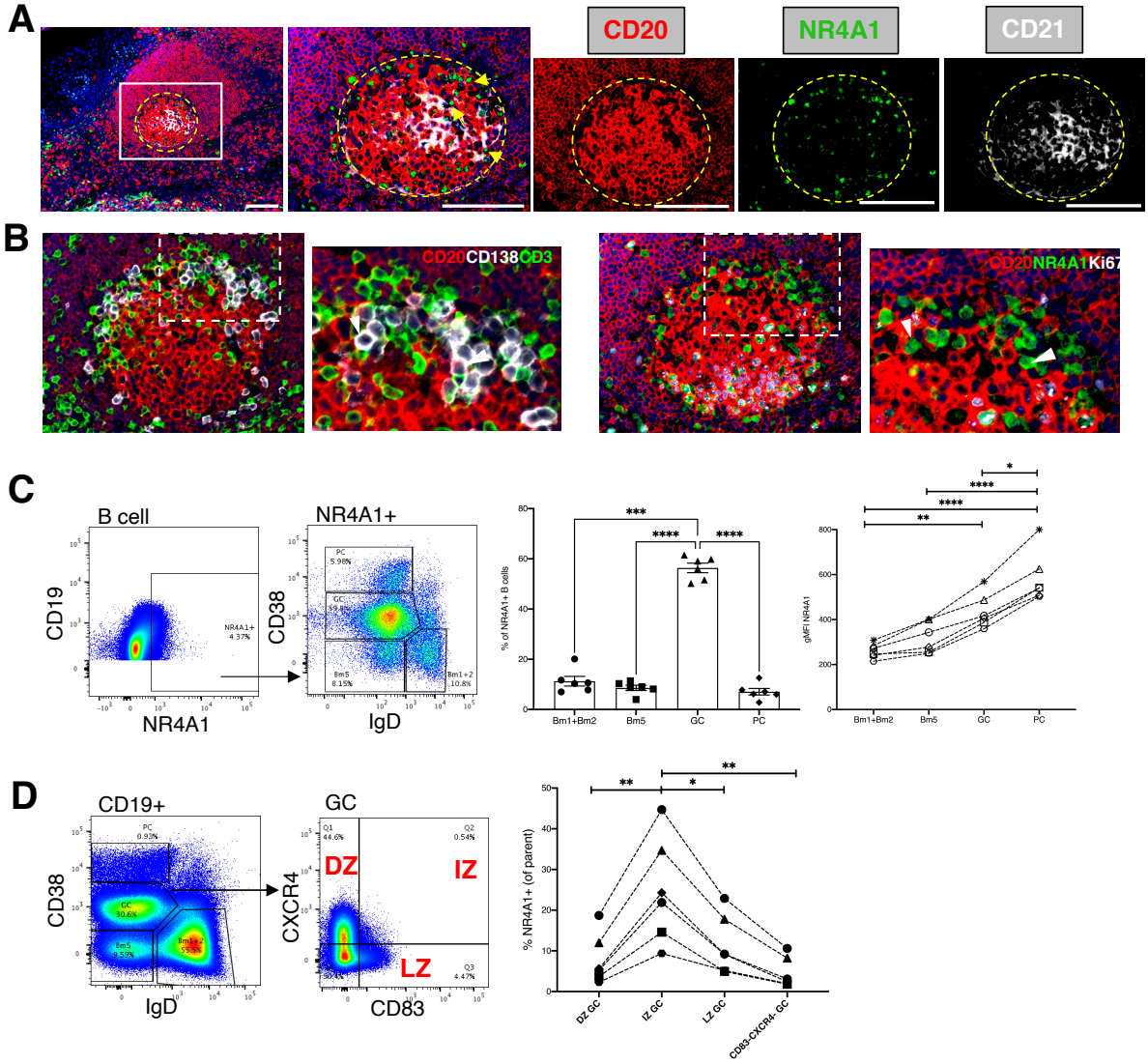
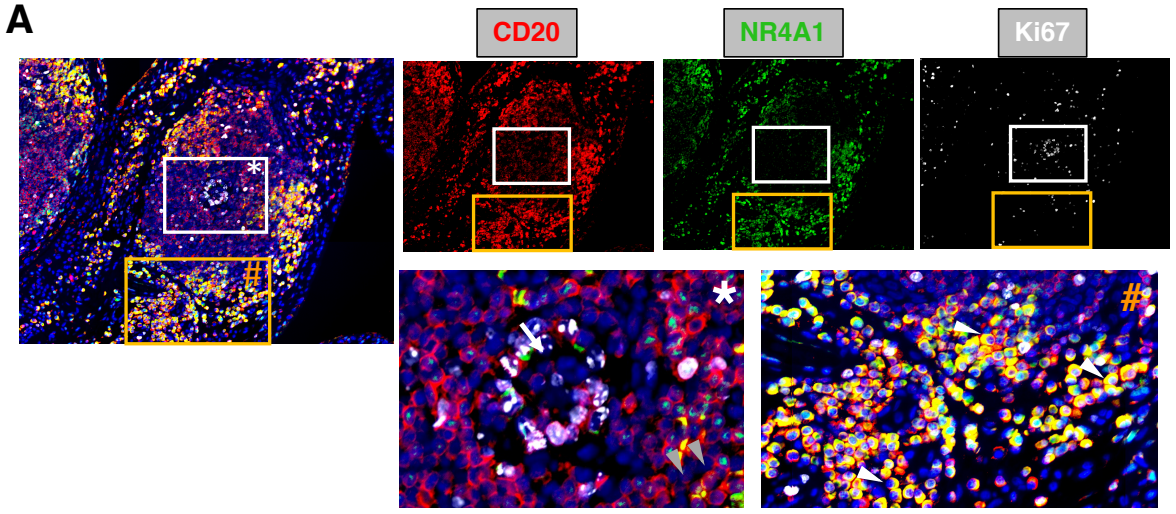
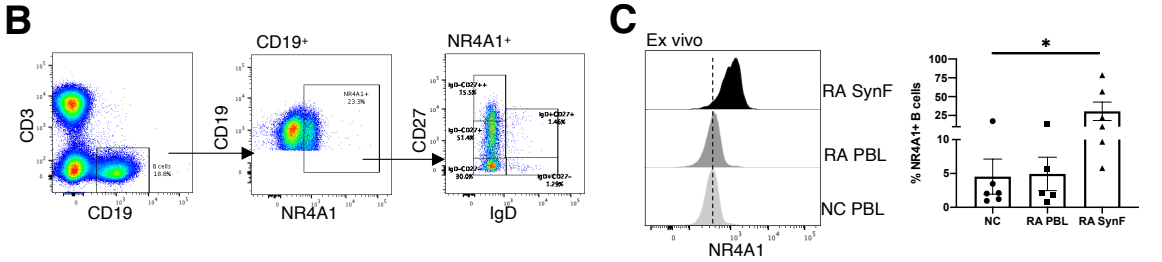

A
CD20
NR4A1
CD21
B
CD20CD138CD3
CD20NR4A1Ki67
C
B cell
NR4A1+
CD19
NR4A1
CD38
IgD
% of NR4A1+ B cells
Bm1+Bm2
Bm5
GC
PC
gMFI NR4A1
D
CD19+
GC
CD38
IgD
CXCR4
CD83
DZ
IZ
LZ
% NR4A1+ (of parent)
DZ GC
IZ GC
LZ GC
CD83-CXCR4- GC

Figure 6


A
CD20
NR4A1
Ki67
*
#


B
CD3
CD19
CD19+
NR4A1
CD27
IgD
NR4A1+
C
Ex vivo
RA SynF
RA PBL
NC PBL
NR4A1
% NR4A1+ B cells
NC
RA PBL
RA SynF
*

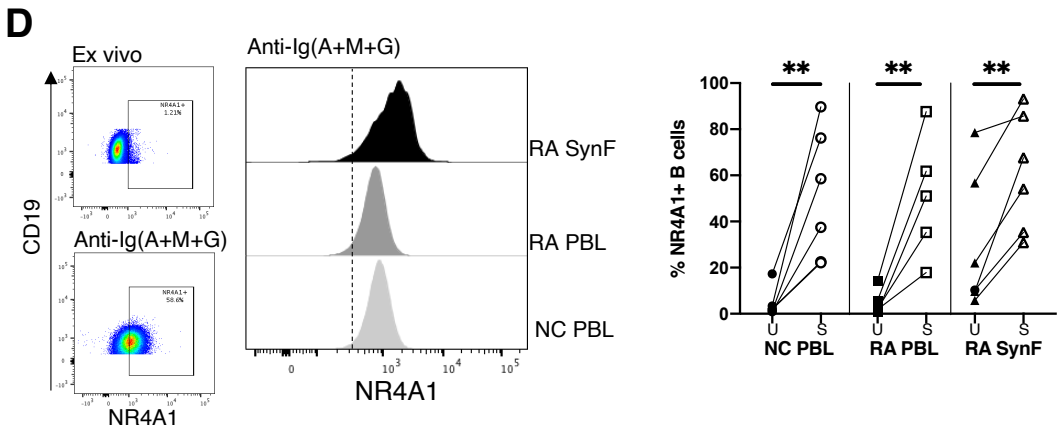

D
Ex vivo
Anti-Ig(A+M+G)
CD19
NR4A1
RA SynF
RA PBL
NC PBL
% NR4A1+ B cells
**
U
S
NC PBL
RA PBL
RA SynF

Figure 7

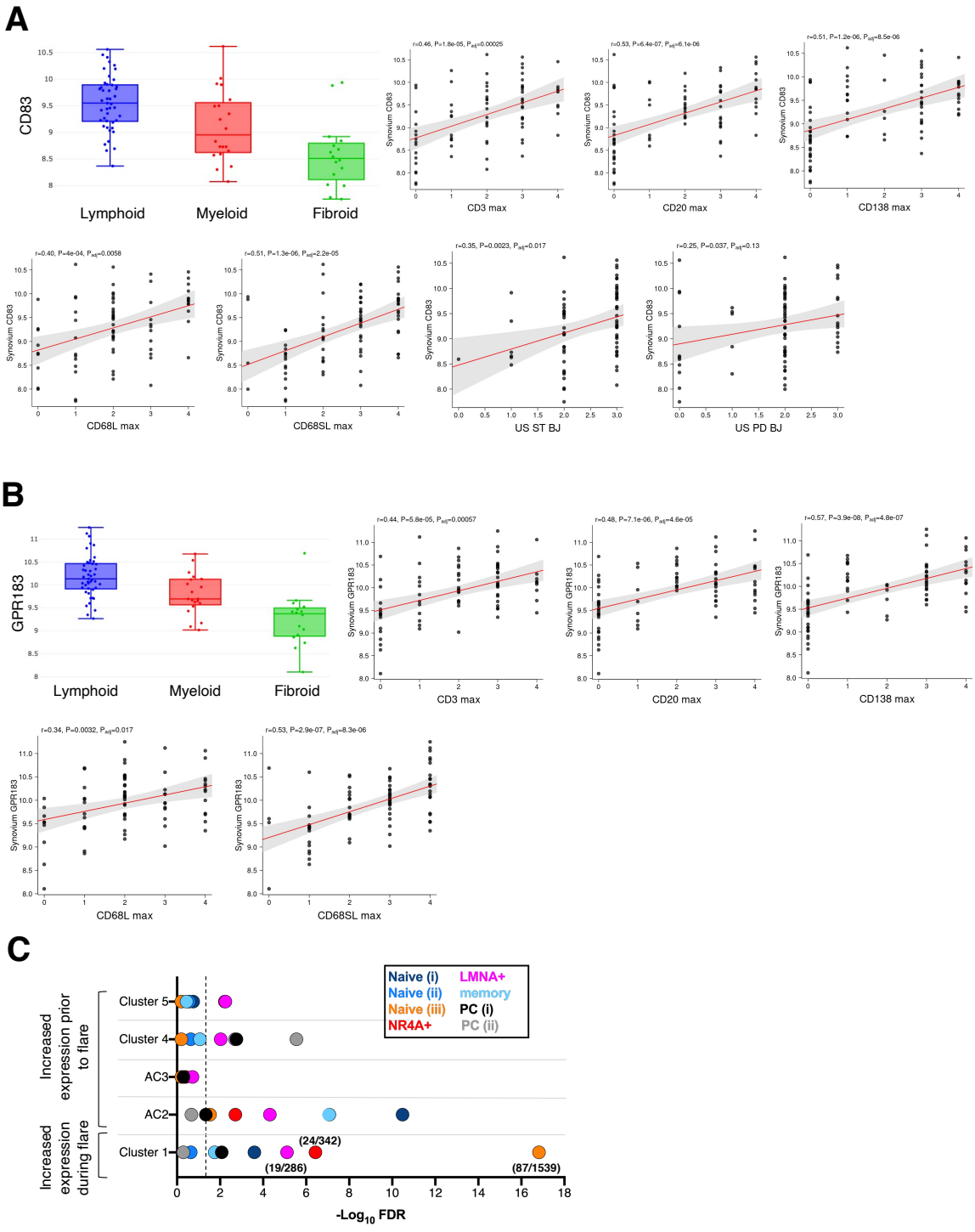

A
CD83
Lymphoid
Myeloid
Fibroid
CD3 max
CD20 max
CD138 max
CD68L max
CD68SL max
US ST BJ
US PD BJ
B
GPR183
Lymphoid
Myeloid
Fibroid
CD3 max
CD20 max
CD138 max
CD68L max
CD68SL max
C
Naive (i)
Naive (ii)
Naive (iii)
NR4A+
LMNA+
memory
PC (i)
PC (ii)
Increased expression prior to flare
Increased expression during flare
Cluster 5
Cluster 4
AC3
AC2
Cluster 1
(24/342)
(19/286)
(87/1539)
-Log10 FDR

Figure 8