Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry

Fan Zhang1,2,3,4,5,27, Kevin Wei5,27, Kamil Slowikowski1,2,3,4,5,27, Chamith Y. Fonseka1,2,3,4,5,27, Deepak A. Rao5,27, Stephen Kelly6, Susan M. Goodman7,8, Darren Tabechian9, Laura B. Hughes10, Karen Salomon-Escoto11, Gerald F. M. Watts5, A. Helena Jonsson5, Javier Rangel-Moreno9, Nida Meednu9, Cristina Rozo12, William Apruzzese3, Thomas M. Eisenhaure4, David J. Lieb4, David L. Boyle13, Arthur M. Mandellin14, Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium15, Brendan F. Boyce16, Edward DiCarlo17, Ellen M. Gravalles17, Peter K. Gregersen18, Larry Moreland19, Gary S. Firestein13, Nir Hacohen4, Chad Nusbaum4, James A. Lederer20, Harris Perlman14, Costantino Pitzalis21, Andrew Filer22,23, V. Michael Holers24, Vivian P. Bykerk7,8, Laura T. Donlin16,8,12,28, Jennifer H. Anolik5,25,28, Michael B. Brenner5,28 and Soumya Raychaudhuri5,1,2,3,4,5,26,28*

To define the cell populations that drive joint inflammation in rheumatoid arthritis (RA), we applied single-cell RNA sequencing (scRNA-seq), mass cytometry, bulk RNA sequencing (RNA-seq) and flow cytometry to T cells, B cells, monocytes, and fibroblasts from 51 samples of synovial tissue from patients with RA or osteoarthritis (OA). Utilizing an integrated strategy based on canonical correlation analysis of 5,265 scRNA-seq profiles, we identified 18 unique cell populations. Combining mass cytometry and transcriptomics revealed cell states expanded in RA synovia: THY1(CD90)+HLA-DRA4i sublining fibroblasts, IL1B+i pro-inflammatory monocytes, ITGA4+TBX21+ autoimmune-associated B cells and PDCD1i peripheral helper T (Tpn) cells and follicular helper T (Tfh) cells. We defined distinct subsets of CD4+ T cells characterized by GZMK+, GZMB+, and GNLY+ phenotypes. We mapped inflammatory mediators to their source cell populations; for example, we attributed IL6 expression to THY1+HLA-DRA4i fibroblasts and IL1B production to pro-inflammatory monocytes. These populations are potentially key mediators of RA pathogenesis.

A is an autoimmune disease with chronic inflammation in the synovium of the joint tissue1-5. This inflammation leads to joint destruction, disability, and shortened life span6. Defining key cellular subsets and their activation states in the inflamed tissue is a critical step in defining new therapeutic targets for RA. CD4+ T cells17, B cells17, monocytes18, and fibroblasts19,20 have established relevance to RA pathogenesis. Here, we use single-cell technologies to view all of these cell types simultaneously across a large collection of samples...
of samples from inflamed joints. We believe a global single-cell portrait of how different cell types work together would help identify new pathways in RA and eventually new therapeutics.

Application of transcriptomic and cellular profiling technologies to whole synovial tissue has already identified specific cell populations associated with RA\textsuperscript{12,13,14}. However, most studies have focused on a preselected cell type, surveyed whole tissues rather than disaggregated cells, or used only a single technology platform. The latest advances in single-cell technologies offer an opportunity to identify disease-associated cell subsets in human tissues at high resolution in an unbiased fashion\textsuperscript{15,16}. These technologies have already been used to discover roles for T peripheral helper (TPH) cells\textsuperscript{18} and HLA-DR$^+$CD27$^-$ cytotoxic T cells\textsuperscript{19} in RA pathogenesis. Studies using scRNA-seq have defined myeloid cell heterogeneity in human blood\textsuperscript{20} and identified overabundance of PDPN$^+$CD34$^+$THY1$^+$ (THY1, also known as CD90) fibroblasts in RA synovial tissue\textsuperscript{21,22}.

To generate high-dimensional multimodal single-cell data from synovial tissue samples collected across a collaborative network of research sites, we developed a robust pipeline\textsuperscript{23} in the Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) consortium. We collected and disaggregated tissue samples from patients with RA and OA and then subjected constituent cells to scRNA-seq, sorted-population bulk RNA-seq, mass cytometry, and flow cytometry. We developed a unique computational strategy based on canonical correlation analysis (CCA) to integrate multimodal transcriptomic and proteomic profiles at the single-cell level. A unified analysis of single cells across data modalities can precisely define contributions of specific cell subsets to pathways relevant to RA and chronic inflammation.

**Results**

**Generation of parallel mass cytometric and transcriptomic data from synovial tissue.** In phase 1 of AMP RA/SLE, we recruited 36 patients with RA who met the 1987 American College of Rheumatology (ACR) classification criteria and 15 patients with OA from ten clinical sites over 16 months (Supplementary Table 1) and obtained synovial tissues from ultrasound-guided biopsies or joint replacements (Methods and Fig. 1a). We required that all tissue samples included had synovial lining documented by means of histology. Synovial tissue disaggregation yielded an abundance of viable cells for downstream analyses (362,190 ± 7,687 (mean ± s.e.m.) cells per tissue). We used our validated strategy for cell sorting\textsuperscript{24} (Fig. 1a) to isolate B cells (CD45$^+$CD3$^-$CD19$^+$), T cells (CD45$^+$CD3$^+$), monocytes (CD45$^+$CD14$^+$), and stromal fibroblasts (CD45$^-$CD31$^-$PDPN$^+$) (Supplementary Fig. 1a). We applied bulk RNA-seq to all four sorted subsets for all 51 samples. For samples with sufficient cell yield (Methods), we also measured single-cell protein expression using a 34-marker mass cytometry panel (n = 26; Supplementary Table 2) and single-cell RNA expression in sorted cell populations (n = 21; Fig. 1b).

**Summary of computational data integration strategy to define cell populations.** To confidently define RA-associated cell populations, we integrated multiple data modalities (Fig. 1b,c). We used bulk RNA-seq data as the reference point, because these data were available for almost all the donors for all the cell types, had the highest dimensionality, and were least sensitive to technical artifacts (Fig. 1b).

Integrating scRNA-seq with bulk RNA-seq data ensures robust discovery of cell populations. Here, we used CCA to find linear combinations of bulk RNA-seq samples and scRNA-seq cells (Fig. 1c,d) to create gene expression profiles that were maximally correlated. These linear combinations captured sources of shared variation between the two datasets and allowed us to identify individual cell populations that drive variation in the bulk RNA-seq data. We analyzed the scRNA-seq data by using the canonical variate coefficients for each cell to compute a nearest neighbor network, identifying clusters with a community detection algorithm, and evaluating the separation between clusters with Silhouette analysis (Methods and Supplementary Fig. 2b).

We identified cell clusters in mass cytometry data using density-based clustering\textsuperscript{25}. Next, we used CCA to identify linear combinations of bulk RNA-seq genes and mass cytometry cluster abundances that maximized correlation across patients. These canonical variations offered a way to visualize genes and mass cytometry clusters together. We then queried this CCA result with the best marker genes from scRNA-seq to establish a relationship between each scRNA-seq cluster and each mass cytometry cluster (Methods). We also used CCA to associate bulk gene expression in each sample with proportions of cells in different flow cytometry gates.

**Flow cytometry features define a set of RA synovia that are leukocyte rich.** Histology of RA synovial tissues revealed heterogeneous tissue composition with variable lymphocyte and monocyte infiltration (Fig. 2a,b and Supplementary Fig. 2c,d). This heterogeneity was expected, because variation in tissue immune cell infiltration reflects local disease activity in the source joint. Consequently, we employed a data-driven approach to separate samples on the basis of flow cytometry of lymphocyte and monocyte infiltration in each tissue sample (Supplementary Fig. 1b,c). We calculated a multivariate normal distribution of these parameters based on OA samples as a reference, and for each RA sample, we calculated the Mahalanobis distance from OA\textsuperscript{26}. We defined the maximum OA distance (4.5) as the threshold for defining leukocyte-rich RA (r > 4.5, n = 19) or leukocyte-poor RA (< 4.5, n = 17) samples (Methods and Supplementary Fig. 1d). Whereas leukocyte-rich RA tissues had substantial infiltration of synovial T cells and B cells, leukocyte-poor RA tissues had cellular compositions more similar to those of OA samples (Fig. 2c). Synovial monocyte abundances were similar between RA and OA samples (Fig. 2c).

To test whether our classification indicated inflammation, we assessed tissue histology and assigned each sample a Krenn inflammation score\textsuperscript{17}. Samples that we classified as leukocyte-rich RA had significantly higher Krenn inflammation scores than those of leukocyte-poor RA or OA samples (Fig. 2d). In contrast, synovial lining membrane hyperplasia was not significantly different between leukocyte-rich RA, leukocyte-poor RA, and OA samples (Fig. 2d).

We observed significant correlation between synovial leukocyte infiltration, as measured via flow cytometry and the histological Krenn inflammation score (Fig. 2e). Mass cytometry in 26 synovial tissues was consistent with flow cytometry and histology. OA and leukocyte-poor RA samples were characterized by a high abundance of fibroblasts and endothelial cells; whereas, leukocyte-rich RA tissues were characterized by a high abundance of CD4$^+$ T cells, CD8$^+$ T cells, and B cells (Fig. 2f and Supplementary Fig. 3a).

**Single-cell RNA-seq analysis reveals distinct cell subpopulations.** Next, we analyzed 5,265 scRNA-seq profiles passing quality control (Methods), including 1,142 B cells, 1,844 fibroblasts, 750 monocytes, and 1,529 T cells. We used canonical variates (from CCA with technical plate effects \(n = 17\) samples (Methods and Supplementary Fig. 4b)). We defined the maximum OA inflammation score as the threshold for defining leukocyte-rich RA (\(r > 4.5\), \(n = 19\)) or leukocyte-poor RA (\(r < 4.5\), \(n = 17\)) samples (Methods and Supplementary Fig. 1d). Whereas leukocyte-rich RA tissues had substantial infiltration of synovial T cells and B cells, leukocyte-poor RA tissues had cellular compositions more similar to those of OA samples (Fig. 2c). Synovial monocyte abundances were similar between RA and OA samples (Fig. 2c).

To test whether our classification indicated inflammation, we assessed tissue histology and assigned each sample a Krenn inflammation score\textsuperscript{17}. Samples that we classified as leukocyte-rich RA had significantly higher Krenn inflammation scores than those of leukocyte-poor RA or OA samples (Fig. 2d). In contrast, synovial lining membrane hyperplasia was not significantly different between leukocyte-rich RA, leukocyte-poor RA, and OA samples (Fig. 2d).

We observed significant correlation between synovial leukocyte infiltration, as measured via flow cytometry and the histological Krenn inflammation score (Fig. 2e). Mass cytometry in 26 synovial tissues was consistent with flow cytometry and histology. OA and leukocyte-poor RA samples were characterized by a high abundance of fibroblasts and endothelial cells; whereas, leukocyte-rich RA tissues were characterized by a high abundance of CD4$^+$ T cells, CD8$^+$ T cells, and B cells (Fig. 2f and Supplementary Fig. 3a).

Single-cell RNA-seq analysis reveals distinct cell subpopulations. Next, we analyzed 5,265 scRNA-seq profiles passing quality control (Methods), including 1,142 B cells, 1,844 fibroblasts, 750 monocytes, and 1,529 T cells. We used canonical variates (from CCA with bulk RNA-seq) to define 18 cell clusters that were independent of donor \((n = 21)\) and technical plate effects \((n = 24)\) (Fig. 3a, Supplementary Figs. 2c and Fig. 4a). In contrast, conventional principal component analysis (PCA)-based clustering led to clusters that were confounded by batch effects (Supplementary Fig. 4b). All of the clusters in the PCA-based clustering, excluding clusters confounded by batch, were identified in CCA-based clustering. Next, we compared expression values between cells in the cluster and all other cells to select cluster marker genes (Methods and Supplementary Table 4). For selected genes, expression values in each cell positioned in a t-Distributed Stochastic Neighbor Embedding (tSNE\textsuperscript{29}) are shown in Fig. 3c–f. Among fibroblasts, we identified four
putative subpopulations (Fig. 3c): CD34⁺ sublining fibroblasts (SC-F1), HLA-DRAhi sublining fibroblasts (SC-F2), DKK3⁺ sublining fibroblasts (SC-F3), and CD55⁺ lining fibroblasts (SC-F4). In monocytes (Fig. 3d), we identified IL1B⁺ pro-inflammatory monocytes (SC-M1), NUPR1⁺ monocytes (SC-M2), C1QA⁺ monocytes (SC-M3), and interferon (IFN) activated monocytes (SC-M4). In T cells (Fig. 3e), we identified three CD4⁺ clusters: CCR7⁺ T cells (SC-T1), FOXP3⁺ regulatory T cells (Treg cells) (SC-T2), and PDCD1⁺ T cells (SC-T3); and three CD8⁺ clusters: GZMK⁺ T cells (SC-T4), GNL1⁺GZMB⁺ cytotoxic lymphocytes (CTLs) (SC-T5), and GZMK⁺GZMB⁺ T cells (SC-T6). Within B cells (Fig. 3f), we identified four cell clusters, including naïve IGHD⁺CD27⁺ (SC-B1) and IGHG3⁺CD27⁺ memory B cells (SC-B2). We identified an auto-immune-associated B cell (ABC) cluster (SC-B3) with high expression of ITGAX (also known as CD11c) and a plasmablast cluster (SC-B4) with high expression of immunoglobulin genes and XBP1, a transcription factor for plasma cell differentiation. We assessed protein fluorescence measurements of typical cell type markers, which were consistent with our identified scRNA-seq clusters (Supplementary Fig. 2e). Cell density quantified from ten histology samples correlated with the lymphocyte flow cytometric cell yields, suggesting that samples with the most single-cell measurements were those with the best yields and the most inflammation (Supplementary Fig. 5).

Distinct synovial fibroblasts defined by cytokine activation and MHC II expression. To identify the fibroblast subpopulations that are overabundant in leukocyte-rich RA synovia, we selected marker genes for each cluster and assessed their expression levels in bulk RNA-seq from sorted fibroblasts (CD45⁻PDNP⁺) of samples from patients with RA and OA. For example, genes associated with HLA-DRAhi (SC-F2) fibroblasts were more highly expressed in bulk RNA-seq from leukocyte-rich RA than in OA samples (t-test P < 1 × 10⁻³ for HLA-DRA, IFI30, and IL6) (Fig. 4a). Because the expression profile of a bulk tissue sample is an aggregate of the expression profiles of its constituent cell populations, this result suggests expansion of HLA-DRAhi (SC-F2) fibroblasts in RA tissues. Genes associated with CD55⁺ fibroblasts (SC-F4) were significantly more highly expressed in bulk RNA-seq from OA than in those from leukocyte-rich RA (t-test P < 1 × 10⁻³ for HBEFG, CLICS, HTRA4, and DNAE1L3) (Fig. 4a). CD55⁺ fibroblasts (SC-F4) were the most transcriptionally distinct subset from the three THY1⁺ clusters (SC-F1-3), including the highest expression of lubricin (PRG4), suggesting that these cells represent synovial lining fibroblasts and THY1⁺ fibroblasts (SC-F1-3) represent sublining (Fig. 4a). Next, we use the averaged expression level of the best marker genes for each scRNA-seq cluster (AUC > 0.7) and tested for differential expression in bulk RNA-seq fibroblast samples from leukocyte-rich RA and OA synovia. The gene averages for HLA-DRAhi sublining...
fibroblasts (SC-F2) and CD34+ sublining fibroblasts (SC-F1) were higher in leukocyte-rich RA compared with those in OA (t test $P = 2 \times 10^{-4}$ and $P = 2 \times 10^{-3}$, respectively), whereas the gene averages for CD55+ lining fibroblasts (SC-F4) were higher in OA than in leukocyte-rich RA (t test $P = 5 \times 10^{-4}$) (Fig. 4b).

Consistent with the role of synovial fibroblasts in matrix remodeling, the sublining fibroblast subsets (SC-F1-3) expressed genes encoding extracellular matrix constituents (Fig. 4c). HLA-DRAhi sublining fibroblasts (SC-F2) expressed genes related to MHC class II presentation and the interferon-$\gamma$-mediated signaling pathway (IFI30) (Fig. 4a,c), suggesting upregulation of MHC class II in response to interferon-$\gamma$ signaling in these cells. We identified a novel sublining fibroblast subtype (SC-F3) that is characterized by high expression of DKK3, CADM1, and COLR2A (Fig. 4a).

To independently confirm the presence of the four fibroblast subpopulations discovered by means of scRNA-seq, we analyzed CD45+PDNP+ cells in mass cytometry data and found eight putative cell clusters with differential protein levels of THY1, HLA-DR, CD34, and cadherin-11 with no obvious batch effects (Fig. 4d–g and Supplementary Fig. 3b). CCA revealed that greater abundance of THY1+CD34+HLA-DRhi fibroblasts measured via mass cytometry assessed by cytometry with histologic inflammation score (n = 44 biologically independent samples). Student’s one-sided t test $P = 3 \times 10^{-3}$, t value = 7.15, df = 46.51. f, tSNE visualization of synovial cell types in OA, leukocyte-poor RA, and leukocyte-rich RA by mass cytometry density plot.
expression analysis of bulk RNA-seq, THY1+CD34+HLA-DRhi cells in the mass cytometry data were found to be overabundant in leukocyte-rich RA relative to leukocyte-poor RA and OA controls (36% versus 2% of fibroblasts, MASC (mixed-effects modeling of associations of single cells) OR = 33.8 (95% CI: 11.7–113.1), one-sided MASC [P = 1.9 × 10−5] (Table 1).

To validate that the protein surface markers from mass cytometry were capturing the same transcriptional populations from scRNA-seq, we isolated fibroblasts from ten synovial tissue samples on the basis of surface protein levels of THY1 and HLA-DR and applied bulk RNA-seq (Supplementary Fig. 6a). We trained a linear discriminant analysis (LDA) classifier on fibroblast scRNA-seq data and used it to determine the most similar scRNA-seq cluster for each bulk RNA-seq sample. The sorted THY1+HLA-DR+ fibroblast population was similar to THY1+HLA-DRAhi (SC-F2), and the THY1+HLA-DR− population was similar to THY1+ (SC-F4) (Supplementary Fig. 7a–d). Genes upregulated in the sorted THY1+HLA-DR+ fibroblasts included IL6 and CXCL12, consistent with the scRNA-seq data.

Activation states define heterogeneity among synovial monocytes. We identified four transcriptionally distinct monocyte subsets in the scRNA-seq data: IL1B+ pro-inflammatory monocytes (SC-M1), NUPR1+ monocytes (SC-M2), C1QA+ monocytes (SC-M3), and IFN-activated SPP1+ monocytes (SC-M4) (Fig. 5a). In bulk RNA-seq monocyte samples from individuals with leukocyte-rich RA and OA, we found that genes associated with IL1B+ monocytes (SC-M1), including NRAA2, HBEGF, PLAUR, and the IFN-activated gene IFITM3 were significantly upregulated in leukocyte-rich RA samples (t test \( P < 1 \times 10^{-3} \)). In contrast, marker genes associated with NUPR1+ monocytes (SC-M2) were downregulated in leukocyte-rich RA relative to OA (Fig. 5a). Next, we took the average of the top marker genes (AUC = 0.7) for each monocyte scRNA-seq subset and tested for differential expression of these averages in the bulk RA versus OA RNA-seq data. This analysis showed that leukocyte-rich RA synovia have a greater abundance of IL1B+ monocytes (t test \( P = 6 \times 10^{-3} \)) and IFN-activated monocytes (t test \( P = 6 \times 10^{-3} \)) than OA, but lower abundance of NUPR1+ monocytes.
monocytes ($t$ test $P = 2 \times 10^{-4}$) (Fig. 5b). These data suggest that cytokine activation drives expansion of unique monocyte populations in active RA synovia.

Using Gene Set Enrichment Analysis (GSEA), we tested MSigDB (the Molecular Signatures Database) immunologic gene sets and found $IL1B$ monocytes (SC-M1) have relatively high expression levels of genes defining the lipopolysaccharide response in monocytes and macrophages (Fig. 5b). This finding suggests that $IL1B$ monocytes (SC-M1) are similar to TLR-activated IL-1-producing pro-inflammatory monocytes. Among Gene Ontology gene sets, we found that $SPP1$ monocytes (SC-M4) express genes induced by type I and type II IFN (Supplementary Fig. 8a), including $IIFTM3$ and $IFI6$ (Fig. 5a). The transcriptional profiles of monocytes in SC-M2 and SC-M3 do not align with known activation states, possibly indicating that these clusters represent cell phenotypes tailored to the unique homeostatic needs of the synovium. Immunofluorescence staining confirmed the presence of CD14$^+$ monocytes in six tissue samples, consistent with an $\beta$-positive cells in six tissue samples, consistent with an $\beta$-positive phenotype in RA synovium (Fig. 5d and Supplementary Fig. 9a,b).

In the mass cytometry data, we identified five CD14$^+$ sublining fibroblast clusters with top ten canonical variates in the low-dimensional CCA space. We computed the spearman correlation between each pair of scRNA-seq cluster and mass cytometry cluster in the CCA space and performed permutation test 10$^4$ times. $Z$ score is calculated on the basis of permutation $P$ value. We observed HLA-DRA$^+$ sublining fibroblasts by scRNA-seq are strongly correlated with THY1$^+$ fibroblasts by mass cytometry.

**Fig. 4** | Distinct synovial fibroblast subsets defined by cytokine activation and MHC II expression. a, scRNA-seq analysis identified three sublining subsets, CD34$^+$ (SC-F1), HLA-DRA$^+$ (SC-F2), and DKK3$^+$ (SC-F3) and one lining subset (SC-F4). Differential analysis between leukocyte-rich RA (n = 16) and OA (n = 12) bulk RNA-seq fibroblast samples shows marker genes upregulated or downregulated in leukocyte-rich RA. Fold changes with 95% confidence interval (CI). b, By querying the leukocyte-rich RA (n = 16) and OA (n = 12) fibroblast bulk RNA-seq samples, scRNA-seq cluster HLA-DRA$^+$ (SC-F2) and CD34$^+$ (SC-F1) fibroblasts are significantly overabundant (two-sided Student’s $t$ test $P = 2 \times 10^{-4}$, $t$ value = 6.2, df = 23.91) and $P = 2 \times 10^{-6}$, $t$ value = 3.20, df = 24.14, respectively) in leukocyte-rich RA relative to OA. Lining fibroblasts (SC-F4) are overabundant (two-sided Student’s $t$ test $P = 5 \times 10^{-2}$, $t$ value = −5.31, df = 21.97) in OA samples. Fold changes with 95% CI. c, Pathway enrichment analysis for each cluster. Two-sided Kolmogorov-Smirnov test with 10$^5$ permutations; Benjamin-Hochberg FDR is shown. d, e, Identified subpopulations from fibroblasts (n = 25,161) and disease status from six leukocyte-rich RA samples, nine leukocyte-poor RA samples, and eight OA samples by mass cytometry on the same gating with scRNA-seq. f, g, Normalized intensity of distinct protein markers shown in tSNE visualization and averaged for each cluster heatmap. h, i, Integration of mass cytometry clusters with scRNA-seq clusters based on the top markers (AUC > 0.7) for each scRNA-seq cluster using top ten canonical variates in the low-dimensional CCA space. We computed the spearman correlation between each pair of scRNA-seq cluster and mass cytometry cluster in the CCA space and performed permutation test 10$^4$ times. $Z$ score is calculated on the basis of permutation $P$ value. We observed HLA-DRA$^+$ sublining fibroblasts by scRNA-seq are strongly correlated with THY1+CD34+HLA-DR$^+$ fibroblasts by mass cytometry.
Table 1 | Connection between cell populations determined by mass cytometry and scRNA-seq clusters and disease associations

| scRNA-seq cluster | Mass cytometry cluster | Leukocyte-poor RA and OA | Leukocyte-rich RA | One-sided MASC P value | Leukocyte-rich OR (CI) |
|-------------------|------------------------|--------------------------|------------------|-----------------------|-----------------------|
| Lining fibroblasts (SC-F4) | THY1+ Cadherin-11+ | 21% | 4% | 1.00 | 0.04 (0.0–0.2) |
| | THY1+ Cadherin-11+ | 18% | 2% | 1.00 | 0.01 (0.0–0.3) |
| | THY1+ CD34+ HLA-DR+ | 7% | 3% | 0.87 | 0.5 (0.3–1.2) |
| | THY1+ CD34+ HLA-DR+ | 17% | 15% | 0.48 | 1.2 (0.3–4.4) |
| HLA-DRA* sublining fibroblasts (SC-F2) | THY1+ CD34+ HLA-DR+ | 2% | 36% | 1.9 × 10^−5 | 33.8 (11.7–113.1) |
| DKK3* sublining fibroblasts (SC-F3) | THY1+ CD34+ HLA-DR+ | 16% | 15% | 0.66 | 0.8 (0.3–1.8) |
| CD34* sublining fibroblasts (SC-F1) | THY1+ CD34+ HLA-DR+ | 18% | 4% | 1.00 | 0.2 (0.1–0.4) |
| NUPRI* (SC-M2) | CD11c− | 30% | 4% | 1.00 | 0.01 (0.0–0.4) |
| IL18* (SC-M1), IFN-activated (SC-M4) | CD11c+ CCR2+ | 34% | 40% | 0.23 | 1.6 (0.7–3.6) |
| | CD11c+ CD38− | 13% | 2% | 1.00 | 0.1 (0.0–0.3) |
| | CD11c+ CD38− CD64+ | 13% | 3% | 0.93 | 0.3 (0.1–1) |
| IL18* (SC-M1), IFN-activated (SC-M4), CIQA+ (SC-M3) | CD11c+ CD38+ | 15% | 51% | 6.7 × 10^−5 | 7.8 (3.6–17.2) |

Bold mass cytometry clusters are significantly enriched in leukocyte-rich RA (one-sided Benjamini–Hochberg FDR q value < 0.05). Two significant digits are given to the one-sided F tests conducted on nested models with MASC. 95% confidence interval (CI) for the odds ratio (OR) is given for each mass cytometry cluster. Where possible, we have identified the most similar scRNA-seq clusters for each cluster found by mass cytometry. The mass cytometry analysis is performed on downsampled datasets of 25,161 fibroblasts from 23 patients, 15,298 monocytes from 26 patients, 19,985 T cells from 26 patients and 8,179 B cells from 23 patients.

To confirm that putative populations from mass cytometry correspond to those identified by scRNA-seq clusters, we sorted CD14+ monocytes from four synovial tissue samples using CD11c and CD38 protein markers and assayed the cells via RNA-seq (Supplementary Fig. 6c). Importantly, we found that CD14+ synovial cells showed higher expression of both CD11c and CD38, particularly in the RA samples. The CD14+CD11c+CD38+ and CD14+CD11c+CD38− sorted cells were consistent with IL1B+ pro-inflammatory (SC-M1) and NUPRI* (SC-M2) cells, respectively (Supplementary Fig. 7e–h). These data, alongside the mass cytometry data, support the findings of greater abundance of IL1B+ pro-inflammatory (SC-M1) monocytes and lower abundance of NUPRI* (SC-M2) monocytes in leukocyte-rich RA samples.
**Fig. 5 | Unique activation states define synovial monocytes heterogeneity.** a, scRNA-seq analysis identified four subsets: IL1B+ pro-inflammatory monocytes (SC-M1), NUPR1+ monocytes (SC-M2) with a mixture of leukocyte-poor RA and OA cells, C1QA+ (SC-M3), and IFN-activated monocytes (SC-M4). Differential analysis by bulk RNA-seq on leukocyte-rich RA samples (n = 17) and OA samples (n = 13) revealed upregulation and downregulation of cluster marker genes. Effect sizes with 95% CI are given. b, By querying the bulk RNA-seq, we found scRNA-seq cluster IL1B+ pro-inflammatory monocytes (two-sided Student’s t test P = 6 × 10−4, t value = 4.56, df = 26.33) and IFN-activated monocytes (two-sided Student’s t test P = 6 × 10−5, t value = 3.28, df = 23.68) are upregulated in leukocyte-rich RA (n = 17) compared to OA (n = 13), while SC-M2 is depleted (two-sided Student’s t test P = 2 × 10−4, t value = −5.62, df = 26.81) in leukocyte-rich RA. Error bars indicate mean and 95% CI. c, Pathway enrichment analysis indicates the potential pathways for each subset. Two-sided Kolmogorov–Smirnov test with 10 5 times permutation was performed; Benjamini–Hochberg was used to control the FDR of multiple tests. The standard names for the immunological gene sets from top to bottom are as follows: genes downregulated in neutrophils versus monocytes (GSE22886); genes downregulated in healthy myeloid cells versus SLE myeloid cells (GSE10325); genes downregulated in control microglia cells versus those 24 h after stimulation with IFNG (GSE1432); genes upregulated in monocytes treated with lipopolysaccharide versus monocytes treated with control IgG (GSE9988); genes upregulated in monocytes versus myeloid dendritic cells (mDC) (GSE29618); genes upregulated in monocytes versus plasmacytoid dendritic cells (pDC) (GSE29618). d, Detection of pro-inflammatory IL-1β in inflamed synovium by multicolor immunofluorescent staining with antibodies to CD14 (red), IL-1β (green), and counterstained with DAPI (blue) identified CD14+IL-1β+ cells (white arrow). The experiment was repeated >5 times with staining of six independent leukocyte-rich RA samples with similar results. Image was acquired at ×200 magnification. Scale bar, 50 μm. e, f, Identified subpopulations from monocytes (n = 15,298) and disease status from six leukocyte-rich RA sample, nine leukocyte-poor RA samples, and 11 OA samples by mass cytometry on the same gating with scRNA-seq. g, h, Normalized intensity of distinct protein markers by tSNE visualization, averaged for each cluster in the heatmap. i, Integration of identified mass cytometry clusters with bulk RNA-seq reveals genes that are associated with CD11c−CD83+ and CD11c−CCR2+, like ITIM3, CD38, HBEGF, ATF3, and HLA-A. j, Integration of mass cytometry clusters and scRNA-seq clusters revealed that CD11c−CD83+ cells by mass cytometry are significantly associated with IL1B+ pro-inflammatory (SC-M1) monocytes.
Fig. 6 | Synovial T cells display heterogeneous CD4+ and CD8+ T cell subpopulations in RA synovium. a, scRNA-seq analysis identified three CD4+ subsets: CCR7+ (SC-T1), Treg cells (SC-T2), and Treg (SC-T3); and three CD8+ subsets: GZMK+ (SC-T4), CTLs (SC-T5), and GZMK* GZMB* (SC-T6). Differential expression analysis on leukocyte-rich RA samples (n = 18) compared with OA samples (n = 13) on sorted T cell bulk RNA-seq samples revealed that CXCL13 is most significantly enriched in leukocyte-rich RA compared with OA. Effect sizes with 95% CIs are given. b, Disease association of scRNA-seq clusters by aggregating top markers (AUC = 0.7) for each scRNA-seq cluster in the top ten canonical variates. c, Pathway analysis based on immunologic gene set enrichment indicates the potential enriched T cell state pathways. Two-sided Kolmogorov–Smirnov test with 10^5 permutation was performed; Benjamini–Hochberg procedure was used to control the FDR of multiple tests. The brief description of the standard names from top to bottom are: genes upregulated in CD4hi cells from thymus: Treg versus conventional T (GSE42021); genes upregulated in central memory T cells from peripheral blood mononuclear cells (PBMCs) (GSE11057); genes upregulated in comparison of effective memory CD4+ T cells versus T1 cells (GSE3982); genes upregulated in comparison of Treg cells versus T17 cells (GSE11924). d, Detection of CD3+CD8+IFNγ+ (white arrow) in inflamed RA synovium via multicolor immunofluorescent staining with antibodies to CD3 (green), CD8 (red), IFNγ (white), and counterstained with DAPI (blue). The experiment was repeated > 5 times with staining of six independent leukocyte-rich RA samples with similar results. Image was acquired at x200 magnification. Scale bar, 50 μm. e,f, Identified subpopulations from T cells (n = 19,985) and disease status from six leukocyte-rich RA samples, nine leukocyte-poor RA samples, and 11 OA samples with mass cytometry. g,h, Distinct patterns of protein markers with tSNE and heat map that define these clusters. i, Integration of identified mass cytometry clusters with bulk RNA-seq using CCA reveals bulk genes that are associated with CD4+PD-1+ICOS+ and CD8+PD-1-HLA-DR+ by mass cytometry. j, Integration of mass cytometry clusters with scRNA-seq clusters on the top markers (AUC > 0.7) for each scRNA-seq cluster in the top ten canonical variates. Z score based on permutation test reveals that CD4+PD-1+ICOS+ and CD8+PD-1-HLA-DR+ by mass cytometry are highly associated with Treg, and Treg (SC-T3) by scRNA-seq; CD8+PD-1-HLA-DR+ T cells by mass cytometry are highly associated with CD8+ T cells (SC-T5 and SC-T6).

(SC-T6) also expressed HLA-DPA1 and HLA-DRB1 and other genes suggestive of an effector phenotype (Fig. 6a,c).

To confirm these findings, we applied intracellular staining to tissues from RA samples and RNA-seq to sorted CD8+ T cells.
express both GZMB and GZMK by intracellular protein staining (Supplementary Fig. 10b). In a comparison of seven synovial tissue samples, CD8+ T cells had a higher proportion of IFNγ+ cells than CD4+ T cells from the same sample (Supplementary Fig. 10c,d). We also applied immunofluorescence to six synovial tissue samples and found that IFNγ+CD3+CD8+ T cells were higher in frequency in RA than OA (Fig. 6d and Supplementary Fig. 9c,d). Overall, these results closely mirror the findings from the scRNA-seq clusters.

Using mass cytometry, we identified nine putative T cell clusters among the synovial T cells (CD45+CD14+CD3+)
(Fig. 6e–h and Supplementary Fig. 3d). By integrating bulk RNA-seq with mass cytometry cluster abundances, we found that higher gene expression of CXCL13 and inhibitory receptors TIGIT and CTLA4 was associated with a greater abundance of the CD4+PD-1+ICOS+ mass cytometry cluster. Greater abundance of CD8+PD-1+HLA-DR+ cells was associated with greater expression of IFNγ (Fig. 6i). We found correspondence between T
and T
clusters (SC-T3) and CD4+PD-1+ICOS- T cells (z score = 3.4). CD8+ subsets including GZMK+GZMB+ cells (SC-T6) and CTLs (SC-T5) tracked with CD8+PD-1+HLA-DR+ T cells by means of mass cytometry (Fig. 6f) and Table 1. Additionally, CD4+PD-1+ICOS+ cells were significantly overabundant in leukocyte-rich RA (MASC OR = 3 (95% CI: 1.7–5.2), one-sided MASC P = 2.7 × 10^-4) (Table 1).

**Autoimmune-associated B cells expanded in RA synovium by single-cell RNA-seq.** We identified four synovial B cell clusters with scRNA-seq: naive B cells (SC-B1), memory B cells (SC-B2), ITGAX+ ABC cells (SC-B3), and plasmablasts (SC-B4) (Fig. 7a). Findings from GSEA with Gene Ontology pathways suggested that SC-B1, SC-B2, and SC-B3 clusters represent activated B cells (Supplementary Fig. 8b). GSEA with MsigDB immunological gene sets revealed that SC-B1 cells expressed naive B cell genes, whereas SC-B2 and SC-B3 cells expressed IgM and IgG memory B cell genes (Fig. 7b). SC-B3 cells expressed high levels of ITGAX and TBF21 (T-bet), which are markers of autoimmunity-associated B cells (Fig. 7c and Table 7a), as well as markers of recently activated B cells, including ACTB. High expression of AICDA is consistent with the recently reported transcriptomic analysis of CD11c+ B cells from peripheral blood of systemic lupus erythematosus (SLE) samples.

Interferon-stimulated genes (GBP1 and ISG15) were also expressed in ABCs (SC-B3) and upregulated in leukocyte-rich RA (Fig. 7a). Although ABCs (SC-B3) constitute a relatively small proportion of all B cells, they were almost exclusively derived from two patients with leukocyte-rich-RA (Fig. 3b). To confirm the presence of ABCs in human tissues, we applied immunofluorescence staining to six synovial tissue samples. RA synovium had increased numbers of CD20+T-bet+CD11c+ B cells compared with OA synovium. Specifically, we observed ABC cells in tissue sections from the same inflamed tissue samples that had a high proportion of ABCs by scRNA-seq analysis (Fig. 7c and Supplementary Fig. 9e,f).

We identified ten putative B cell clusters in the mass cytometry data (CD45+CD3-CD14+CD19+) (Fig. 7d–g and Supplementary Fig. 3e). CCA analysis showed that samples with higher gene expression of CD38, MZB1, and plasma cell differentiation factor XBP1 had greater abundance of CD38+CD20+IgM+IgD- plasmablasts (Fig. 7h). Plasmablasts (SC-B4) corresponded with CD38+CD20+IgM+IgD- B cells (z score = 2.7) (Fig. 7i and Table 1). ABCs (SC-B3) corresponded with the IgM+IgD+HLA-DR++ CD20+ CD11c+ mass cytometry cluster (z score = 1.6), which is significantly overabundant in leukocyte-rich RA (OR = 5.7 (95% CI: 1.8–22.3), one-sided MASC P = 2.7 × 10^-3) (Fig. 7i and Table 1). Mass cytometry analysis further identified three putative subsets within CD11c+ cells: IgM+IgD+HLA-DR++CD20+CD11c+, CD38+ HLA-DR++CD20+CD11c+, and IgM+IgD+CD11c+. This finding is suggestive of additional heterogeneity within ABCs.

To demonstrate that CD19+CD11c+ cells by surface protein markers correspond to SC-B3 (ABCs), we flow sorted CD19+CD11c+ cells from an independent cohort of six RA synovial samples and applied RNA-seq (Supplementary Fig. 6b). We show that these RNA-seq profiles are most consistent with that of ABC cells (Supplementary Fig. 7i–k). In these sorted samples, we found more putative marker genes (for example, ZEB2 and CHIA) and interferon-induced genes (IFITM3 and IFI27) for the ABC population (Supplementary Fig. 7i).

**Inflammatory pathways and effector modules revealed by global single-cell profiling.** We used bulk and single-cell transcriptomics of sorted synovial cells to examine pathologic molecular signaling pathways. First, PCA on post–quality control OA and RA bulk RNA-seq samples (Supplementary Fig. 11a,b) showed that cell type accounted for most of the data variance. Each cell type expressed specific marker genes: PDGFRα for fibroblasts, C1QA for monocytes, CD3D for T cells, and CD19 for B cells (Supplementary Fig. 11c). Within each cell type, PCA showed that leukocyte-rich RA samples separated from OA and leukocyte-poor RA samples (Supplementary Fig. 11d–g). Differential gene expression analysis between leukocyte-rich RA and OA (FC > 2 and FDR < 0.01) revealed genes upregulated in leukocyte-rich RA tissues: 173 in fibroblasts, 159 in monocytes, ten in T cells, and five in B cells. To define the pathways relevant to leukocyte-rich RA, we used GSEA weighted by gene effect sizes on Gene Ontology pathways and identified type I interferon response and inflammatory response (monocytes and fibroblasts) (Supplementary Fig. 11h,i), Fc receptor signaling (monocytes), NF-kappa B signaling (fibroblasts), and interferon gamma (T cells) (Fig. 8a).

Leukocyte-rich RA samples had significantly higher expression of genes in fibroblasts and monocytes: inflammatory response genes (PTGS2, PTGER3, and ICA1M), interferon response genes (IFIT2, RASAD2, STAT1, and XAF1), and chemokine or cytokine genes (CCL2 and CXCL9) (Fig. 8b), consistent with a coordinated chemotactic response to interferon activation. T cells had upregulation of interferon regulatory factors (IRFs), including IRF7 and IRF9, and monocytes had upregulation of IRF7, IRF8, and IRF9. Taken together, results from the pathway analysis suggests cross-talk between immune and stromal cells in leukocyte-rich RA synovia. Inflammatory response genes upregulated in leukocyte-rich RA had comparable expression levels between leukocyte-poor RA and OA synovial cells (Fig. 8b).

Next, we asked whether inflammatory cytokines upregulated in leukocyte-rich RA are driven by global upregulation within a single synovial cell type or specific upregulation within a discrete cell subset defined by scRNA-seq. Whereas TNF was produced at a high level by multiple monocyte, B cell and T cell populations; IL6 expression was restricted to HLA-DRA++ sublining fibroblasts (SC-F2) and a subset of B cells (SC-B1) (Fig. 8c); CD8+ T cells, rather than CD4+ T cells, were the dominant source of IFNγ transcription in leukocyte-rich synovia.

We also observed cell subset–specific responses to inflammatory pathways. Toll-like receptor signaling pathway was enriched in B cells and monocytes in leukocyte-rich RA tissues (Fig. 8a). At the single-cell level, TLR10 was only expressed by activated B cells (Fig. 8c), indicating that TLR10 has a functional role within the B cell lineage. In contrast, TLR8 was elevated in all RA monocyte subsets. The hematopoietic cell–specific transcription factor IRF8 was expressed in a significant fraction of monocytes and B cells that cooperatively regulate differentiation of monocytes and activated B cells in RA synovium. SLAMF7 is highly expressed by pro-inflammatory monocytes (SC-M1), IFN-activated monocytes (SC-M4), CD8+ T cells, and plasmablasts (SC-B4). Furthermore, mass cytometry analysis across all identified cell clusters revealed that patients with leukocyte-rich RA showed...
**Fig. 7 | Synovial B cells display heterogeneous subpopulations in RA synovium.** a, scRNA-seq analysis identified naïve B cells (SC-B1), memory B cells (SC-B2), autoimmune-associated B cells (ABCs) (SC-B3), and plasmablasts (SC-B4). Differential expression analysis is given by comparing leukocyte-rich RA (n = 16) with OA (n = 7) using bulk RNA-seq B cell samples. Effect sizes with 95% CI are given, b, Pathway enrichment analysis using immunologic gene sets indicates the distinct enriched pathways for each scRNA-seq cluster. Two-sided Kolmogorov-Smirnov test with 1000 times permutation was performed; Benjamin-Hochberg was used to control the FDR of multiple tests. The standard names for the immunologic gene sets from top to bottom are as follows: genes up-regulated in plasma cells versus memory B cells (GSE12366); genes up-regulated in B cells versus plasmacytoid dendritic cells (pDC) (GSE29618); genes up-regulated in B lymphocytes: naïve versus plasmablasts (GSE42724); genes upregulated in B lymphocytes: human germinal center light zone versus dark zone (GSE38697); genes up-regulated in comparison of memory IgM B cells versus plasma cells from bone marrow and blood (GSE22886); genes up-regulated in comparison of memory IGG and IGA B cells versus plasma cells from bone marrow and blood (GSE22886). c, Detection of CD20+T-bet+CD11c+ (white arrow) in inflamed synovium by multicolor immunofluorescence. Immunofluorescent staining with antibodies CD20 (red), CD11c (white), T-bet (green), and counterstained with DAPI (blue). The experiment was repeated >5 times with staining of 6 independent leukocyte-rich RA samples with similar results. Image was acquired at 200 magnification. Scale bar is 50 μm. d-e, Identified subpopulations of B cells (n=8) 179) and disease status from 6 leukocyte-poor RA, 9 leukocyte-poor RA, and 8 OA by mass cytometry. f-g, Distinct expression patterns of protein markers by tSNE and averaged for each cluster in heatmap. h, Integrating mass cytometry clusters with bulk RNA-seq data using CCA shows that CD38+CD20+IgM+HLA-DR+ (SC-B4) population is highly associated with gene expression of plasma cells makers, like XBP1. i, Integration of mass cytometry clusters with scRNA-seq clusters suggested that CD38+CD20+IgM+HLA-DR+ and CD38+CD20+IgM–IgD– are significantly associated with plasmablast (SC-B4); IgM–IgD–HLA-DR++CD20+CD11c+ B cells are associated with ABCs (SC-B3).

**Discussion**

Using multi-model, high-dimensional synovial tissue data, we defined stromal and immune cell populations overabundant in RA and described their transcriptional contributions to essential inflammatory pathways. Recognizing the considerable variation in disease duration and activity, treatment types, and joint histology scores, we elected to use a molecular parameter, based on percent leukocytes of the total cellularity, to classify our samples at the local tissue level. We note that differences in leukocyte enrichment of joint replacement samples and biopsy samples were best explained by leukocyte infiltration and not by the histological scores (Supplementary Figs. 1 and 11d–g).
This study and a previous study have highlighted sublining fibroblasts as a potential therapeutic target in RA. Sublining fibroblasts are a major source of pro-inflammatory cytokines such as IL6 (Fig. 4), and a specific subset of sublining fibroblasts expressing MHC II (SC-F2, THY1+CD34+HLA-DRhi) was >15-fold expanded in RA tissues. Further studies are needed to define molecular mechanisms that regulate sublining fibroblast expansion in RA. T cells, B cells, and monocyte proportions track with expression of individual fibroblast genes (Supplementary Fig. 11). We found DNASE1L3, a gene whose loss of function is associated with RA and systemic lupus erythematosus to be highly expressed in CD55+ lining fibroblasts (SC-F4) (Fig. 4a). We identified a novel fibroblast subset (SC-F3) with high expression of DKK3 (Fig. 4), encoding Dickkopf3, a protein upregulated in OA that prevents cartilage degradation in vitro.

Transcriptional heterogeneity in the synovial monocytes indicated that distinct RA-enriched subsets are driven by inflammatory cytokines and interferons (Fig. 5). This suggests monocytes may be...
differentially polarized by unique cytokine combinations in local microenvironments. These newly identified inflammatory phenotypes align with RA therapeutic targets, including anti-TNF therapies and interferon pathway JAK kinase inhibitors. The NUPR1+ (SC-M2) monocytes were inversely correlated with tissue inflammation, and expressed high levels of monocyte tissue remodeling factors such as MERTK and cathepsin K (CTSK) may indicate a subset of osteoclast progenitors that control bone remodeling. Furthermore, spatial studies—particularly focused on lining versus sublining, perivascular and lymphocyte-associate monocytes—will help understand the functional roles of these subsets.

Single-cell classification of T cell subsets in RA synovium demonstrated CD4+ T cell heterogeneity that is consistent with distinction between the homing capacity and effector functions of these subsets. Consistent with previous studies, we observed expansion of PDCD1+CD4+ Treg cells (SC-T3) within leukocyte-rich RA. We also found CD8+ T cell subsets (SC-T4-6) characterized by a distinct granzyme expression pattern (Fig. 6a). A larger study may be better powered to differentiate the relative expansion of individual subpopulations.

This study is the first to report the presence of autoimmune-associated B cells (SC-B3) by transcriptomic sequencing in human leukocyte-rich synovial RA and, in fact, in any human autoimmune target tissue. This B cell population was first reported in aging mice and subsequently seen in autoimmune mice and peripheral blood of patients with SLE. We observed a heterogeneity of CD11c+ B cells detectable in both IgD+ and switched B cell populations by mass cytometry. The gene expression of other ABCs markers suggests a balance between germinal center (IRF8 and AID) and plasma cell (SLAMF7) differentiation within the RA synovium. We have few B cells from OA synovia (Fig. 2b), which limited our ability to identify RA-associated B cell subsets through case-control comparisons (Fig. 7g).

A critical unmet need in RA is identifying therapeutic targets for patients failing to respond to disease-modifying antirheumatic drugs. We observed upregulation of chemokines (CXCL8, CXCL13, and CXCL14), cytokines (IFNγ and IL15, refs. 42,43), and surface receptors (PDGFRB and SLAMF7) in distinct immune and stromal cell populations, suggesting potential novel targets. This study was enabled by advances in the statistical integration of single-cell data and our recent work optimizing robust methodologies for disaggregation of synovial tissue.

We developed advanced strategies to integrate multiple molecular datasets by modulating technical artifact from single-cell technologies while emphasizing biological signals. CCA has enabled cell populations and discovery of novel cell states, such as the CD8+ T cell states noted here. As an ongoing AMP phase 2 study, we are examining larger numbers of ungated cell populations from ~100 synovial tissue patients with RA by capturing mRNA and protein expression simultaneously with detailed clinical data and ultrasound score evaluation of synovitis. We anticipate that this larger study will enable us not only to discover additional subpopulations, but to better define their link to clinical subphenotypes.

It is essential to interrogate the tissue infiltration of diseases other than RA, including SLE, type I diabetes, psoriasis, multiple sclerosis, and other organ-targeting conditions. Application of multiple single-cell technologies together can help define key novel populations, thereby providing new insights about etiology and potential therapies.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41590-019-0378-1.

Received: 19 June 2018; Accepted: 18 March 2019;
Published online: 6 May 2019

References
1. Gibofsky, A. Epidemiology, pathophysiology, and diagnosis of rheumatoid arthritis: A Synopsis. Am. J. Manag. Care 20, S128–S135 (2014).
2. McInnes, I. B. & Schett, G. The pathogenesis of rheumatoid arthritis. Nat. Rev. Rheumatol. 13, 463–475 (2017).
3. Fernelius, C. & Sklar, J. The mortality of rheumatoid arthritis. Arthritis Rheum. 37, 481–494 (1994).
4. Namekawa, T., Wagner, U. G., Goronzy, J. J. & Weyand, C. M. Functional subsets of CD4+ T cells in rheumatoid synovitis. Arthritis Rheum. 41, 2108–2116 (1998).
5. Aziz, M. A. & Fox, D. A. T cell subsets and their role in the pathogenesis of rheumatic disease. Curr. Opin. Rheumatol. 26, 204–210 (2014).
6. Reponar-Schuijt, C. C. et al. Secretion of anti-citrulline-containing peptide antibody by lymphocytes in rheumatoid arthritis. Arthritis Rheum. 44, 41–47 (2001).
7. Meller-Ladner, U. et al. Sympathetic hypactivity affects the expression of the major histocompatibility complex in rheumatoid arthritis. Arthritis Rheum. 43, 1607–1615 (1996).
8. Kinne, R. W., Brauer, R., Stuhlmüller, B., Palombo-Kinne, E. & Burmester, G. R. Macrophages in rheumatoid arthritis. Arthritis Res. 2, 189–202 (2000).
9. Meller-Ladner, U. et al. Sympathetic hypactivity affects the expression of the major histocompatibility complex in rheumatoid arthritis. Arthritis Rheum. 49, 115–124 (1996).
10. Aziz, M. A. & Fox, D. A. T cell subsets and their role in the pathogenesis of rheumatic disease. Curr. Opin. Rheumatol. 26, 204–210 (2014).
11. Meller-Ladner, U. et al. Sympathetic hypactivity affects the expression of the major histocompatibility complex in rheumatoid arthritis. Arthritis Rheum. 43, 1607–1615 (1996).
12. Pap, T., Müller-Ladner, U., Gay, R. E. & Gay, S. Fibroblast biology. Role of fibroblasts in the pathogenesis of rheumatoid arthritis. Nature Rev. Rheumatol. 10, R90 (2014).
13. Orman, D. E. et al. Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. Arthritis Rheumatol. 70, 690–701 (2018).
14. Lindberg, J. et al. Variability in synovial inflammation in rheumatoid arthritis investigated by microarray technology. Arthritis Res. Ther. 8, R47 (2006).
15. Stephenson, W. et al. Single-cell RNA-seq of rheumatoid arthritis synovial tissue using low-cost microfluidic instrumentation. Nat. Commun. 9, 791 (2018).
16. Papaioxi, E. & Satija, R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat. Rev. Immunol. 1, 35–45 (2018).
17. Meier, M. et al. Estimation of immune cell content in tumour tissue using single-cell RNA-seq data. Nat. Commun. 8, 2032 (2017).
18. Rao, D. A. et al. Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. Nature 542, 110–114 (2017).
19. Fonseka, C. Y. et al. Mixed-effects association of single cells identifies an expanded effector CD4+ T cell subset in rheumatoid arthritis. Sci. Transl. Med. 10, eaag305 (2018).
20. Villani, A. C. et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science 356, eaah4573 (2017).
Nature Immunology

Acknowledgements

This work was supported by the Accelerating Medicines Partnership (AMP) in Rheumatoid Arthritis and Lupus Network. AMP is a public-private partnership (AbbVie Inc., Arthritis Foundation, Bristol-Myers Squibb Company, 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, 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 (U10-AR067676, U10-AR067677, U10-AR067678, U10-AR067681, U10-AR067685, U10-AR067688, U10-AR067689, U10-AR067691, U10-AR067694, and U10-AR067678). K.S. is supported by the Ruth L. Kirschstein National Research Service Award (NIAMS F31AR070582). K.W. is supported by a Rheumatology Research Foundation Scientist Development Award, and KL2/Catalyst Medical Research Investigator Training award (an appointed KL2 award) from Harvard Catalyst, The Harvard Clinical and Translational Science Center (National Center for Advancing Translational Sciences, National Institutes of Health Award KL2 TR002542). D.A.R. is supported by NIAMS K08 AR072791-01. L.T.D. is supported by NIAMS K01 AR066063. J.H.A. is supported by R21 AR071670, and the Bertha and Louis Weinstein research fund. S.R. is supported by NIAMS R01 AR063759-01A1; NIAID U19 AI111224, NIH/GR1 U01 HG009379, and Doris Duke Charitable Foundation Grant #203897. A.H.J. is supported by an Arthritis National Research Foundation Grant. A.E., C.D.B. and J.D.T. were supported by the Arthritis Research UK Rheumatoid Arthritis (#20298), and by the National Institute for Health Research (NIHR)’s Birmingham Biomedical Research Centre programme, supported by the National Institute for Health Research/Wellcome Trust Clinical Research Facility at University Hospitals Birmingham NHS Foundation Trust.

Author contributions

S.K., S.M.G., D.T., L.B.H., K.S.-E., A.M.M., D.L.B., J.H.A., V.P.B., C.P., P., G.S.E., L.M., P.K.G., W.A., and J.L.T.D. were recruited patients and obtained synovial tissues. B.F.B., E.D. and E.M.G. performed histological assessment of tissues. K.W., D.A.R., L.T.D. and M.B.B. designed and implemented tissue processing and the cell sorting pipeline. J.A.L. obtained mass cytometry data from samples. E.Z., K.S., C.Y.E., D.I.L. and S.R. conducted computational and statistical analysis. A.H.J., J.R.-M., N.M., and C.R., designed and performed validation experiments. K.S., T.Z., and J.R.M. implemented the sorting pipeline. J.A.L. obtained mass cytometry data from samples. N.H., C.N., and T.M.F.E. obtained single-cell RNA-seq data from samples. E.Z., K.S., C.Y.E., D.I.L. and S.R. conducted computational and statistical analysis. A.H.J., J.R.-M., N.M., and C.R., designed and performed validation experiments. K.S., T.Z., and J.R.M. implemented the website. J.A., S.L.B., C.D.B., J.H.B., J.M.G., M.G.-A., L.B.I., E.A.J., J.A.J., Y.C.L., M.J.M., M.A.M., E.M., J.P.N., A.N., D.E.O., M.R.-P., C.R., W.H.R., A.S., D.S., J.S., J.D.T., and P.J.U. contributed to the procurement and processing of samples, design of the AMP study. S.R., M.B.B., J.H.A., and L.T.D. supervised the research. E.Z., K.W., and C.S. performed the initial draft. K.S., C.Y.E. and M.B.B. edited the draft, and all the authors participated in writing the final manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-019-0578-1.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to S.R.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2019
Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium

Jennifer Albrecht, S. Louis Bridges Jr., Christopher D. Buckley, Jane H. Buckner, James Dolan, Joel M. Guthridge, Maria Gutierrez-Arcelus, Lionel B. Ivashkiv, Eddie A. James, Judith A. James, Josh Keegan, Yvonne C. Lee, Mandy J. McGeachy, Michael A. McNamara, Joseph R. Mears, Fumitaka Mizoguchi, Jennifer P. Nguyen, Akiko Noma, Dana E. Orange, Mina Rohani-Pichavant, Christopher Ritchlin, William H. Robinson, Anupamaa Seshadri, Danielle Sutherby, Jennifer Seifert, Jason D. Turner, and Paul J. Utz.

29Translational Research Program, Benaroya Research Institute at Virginia Mason, Seattle, WA, USA. 30Department of Arthritis & Clinical Immunology, Oklahoma Medical Research Foundation, Oklahoma City, OK, USA. 31Graduate Program in Immunology and Microbial Pathogenesis, Weill Cornell Graduate School of Medical Sciences, New York, NY, USA. 32David Z. Rosensweig Genomics Research Center, Hospital for Special Surgery, New York, NY, USA. 33Department of Rheumatology, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan. 34The Rockefeller University, New York, NY, USA. 35Division of Immunology and Rheumatology, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. 36The Institute for Immunity, Transplantation, and Infection, Stanford University School of Medicine, Stanford, CA, USA.
Multiclor immunofluorescent staining of paraffin synovial tissue. Briefly, 5mm-thick formalin-fixed paraffin sections were incubated in a 60 °C oven to melt then paraffin. Slides were quickly transferred to xylene to completely dissolve the paraffin, and after 5 min, were transferred to absolute ethanol. Slides were left in absolute ethanol for 5 min and then transferred to 95% ethanol. At the end of the 5min immersion in 95% ethanol, slides were rinsed several times with distilled water and transfer to a plastic cup folllow filled with 1x DAKO retrieval solution (S1699, DakoCytomation). Antigens were unmasked by immersing a plastic cup in boiling water for 30 min. Slides were cooled to 10 min at room temperature and washed several times with distilled water. Nonspecific binding was blocked with 5% normal donkey serum (017-000-121, Jackson ImmunoResearch Laboratories) dissolved in PBS containing 0.1% Tween 20 and 0.1% Triton X-100. Without washing, blocking solution was removed from slides, and combinations of primary antibodies were added to PBS containing 0.1% Tween 20 and 0.1% Triton X-100. Primary antibodies to detect IFNγ (MAB1279, R&D Systems), mouse anti-human CD3 (clone 144B, GeneTex), and rabbit anti-human IFNγ (Biorbyt, orb241082). To visualize ABCs, we incubated slides with goat anti-human CD20 (LifeSpan Biosciences, LS-B1144), rabbit anti-Tbet (H-210, Santa Cruz Biotechnology) and biotinylated mouse anti-human CD11c (clone 118/A5, Thermo Fisher Scientific). To identify IIB+ monocytes, we used a mixture of goat anti-human CD14 (119–13402, RayBiotech), biotinylated rabbit anti-human IILb (OABF00305-Biotin, Aviva Systems Biology), and mouse anti-human CD16 (clone D1J30c, LifeSpan Biosciences). Finally, slides were probed with biotinylated monoclonal rabbit CD68 (M7649, Dako), rat anti-human HLADR (cloneYE2/36 HLK, LifeSpan Biosciences), and monoclonal mouse CD45 (clone F10–89–4, abcam) to detect fibroblasts, class II-expressing cells, and hematopoietic cells, respectively. Slides with primary antibodies were incubated in a humid chamber at room temperature overnight. The following morning, primary antibodies for triple T cell stain and for detecting ABCs were revealed with Alexa Fluor 568 donkey anti-goat IgG (A-11057, Thermo Fisher Scientific), Alexa Fluor 488 donkey anti- rabbit (771-546-152, Jackson ImmunoResearch Laboratories), and Alexa Fluor 647 donkey anti-mouse (710-606-151, Jackson ImmunoResearch Laboratories). Primary antibodies in the stain for monocytes were revealed with Alexa Fluor 568 donkey anti-goat IgG, Alexa Fluor 488 streptavidin (S1223, Thermo Fisher Scientific), and Alexa Fluor 647 donkey anti-mouse IgG. Antibodies in the stain for fibroblasts and hematopoietic cells were detected with Cy3 donkey anti-rabbit (711-166-152, Jackson ImmunoResearch Laboratories), Alexa Fluor 488 donkey anti-rat IgG (A-21208, Thermo Fisher Scientific), and Alexa Fluor 647 donkey anti-mouse IgG. After 2 h of incubation, slides were washed and mounted with Vectashield mounting media with DAPI (H-1200, Vector Laboratories). Pictures were taken with an Axiosplan Zeiss microscope and recorded with a Hamamatsu camera. Double immunofluorescence images were obtained by merging individual channels in NIH Image J software.

Estimation of number of cells by counting nuclei. To estimate number of cells, we counted the number of nuclei in five random ×200 magnification fields that showed synovial lining with Image J NIH software. Briefly, original color TIFF files were first transformed into 8-bit grayscale images. We used similar settings to adjust threshold in 8-bit images (lower threshold level: 0, upper threshold level: 1). Next, we used process:binary:watershed to separate nuclei. In the analyze image icon, we selected 'analyze particles', and we used equal settings to count particles in our images (size [pixel]: 50–infinity, circularity 0.00–1.00, show outlines), and we selected to display results. We visually confirmed that individual nuclei were outlined in the final image, and calculate the average number of cells/×200 field in individual samples.

Tissue sample classification by leukocyte infiltration. We classified RA tissue samples into leukocyte-poor RA and leukocyte-rich RA on the basis of the Mahalanobis distance from OA samples computed on leukocyte abundance measured via flow cytometry. We first took OA samples as a reference and calculated a multivariate normal distribution of the periarticular leukocytes T cells, B cells, monocytes, and eosinophils. To identify the maximum value of all OA samples (4.5) as a threshold to define 19 leukocyte-rich RA (≥4.5) and 17 leukocyte-poor RA (<4.5) samples in our cohort (Supplementary Fig. 1d).

Bulk RNA-seq gene expression quantification. We sorted cells into the major immune and stromal cell populations: T cells, B cells, monocytes, and synovial fibroblasts. We then performed RNA sequencing. Full-length cDNA and sequencing libraries were performed using Illumina Smart-seq2 protocol.
Single-cell RNA-seq quality control. For quality control of single-cell RNA-seq data, we filtered out molecules that are likely to be contamination derived from other cells. We developed a simple algorithm to set a threshold for the minimum number of reads per molecule, and we ran it separately for each quartet of 96 wells in each 384-well plate. We used 2 marker genes expected to be exclusively expressed in each of the four cell types: PDGFRα and ISLR for fibroblasts, CD2 and CD3D for T cells, C1QA and CD79A for B cells, and CD14 and CD1A for monocytes. We counted nonzero expression of these genes in the correct cell type as a true positive and nonzero expression in the incorrect cell type as a false positive. Then we tried each threshold for reads per molecule from 1 to 20 and chose the threshold that maximizes the ratio of true positive to false positive (Supplementary Fig. 14). This left us with 7,127 cells and 32,391 genes. Next, we discarded cells with fewer than 0.5% of common genes detected in that sample. We also discarded cells that had more than 25% of molecules coming from mitochondrial genes. This left us with 5,265 cells. We discarded genes that had nonzero expression in fewer than 10 cells. We show all post-QC single cells based on the number of genes detected and percent of molecules from mitochondrial genes for each identified cluster (Supplementary Fig. 15).

Single-cell RNA-seq gene expression quantification. Single-cell RNA-seq was performed using the Cel-Seq2 method \(^{22}\) with the following modifications. Single cells were sorted into 384-well plates containing 0.6 µl 1% NP-40 buffer in each well. Then, 0.6 µl dNTPs (10 mM each; NEB) and 5 µl of barcoded reverse transcription primer (1 µg/µl) were added to each well along with 20 µl of ERCC spike-in (diluted 1:8,000,000). Reactions were incubated at 65°C for 5 min, and then moved immediately to ice. Reverse transcription reactions were carried out, as previously described (Hashimshony et al., 2016), and cDNA was purified using 0.8X volumes of Agencourt RNAClean XP beads (Beckman Coulter). In vitro transcription reactions (IVT) were performed, as described following Scheme E. Amplified RNA (arRNA) was fragmented at 80°C for 3 min and purified using Agencourt RNAClean XP beads (Beckman Coulter). The purified arRNA was converted to cDNA using an anchored random primer and Illumina adaptors. cDNAs were analyzed by PCR. The final cDNA library was purified using Agencourt RNAClean XP beads (Beckman Coulter). Paired-end sequencing was performed on the HiSeq 2500 in High Output Run Mode with a 5% PhiX spike-in using 15 bases for Read 1, 6 bases for the Illumina barcode and 36 bases for Read 2. We mapped Read2 to human reference genome hg19 using STAR 2.5.2b, using 15 bases for Read 1, 6 bases for the Illumina barcode and 36 bases for Read 2. We mapped Read2 to human reference genome hg19 using STAR 2.5.2b, using 15 bases for Read 1, 6 bases for the Illumina barcode and 36 bases for Read 2.
following parameters: perplexity = 30 and theta = 0.5. We used all markers except those used to gate each population in the SNE model. To identify high-dimensional populations, we used a modified version of DensVM. We modified the DensVM code to increase the range of potential bandwidths searched during the density estimation step and to return the SNE model generated from the TSNE projection. We summarized the details of the clusters with proportion of cells from each disease cohort in Supplementary Table 3.

**Disease association test of cell populations.** We tested whether abundances of individual populations were altered in RA case samples compared to OA controls using two ways. First, we assessed whether marker genes (AUC > 0.7, 20 < n < 100) characteristic of each scRNA-seq cluster were differentially expressed in the same direction in scRNA-seq and bulk RNA-seq datasets. Second, we applied MASC, a single cell association method for testing whether case-control status influences the membership of single cells in any of multiple cellular subsets while accounting for technical confounds and biological variation. We specified donor identity and batch as random-effect covariates.

**Integration of bulk RNA-seq with mass cytometry.** We used CCA to associate the abundances of mass cytometry clusters with gene expression in bulk RNA-seq. We started by selecting the samples that had both data types. The mass cytometry data matrix has samples and clusters, where the values represent proportions of cells from each sample in each cluster. The bulk RNA-seq data matrix has samples and genes, where the values represent proportions of gene abundance from each sample in each gene. CCA identifies canonical variates (a linear combination of bulk RNA-seq genes and a linear combination of mass cytometry cluster proportions) that maximize correlation of samples along each canonical variate. In other words, it tries to arrange samples from each dataset in a similar order along each canonical variate. We ran CCA separately for fibroblasts, monocytes, T cells, and B cells. For fibroblasts, we associated 2,299 genes with 8 mass cytometry clusters on 22 samples. For monocytes, we associated 2,161 genes with 5 mass cytometry clusters on 25 samples. For T cells, we associated 2,255 genes with 9 mass cytometry clusters on 26 samples. For B cells, we associated 2295 genes with 10 mass cytometry clusters on 17 samples.

**Finding correspondence between scRNA-seq clusters and mass cytometry clusters.**

1. For each cell type, we ran CCA with mass cytometry clusters with bulk RNA-seq. Each gene is correlated with each canonical variate (CV). Also, each mass cytometry cluster is correlated with each CV. By visualizing these correlations, we can see the positions of bulk RNA-seq genes and mass cytometry clusters in the same space (Fig. 4b).
2. We then associated single-cell RNA-seq clusters with mass cytometry clusters by projecting cluster marker genes (AUC > 0.7) for each single-cell RNA-seq cluster in the CCA space acquired from step 1).
3. We took the average across the cluster marker genes for each single-cell RNA-seq cluster for each CV and obtained an “average CV” matrix.
4. Based on the “average CV” matrix, we computed Spearman correlation between the scRNA-seq average CV and the CV for mass cytometry clusters.
5. Next, we generated a null distribution for the Spearman correlations by shuffling the scRNA-seq gene names and then repeating steps 2–4 10,000 times.
6. For the 10,000 replicates of CCA matrix, we repeated step 2 to step 5. Then, we counted how many times the correlation of each pair was greater than the observed value from step 4). Permutation p = 1 + a sum(cor > cor) + 10^6
7. Finally, we converted the permutation p to a z score.

**Differential expression analysis with bulk RNA-seq.** We classified all the samples into OA, leukocyte-poor RA, and leukocyte-rich RA synovial tissues on the basis of the quantitative analysis of T cells, B cells, and monocytes by flow cytometry. PCA on bulk RNA-seq samples showed separation of leukocyte-rich and leukocyte-poor RA on the first or second principal components. For differential analysis, we used the limma R package to identify significantly differentially expressed genes. We used the Benjamini–Hochberg method to estimate false discovery rate (FDR).

**Identification of markers for distinct scRNA-seq clusters.** On the basis of single-cell RNA-seq clusters, we identified cluster marker genes by comparing the cells in one cluster with all other clusters from the same cell type, based on log(CPM + 1). We prioritized cluster marker genes using three criteria: (1) percent of non-zero expressing cells > 60%; (2) are under the receiver-operator curve (AUC) > 0.7; and (3) fold change (FC) > 2.

**Intracellular flow cytometry of synovial tissue T cell stimulation.** Disaggregated synovial tissue cells were incubated with Fixable Viability Dye (eBioscience) and Fc-blocking antibodies (eBioscience), then stained for surface markers in Brilliant Stain Buffer (BD Bioscience). Cells were then fixed and permeabilized using an intracellular staining kit (eBioscience), then subjected to intracellular staining for granzymes or cytokines. Antibodies used in this study include anti-CD45 (clone HI30) from BD Biosciences; anti-CD3 (clone UCHT1), anti-CD8 (clone SK1), anti-CD14 (clone M5E2); anti-CD4 (clone RPA-T4), anti-CD8 (clone M5E2), anti-IFNG (clone 4 S.B3) and anti-TNF (clone MAB11) from eBioscience. Data were collected on a BD Fortessa flow cytometer and analyzed using FlowJo 10.5 software. Disaggregated synovial tissue cells were incubated with a cell stimulation cocktail containing PMA and ionomycin (eBioscience) in RPMI with 10% fetal calf serum (Gibco). After 15 min, brefeldin A (eBioscience) was added. The cells were incubated at 37 °C 5% CO2 for an additional 2 h. The cells were then collected and stained for intracellular cytokines following the protocol described above. Data are shown in Supplementary Fig. 10.

**Statistics.** Results are shown as mean with 95% confidence intervals. The statistics tests used were t test and Kolmogorov–Smirnov test, unless otherwise stated, as described with one-sided or two-sided in the figure legends. Benjamini–Hochberg FDR < 0.01 and fold change > 2 were considered to be statistically significant when appropriate.

**Reporting Summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

**Data availability**

The single-cell RNA-seq data, bulk RNA-seq data, mass cytometry data, and the clinical and histological data for this study are available at dbGaP (https://www.ncbi.nlm.nih.gov/projects/gds/cgi-bin/study.cgi?study_id=phs001457.v1.p1). The source code repository of the computational and statistical analysis is located at https://github.com/immunogenomics/amp_phase1_ra. Data can also be viewed on three different websites at https://immunogenomics.io/ampra, https://immunogenomics.io/cellbrowser/, and https://portals.broadinstitute.org/single_cell/study/amp-phase1.

**References**

51. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171–181 (2014).
52. Finck, R. et al. Normalization of mass cytometry data with bead standards. Cytometry A 83, 483–494 (2013).
53. González, I., Déjean, S., Martin, P. & Baccini, A. CCA: an R package to extend canonical correlation analysis. J. Stat. Software 9, 171–181 (2014).
54. Sing, T., Sandor, O., Beezerwinkel, N. & Lengauer, T. ROCPR: visualizing classifier performance in R. Bioinformatics 21, 3940–3941 (2005).
55. Subramanian, A. et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. 102, 15545–15550 (2005).
56. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25, 29–29 (2000).
57. Libezron, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 1, 417–423 (2015).
58. Reynolds, A. P., Richards, G., De La Iglesia, B. & Rayward-Smith, V. J. Clustering rules: a comparison of partitioning and hierarchical clustering algorithms. J. Math. Model. Algorithms 5, 475–504 (2006).
59. Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J. Comput. Appl. Math. 20, 53–65 (1987).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referrees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a
- Confirmed

☐ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement

☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ The statistical test(s) used AND whether they are one- or two-sided

☐ Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ A description of all covariates tested

☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted

☒ Give P values as exact values whenever suitable.

☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☒ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection We used R version 3.5.1. The source code repository is located at https://github.com/immunogenomics/amp_phase1_ra.

Data analysis We used R version 3.5.1. The source code repository is located at https://github.com/immunogenomics/amp_phase1_ra.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data presented in the manuscript are available through NIH IMMPORT (accession: SDY998 and SDY999) and dbGAP (study accession: phs001457.v1.p1).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf
Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
No sample size was calculated. This study represents a "feasibility" study where single cell analyses were applied to a cohort of patients. Sample size was determined based on the total number of patients recruited during the time period over phase 1 of this study. This study is a proof of principle, to demonstrate that single cell analyses can be applied to samples taken from a large cohort of patients from multiple research sites. Since the goal of the study was to test the feasibility of applying high-dimensional analysis, the total number of patients recruited here was considered sufficient for the sample size.

Data exclusions
Data were excluded from analyses based on specific quality control criteria as described in detail in the manuscript for each data set. For synovial tissues that did not pass standard histologic QC (i.e. lack of identifiable lining structure) were excluded from main pipeline analysis. For single cell data, we discarded cells with fewer than 1,000 genes detected with at least one fragment. We also discarded cells that had more than 25% of molecules coming from mitochondrial genes. For bulk RNA-seq experiments, samples with low quality as determined by gene reads were excluded from subsequent analysis.

Replication
No experimental replication were performed in this study due to the nature of the study design

Randomization
No randomization was performed due to the cross-sectional nature of the study

Blinding
No blinding was performed in this study due to the cross-sectional nature of the study

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems
- n/a Involved in the study
- Antibodies
- Eukaryotic cell lines
- Palaeontology
- Animals and other organisms
- Human research participants
- Clinical data

Methods
- n/a Involved in the study
- ChIP-seq
- Flow cytometry
- MRI-based neuroimaging

Antibodies
Antibodies used for flow cytometry and cell sorting:
- antibody clone vendor catalog number Dilution
  - anti-CD45-FITC (Biolegend, HI30) H130 Biolegend 304006 1:400
  - anti-CD90-PE "5E10" Biolegend 328110 1:500
  - anti-Pdpn-PerCP eF710 NZ:1.3 ebioscience 46-9381-42 1:50
  - anti-CD3-PE-Cy7 UCHT1 Biolegend 300420 1:100
  - anti-CD19-BV421 HIB19 Biolegend 302233 1:20
  - anti-CD34-BV605-A (eBioscience, 4H11) 581 Biolegend 343529 1:400
  - anti-CD14-BV510 M5E2 Biolegend 301842 1:100
  - anti-CD33-BV605-A (eBioscience, 4H11) 581 Biolegend 343529 1:400
  - anti-CD4-BV650 (Biolegend, RPA-T4) RPA-T4 Biolegend 300536 1:50
  - anti-CD8a-BV711A RPA-T8 Biolegend 301044 1:100
  - anti-CD31-AF700 WM59 Biolegend 303134 1:100
  - CD27-APC M-T271 Biolegend 356410 1:100
  - anti-CD235a-APC-750 11E4B-7-6 Beckman Coulter A89314 1:10

Antibodies used for immunofluorescent microscopy studies:
- antibody clone vendor catalog number Dilution
  - mouse anti-human CD8 CB/144B Genetex GTX72053 1:50 [3ug/ml]
  - rabbit anti-human IFNg polyclonal biorbyt orb214082 1:100 [10ug/ml]

Alexa Fluor 568 donkey anti-goat Ig G N/A Thermo Fisher Scientific Cat#A-11075 1:200 [10ug/ml]
Alexa Fluor 488 donkey anti-rabbit N/A Jackson ImmunoResearch Laboratories Cat#711-546-152 1:200 [6ug/ml]

Antibodies used for mass cytometry:
- antibody clone metal dilution
  - CD45 HI30 141Pr 1:100
  - CD19 HIB19 142Nd 1:100
| Antibody Name                  | Binding | Concentration |
|-------------------------------|---------|---------------|
| RANKL MIH24                   | 1:50    |               |
| CD64 10.1                     | 1:100   |               |
| CD16 3G8                      | 1:100   |               |
| CD8a RPA T8                   | 1:100   |               |
| FAP Poly                      | 1:50    |               |
| CD20 2H7                      | 1:100   |               |
| CD45RO UCHL1                  | 1:100   |               |
| CD38 HIT2                     | 1:100   |               |
| CD69 FN50                     | 1:100   |               |
| CD185/CXCR5 J252D4            | 1:100   |               |
| CD4 RPA T4                    | 1:100   |               |
| Podoplanin NC-08               | 1:100   |               |
| CD3 UCHT1                     | 1:100   |               |
| CD146/MCAM SHM-S7             | 1:50    |               |
| IgA 9H9H11                    | 1:100   |               |
| ICOS C398.4A                  | 1:100   |               |
| CD65b G10FS                   | 1:100   |               |
| IgM MHM-B8                    | 1:200   |               |
| CD144/VE-Cadherin BV9         | 1:100   |               |
| HLA-DR L243                   | 1:100   |               |
| IgD IA6-2                     | 1:100   |               |
| CD106/V/CAM-1 STA             | 1:100   |               |
| CD45 HI30                     | 1:100   |               |
| CD64 10.1                     | 1:100   |               |
| RANKL MIH24                   | 1:50    |               |
| CD69 FN50                     | 1:100   |               |
| CD185/CXCR5 J252D4            | 1:100   |               |
| CD4 RPA T4                    | 1:100   |               |
| Podoplanin NC-08               | 1:100   |               |
| CD3 UCHT1                     | 1:100   |               |
| CD146/MCAM SHM-S7             | 1:50    |               |
| IgA 9H9H11                    | 1:100   |               |
| ICOS C398.4A                  | 1:100   |               |
| CD65b G10FS                   | 1:100   |               |
| IgM MHM-B8                    | 1:200   |               |
| CD144/VE-Cadherin BV9         | 1:100   |               |
| HLA-DR L243                   | 1:100   |               |
| IgD IA6-2                     | 1:100   |               |
| CD106/V/CAM-1 STA             | 1:100   |               |

Validation

All commercial antibodies used for flow cytometry and cell sorting experiments were validated for flow cytometric analysis of human cells according to manufacturer’s production information. Additional validation on synovial cells for cell type specificity were performed as described in Donlin and Rao et al., Methods for high-dimensional analysis of cells dissociated from cryopreserved synovial tissue. Arthritis Res. Ther. 20, 139 (2018). For antibodies used in mass cytometry experiments, cell type specificity in synovial cells were tested and described in Donlin and Rao et al. For antibodies used in immunofluorescence microscopy experiments, all antibodies were tested for IF studies on human tissues and cells based on manufacturer’s product information.
Human research participants

Population characteristics

| Clinical characteristics of 51 recruited patients. | OA leukocyte-poor RA leukocyte-rich RA |
|--------------------------------------------------|----------------------------------------|
| n=15                                             | n=17                                   |
| n=19                                             |                                        |
| Demographic variables                            |                                        |
| Mean age, years (Range)                          | 71 (64-81)                             |
| (n=15)                                           | (n=17)                                 |
| Age, mean ±SD (Range)                            | 64.2 ±1.0 (42-79)                      |
| Females, n (%)                                   | 10 (66.7)                              |
| Mean age, years (Range)                          | 57.3 ±1.7 (36-71)                      |
| (n=19)                                           |                                        |
| RF positive, n (%)                               | 8 (47.1)                               |
| Mean years of disease duration, years (Range)    | 15.7 ±5.5*                             |
| CCP positive, n (%)                              | 10 (58.8)                              |
| (n=15)                                           | (n=17)                                 |
| DMARDs Prednisone, n (%)                         | 10 (55.6)                              |
| (n=15)                                           | (n=17)                                 |
| Methotrexate, n (%)                              | 7 (41.2)                               |
| TNF, n (%)                                       | 4 (23.5)                               |
| Rituximab, n (%)                                 | 0 (0)                                  |
| DMARDs = Disease-Modifying Antirheumatic Drugs.  |                                        |
| TNF = TNF inhibitors (infliximab, etanercept, adalimumab, Golimumab). |}

Recruitment

The study was performed in accordance with protocols approved by the institutional review board. A multicenter, cross-sectional study of individuals undergoing elective surgical procedures and a prospective observational study of synovial biopsy specimens from RA patients ≥ age 18, with at least one inflamed joint, recruited from 10 contributing sites in the network. Subjects in the biopsy portion were being asked to undergo a research procedure to obtain synovial tissue.

Ethics oversight

We have been approved by all relevant ethical regulations and the study protocol. Protocols were approved by University of Rochester Medical Center, Hospital for Special Surgery, University of Pittsburgh Medical Center, University of California San Diego, University of Colorado: Denver, Northwestern University, University of Birmingham UK, Queen Mary University of London, University of Alabama Birmingham, University of Massachusetts Medical Center.

Flow Cytometry

Plots

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Synovial T cells, B cells, monocytes, and fibroblasts were isolated from disaggregated synovial tissue. Briefly, disaggregated synovial cells were stained with antibodies against CD45 (HIS30), CD90 (5E10), podoplanin (NZ1.3), CD3 (UCHT1), CD19 (HIB19), CD14 (M5E2), CD34 (4H11), CD4 (RPA-T4), CD8 (SK1), CD31 (WM59), CD27 (M-T271), CD235a (KC16), using human TruStain FcX in 1% BSA in Hepes-Buffered Saline (HBS, 20 mM HEPES, 137 mM NaCl, 3mM Kcl, 1mM CaCl2) for 30 minutes. For validation experiments, RA and OA synovial tissue were disaggregated and synovial cells were stained with cell-type specific antibody panels. For each cell subset, up to 1000 cells were collected directly into buffer TCL (Qiagen). Antibody panels used to define cell subsets are fibroblasts: CD90 (5E10), podoplanin (NZ1.3), HLA-DR (G46-6); B cell subsets: HLA-DR (G46-6), CD11c (3.9), CD19 (SJ25C1), CD27 (M-T271), IgD (IA6-2), CD3 (UCHT1), CD14 (M5E2), CD38 (HIT2); Monocyte subsets: CD14-BV421 (M5E2), CD38-APC (HB-7), and CD11c-PECy7 (B-ly6). Immediately prior to sorting, DAPI or LIVE/DEAD viability dye was added to cell suspensions and cells were passed through a 100μm filter.

Instrument

T cells (CD45+, CD3+, CD14-), monocytes (CD45+, CD3-, CD14+), B cells (CD45+, CD3-, CD14-), and synovial fibroblasts (CD45-, CD31-, PDPN+) were collected by fluorescence-activated cell sorting (BD FACSAria Fusion).

Software

Flowjo (version 10) was used for analysis.

Cell population abundance

95% purity were achieved during sorting of synovial cells based on flow cytometry analysis during single cell sorting (second sort).
Gating strategy

Synovial cells were gated based on the following schemes: T cells (CD45+, CD3+, CD14-), monocytes (CD45+, CD3-, CD14+), B cells (CD45+, CD3-, CD14-, CD19+), and synovial fibroblasts (CD45-, CD31-, PDPN+)