Defining Inflammatory Cell States in Rheumatoid Arthritis Joint Synovial Tissues by Integrating Single-cell Transcriptomics and Mass Cytometry

Methods

Study design and patient 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. Synovial tissues were obtained from joint replacement procedures or ultrasound-guided biopsies, followed by cryopreservation in cryopreservation media Cryostor CS10 (Sigma-Aldrich) and transit to a central technology site[24].

Histological assessment of synovial tissue and quality control

Synovial tissue quality and grading of synovitis were evaluated in formalin-fixed, paraffin-embedded sections by histologic analysis (H and E staining). Specimens were identified as synovium by the presence of a lining layer or by characteristic histologic features of synovium, including the presence of loose fibrovascular or fatty tissue lacking a lining layer. Samples consisting of dense fibrous tissue, joint capsule or other tissues were determined not to be synovium. Krenn lining scores (0-3) and inflammation scores (0-3) for each tissue sample were determined independently by three pathologists[28].

Tissue disaggregation for mass cytometry and RNA-sequencing

For pipeline analysis, synovial tissue samples stored in cryovials were disaggregated into single cell suspension as described[24]. Briefly, synovial tissue fragments were separated mechanically and enzymatically in digestion buffer (Liberase TL (Sigma-Aldrich) 100 ug/mL and DNase I

(Roche) 100 ug/ml in RPMI) in a 37°C water bath for 30 minutes. Single cell suspensions from disaggregated synovial tissues were assessed for cell quantity and cell viability by trypan Blue. For samples with greater than 200,000 viable synovial cells, 50% of all synovial cells were allocated for analysis by mass cytometry and the remaining cells were allocated for RNA-seq. For samples with less than 200,000 viable synovial cells, all synovial cells were utilized for RNA-seq analysis.

Synovial cell sorting strategy for RNA sequencing

Synovial T cells, B cells, monocytes, and fibroblasts were isolated from disaggregated synovial tissue, as described [24]. Briefly, disaggregated synovial cells were stained with antibodies against CD45 (HI30), 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. 1000 viable (PI-) T cells ($CD45^+$, $CD3^+$, $CD14^-$), monocytes ($CD45^+$, $CD3^-$, $CD14^+$), B cells ($CD45^+$, $CD3^-$, $CD14^-$, $CD19^+$), and synovial fibroblasts ($CD45^-$, $CD31^-$, $PDPN^+$) were collected by fluorescence-activated cell sorting (BD FACSAria Fusion) directly in buffer RLT (Qiagen) for bulk RNA-seq. For single cell RNA-seq, live cells of each population were re-sorted into 384-well plates single cells with a maximum of 144 cells for each cell type, per patient sample.

Tissue samples classification based on leukocyte infiltration

We classified RA tissue samples into leukocyte-poor RA and leukocyte-rich RA based on leukocyte infiltrations by flow cytometry gates using Mahalanobis distance. We first took OA samples as a reference and calculated a multivariate normal distribution of T cells, B cells, and monocytes by flow cytometry. Here we use mahalanobis R function: data x = a matrix of all 51 samples by flow gates of T cells, B cells, and monocytes; center μ = mean of T cells, B cells, and monocytes for all OA samples; covariance Σ = covariance of T cells, B cells, and

monocytes for all OA samples. We calculated the square root to get Mahalanobis distance for each sample,

$$mah = \sqrt{(x-\mu)'\Sigma^{-1}(x-\mu)}\,.$$

We then defined the maximum mah value of all OA samples (4.5) as a threshold to measure the distance of each RA sample from the reference OA and defined 19 leukocyte-rich RA ($mah > 4.5$) and 17 leukocyte-poor RA samples ($mah < 4.5$) in our cohort (**Supplemental 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. Libraries were sequenced on MiSeq from Illumina to generate 35 base paired-end reads. Reads were mapped to transcriptome using kallisto 0.42.4 and transcriptional levels of genes were quantified with the Log_2(TPM+1) (Transcripts Per Kilobase Million) metric.

Bulk RNA-seq quality control

For quality control of bulk RNA-seq data, we began by defining *common genes* as the set of genes detected with at least 1 mapped fragment in 95% of the samples. Then, for each sample, we computed the percent of common genes detected in that sample. Low quality samples are those that have less than 99% of common genes detected, and these were discarded. We found that the low-quality samples also had low cell counts (**Supplemental Fig. 7a**). After discarding 25 low quality samples, we used 167 good quality samples, including 45 fibroblast samples, 46 monocyte samples, 47 T cell samples, and 29 B cell samples in all bulk RNA-seq analyses. Isolated synovial cells from 51 donors revealed distinct transcriptomic profiles between different cell types (**Supplemental Fig. 7b**). Cell lineage markers, *TCRB*, *CD20*, *CD14*

and *PDGFRA,* are expressed selectively by T cells, B cells, monocytes, and fibroblasts, respectively (**Supplemental Fig. 7c**).

Single-cell RNA-seq gene expression quantification

Single-cell RNA-seq was performed using the CEL-Seq2 method[70] 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 (10mM each; NEB) and 5 nl of barcoded reverse transcription primer (1 µg/µL) were added to each well along with 20 nL of ERCC spike-in (diluted 1:800,000). 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 AMPure XP beads (Beckman Coulter). *In vitro* transcription reactions (IVT) were performed, as described followed by EXO-SAP treatment. Amplified RNA (aRNA) was fragmented at 80°C for 3 min and purified using RNAClean XP beads (Beckman Coulter). The purified aRNA was converted to cDNA using an anchored random primer and Illumina adaptor sequences were added by PCR. The final cDNA library was purified using AMPure XP beads (Beckman Coulter). Paired-end sequencing was performed on the HiSeq 2500 in Rapid 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, and removed samples with outlier performance using Picard. We quantified gene levels with the Log_2(CPM+1) (Counts Per Million) metric and also UMIs (Unique Molecular Identifiers).

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 between cells, and we used several metrics to exclude poor quality cells. We identified molecules that are likely to represent cell-to-cell cross-contamination. All single-cell RNA-seq library preparation protocols include pooling and amplification of cDNA molecules from

a large number of cells. This introduces cell-to-cell contamination. We found that molecules represented by a small number of reads are more likely to be contaminant molecules derived from other cells. We developed an algorithm to set a threshold for the minimum number of reads per molecule, and we repeated this separately for each quadrant of 96 wells in 384-well plates. We used 2 marker genes expected to be exclusively expressed in each of the 4 cell types: *PDGFRA* and *ISLR* for fibroblasts, *CD2* and *CD3D* for T cells, *CD79A* and *RALGPS2* for B cells, and *CD14* and *C1QA* 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 1-20 and chose the threshold that maximizes the ratio of true positive to false positive (**Supplemental Fig. 9**). This left us with 7,127 cells and 32,391 genes. Next, 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. This left us with 5,452 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 (**Supplemental Fig. 10**).

Mass cytometry sample processing and quality control

We collected 6 leukocyte-rich, 9 leukocyte-poor RA, and 11 OA samples for mass cytometry analysis, and processed the samples, as described in[24]. Briefly, we analyzed samples on a Helios instrument (Fluidigm) after antibody staining and fixation (**Supplemental Table 2**). Mass cytometry data were normalized using EQ™ Four Element Calibration Beads (Fluidigm), as previously described [76]. Cells were first gated to live DNA+ cells prior to gating for specific cell populations using the following scheme: B cells ($CD3^{-}CD14^{-}CD19^{+}$), fibroblasts ($CD45^{-}PDPN^{+}$), monocytes ($CD3^{-}CD14^{+}$), and T cells ($CD3^{+}CD14^{-}$). All biaxial gating was performed using FlowJo 10.0.7.

Integrative computational pipeline for scRNA-seq clustering

We developed a graph-based unbiased clustering based on canonical correlation analysis to identify the shared cell type subset variations from both single-cell RNA-seq and bulk RNA-seq. We used this computational pipeline to analyze single cells from each cell type. The overall flowchart of graph-based clustering driven by data integration is shown in **Supplemental Fig. 2a**. We describe the details of each step as follows:

1) We first selected the highly variable genes such that the mean and standard deviation are in the top 80% of the density distributions from the single-cell RNA-seq matrix C(g genes by m cells, $c_{1,\ldots,m}$) and bulk RNA-seq matrix S (g genes by n samples, $s_{1,\ldots,n}$), respectively. We focused on the highly variable genes from both sides that we can maintain the distinction between dataset-specific variation.

2) Based on the shared highly variable genes, we integrated single-cell RNA-seq with bulk RNA-seq by finding a linear projection of bulk samples and single cells such that the correlation between the genes are maximized using the CCA method[77]. CCA finds two basic vectors $a_{1,\ldots,n}$ and $b_{1,\ldots,n}$ that maximize the linear correlations $cor(CV_{s1}, CV_{c1})$, where $CV_{s1} = a_1 s_1 + a_2 s_2 + \ldots + a_n s_n$ and $CV_{c1} = b_1 c + b_2 c_2 + \ldots + b_m c_m$. Each bulk sample s_i gets a coefficient a_i and each cell c_i gets a coefficient b_i. The linear combination of all samples $s_{1,\ldots,n}$ arranges bulk genes along the canonical variate CV_{s1} and the linear combination of all cells $c_{1,\ldots,m}$ arranges single-cell genes along CV_{c1}. CCA defines the coefficients $a_{1,\ldots,n}$ and $b_{1,\ldots,n}$ that arrange the genes from the two data sets in such a way that the correlation between the genes is maximized. After CCA finds the first pair of canonical variates, the next pair is computed on the residuals, and so on.

3) We calculated the cell-to-cell similarity matrix using Euclidean distance on the top ten CCA canonical variates.

4) We built up a K-nearest neighbors (KNN) graph based on the cell-to-cell similarity matrix based on local ordinal embedding (LOE), a graph embedding method. We then converted the KNN neighbor relation matrix into an adjacency matrix using the graph.adjacency function from igraph R package;

5) We clustered the cells using the Infomap algorithm (Rosvall and Carl, 2008) for community detection by applying a cluster_infomap function from igraph R package to decompose the cell-to-cell adjacency matrix into major modules by minimizing a description of the information flow;

6) We then constructed a low dimensional embedding using tSNE based on the cell-to-cell distance matrix using the following parameters: perplexity = 50 and theta = 0.5;

7) We identified and prioritized significantly differentially expressed genes for each distinct cluster based on percent of non-zero expressing cells, AUC score [78], and Log_2FC;

8) For pathway analysis, we downloaded gene sets from Gene Ontology (GO) terms on April 2017[79,80]. This included 9,797 GO terms and 15,693 genes. We also used the immunological signatures from 4872 hallmark gene sets from MSigDB[81] to test enrichment of all the tested genes sorted by decreased AUC scores for each cluster by 10^5 permutation tests[79]. The liger function (https://github.com/JEFworks/liger) was used to visualize the statistics permutation distribution.

To identify the most reasonable and stable clusters, we ran this pipeline by tuning the number of top canonical variates (4, 8, 12, 16, and 20) that were incorporated for the cell-to-cell similarity matrix, and the number of k (50, 100, 150, 200, 250, and 300) to build up the K-nearest neighbors graph. We decided the robust cell type clusters based on the ones that yielded the most differentially expressed genes. To evaluate the robustness of our clustering results, we tested the clustering performance using the Silhouette metric on the cell-to-cell

similarity matrix [82,83], which is an Euclidean distance in our study (**Supplemental Fig. 11**). For each cell i, the silhouette width $s(i)$ is defined as follows:

$$s(i) = \frac{b(i) - a(i)}{max(a(i),b(i))},$$

where $a(i)$ is the average dissimilarity between cell i and all the other cells in the same cluster. We let $b(i)$ be the lowest average distance of cell i to all cells in any other cluster, which can be seen as the nearest one to which cell i does not belong. Large $s(i)$ means the cell is well assigned to its own cluster but improperly matched to other clusters.

Thus, for each pair of single-cell RNA-seq and bulk RNA-seq, we ran our pipeline on the shared samples that have both datasets for each cell type. For integrating fibroblast data, we used 45 bulk RNA-seq samples, 1,844 single cells on 7,016 shared highly variable genes; for integrating monocyte data, we used 47 bulk RNA-seq samples, 750 single cells on 7,016 shared highly variable genes; for integrating T cell data, we used 47 bulk RNA-seq samples, 1,716 single cells on 7,003 shared highly variable genes; for integrating B cell data, we used 29 bulk RNA-seq samples, 1,142 single cells on 7,023 shared highly variable genes.

Mass cytometry clustering and disease association test

We created mass cytometry datasets for analysis by concatenating cells from all individuals for each cell type. For donors with more than 1,000 cells, we randomly selected 1,000 cells to ensure that samples were equally represented. In this way, we created 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 for analysis. We then applied the tSNE algorithm (Barnes-Hut implementation) to each dataset using the following parameters: perplexity = 30 and theta = 0.5. We used all markers except those used to gate each population in the SNE clustering. To identify high-dimensional populations, we used a modified version of DensVM[25]. DensVM performs kernel density estimation across the dimensionally reduced SNE map to build

a training set, then assigns cells to clusters by their expression of all markers using an SVM classifier. We modified the DensVM code to increase the range of potential bandwidths searched during the density estimation step and to return the SVM model generated from the tSNE projection. We summarized the details of the clusters with proportion of cells from each disease cohort in **Supplemental Table 3**.

After clustering, we identified cell subsets expanded in leukocyte-rich RA using 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[21]. We specified donor identity and batch as random-effect covariates.

Integration of bulk RNA-seq with mass cytometry

We used the identified clusters of cell populations by DensVM to query bulk RNA-seq data about the gene signatures that are highly enriched in the same clusters of cells from mass cytometry. To complement the proteomic data, we integrated identified mass cytometry clusters with bulk RNA-seq using our CCA-based framework. We have two data matrices from the same tissue samples. The mass cytometry data matrix has samples and clusters, and the values represent proportions of cells from each sample in each cluster. The bulk RNA-seq data matrix has samples and genes, and the values represent proportions of gene expression 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. For each pair of datasets, we ran our pipeline on the post-QC data from all samples that had both datasets. For fibroblasts data integration, we associated 22,99 genes with 8 mass cytometry clusters on 22 shared samples; we associated 2,161 genes with 5 mass cytometry clusters on 25 shared monocyte samples;

we associated 2,255 genes with 9 mass cytometry clusters on 26 shared T cell samples; we associated 22,95 genes with 10 mass cytometry clusters on 17 shared B cell samples.

Integration of scRNA-seq clusters with mass cytometry clusters

1) For each cell type data, we first integrated mass cytometry clusters identified by DensVM with bulk RNA-seq on the shared samples using CCA framework. Each gene is correlated with each canonical variate (CV). Also, each mass cytometry cluster is correlated with each CV.

2) We then associated single-cell RNA-seq clusters with mass cytometry clusters by projecting cluster markers (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 cor between the scRNA-seq average CV and the CV for mass cytometry clusters.

5) Next, we obtained $permutation\ p$ by shuffling all the tested gene names 1e4 times to generate replicates of CCA matrix from step 1).

6) For the 1e4 replicates of CCA matrix, we repeated from step 2 to step 5. Then, we counted how many times the correlation of each pair cor_{perm} was greater than the observed value cor from step 4).

$$permutation\ p\ = \frac{1 + sum(cor_{perm} > cor)}{1 + 1e4}.$$

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 and leukocyte-rich RA synovial tissues based on the quantitative analysis of T cells, B cells, and monocytes by flow cytometry.

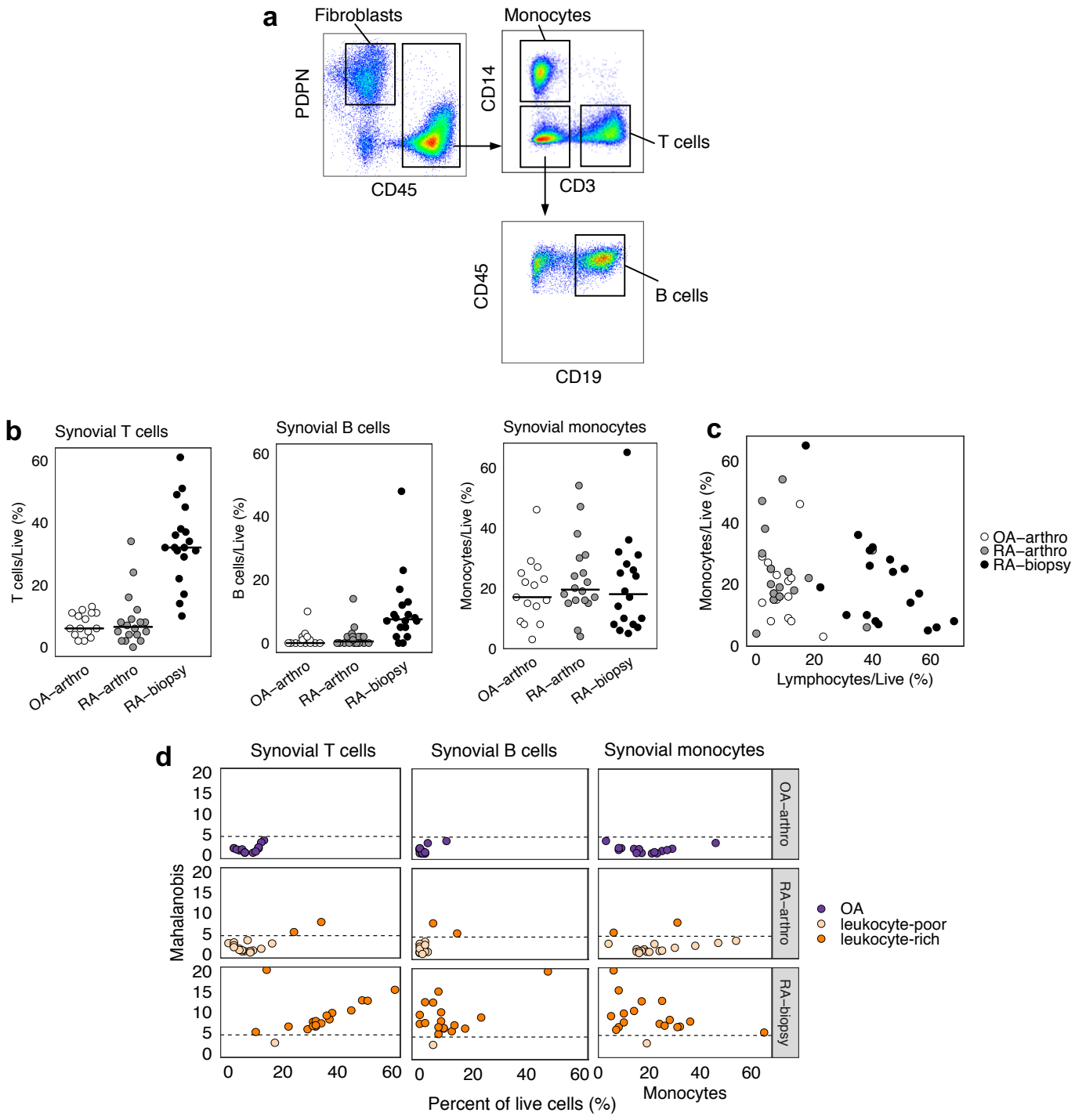
Consistently, our PCA analysis on bulk RNA-seq samples showed the same classification of leukocyte-rich and leukocyte-poor RA. By comparing leukocyte-rich RA samples with OA samples, we were able to enhance the power to detect molecular pathways associated with tissue inflammation. For differential analysis, we use Limma R function to identify significantly differentially expressed genes. Benjamini-Hochberg (BH) was used for false discovery rate (FDR) control.

Identification of markers for distinct scRNA-seq clusters

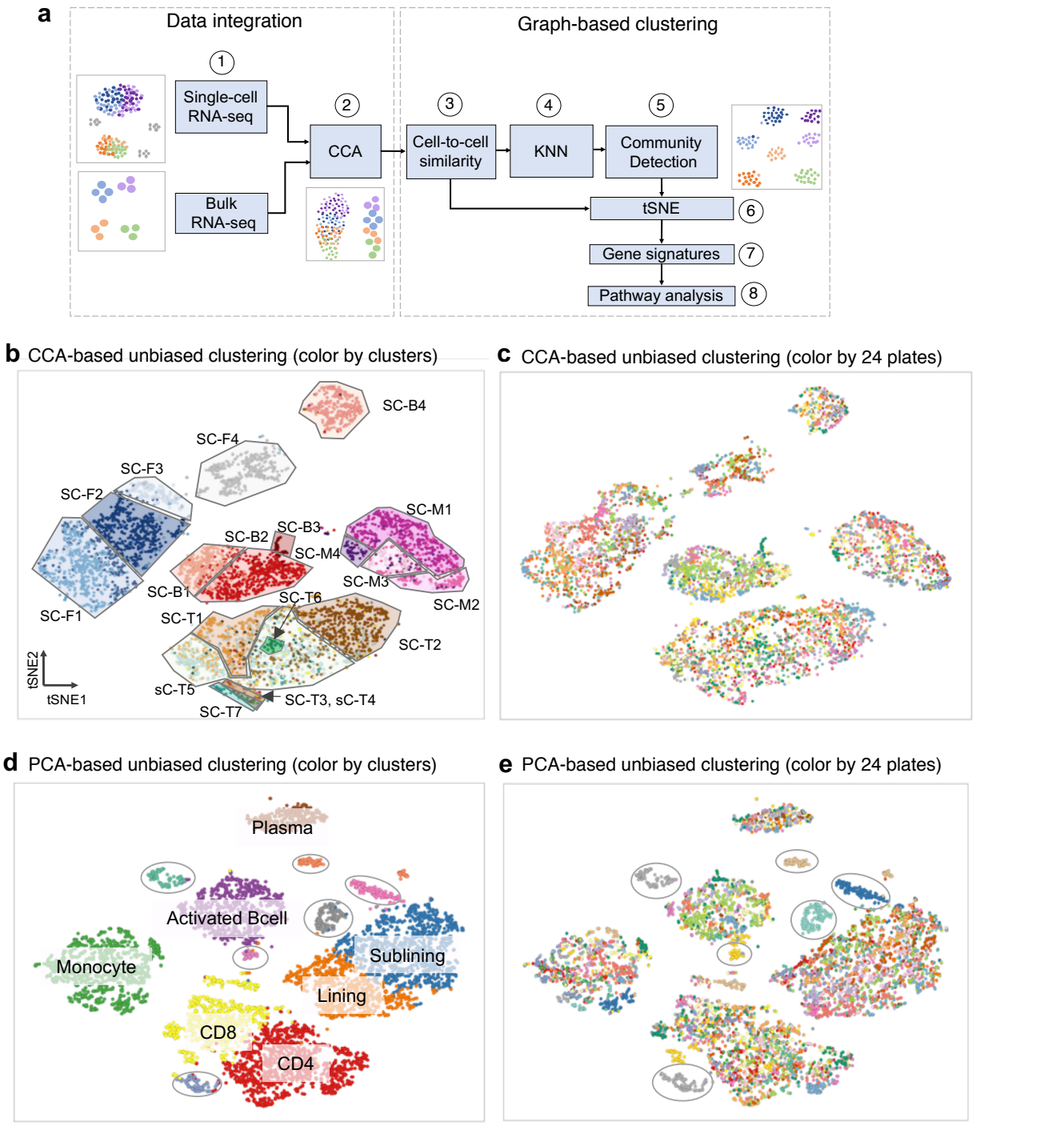
Based on the single-cell RNA-seq clusters, we identified cluster marker genes by comparing the cells within one cluster with the cells from all the other clusters from the same cell type, based on Log_2(CPM+1). We prioritized cluster marker genes using three criteria: 1) percent of non-zero expressing cells > 60%; 2) AUC score > 0.7; and 3) FC > 2. We display top 20 cluster marker genes in **Supplemental Table 4.** We noticed that some clusters showed fewer significantly differentially expressed genes. For example, fewer genes were distinctly expressed in sublining subset SC-F1, since subset SC-F1 and subset SC-F2 shared the similar inflammatory genes. Marker genes for SC-M3 cells are common to all *CD14*$^+$ monocytes, which makes it challenging to identify the distinct genes compared to other monocytes subsets.

Data and code availability

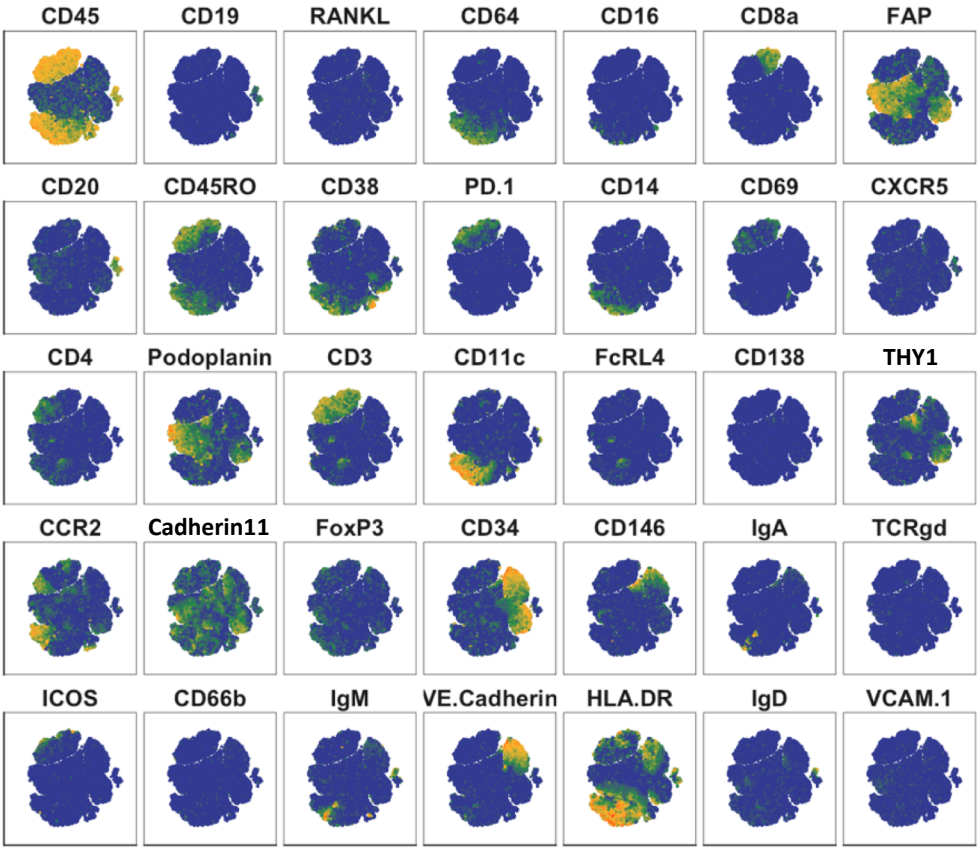
The data of this study are available at ImmPort http://immport.org with the study accession code SDY998 and SDY999. The source code repository of the computational pipeline for data analysis and integration is located at https://github.com/immunogenomics/amp_phase1_ra.



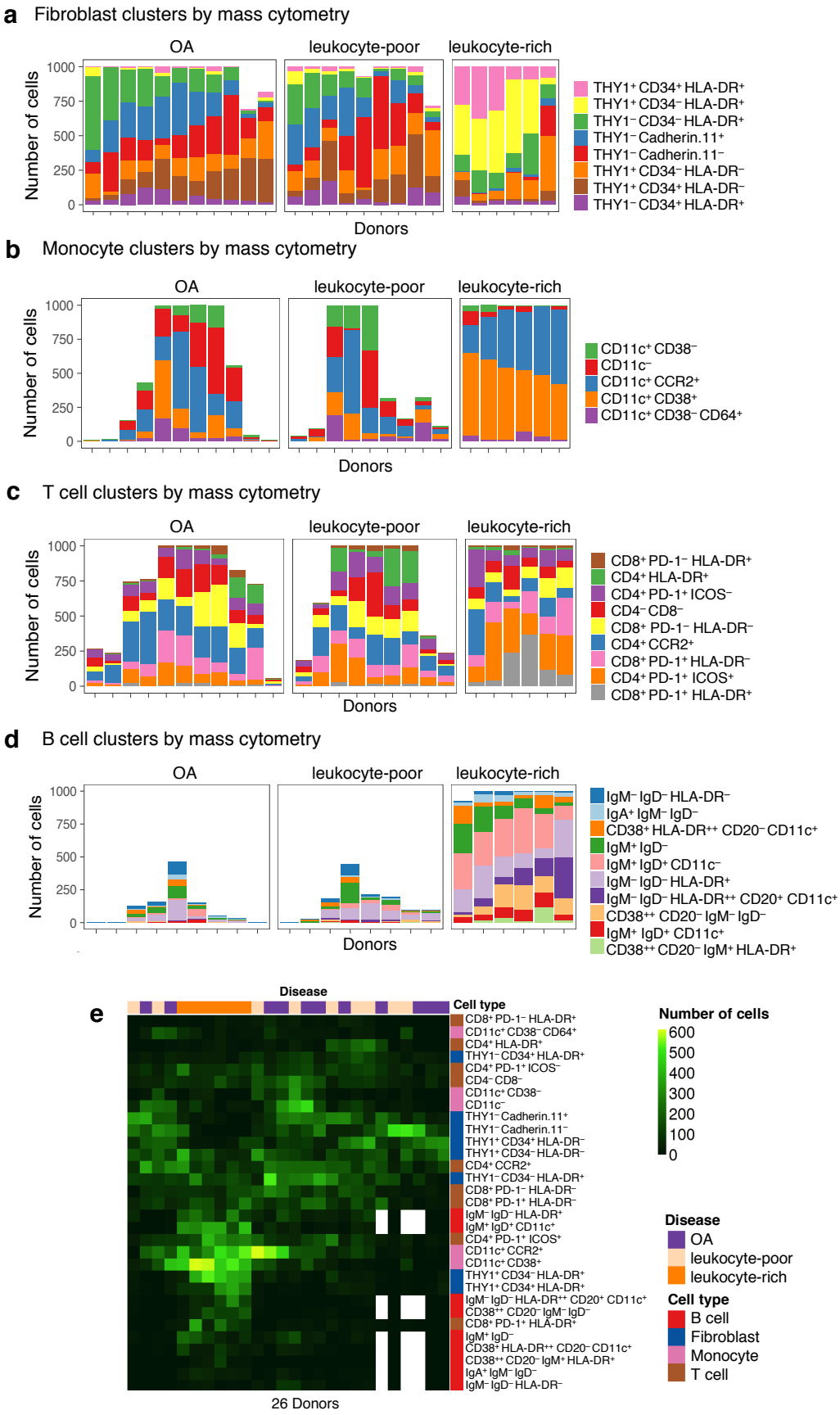
Supplemental Fig. 1. Flow cytometry gating scheme and a data-driven approach to separate samples based on flow cytometry data.
a. Flow cytometry gating: stromal fibroblasts ($CD45^{-}PDPN^{+}$), monocytes ($CD45^{+}CD14^{+}$), T cells ($CD45^{+}CD3^{+}$), and B cells ($CD45^{+}CD3^{-}CD19^{+}$). **b.** Synovial T cells, B cells, and monocytes for OA-arthro (OA arthroplasty), RA-arthro (RA arthroplasty), and RA-biopsy (RA biopsy) by flow cytometry. **c**. Association between lymphocytes percent with monocytes percent by flow cytometry. **d**. Mahalanobis distance from OA samples by T cells, B cells, and monocytes for OA-arthro, RA-arthro, and RA-biopsy samples by flow cytometry. Leukocyte-rich RA samples are defined with Mahalanobis distance from OA greater than 4.5 (dashed line). We identified 19 leukocyte-rich RA, 17 leukocyte-poor RA, and 15 OA samples in our cohort.



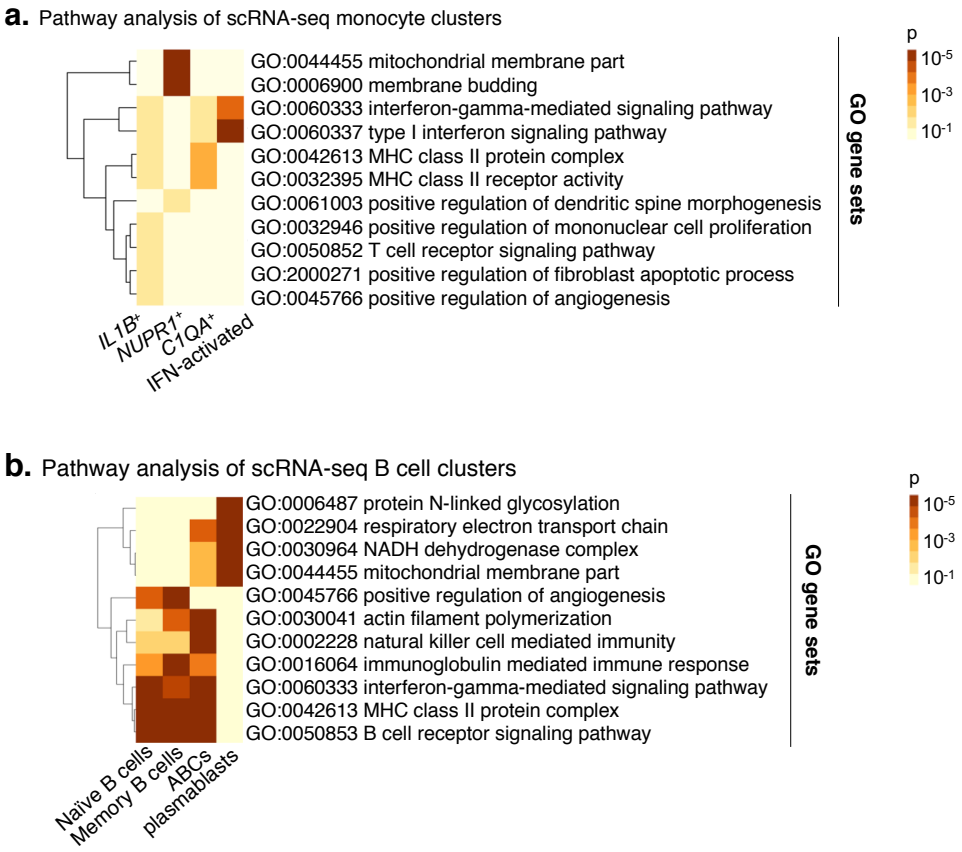
Supplemental Fig. 2. CCA-based integrative pipeline of scRNA-seq analysis and method comparison.
a. Steps of computational pipeline using CCA for analyzing scRNA-seq data: 1) We first select the highly variable genes from both scRNA-seq and bulk RNA-seq; 2) based on the selected genes from both sides we integrate single cells with bulk samples that by learning a linear projection that the correlation between them are maximized using CCA; 3) we then calculate a cell-to-cell similarity matrix based on the canonical variates from CCA; 4) based on the cell-to-cell similarity matrix, we built a K-nearest neighbors (KNN) and then convert it into an adjacency matrix; 5) we cluster the cells using community detection unbiased clustering algorithm, Infomap, to identify major groups based on the cell-to-cell adjacency matrix; 6) project the cells with identified clusters on to tSNE space; 7) Based on the identified cell type clusters, we do gene expression differential analysis using AUC and Wilcox text; 8) finally, we perform gene set enrichment analysis to find out the upregulated pathways. **b**. Identified 19 scRNA-seq clusters using the CCA-based integrative pipeline. **c**. Source of 24 plates for cells by CCA-based unbiased clustering. **d**. Identified cell clusters using PCA-based unbiased clustering by Seurat. **e**. Sources of 24 plates for cells for PCA-based unbiased clustering by Seurat. Clusters of batch effect plates are highlighted using circles.



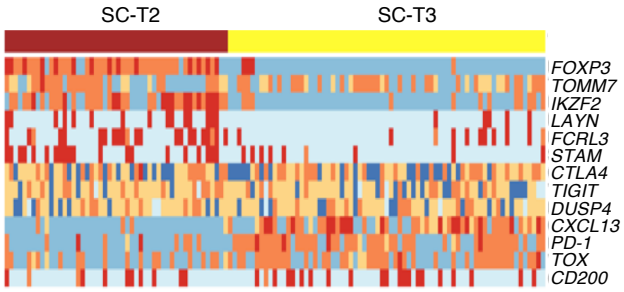
Supplemental Fig. 3. Protein markers of viable and DNA+ synovial cells (3,000 downsampled) from all donors by mass cytometry.



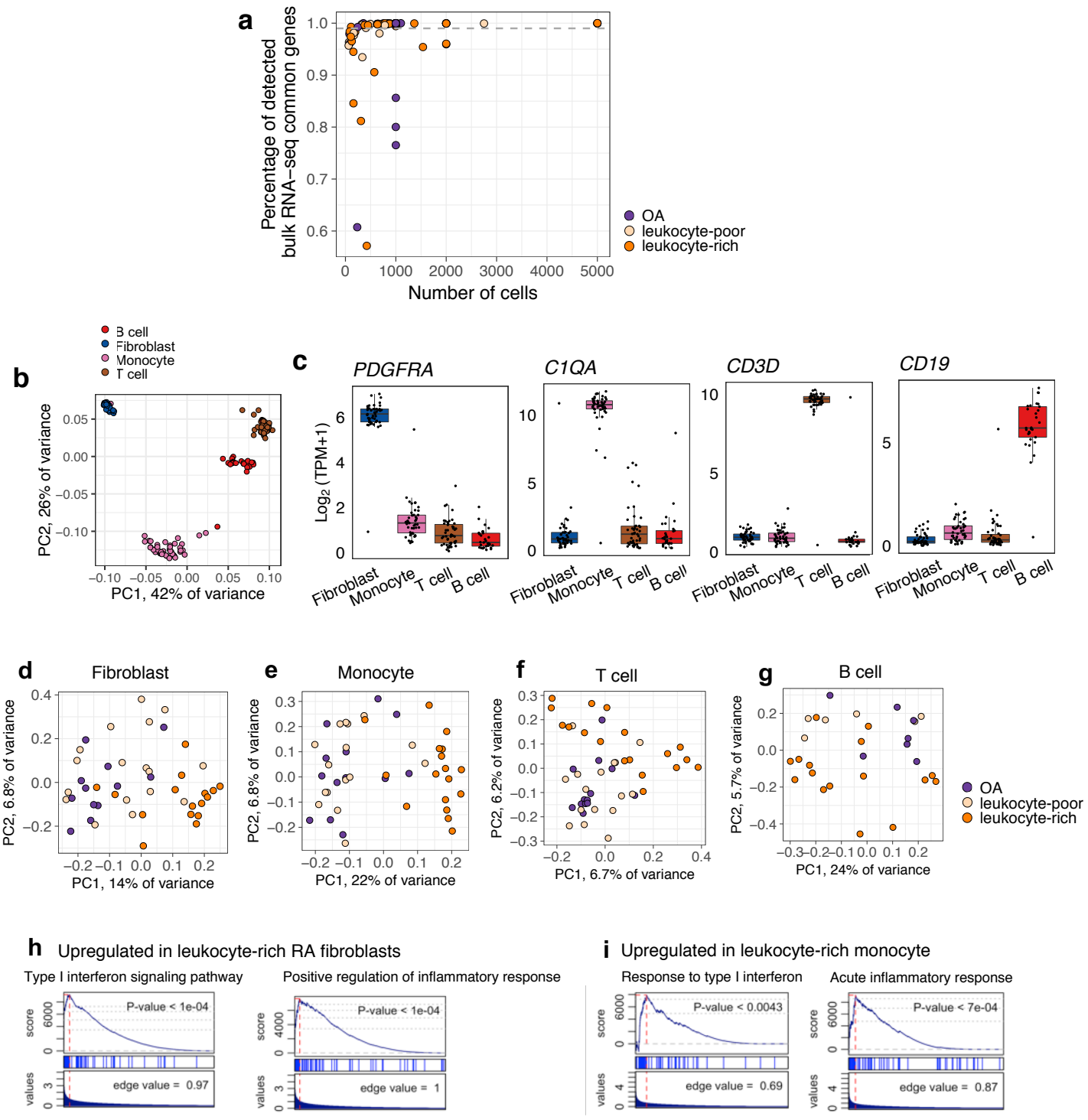
Supplemental Fig. 4. Distribution of identified cell type clusters for each donor by mass cytometry. **a-d**. Distribution of mass cytometry clusters for each cell type confirms that the identified clusters are not confounded by obvious batch effects. **e**. Cell counts of all clusters by comparing all the 26 donors reveal that leukocyte-rich donors show high cell abundance of HLA-DR+ fibroblasts (THY1+ CD34− HLA-DR+ and THY1+ CD34+ HLA-DR+), Tph cells (CD4+ PD-1+ ICOS+), two populations of CD14+ monocytes (CD11c+ CCR2+ and CD11c+ CD38+), and a B cell population (IgM+ IgD+ CD11c+).



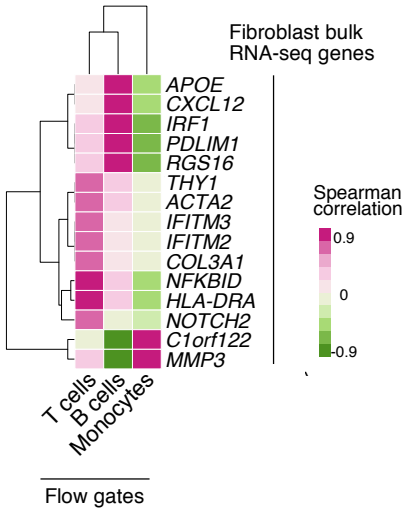
Supplemental Fig. 5. Pathway enrichment analysis on GO gene sets for identified scRNA-seq clusters from monocytes and B cells.



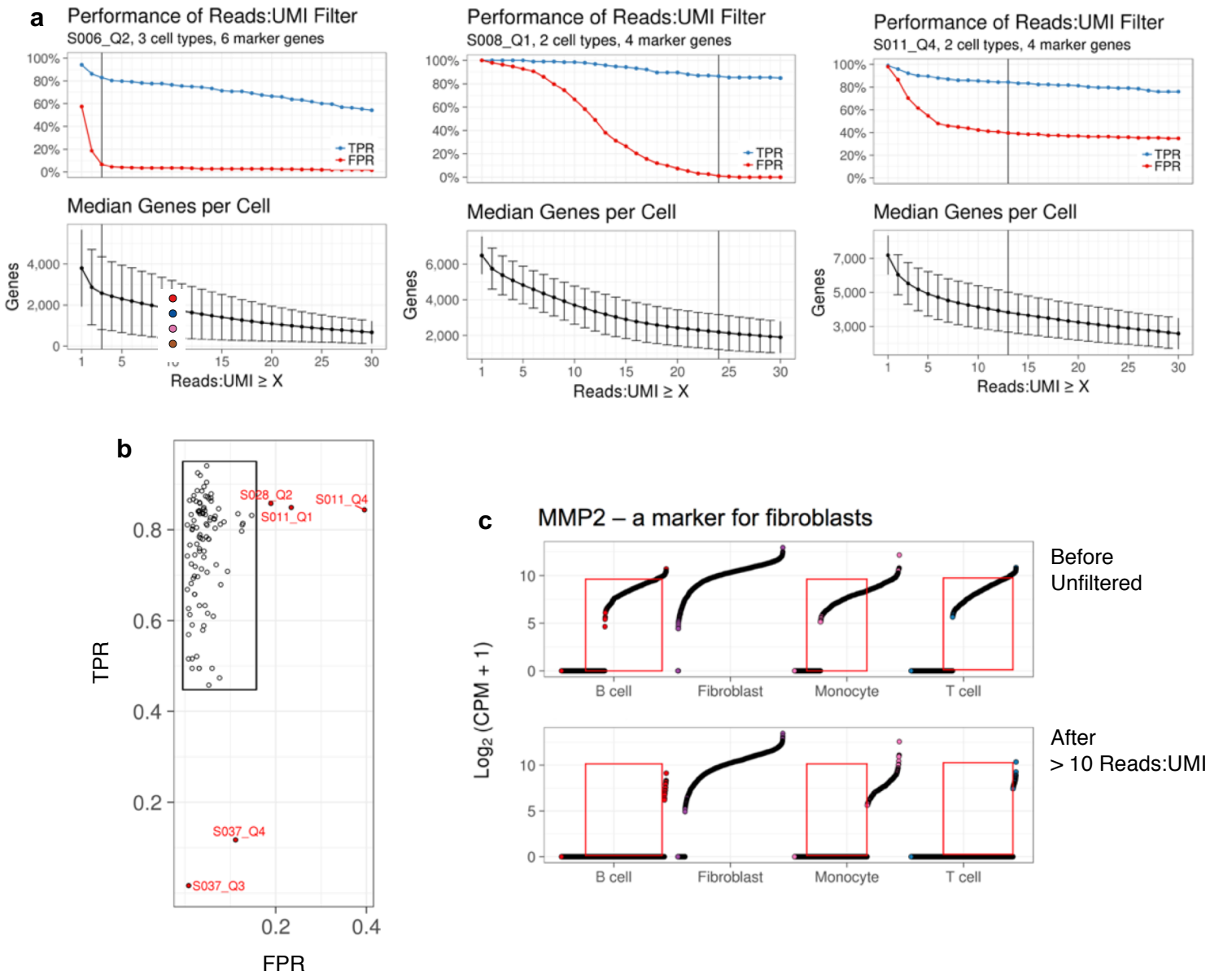
Supplemental Fig. 6. scRNA-seq clusters Tregs (SC-T2) and Tph (SC-T3) that separated based on the most informative markers from (Rao et al. 2017). We use hierarchical clustering with R function hclust() and then cutree(k=2) to pinpoint previously characterized rare cell populations, Tregs and Tph cells.



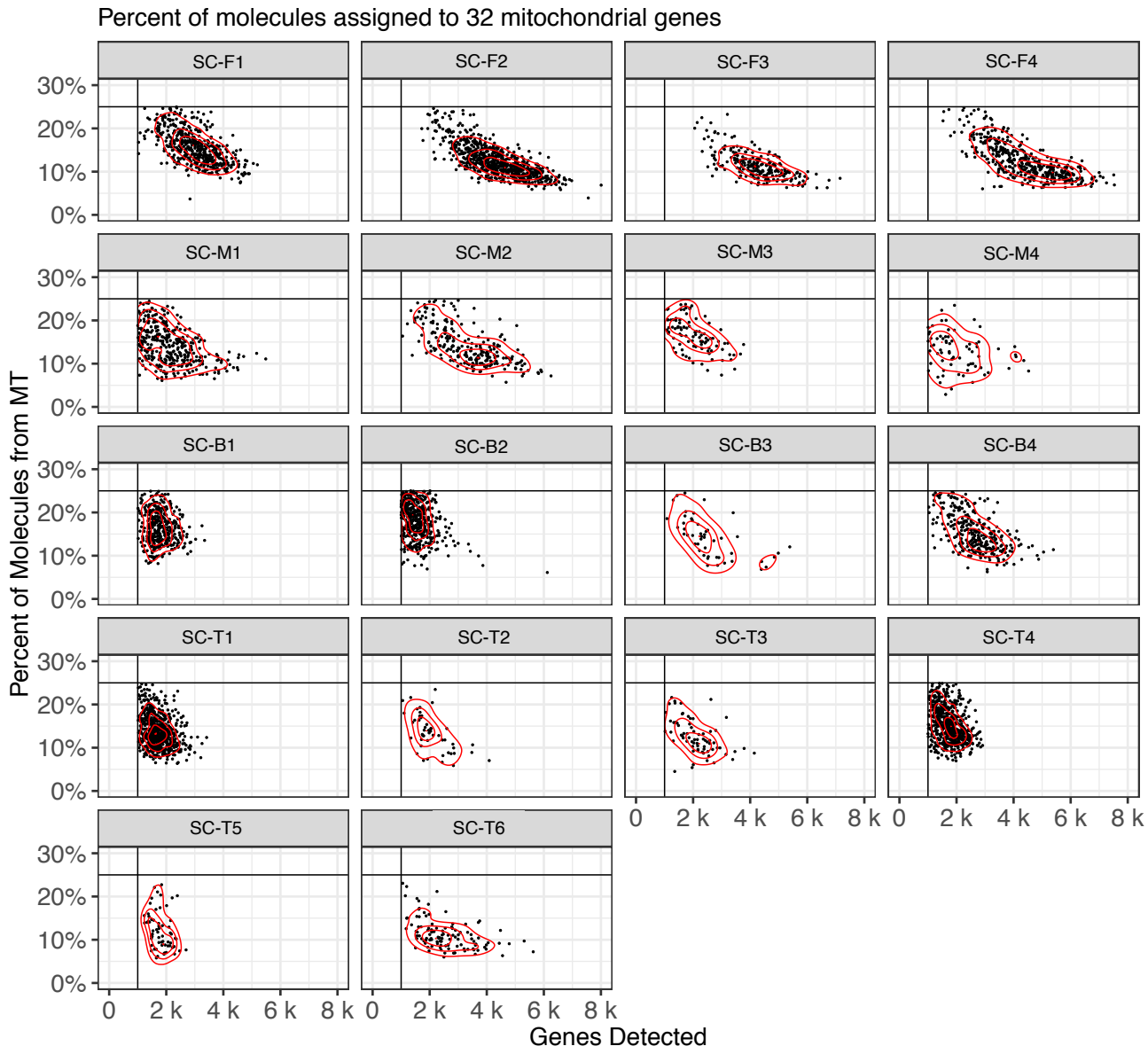
Supplemental Fig. 7. Quality control of bulk RNA-seq and PCA analysis for each cell type samples. **a**. Quality control of bulk RNA-seq samples. Common genes are defined as the set of genes detected with at least 1 mapped fragment in 95% of the samples (13,041 genes). X-axis is the number of cells for each bulk RNA-seq sample. Y-axis is the percentage of detected common genes for each sample. We discarded 25 low quality samples that have less than 99% (dashed line) of common genes detected, resulting 167 post-QC samples in all. **b**. PCA analysis on all the samples shows that most of the variance in the bulk RNA-seq data is due to cell type. **c**. Cell type marker genes show that there is no obvious contamination in the bulk RNA-seq data. **d-g**. PCA analysis on samples from each cell type. The samples from leukocyte-rich RA appear distinct from leukocyte-poor RA and OA samples. **h-i.** Distribution of significantly enriched GO terms in leukocyte-rich RA by GSEA. Leukocyte-rich fibroblasts and monocytes share the common pathways of Type I interferon and inflammatory response.



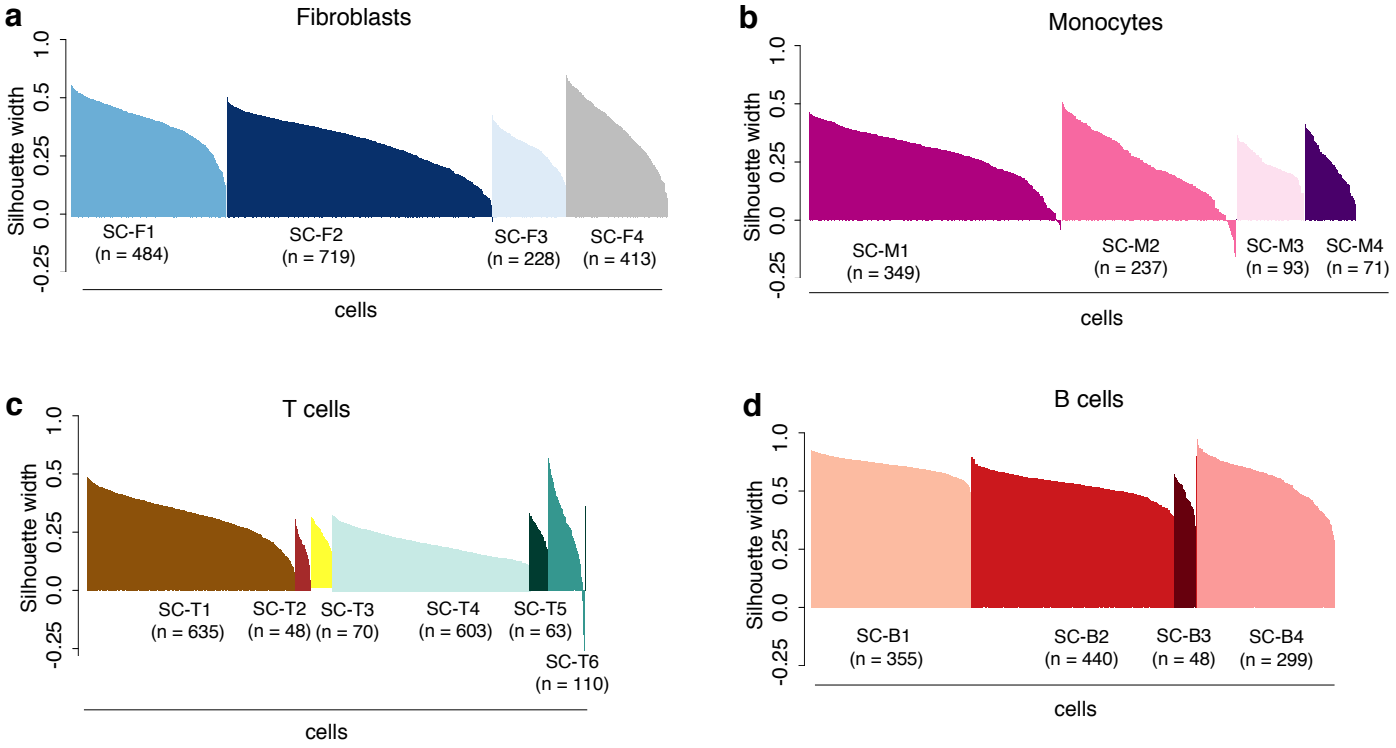
Supplemental Fig. 8. Correlation between bulk RNA-seq genes and immune cell type abundances in RA synovial fibroblasts. Integrating bulk RNA-seq samples from fibroblasts with multiple cell type flow gates reveals that T cells, B cells, and monocytes that are abundant in RA synovial tissue directly influence the expression of fibroblasts in the RA synovium.



Supplemental Fig. 9. Dynamic filtering strategy for scRNA-seq quality control. **a**. Estimated optimal threshold for three example quadrants of scRNA-seq plates. **b**. Select optimal threshold that maximizes the ratio of TPR (true positive rate) to FPR (false positive rate). We selected 2 marker genes expected to be exclusively expressed in each of the 4 cell types: *PDGFRA* and *ISLR* for fibroblasts, *CD2* and *CD3D* for T cells, *CD79A* and *RALGPS2* for B cells, and *CD14* and *C1QA* 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. We discard the cells in the red outlier quadrants. **c**. An example of fibroblast gene *MMP2*. After dynamic filtering strategy, no B cells, monocytes, or T cells express gene *MMP2*.



Supplemental Fig. 10. All post-QC scRNA-seq data for each identified cluster based on number of genes detected and percent of molecules from 32 mitochondrial genes.



Supplemental Fig. 11. Stability test of identified scRNA-seq clusters by Silhouette. **a**. Silhouette widths for fibroblast subsets. **b**. Silhouette widths for monocyte subsets. **c**. Silhouette widths for T cell subsets. **d**. Silhouette widths for B cell subsets. Cells with a large Silhouette width are well clustered; cells with a negative Silhouette width are unstably clustered.

Supplemental Table 1. Clinical characteristics of 51 recruited patients.

		OA (n=15)	**leukocyte-poor RA** (n=17)	**leukocyte-rich RA** (n=19)
Demographic variables	Age, mean	71	64.2	57.3
	(Range)	(64-81)	(42-79)	(36-71)
	Females, n (%)	10 (66.7)	15 (82.4)	14 (73.7)
RA-related variables	Mean years of disease duration		15.7	5.5*
	(range)		(<1-51)	(<1-29)
	RF positive, n (%)		8 (47.1)	16* (84.2)
	CCP positive, n (%)		10 (58.8)	14 (73.7)
DMARDs	Prednisone, n (%)		10 (55.6)	4* (22.2)
	Methotrexate, n (%)		7 (41.2)	3 (15.8)
	TNFi, n (%)		4 (23.5)	2 (10.5)
	Rituximab, n (%)		0 (0)	1 (5.3)
	Abatacept, n (%)		1 (5.9)	1 (5.3)
	Tofacitinib, n (%)		2 (11.8)	1 (5.3)

DMARDs = Disease-Modifying Antirheumatic Drugs.
TNFi = TNF inhibitors (infliximab, etanercept, adalimumab, Golimumab).
RhF = Rheumatoid Factor. CCP = Cyclic Citrullinated Peptide.
*Significant p-value between leukocyte-poor RA and leukocyte-rich RA.

Supplemental Table 2. Antibody staining and fixation of mass cytometry panel.

marker	clone
CD45	HI30
CD19	HIB19
RANKL	MIH24
CD64	10.1
CD16	3G8
CD8a	RPA T8
FAP	Poly
CD20	2H7
CD45RO	UCHL1
CD38	HIT2
CD279/PD-1	EH12.2H7
CD14	M5E2
CD69	FN50
CD185/CXCR5	J252D4
CD4	RPA T4
Podoplanin	NC-08
CD3	UCHT1
CD11c	Bu15
CD307d/FcRL4	413D12
CD138	MI15
CD90	5.00E+10
CCR2	K036C2
Cadherin 11	3C10
FoxP3	PCH101
CD34	581
CD146/MCAM	SHM-57
IgA	9H9H11
ICOS	C398.4A
CD66b	G10F5
IgM	MHM-88
CD144/VE-Cadherin	BV9
HLA-DR	L243
IgD	IA6-2
CD106/VCAM-1	STA

Supplemental Table 3. Identified mass cytometry populations with proportion of cells from each disease cohort and on tailed FDR q value.

Cell type	mass cytometry cluster	# Cells	Cells from OA (%)	Cells from leukocyte-poor RA (%)	Cells from leukocyte-rich RA (%)	One tailed FDR q value
Fibroblast	$THY1^-$ $Cadherin11^-$	4,333	48.3	45.7	6.0	1.00
	$THY1^-$ $Cadherin11^+$	3,752	55.8	40.9	3.3	1.00
	$THY1^-$ $CD34^+$ $HLA\text{-}DR^{high}$	1,474	42.5	44.2	13.3	1.00
	$THY1^-$ $CD34^-$ $HLA\text{-}DR^{high}$	4,171	50.2	28.3	21.5	1.00
	$THY1^+$ $CD34^-$ $HLA\text{-}DR^{high}$	2,577	7.1	9.4	83.4	5.8E-04
	$THY1^+$ $CD34^-$ $HLA\text{-}DR^{low}$	3,794	38.1	38.6	23.3	1.00
	$THY1^+$ $CD34^+$ $HLA\text{-}DR^{low}$	3,514	52.0	40.9	7.1	1.00
	$THY1^+$ $CD34^+$ $HLA\text{-}DR^{high}$	1,546	9.8	10.2	80.0	1.2E-03
Monocyte	$CD11c^-$	2,801	58.0	33.4	8.6	1.00
	$CD11c^+$ $CCR2^+$	5,602	32.6	24.2	43.2	0.72
	$CD11c^+CD38^-$	1,313	37.0	54.8	8.2	1.00
	$CD11c^+CD38^-$ $CD64^+$	1,031	37.9	43.7	18.3	1.00
	$CD11c^+$ $CD38^+$	4,551	20.0	13.1	66.9	9.5E-04
T cell	$CD4^-$ $CD8^-$	2,475	44.6	34.5	20.9	1.00
	$CD4^+$ $CCR2^+$	4,093	51.7	28.7	19.6	1.00
	$CD4^+$ $HLA\text{-}DR^+$	1,178	29.3	62.3	8.4	1.00
	$CD4^+$ $PD\text{-}1^+$ $ICOS^-$	2,368	33.7	37.1	29.1	1.00
	$CD4^+$ $PD\text{-}1^+$ $ICOS^+$	3,138	22.9	28.6	48.5	1.7E-03
	$CD8^+PD\text{-}1^-$ $HLA\text{-}DR^-$	2,535	46.1	31.6	22.3	1.00
	$CD8^+$ $PD\text{-}1^-$ $HLA\text{-}DR^+$	400	42.3	35.5	22.3	1.00
	$CD8^+$ $PD\text{-}1^+HLA\text{-}DR^-$	2,710	40.0	29.3	30.8	1.00
	$CD8^+$ $PD\text{-}1^+$ $HLA\text{-}DR^+$	1,088	9.8	9.4	80.8	9.5E-04
B cell	$CD38^{++}$ $CD20^-$ IgM^- IgD^-	770	3.6	6.5	89.9	0.04
	$CD38^{++}$ $CD20^-$ IgM^+ $HLA\text{-}DR^+$	220	1.8	5.5	92.7	0.04
	IgM^- IgD^- $HLA\text{-}DR^-$	379	42.0	39.1	19.0	1.00
	$CD38^+$ $HLA\text{-}DR^{++}$ $CD20^-$ $CD11c^+$	550	16.7	18.2	65.1	1.00
	IgM^- IgD^- $HLA\text{-}DR^{++}$ $CD20^+$ $CD11c^+$	771	1.0	6.6	92.3	0.01
	IgM^- IgD^- $HLA\text{-}DR^+$	1,774	15.7	25.3	59.0	1.00
	IgA^+ IgM^- IgD^-	336	20.2	11.0	68.8	1.00
	IgM^+ IgD^-	1,126	18.7	22.6	58.8	1.00
	IgM^+ IgD^+CD11c^-	1,762	5.6	7.3	87.2	0.05
	IgM^+ IgD^+CD11c^+	491	8.8	7.5	83.7	0.69

Supplemental Table 4. Top 20 marker genes for each single-cell RNA-seq cluster.

cell type	cluster	gene	auc	wilcox_pvalue	ttest_pvalue	pct_nonzero	log2FC
Fibroblast	*CD34*$^{+}$ sublining (SC-F1)	*DCN*	0.73	9.7E-57	1.5E-47	99.8%	1.01
		IGF1	0.72	6.8E-55	1.2E-54	81.8%	3.73
		FBLN1	0.71	2.3E-48	5.3E-36	82.2%	2.55
		PDGFRL	0.69	2.9E-39	9.5E-21	79.2%	1.76
		SFRP1	0.69	5.1E-41	2.6E-39	74.9%	3.32
		C3	0.68	4.9E-39	1.2E-33	69.5%	2.88
		GAS6	0.67	5.4E-40	3.5E-32	60.8%	2.43
		FBLN5	0.66	1.0E-30	1.6E-24	64.5%	2.07
		RPL37	0.65	3.1E-25	7.2E-06	98.3%	0.32
		VCAN	0.65	6.3E-24	2.0E-31	98.0%	1.48
		RPL35A	0.65	3.4E-23	6.1E-06	98.7%	0.29
		CD34	0.64	7.4E-37	3.7E-28	44.6%	2.16
		GSN	0.64	2.2E-22	3.2E-05	94.2%	0.55
		SERPINF1	0.64	5.1E-22	1.5E-28	95.7%	1.66
		PPAP2B	0.64	1.8E-21	4.0E-04	84.8%	0.61
		MFAP5	0.64	1.5E-36	2.3E-28	43.3%	2.43
		PODN	0.64	2.0E-23	6.3E-19	66.4%	1.82
		PTGFR	0.64	4.5E-22	1.6E-13	71.2%	1.57
		SFRP2	0.64	1.9E-44	1.0E-30	37.0%	2.79
		CFH	0.64	7.9E-21	1.1E-03	91.1%	0.46
Fibroblast	*HLA*^{high} sublining (SC-F2)	*CD74*	0.82	1.5E-122	8.6E-130	96.3%	4.08
		IFITM3	0.80	5.1E-106	8.9E-100	100.0%	1.02
		LGALS3BP	0.79	1.9E-103	1.3E-87	96.8%	2.60
		HLA-DRA	0.78	1.7E-108	3.6E-118	80.7%	4.24
		HLA-B	0.78	8.6E-91	3.7E-73	99.9%	1.16
		CXCL12	0.77	2.0E-90	6.7E-118	99.2%	3.72
		HLA-DPA1	0.77	3.9E-98	1.8E-105	80.2%	3.80
		HLA-DRB1	0.77	4.1E-98	1.5E-103	79.3%	4.03
		PLAU	0.76	1.0E-88	1.2E-99	83.3%	3.77
		HLA-A	0.76	3.6E-81	4.4E-67	99.9%	1.02
		IFI27	0.76	1.1E-81	8.6E-72	93.1%	2.74
		B2M	0.76	1.7E-80	8.2E-80	100.0%	0.72
		HLA-DPB1	0.76	1.1E-88	1.1E-89	79.0%	3.60
		MDK	0.76	1.7E-87	1.5E-114	85.5%	3.96
		ZFP36L1	0.75	9.5E-77	3.0E-75	100.0%	1.55
		IFI30	0.75	3.1E-85	1.0E-87	73.8%	3.37
		STAT1	0.75	9.6E-74	1.6E-72	90.9%	2.76
		LOXL1	0.74	7.1E-77	5.2E-91	84.0%	3.37
		BGN	0.74	8.5E-70	4.4E-74	99.7%	2.06
		C1S	0.74	9.5E-70	1.2E-55	99.9%	0.75
Fibroblast	*DKK3*$^{+}$ sublining (SC-F3)	*ASPN*	0.86	7.0E-75	2.1E-68	91.7%	5.58
		DKK3	0.82	3.4E-83	4.0E-56	80.3%	4.94
		MAMDC2	0.78	1.8E-46	1.4E-46	88.6%	3.62
		ABHD2	0.78	4.4E-43	1.8E-32	89.0%	2.77
		PRELP	0.77	3.2E-41	1.7E-41	94.3%	3.38
		COL5A1	0.76	2.3E-37	9.2E-44	98.2%	2.18
		EMP1	0.76	8.1E-37	1.6E-30	96.5%	1.98
		ANGPTL1	0.76	1.8E-36	3.0E-26	91.2%	2.38
		CADM1	0.75	2.8E-60	2.0E-34	65.8%	3.90
		FMOD	0.75	1.2E-36	8.2E-30	84.2%	3.11
		OGN	0.75	8.4E-35	6.8E-35	92.1%	3.07
		PDGFRB	0.75	3.3E-34	7.1E-48	96.1%	2.53
		CD109	0.74	2.4E-37	1.4E-34	82.9%	2.94
		MYADM	0.74	8.6E-33	2.4E-25	96.9%	1.77
		AHNAK	0.74	1.3E-32	1.4E-35	100.0%	0.87
		VIM	0.74	3.1E-32	1.2E-35	100.0%	0.60
		COL8A2	0.73	2.9E-46	6.2E-29	64.9%	3.22
		CAMK1D	0.73	2.0E-31	1.6E-48	94.3%	2.83
		DPT	0.73	1.9E-30	4.1E-45	98.7%	2.22
		COMP	0.73	3.8E-40	2.1E-30	71.9%	4.22
		PRG4	0.96	1.1E-181	5.1E-256	100.0%	3.33
		FN1	0.94	9.5E-169	3.4E-291	100.0%	2.12
		ITGB8	0.94	5.4E-173	1.3E-242	98.3%	5.17
		CRTAC1	0.92	3.0E-154	4.0E-227	99.8%	4.30
		HTRA1	0.92	2.0E-151	1.7E-175	100.0%	2.75
		PCSK6	0.91	1.2E-178	3.9E-205	92.9%	5.76
		HBEGF	0.91	4.9E-189	1.2E-162	89.8%	7.14

Fibroblast	Lining (SC-F4)	*THBS4*	0.90	2.8E-142	6.0E-151	99.0%	3.52
		TIMP3	0.90	5.3E-140	7.7E-191	100.0%	2.89
		APLP2	0.90	1.3E-139	3.9E-146	100.0%	1.75
		SMIM14	0.89	1.5E-134	2.9E-118	98.8%	2.51
		NTN4	0.89	1.3E-153	4.1E-150	90.7%	5.18
		CD55	0.89	5.7E-137	1.8E-157	96.0%	4.69
		CLIC5	0.89	2.1E-238	1.4E-137	81.2%	6.26
		C10orf105	0.88	1.3E-247	4.1E-134	79.5%	5.90
		ERRFI1	0.88	7.4E-134	1.2E-139	93.6%	5.14
		CREB5	0.88	6.3E-126	1.0E-147	99.8%	2.61
		BCAT1	0.88	8.3E-126	1.4E-146	97.9%	3.75
		FAM49A	0.87	1.9E-135	1.3E-125	89.5%	4.50
		ITGBL1	0.87	1.1E-122	4.0E-157	98.3%	3.93
Monocyte	*IL1B*$^+$ pro-inflammatory (SC-M1)	*NR4A2*	0.87	1.6E-86	1.5E-76	97.4%	4.71
		HSPA1A	0.82	2.1E-64	3.4E-53	95.2%	3.96
		ATF3	0.82	1.2E-64	9.1E-71	91.2%	5.05
		FOSB	0.80	3.0E-58	4.8E-43	98.8%	2.85
		JUND	0.79	2.0E-54	5.8E-42	100.0%	1.91
		RGS2	0.79	1.4E-56	4.7E-54	85.3%	4.58
		JUNB	0.78	2.0E-49	2.7E-27	96.2%	2.06
		NR4A1	0.78	5.0E-51	1.4E-53	87.4%	4.22
		ZFP36	0.78	5.7E-49	1.8E-32	92.9%	2.81
		HSP90AA1	0.78	2.7E-48	1.6E-41	99.8%	1.81
		CD83	0.78	3.2E-49	4.3E-47	90.7%	3.77
		HSPH1	0.78	2.5E-49	1.6E-39	83.4%	3.75
		NAMPT	0.77	2.8E-47	8.4E-37	87.6%	3.33
		PNRC1	0.76	4.0E-43	6.2E-20	91.2%	2.03
		DUSP1	0.76	3.7E-42	2.1E-35	97.1%	2.68
		NR4A3	0.75	1.3E-51	1.2E-51	64.6%	4.35
		KDM6B	0.75	3.1E-44	2.6E-39	75.3%	3.70
		HSPA1B	0.75	1.2E-42	1.5E-40	81.0%	3.94
		FOS	0.74	6.1E-38	3.9E-20	97.6%	1.64
		DNAJB1	0.74	1.4E-39	3.6E-33	79.3%	3.40
Monocyte	*NUPR1*$^+$ (SC-M2)	*GPNMB*	0.81	1.1E-52	2.1E-62	97.1%	3.76
		VSIG4	0.80	8.0E-49	9.4E-45	96.8%	3.01
		HTRA1	0.79	1.4E-70	1.0E-55	69.3%	4.64
		PLEC	0.79	1.2E-49	1.1E-58	88.6%	4.18
		CD84	0.79	2.0E-45	9.4E-54	97.1%	3.21
		CTD-2337J16.1	0.78	8.2E-53	1.0E-50	76.4%	4.34
		LGMN	0.77	8.5E-41	1.0E-46	91.8%	3.86
		ABCC3	0.77	2.3E-45	2.8E-48	81.1%	4.07
		ITGB5	0.76	1.9E-45	2.4E-44	74.3%	3.91
		DAB2	0.75	1.3E-33	3.7E-48	98.6%	3.05
		TNS1	0.75	5.2E-36	5.5E-44	85.4%	3.73
		MERTK	0.75	4.0E-46	1.6E-40	66.8%	3.79
		TGFBI	0.75	6.9E-33	1.0E-34	98.6%	2.28
		ADORA3	0.75	8.0E-48	1.5E-41	65.7%	3.93
		MCOLN3	0.74	3.3E-66	6.3E-41	53.9%	3.76
		LILRB5	0.74	3.4E-38	5.0E-38	75.7%	3.81
		ANXA2	0.74	3.8E-32	1.5E-30	96.8%	2.32
		TIMP2	0.74	8.2E-32	2.7E-38	96.1%	2.69
		EMP1	0.74	1.6E-33	1.1E-36	82.9%	3.73
		FOLR2	0.73	2.3E-31	1.6E-51	92.5%	3.78
Monocyte	*C1QA*$^+$ (SC-M3)	*C1QB*	0.78	9.2E-24	9.1E-25	98.4%	2.17
		C1QA	0.77	2.3E-23	1.1E-25	99.2%	1.99
		B2M	0.73	1.7E-17	5.3E-18	100.0%	0.60
		CD74	0.73	2.8E-16	1.1E-19	100.0%	0.52
		HLA-C	0.72	6.3E-16	1.4E-06	96.9%	1.03
		TMEM176B	0.72	2.8E-16	6.0E-13	87.4%	2.59
		RNASE1	0.72	1.1E-15	3.2E-19	99.2%	1.68
		PLTP	0.72	1.5E-15	4.0E-16	96.9%	2.22
		TMEM176A	0.70	1.3E-14	3.0E-13	81.1%	2.87
		IFI27	0.70	3.6E-18	2.4E-12	62.2%	3.55
		C1QC	0.70	3.6E-13	2.4E-18	96.9%	2.16
		MARCO	0.70	3.1E-13	1.4E-12	86.6%	2.85
		CD14	0.69	3.4E-12	7.5E-08	96.9%	1.25
		HLA-B	0.68	2.3E-11	5.1E-03	98.4%	0.43
		TMSB4X	0.68	3.0E-11	1.9E-16	100.0%	0.71
		C2	0.68	2.3E-11	4.4E-09	83.5%	2.21
		RNA28S5	0.67	2.1E-10	1.1E-08	75.6%	2.70
		IFITM3	0.66	1.6E-09	3.8E-07	85.8%	1.84

		SERPING1	0.66	3.6E-10	4.0E-07	70.1%	2.09
		SDC3	0.66	1.9E-10	2.5E-08	68.5%	2.31
Monocyte	IFN-activated (SC-M4)	*SPP1*	0.77	6.6E-37	1.5E-18	63.4%	6.31
		S100A10	0.75	3.6E-18	6.3E-06	92.0%	1.53
		TMSB10	0.74	5.0E-16	2.2E-02	93.8%	0.63
		MT2A	0.73	1.1E-20	4.5E-13	63.4%	3.87
		GAPDH	0.73	2.4E-15	6.2E-07	82.1%	2.11
		B2M	0.73	5.1E-15	4.0E-14	100.0%	0.59
		SH3BGRL3	0.73	7.1E-15	7.3E-05	91.1%	1.33
		FN1	0.72	5.6E-14	2.4E-06	79.5%	2.36
		ENO1	0.71	1.6E-13	2.7E-05	82.1%	1.72
		CSTB	0.71	9.2E-13	6.3E-04	81.3%	1.45
		MALAT1	0.70	1.7E-11	1.7E-06	100.0%	0.57
		SERPINA1	0.69	1.7E-11	5.2E-08	75.0%	2.44
		HLA-B	0.69	1.2E-10	2.0E-04	99.1%	0.52
		VCAN	0.69	8.2E-12	5.2E-08	68.8%	2.75
		IFI6	0.68	6.9E-14	5.3E-09	56.3%	3.07
		TIMP1	0.68	3.2E-10	3.1E-04	83.0%	1.47
		LCP1	0.68	3.8E-10	2.9E-04	81.3%	1.43
		LY6E	0.68	8.5E-17	8.9E-10	49.1%	2.93
		ALDOA	0.67	1.9E-10	6.3E-05	63.4%	1.87
		S100A11	0.67	2.0E-09	3.4E-02	92.9%	0.61
T cell	*CCR7*+ (SC-T1)	*PABPC1*	0.78	1.4E-93	4.4E-53	98.8%	1.50
		YPEL5	0.78	9.6E-92	1.6E-92	95.0%	3.37
		ZNF331	0.75	1.4E-78	2.9E-97	91.1%	4.16
		TPT1	0.75	7.5E-76	6.3E-59	100.0%	0.94
		SLC2A3	0.75	2.2E-74	8.7E-66	89.0%	3.21
		TMEM66	0.74	3.2E-67	3.2E-55	97.3%	2.03
		RPS8	0.73	1.9E-61	7.1E-46	99.2%	1.30
		RPLP2	0.72	1.7E-59	2.0E-37	99.5%	0.83
		RPS6	0.72	5.5E-58	1.1E-46	100.0%	0.78
		CCR7	0.72	2.2E-73	5.9E-67	65.2%	3.86
		IRS2	0.72	6.0E-83	1.1E-71	58.7%	3.77
		IL7R	0.72	4.5E-57	3.3E-81	94.1%	3.57
		SC5D	0.71	8.2E-72	1.1E-62	62.3%	3.58
		GPR183	0.71	5.0E-65	4.3E-60	68.6%	3.67
		CREM	0.71	1.7E-55	2.7E-74	89.9%	3.47
		RPL11	0.71	1.3E-51	2.1E-36	99.7%	0.89
		RPL34	0.71	2.0E-51	5.1E-40	99.8%	0.77
		ZBTB10	0.70	2.1E-71	3.7E-62	58.3%	3.55
		RPS12	0.70	6.1E-49	2.1E-45	100.0%	0.91
		JUND	0.70	1.1E-47	6.5E-46	99.4%	1.35
T cell	Treg (SC-T2)	*FOXP3*	0.92	5.3E-115	5.4E-22	88.0%	7.58
		TIGIT	0.81	2.4E-18	9.2E-17	90.0%	5.23
		IKZF2	0.81	1.8E-30	1.6E-12	74.0%	5.48
		DUSP4	0.80	1.6E-17	4.7E-12	82.0%	5.19
		CTLA4	0.77	1.3E-18	3.8E-10	72.0%	4.71
		ARID5B	0.76	3.8E-11	1.2E-11	90.0%	3.85
		STK17B	0.73	1.4E-08	2.7E-08	96.0%	2.08
		PELI1	0.73	7.0E-14	6.5E-08	66.0%	3.89
		F5	0.73	6.4E-37	1.1E-07	50.0%	3.87
		RGS1	0.72	3.3E-08	6.7E-25	100.0%	3.93
		PKM	0.72	2.7E-08	5.9E-07	84.0%	2.93
		CD4	0.71	4.6E-08	1.2E-09	88.0%	3.53
		ENTPD1	0.71	6.5E-26	6.8E-07	48.0%	3.58
		YWHAB	0.70	8.7E-07	8.0E-18	100.0%	1.68
		IL32	0.70	9.2E-07	3.3E-06	90.0%	2.50
		IKZF4	0.70	2.1E-27	1.2E-06	46.0%	3.17
		BIRC3	0.69	2.6E-06	4.8E-10	96.0%	2.86
		AC013394.2	0.69	2.4E-06	1.2E-05	86.0%	2.37
		GBP2	0.69	8.3E-07	3.0E-05	74.0%	2.74
		SNX9	0.69	3.7E-10	8.7E-06	56.0%	3.14
		CXCL13	0.85	7.1E-126	2.0E-19	73.2%	7.50
		RNF19A	0.81	1.7E-20	3.8E-22	93.0%	4.43
		DUSP4	0.79	8.1E-23	2.6E-16	81.7%	5.08
		PDCD1	0.79	3.3E-37	7.1E-16	71.8%	4.95
		TOX2	0.78	4.8E-46	2.8E-14	64.8%	4.71
		MAF	0.77	2.1E-16	2.5E-11	81.7%	4.01
		CTLA4	0.76	3.7E-24	1.3E-12	69.0%	4.59
		TSHZ2	0.76	1.8E-27	4.0E-13	67.6%	4.57
		TIGIT	0.74	3.4E-15	4.4E-12	77.5%	3.96

T cell	Tph/Tfh (SC-T3)	*SNX9*	0.74	1.2E-20	5.3E-11	64.8%	3.93
		PTPN13	0.73	1.5E-44	1.1E-10	52.1%	4.19
		NR3C1	0.73	5.6E-12	5.5E-12	88.7%	3.25
		ITM2A	0.73	1.4E-11	1.8E-12	88.7%	3.23
		FKBP5	0.72	2.1E-12	2.0E-10	78.9%	3.45
		TP53INP1	0.72	8.9E-28	4.6E-10	54.9%	3.76
		ARID5B	0.72	2.9E-11	1.5E-13	88.7%	3.53
		TOX	0.71	3.1E-17	3.5E-10	63.4%	3.58
		COTL1	0.71	1.9E-10	1.3E-11	85.9%	3.15
		AHI1	0.71	1.5E-13	5.7E-09	66.2%	3.54
		ZNRF1	0.71	9.4E-21	3.8E-09	54.9%	3.47
T cell	*GZMK*$^+$ (SC-T4)	*CCL5*	0.86	1.9E-174	0.0E+00	98.7%	7.58
		GZMA	0.77	7.9E-107	4.9E-118	76.9%	4.93
		GZMK	0.77	8.1E-113	1.8E-109	70.3%	5.07
		CST7	0.72	3.5E-67	2.4E-64	69.1%	3.63
		CCL4	0.71	1.2E-73	2.7E-68	62.7%	4.44
		RUNX3	0.70	4.6E-48	1.0E-58	87.9%	3.04
		HLA-B	0.69	6.8E-42	3.0E-35	100.0%	0.46
		HLA-C	0.68	1.1E-40	5.5E-28	99.6%	0.71
		NKG7	0.67	1.2E-56	4.7E-50	51.0%	3.26
		HCST	0.66	2.8E-33	4.1E-27	74.0%	2.16
		CTSW	0.66	2.4E-43	1.5E-38	51.0%	2.68
		CD8A	0.65	1.6E-49	1.4E-41	45.4%	2.91
		CD74	0.65	3.1E-27	3.8E-29	94.8%	1.73
		APOBEC3G	0.64	4.0E-35	1.7E-31	51.9%	2.41
		KLRG1	0.64	1.8E-48	8.5E-38	39.4%	2.44
		HLA-A	0.64	4.2E-24	1.4E-19	99.9%	0.50
		SAMD3	0.64	1.7E-38	1.2E-32	44.2%	2.29
		LITAF	0.63	7.6E-24	3.5E-22	72.5%	1.99
		PIP4K2A	0.63	3.5E-24	2.1E-22	69.7%	1.96
		ITGB2	0.63	1.7E-22	7.4E-23	84.2%	1.79
T cell	CTLs (SC-T5)	*NKG7*	0.91	1.2E-44	1.1E-53	98.4%	7.48
		GNLY	0.90	1.4E-105	8.4E-25	84.4%	9.10
		FGFBP2	0.89	1.3E-188	1.5E-22	79.7%	7.58
		EFHD2	0.85	3.1E-30	9.7E-21	87.5%	5.31
		GZMH	0.85	7.1E-48	3.9E-21	82.8%	6.39
		ZEB2	0.83	1.9E-34	1.2E-17	79.7%	5.42
		CCL5	0.83	3.1E-21	1.3E-127	100.0%	5.96
		GZMB	0.81	1.5E-55	5.6E-16	71.9%	5.60
		PRF1	0.81	9.3E-38	9.0E-15	71.9%	5.22
		CCL4	0.80	2.3E-22	7.6E-19	87.5%	6.32
		GPR56	0.80	1.3E-65	8.0E-14	64.1%	5.37
		PLEK	0.79	2.6E-34	6.4E-14	70.3%	4.97
		HLA-B	0.79	3.1E-15	3.5E-19	100.0%	0.64
		CX3CR1	0.77	2.0E-105	3.9E-12	56.3%	5.02
		TGFBR3	0.77	9.5E-23	7.4E-12	70.3%	4.43
		JUN	0.77	1.1E-13	8.0E-13	95.3%	3.11
		ITGAL	0.77	1.3E-13	3.1E-13	93.8%	2.86
		FLNA	0.77	3.8E-14	4.8E-11	87.5%	3.27
		RUNX3	0.77	1.3E-13	1.1E-27	98.4%	3.57
		SYNE1	0.75	5.1E-13	9.5E-10	84.4%	3.20
T cell	*GZMK*$^+$*/GZMB*$^+$ (SC-T6)	*GAPDH*	0.90	1.3E-46	1.6E-39	96.4%	4.09
		CD74	0.86	1.6E-36	4.4E-40	99.1%	3.12
		HLA-DRA	0.85	3.7E-50	7.8E-39	89.2%	5.79
		HLA-DRB1	0.85	8.1E-48	2.4E-36	88.3%	5.81
		CFL1	0.84	6.6E-34	3.7E-25	99.1%	1.68
		ACTB	0.83	1.6E-32	4.8E-53	100.0%	2.05
		PFN1	0.83	3.5E-32	1.1E-28	98.2%	2.35
		HLA-DQA1	0.83	8.0E-69	1.1E-29	77.5%	5.68
		CCL5	0.83	5.2E-35	1.3E-29	90.1%	5.41
		COTL1	0.81	9.1E-32	2.6E-29	91.0%	4.15
		HLA-DRB5	0.81	9.0E-72	1.8E-24	69.4%	5.71
		HLA-DPA1	0.81	1.7E-32	4.6E-25	85.6%	4.64
		APOBEC3G	0.80	7.1E-36	4.8E-28	83.8%	4.89
		HLA-DPB1	0.79	2.1E-30	2.4E-23	82.9%	4.50
		GZMA	0.79	7.7E-30	7.6E-26	85.6%	4.76
		CORO1A	0.78	1.4E-31	9.3E-25	82.0%	4.47
		NKG7	0.78	1.0E-36	2.1E-26	80.2%	5.23
		TMSB4X	0.78	1.2E-22	5.4E-27	100.0%	0.99
		LCP1	0.78	1.1E-22	1.4E-20	97.3%	2.18
		CD8A	0.77	3.7E-37	2.5E-22	73.9%	5.05

B cell	Naïve B cells (SC-B1)	*CD83*	0.86	4.8E-97	8.3E-162	98.9%	5.40
		KDM6B	0.85	7.1E-100	1.0E-116	90.8%	5.39
		CXCR4	0.84	1.7E-82	6.0E-102	99.5%	3.43
		NR4A2	0.83	1.6E-83	5.1E-153	99.7%	5.06
		CD69	0.82	1.4E-77	9.2E-121	96.8%	5.24
		YPEL5	0.81	2.0E-70	1.9E-64	95.1%	3.26
		JUNB	0.80	7.3E-67	1.0E-53	96.8%	2.74
		REL	0.80	4.8E-71	2.7E-96	92.7%	4.69
		BTG1	0.80	1.5E-66	7.4E-80	99.7%	2.46
		ZNF331	0.79	6.9E-71	3.2E-65	81.1%	4.95
		NR4A3	0.79	2.6E-82	2.4E-71	74.3%	5.03
		LY9	0.78	5.0E-61	2.7E-75	89.5%	4.53
		SLC2A3	0.78	6.0E-64	2.7E-73	85.1%	4.56
		CD55	0.77	9.7E-57	8.4E-57	88.9%	3.62
		ZBTB10	0.76	1.6E-63	1.1E-56	73.8%	4.34
		DNAJA1	0.76	1.3E-50	1.6E-57	93.8%	3.22
		CXCR5	0.76	1.3E-65	2.4E-56	71.4%	4.41
		NFKBID	0.76	1.6E-58	1.6E-62	79.5%	4.36
		VPS37B	0.75	7.1E-61	1.5E-48	70.0%	4.04
		CREM	0.75	2.6E-54	2.1E-49	76.8%	4.09
B cell	Memory B cells (SC-B2)	*CD74*	0.80	2.0E-86	3.3E-92	100.0%	1.44
		MS4A1	0.76	3.9E-65	2.0E-87	96.6%	4.46
		HLA-DPB1	0.74	1.1E-56	6.9E-78	97.4%	4.06
		RPL34	0.74	9.8E-55	2.7E-38	99.7%	0.86
		RPL30	0.74	2.5E-53	1.4E-37	99.4%	1.18
		HLA-DPA1	0.73	4.4E-51	1.6E-55	99.1%	2.46
		TMSB4X	0.73	7.0E-49	9.3E-33	96.9%	1.81
		RPS29	0.73	9.0E-49	2.5E-24	98.6%	0.95
		HLA-DRA	0.72	1.4E-48	3.6E-79	99.8%	3.89
		RPS21	0.72	4.1E-47	5.5E-28	99.1%	0.99
		RPL11	0.72	8.5E-47	1.6E-22	99.1%	0.79
		RPL38	0.72	1.3E-45	1.2E-19	97.7%	1.00
		RPLP2	0.72	5.0E-45	4.2E-28	99.4%	0.85
		RPS6	0.71	4.1E-42	2.3E-28	99.8%	0.75
		RPL37	0.70	3.8E-40	1.9E-16	98.0%	0.82
		SELL	0.70	2.1E-42	4.6E-37	75.4%	3.01
		LTB	0.70	2.0E-42	5.0E-44	77.4%	3.46
		RPS12	0.70	3.5E-39	1.9E-20	99.1%	0.80
		1-Mar	0.70	1.2E-42	3.0E-42	68.3%	3.21
		RPS13	0.69	4.4E-36	7.2E-11	96.2%	0.77
B cell	ABCs (SC-B3)	*ITGAX*	0.93	2.1E-111	5.3E-26	90.6%	7.35
		ZEB2	0.85	1.5E-32	3.7E-19	86.8%	5.89
		CD52	0.84	2.3E-17	6.5E-17	96.2%	3.93
		ACTB	0.83	2.0E-16	4.1E-07	96.2%	1.94
		HLA-DPB1	0.83	2.8E-16	1.1E-50	100.0%	3.38
		TMSB4X	0.82	1.3E-15	4.5E-11	98.1%	2.05
		LRMP	0.82	7.8E-18	3.3E-23	96.2%	4.92
		ARPC1B	0.82	6.1E-17	4.2E-24	96.2%	4.44
		ARPC2	0.82	6.2E-15	1.5E-11	98.1%	1.95
		PLEK	0.81	1.2E-19	2.1E-13	83.0%	4.72
		ACTG1	0.81	1.4E-14	1.1E-19	100.0%	2.42
		SH3BGRL3	0.81	1.2E-14	1.2E-10	94.3%	2.84
		CD19	0.80	1.5E-15	1.7E-14	90.6%	4.15
		ARPC5	0.80	2.4E-19	3.3E-12	79.2%	4.48
		COTL1	0.80	5.3E-16	7.8E-15	88.7%	4.37
		SAMD9L	0.80	4.8E-18	1.6E-14	84.9%	4.69
		ARL6IP5	0.79	2.8E-14	4.0E-15	92.5%	3.90
		RFTN1	0.78	5.8E-26	9.1E-11	67.9%	4.56
		IFI44L	0.78	5.3E-53	1.4E-10	60.4%	5.12
		LAPTM5	0.78	2.4E-12	7.8E-18	98.1%	3.26
B cell	plasmablasts (SC-B4)	*MZB1*	1.00	2.4E-249	0.0E+00	99.7%	10.09
		XBP1	0.99	3.1E-235	0.0E+00	99.4%	10.26
		FKBP11	0.99	2.6E-240	0.0E+00	98.3%	9.66
		SSR4	0.99	3.1E-193	0.0E+00	98.6%	8.65
		HSP90B1	0.98	3.7E-164	1.4E-227	99.7%	5.62
		DERL3	0.98	3.5E-223	0.0E+00	97.2%	9.09
		SEC11C	0.97	2.2E-195	0.0E+00	96.9%	8.22
		FKBP2	0.97	8.2E-204	1.8E-297	96.3%	8.15
		RRBP1	0.96	5.6E-229	1.3E-255	95.2%	8.59
		HDLBP	0.96	1.7E-191	3.9E-277	96.1%	7.64
		SEL1L	0.95	1.2E-191	1.1E-240	94.6%	7.85

PRDX4	0.95	3.5E-230	6.1E-196	91.5%	8.37
FNDC3B	0.95	8.9E-234	1.4E-205	92.4%	8.00
SLAMF7	0.95	6.7E-202	5.7E-249	94.6%	8.15
SDF2L1	0.95	2.3E-230	9.7E-195	91.5%	7.88
SDC1	0.95	5.5E-256	2.5E-168	89.6%	8.15
LMAN1	0.94	3.4E-176	3.4E-252	95.5%	7.27
NUCB2	0.94	5.5E-225	1.9E-177	90.1%	7.78
SSR3	0.94	1.4E-186	1.9E-212	93.0%	7.31
ITM2C	0.94	4.9E-174	2.2E-189	92.4%	7.53