Open access • Posted Content • DOI:10.1101/351130

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

Fan Zhang, Kevin Wei, Kamil Slowikowski, Chamith Y. Fonseka ...+32 more authors

Institutions: Brigham and Women's Hospital, Barts Health NHS Trust, Hospital for Special Surgery, University of Rochester Medical Center ...+11 more institutions

Published on: 20 Jun 2018 - bioRxiv (Cold Spring Harbor Laboratory)

Topics: Naive B cell, Mass cytometry, T cell, Flow cytometry and GNLY

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

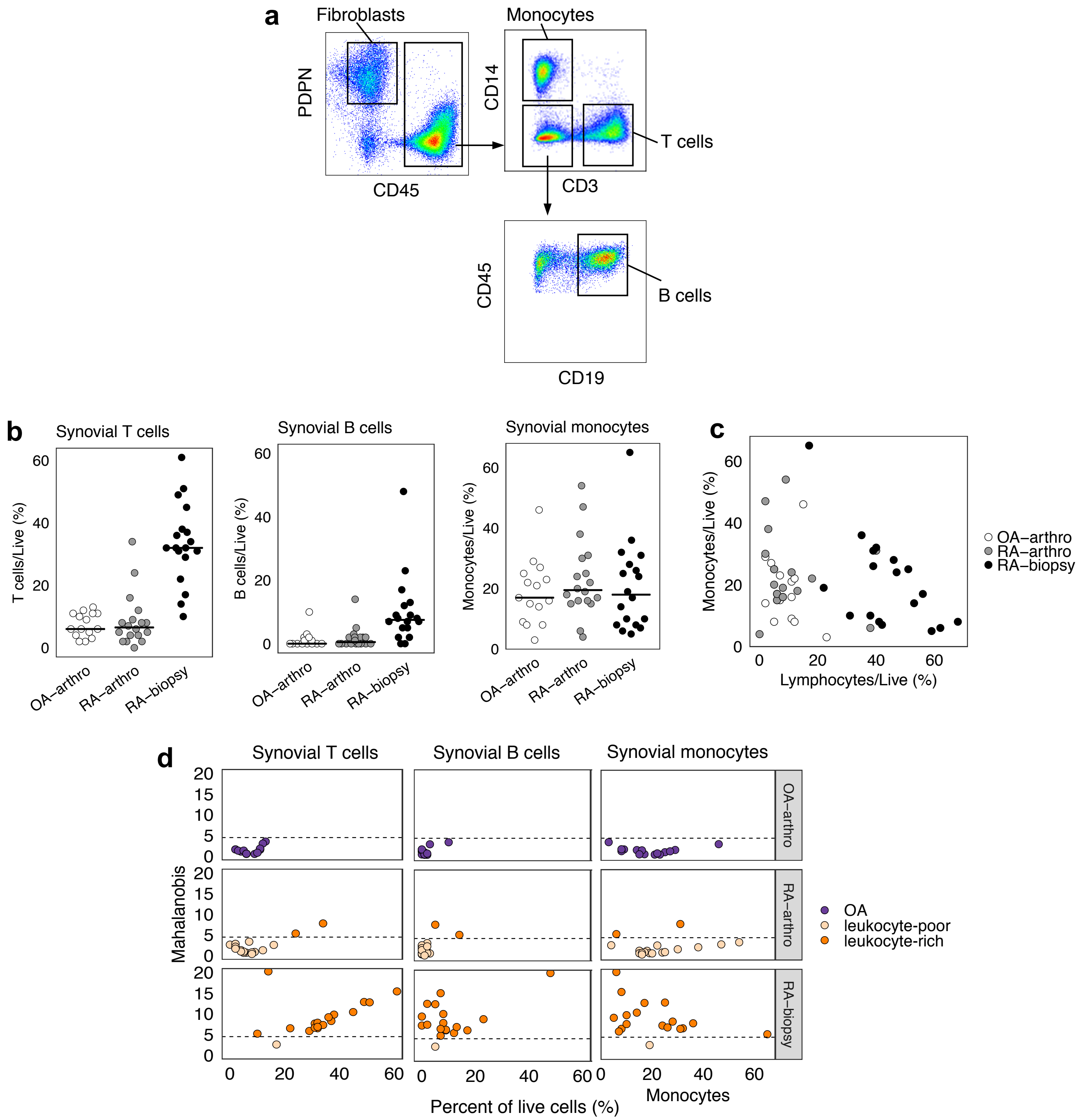
Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium; Filer, Andrew; Buckley, Christopher

DOI:
10.1038/s41590-019-0378-1

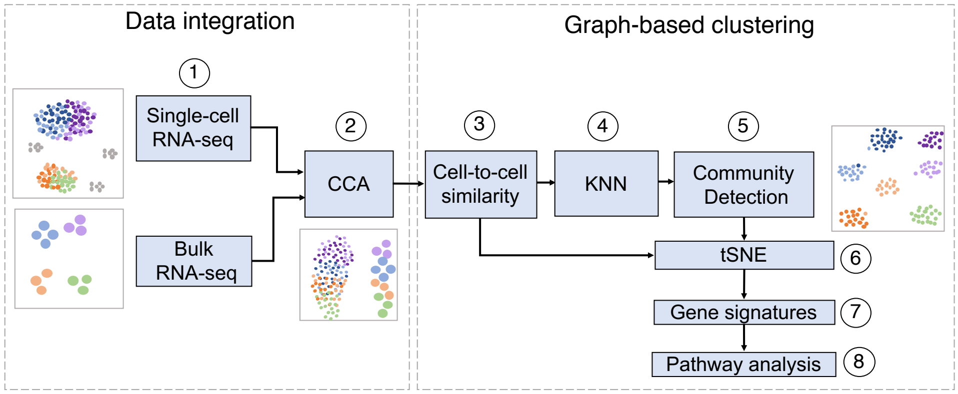
Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry

Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium; Zhang, Fan; Wei, Kevin; Slowikowski, Kamil; Fonseka, Chamith Y; Rao, Deepak A; Kelly, Stephen; Goodman, Susan M; Tabechian, Darren; Hughes, Laura B; Salomon-Escoto, Karen; Watts, Gerald F M; Jonsson, A Helena; Rangel-Moreno, Javier; Meednu, Nida; Rozo, Cristina; Apruzzese, William; Eisenhaure, Thomas M; Lieb, David J; Boyle, David L

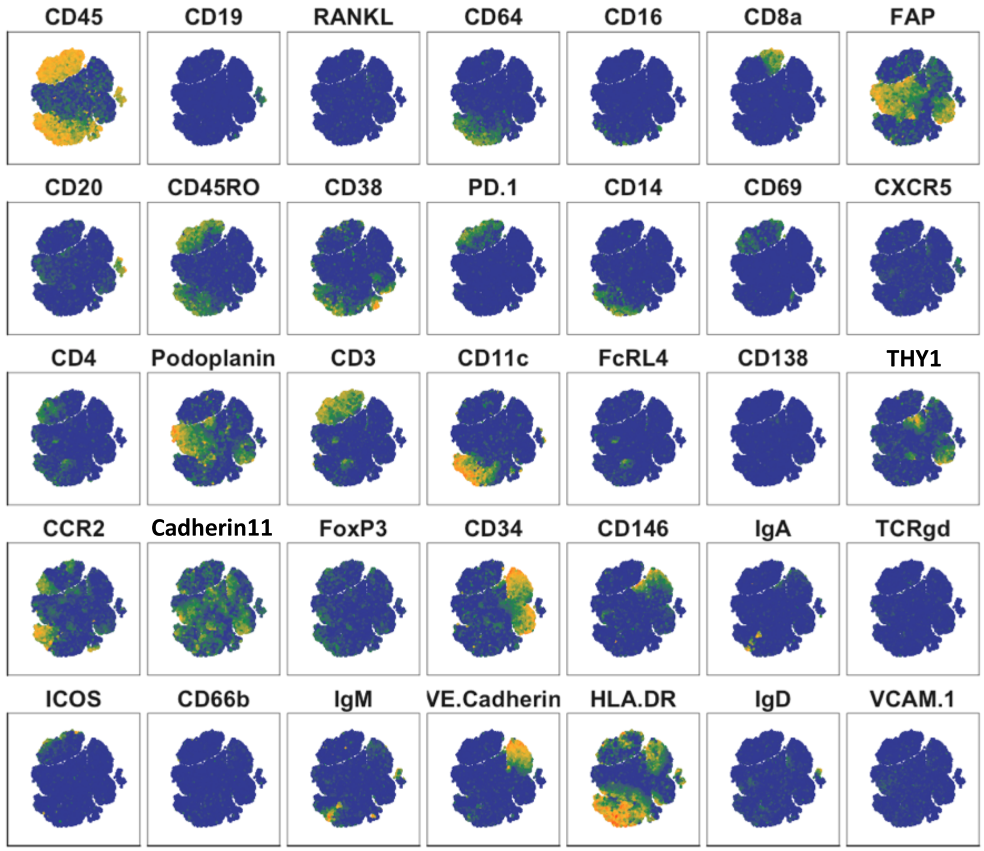
DOI:
10.1038/s41590-019-0378-1



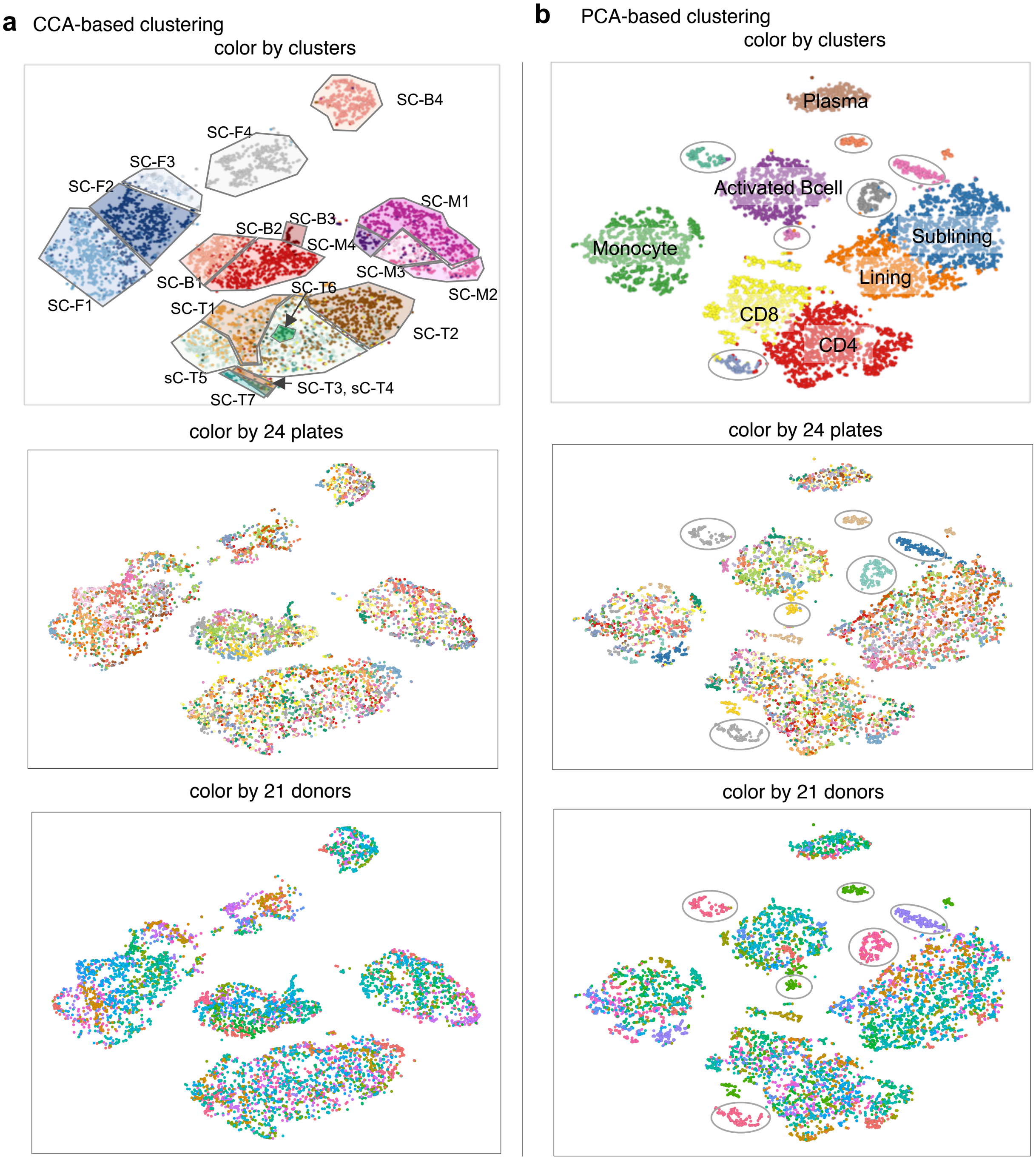
Supplemental Fig. 1. Flow cytometry gating scheme and a data-driven approach to separate samples based on flow cytometry data. **a**. Flow cytometry gating: stromal fibroblasts ($CD45^{-}PDPN^{+}$), monocytes ($CD45^{+}CD14^{+}$), T cells ($CD45^{+}CD3^{+}$), and B cells ($CD45^{+}CD3^{-}CD19^{+}$). **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.



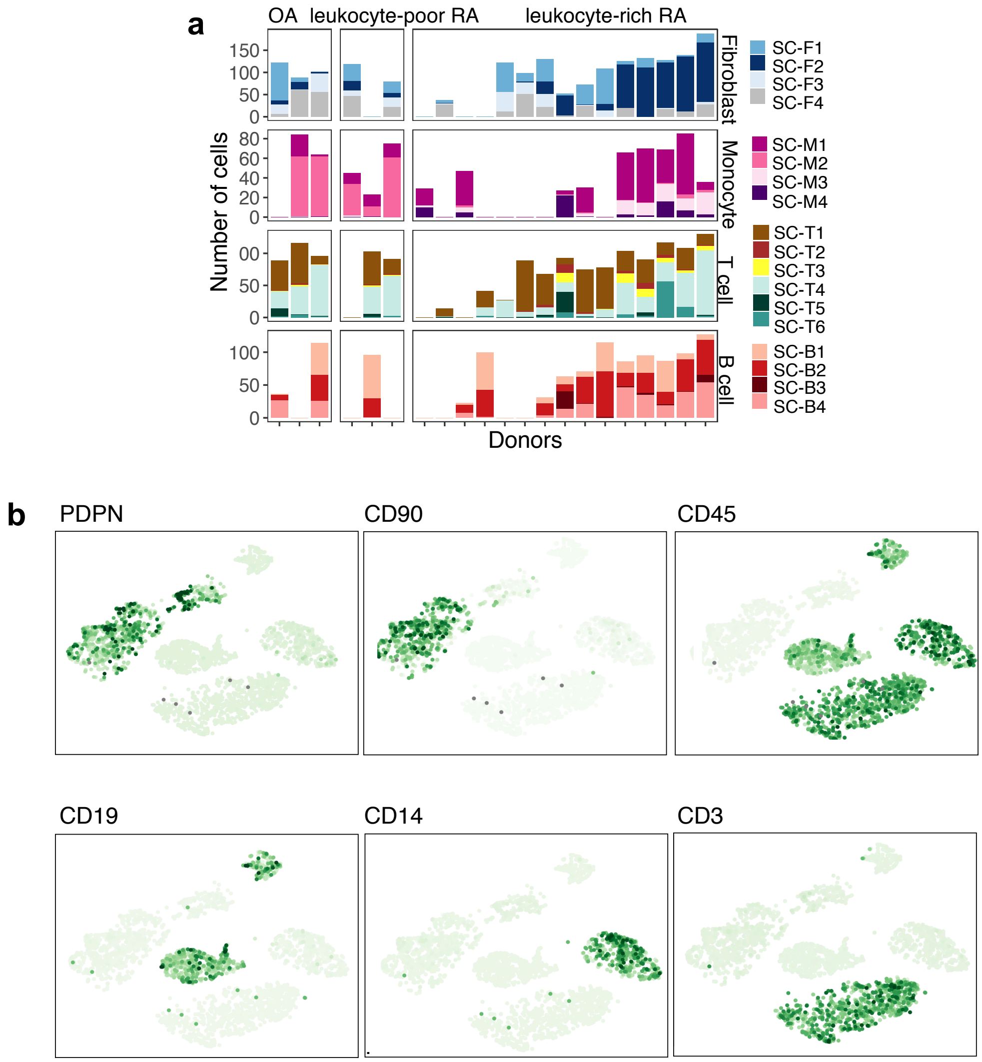
Supplemental Fig. 2. CCA-based integrative pipeline of scRNA-seq analysis: 1) We first select the highly variable genes from both scRNA-seq and bulk RNA-seq; 2) based on the selected genes from both sides we integrate single cells with bulk samples that by learning a linear projection that the correlation 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.



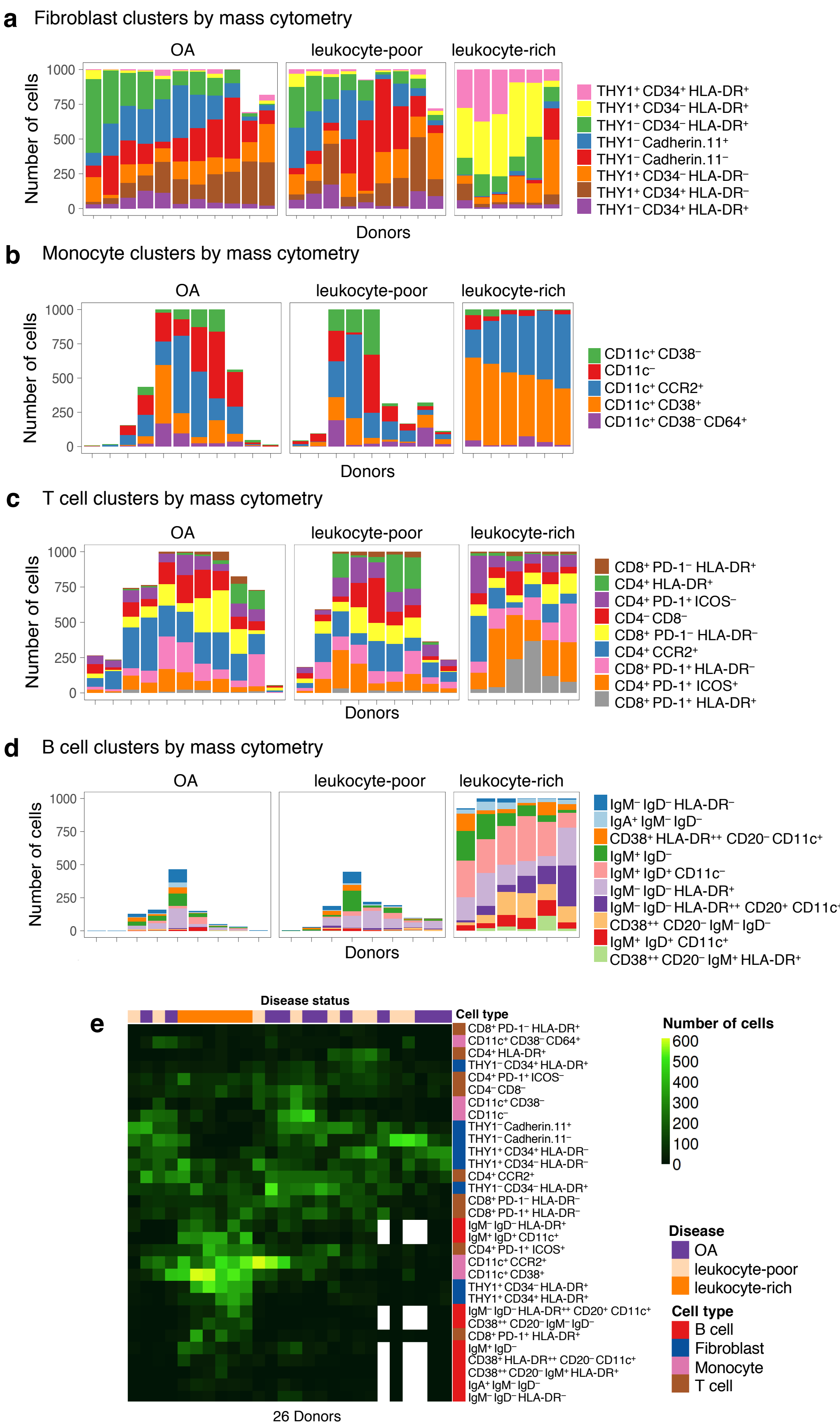
Supplemental Fig. 3. Protein markers of viable and DNA+ synovial cells (3,000 downsampled) from all donors by mass cytometry.



Supplemental Fig. 4. Comparison with PCA-based clustering on batch effect and protein fluorescence validation on each cell from scRNA-seq clusters. **a.** Identified 18 scRNA-seq clusters, source of 24 plates, and source of 21 donors using the CCA-based integrative pipeline. **b**. Identified scRNA-seq clusters, source of plates, and donors using PCA-based clustering by Seurat. Clusters of batch effect plates are highlighted using circles.

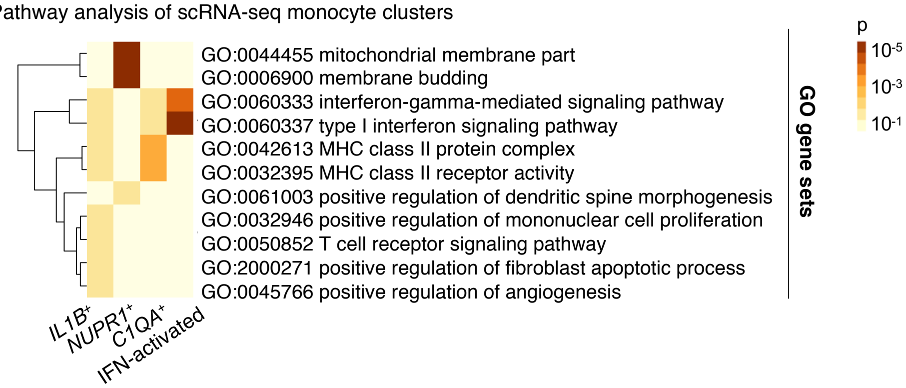


Supplemental Fig. 5. Entropy calculation of mixing of the identified scRNA-seq clusters and protein fluorescence validation on each cell. **a.** Number of cells per donor for each scRNA-seq cluster using CCA-based integrative pipeline**. b.** Flow cytometry protein fluorescence of cell type markers on each single cell.

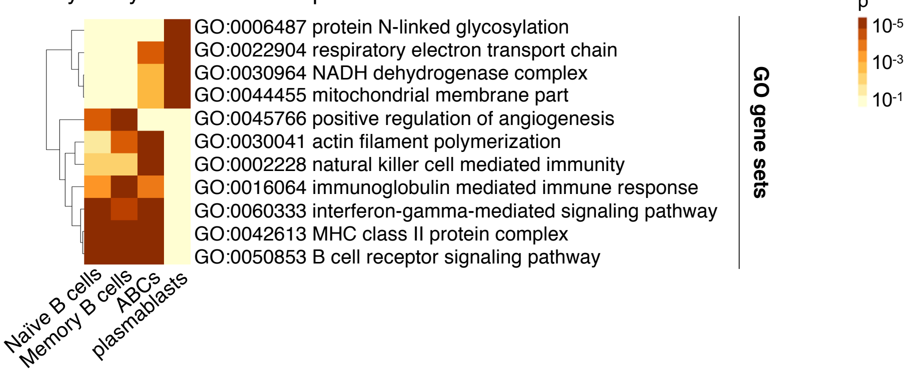


Supplemental Fig. 6. Distribution of identified cell type clusters for each donor by mass cytometry. **a-d**. Distribution of mass cytometry clusters for each cell type confirms that the identified clusters are not confounded by obvious batch effects. **e**. Cell counts of all clusters by comparing all 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⁺).

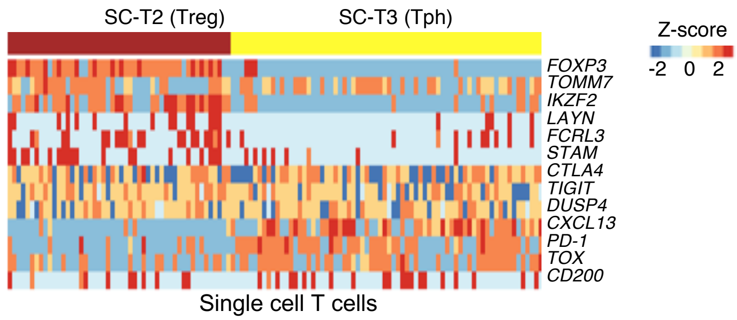
a. Pathway analysis of scRNA-seq monocyte clusters



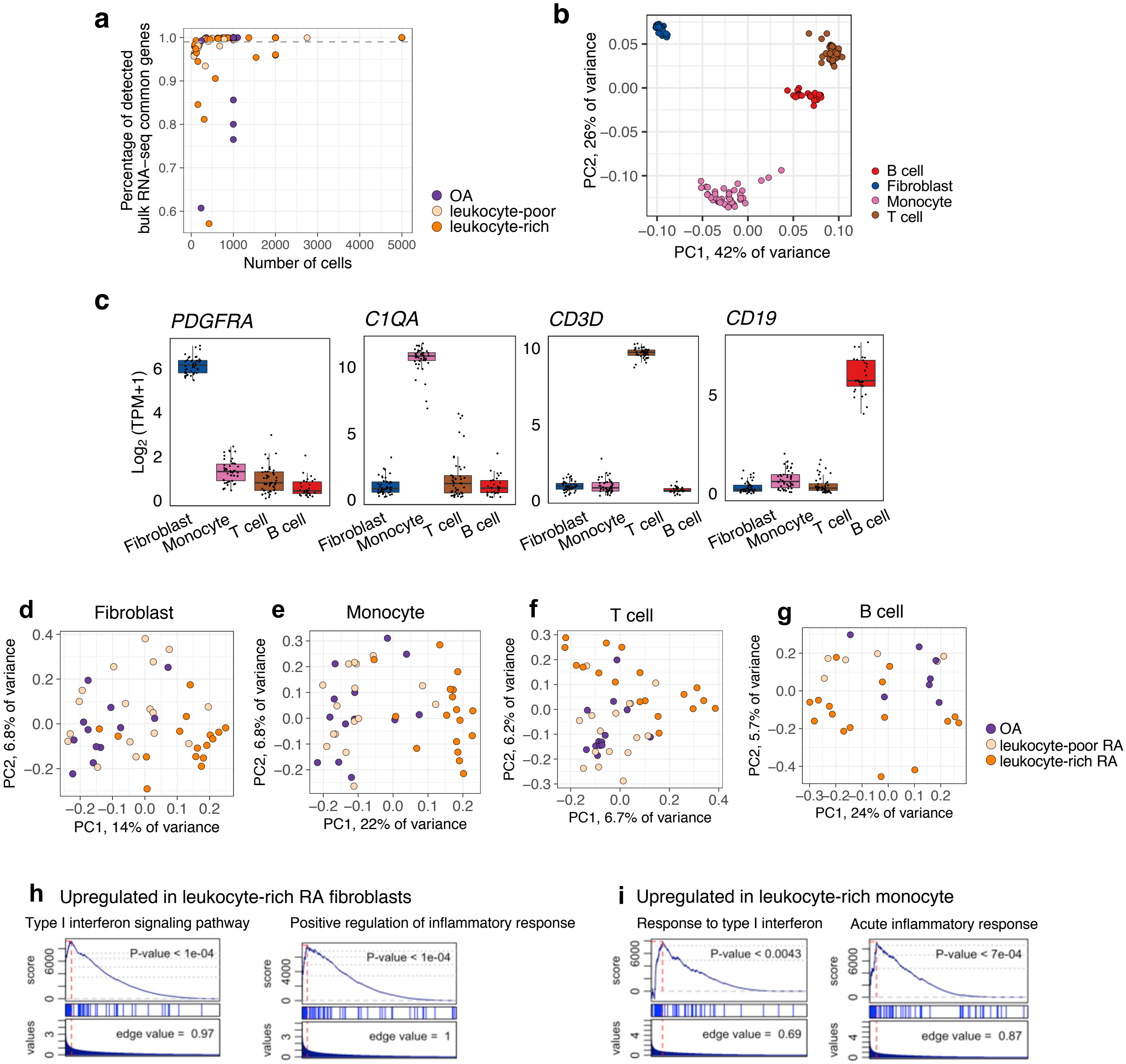
b. Pathway analysis of scRNA-seq B cell clusters



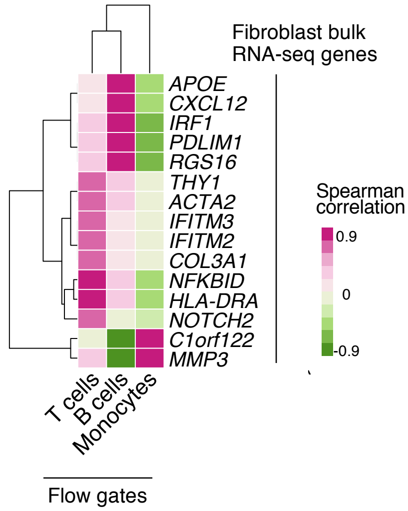
Supplemental Fig.7. Pathway enrichment analysis on GO gene sets for identified scRNA-seq clusters from monocytes and B cells.



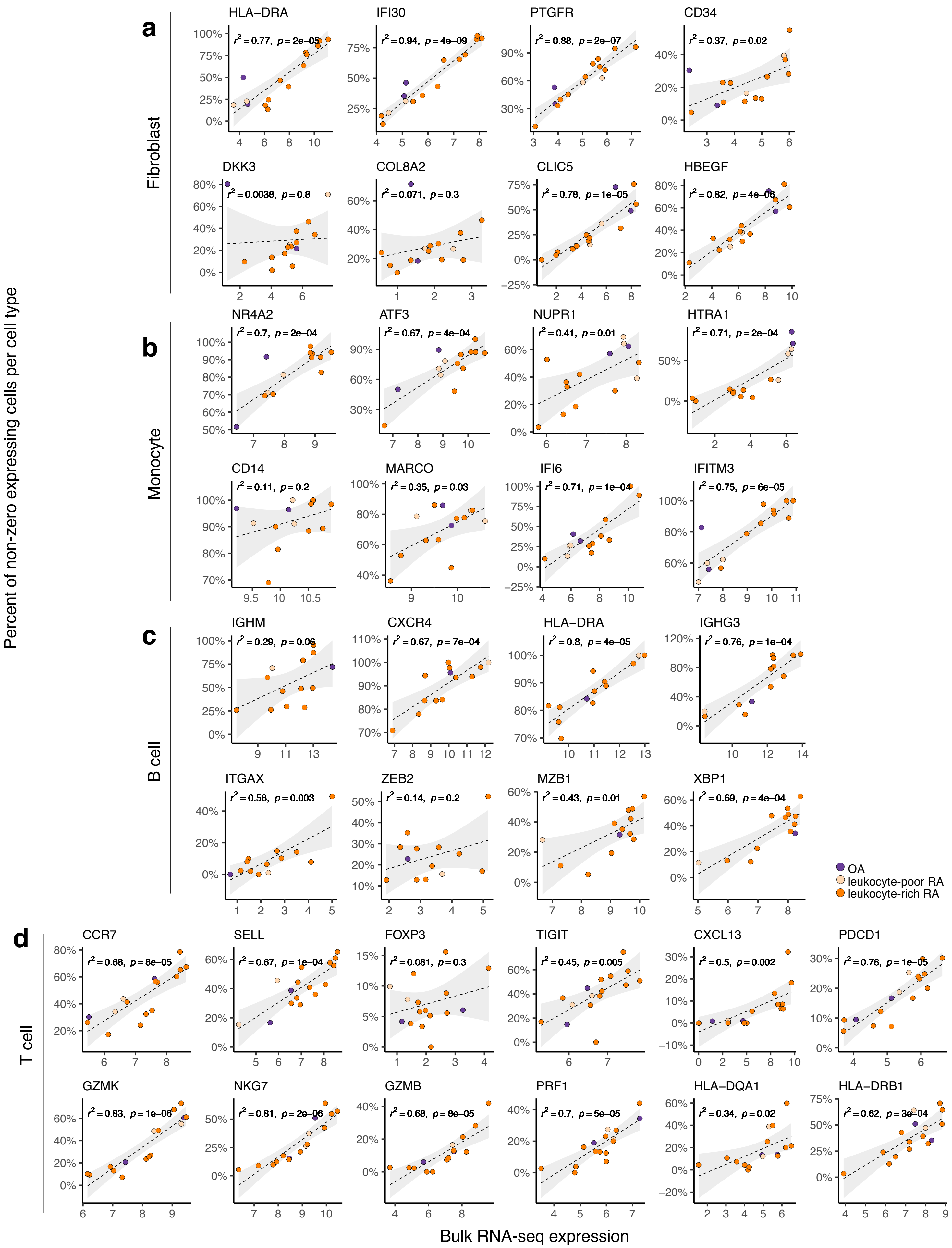
Supplemental Fig. 8. scRNA-seq clusters Tregs (SC-T2) and Tph (SC-T3) that separated based on the most informative markers from (Rao et al. 2017). We use hierarchical clustering with R function hclust() and then cutree(k=2) to pinpoint previously characterized rare cell populations, Tregs and Tph cells.



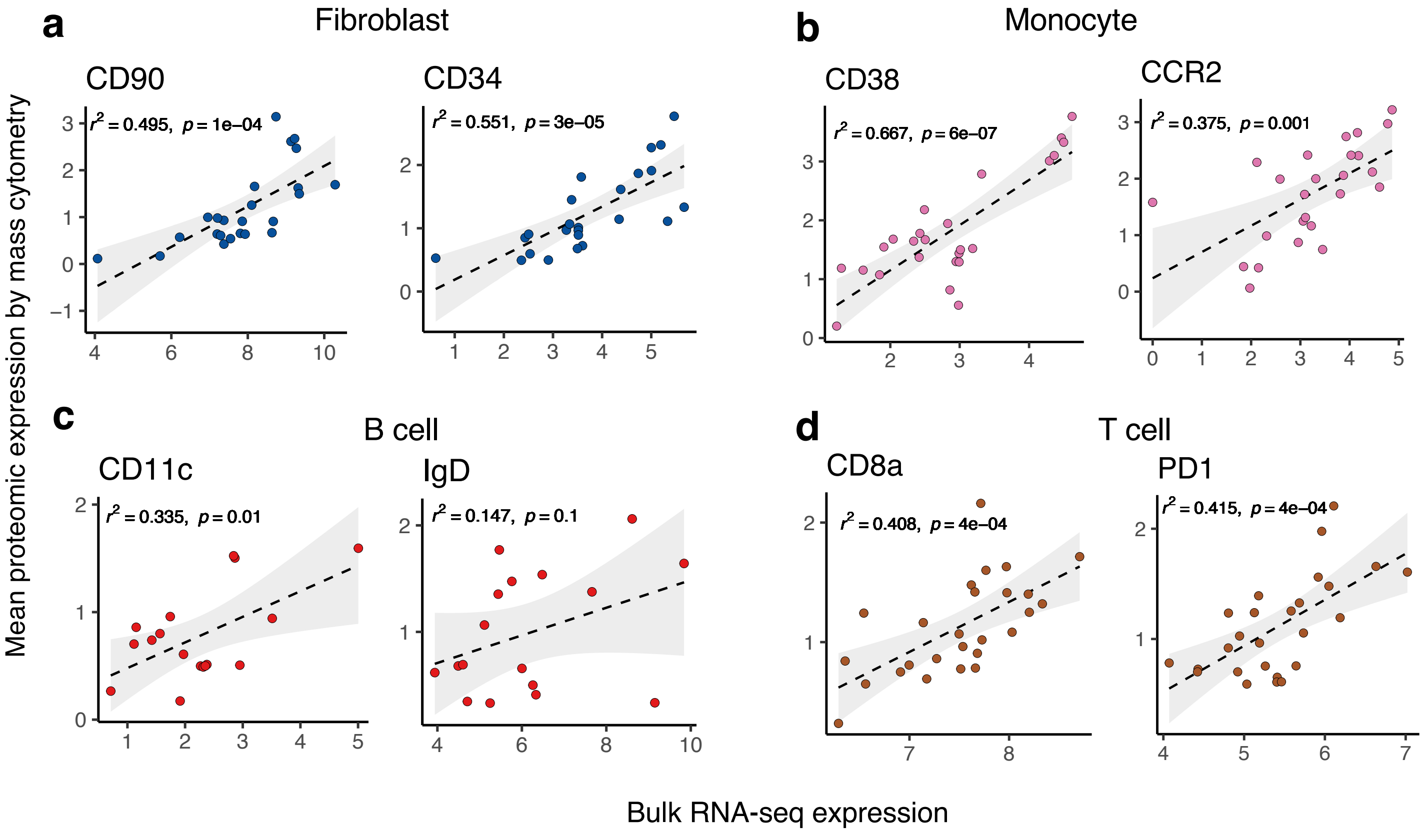
Supplemental Fig. 9. Quality control of bulk RNA-seq and PCA analysis for each cell type samples. **a**. Quality control of bulk RNA-seq samples. Common genes are defined as the set of genes detected with at least 1 mapped fragment in 95% of the samples (13,041 genes). X-axis is the number of cells for 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.



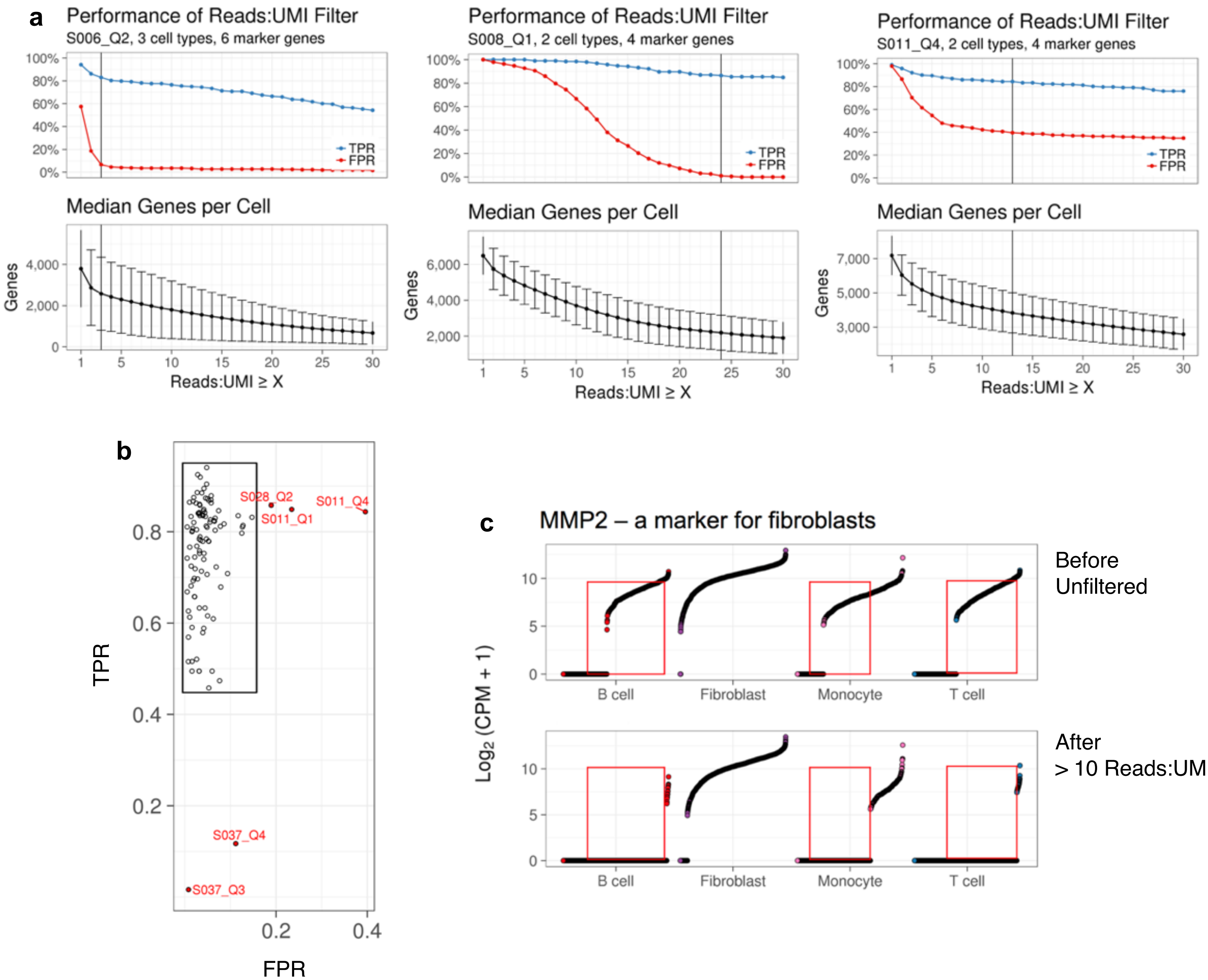
Supplemental Fig. 10. Correlation between bulk RNA-seq genes and immune cell type abundances in RA synovial fibroblasts. Integrating bulk RNA-seq samples from fibroblasts with multiple cell type flow gates reveals that T cells, B cells, and monocytes that are abundant in RA synovial tissue directly influence the expression of fibroblasts in the RA synovium.



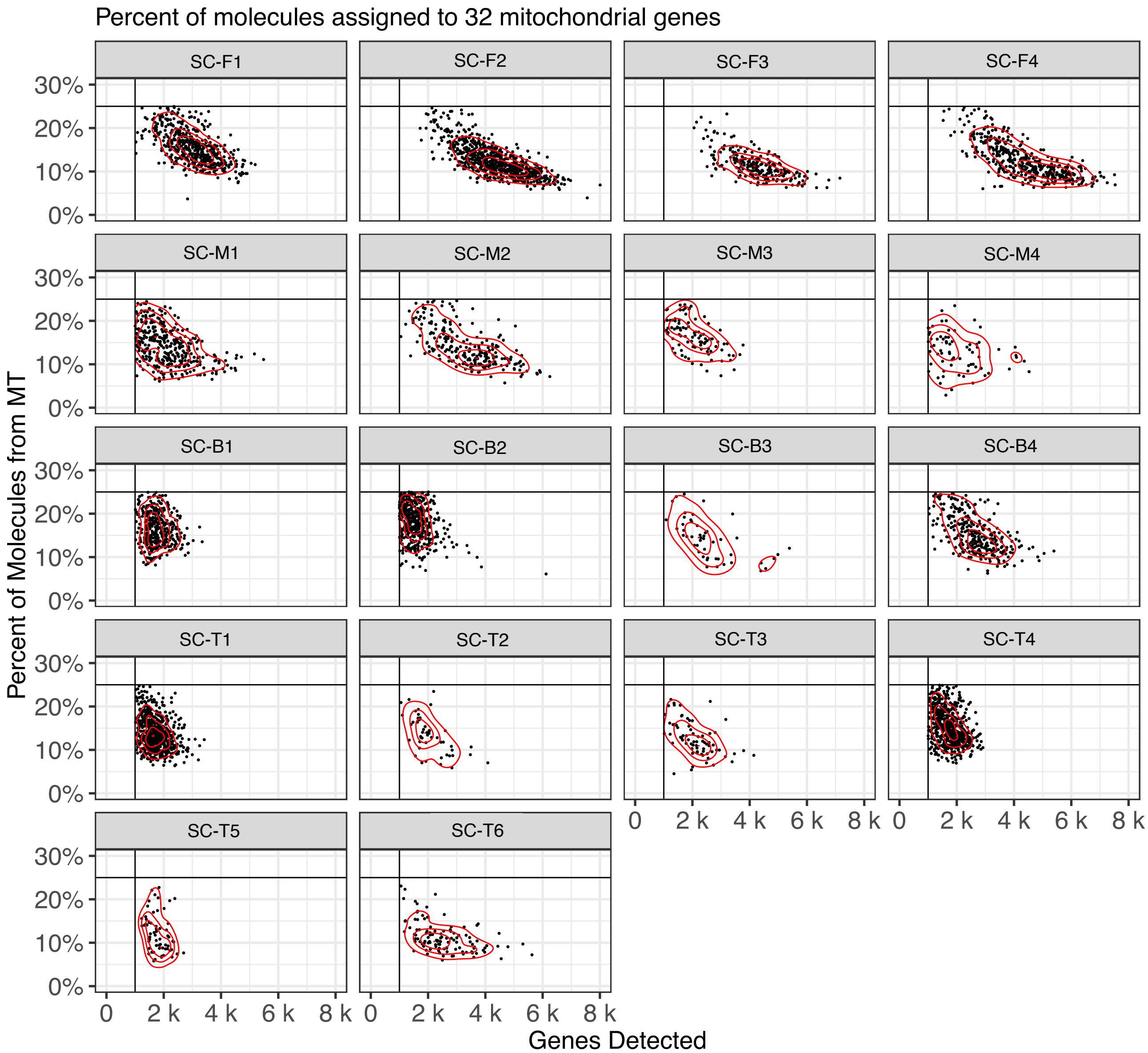
Supplemental Fig. 11. Correlation between bulk RNA-seq expression with proportion of non-zero expressing cells for scRNA-seq cluster markers per cell type. We depict two marker genes per scRNA-seq cluster and show the bulk RNA-seq expression (x-axis) is correlated with the percent of non-zero expressing cells over the total number of cells (y-axis) for the overlapped **a**. fibroblast samples, **b**. monocyte samples, **c**. B cell samples, and **d**. T cell samples. The statistical R-square and p value are given for each correlation.



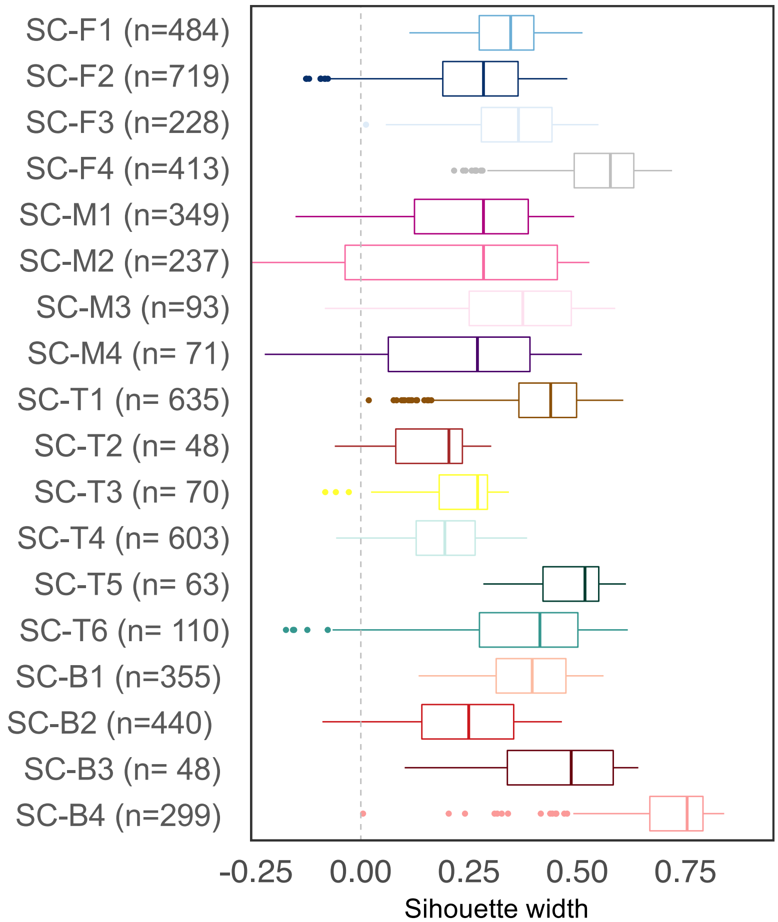
Supplemental Fig. 12. Correlation between mean proteomic expression by mass cytometry and transcriptomic expression by bulk RNA-seq on the overlapped samples. Two typical protein/gene markers per cell type were show for **a**. fibroblast samples, **b**. monocyte samples, **c**. B cell samples, and **d**. T cell samples.



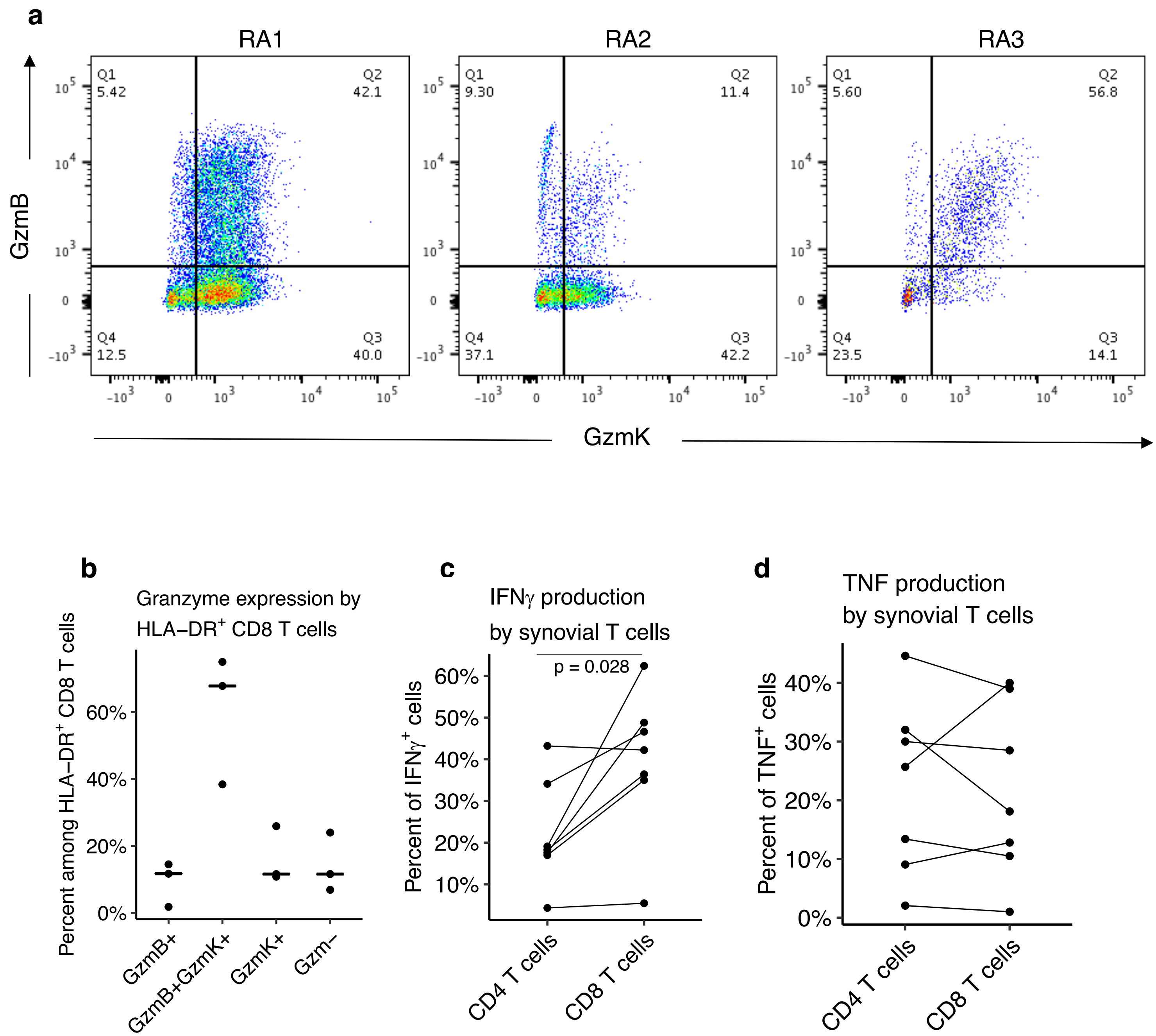
Supplemental Fig. 13. Dynamic filtering strategy for scRNA-seq quality control. **a**. Estimated optimal threshold for three example quadrants of scRNA-seq plates. **b**. Select optimal threshold that maximizes the ratio of TPR (true positive rate) to FPR (false positive rate). We selected 2 marker genes 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*.



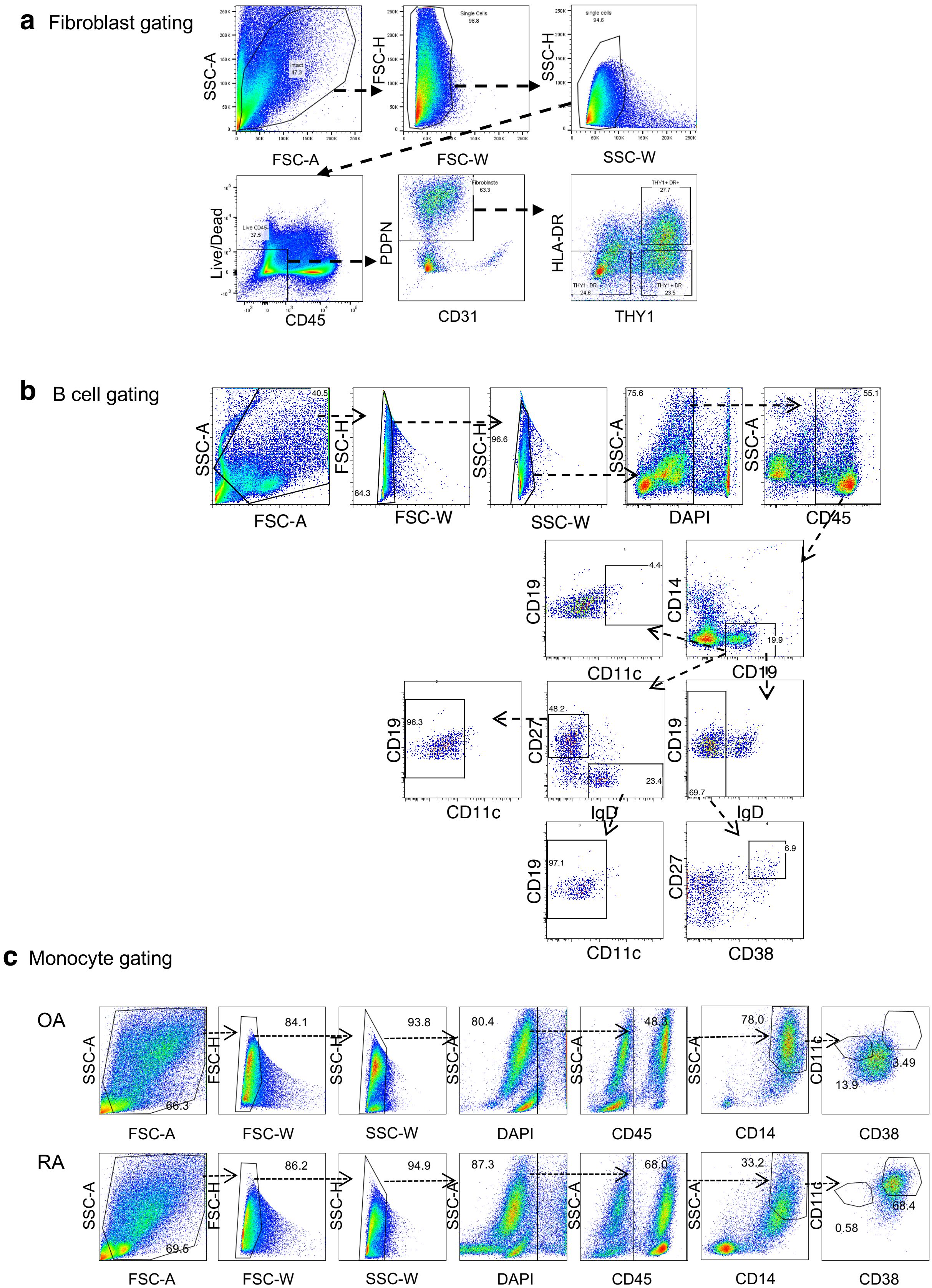
Supplemental Fig. 14. All post-QC scRNA-seq data for each identified cluster based on number of genes detected and percent of molecules from 32 mitochondrial genes.



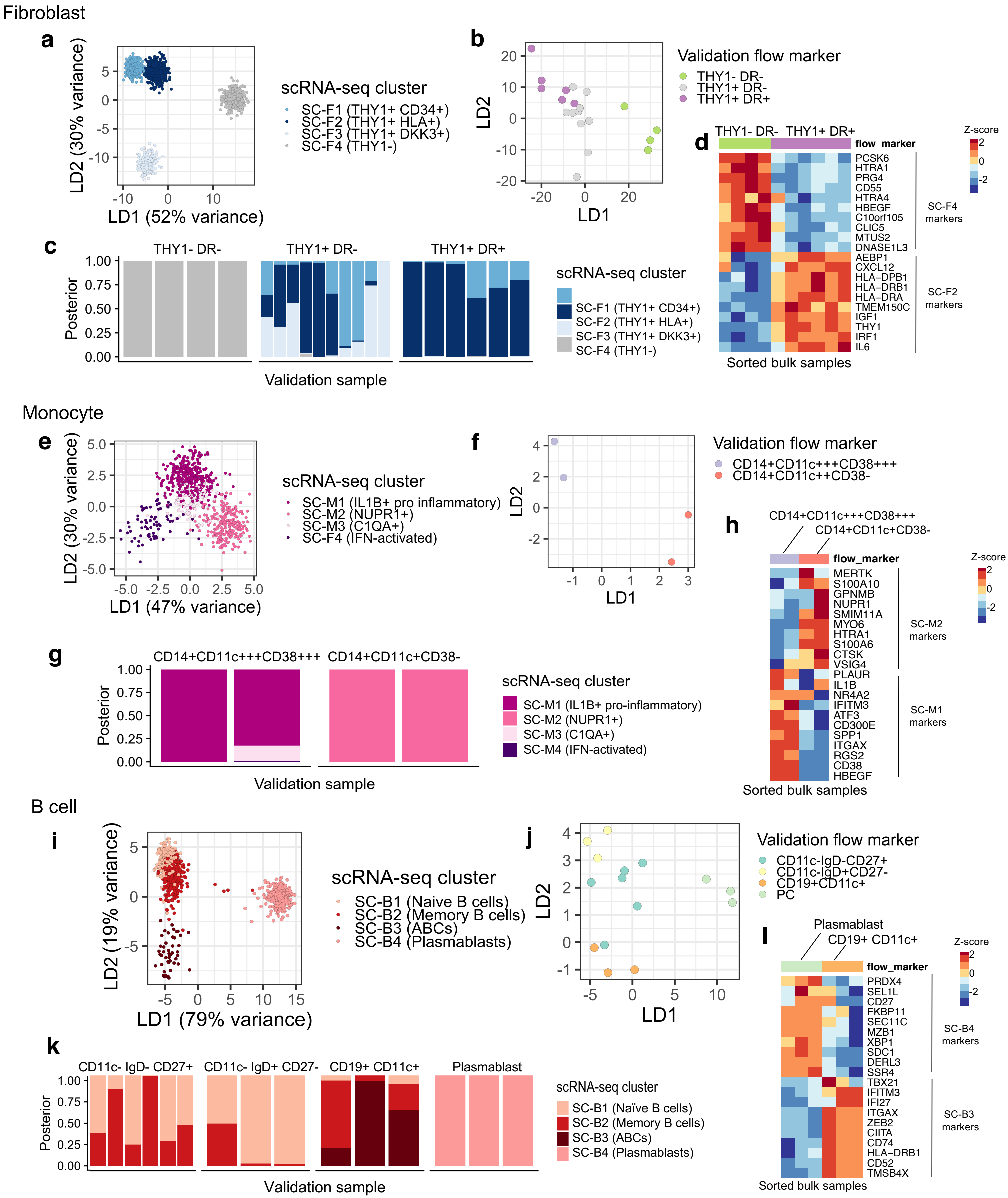
Supplemental Fig. 15. Stability test of identified scRNA-seq clusters by Silhouette.



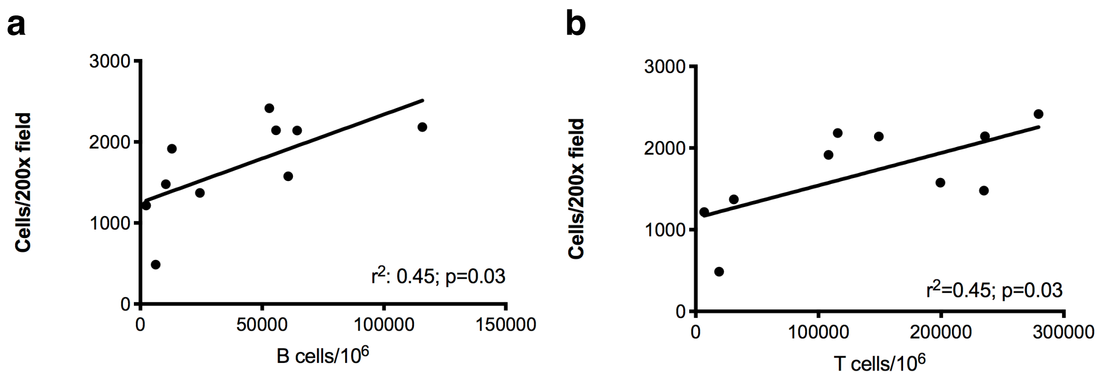
Supplemental Fig. 16. Granzyme expression and cytokine production by synovial tissue CD8 T cells. **a**. RA synovial tissue samples were disaggregated, stained for surface markers and intracellular granzyme B (GzmB) and granzyme K (GzmK), and analyzed by flow cytometry. Shown are plots of GzmB versus GzmK expression by CD8 T cells from three representative tissue specimens. **b**. GzmK and GzmB expression patterns by HLA-DR+ CD8 T cells. **c**. IFNγ production by CD4 and CD8 T cells from RA synovial tissue, measured by intracellular flow cytometry after stimulation with PMA/ionomycin. Cells from the same synovial tissue sample are connected by a line. (one–tailed Student's *t*-test p = 0.028, t-value = 2.1, df = 10.94). **d**. TNF production by CD4 and CD8 T cells from RA synovial tissue.



Supplemental Fig. 17. Flow cytometry gating schema for experimental validation. We sorted synovial cell subsets and disaggregated synovial tissues based on markers emerged from the scRNA-seq in this study. **a**. Flow gating strategy for synovial fibroblasts. **b**. Flow gating strategy for synovial B cells. **c**. Flow gating strategy for synovial monocytes.



Supplemental Fig. 18. Protein flow sorted bulk RNA-seq of subpopulations from synovial fibroblasts, monocytes, and B cells. **a-c**. LDA analysis on single-cell fibroblasts (a), classification on the sorted fibroblast bulk RNA-seq samples (b), and predicted posterior on each validation sample (c). **d**. Markers of fibroblast populations SC-F2 and SC-F4 identified in scRNA-seq data are strongly differentially expressed in bulk. **e-g**. LDA analysis on single-cell monocytes (e), classification on the sorted monocyte bulk RNA-seq samples (f), and predicted posterior on each validation sample (g). **h**. Markers of monocyte populations SC-M2 and SC-M1 identified in scRNA-seq data are strongly differentially expressed in bulk. **i-k**. LDA analysis on single-cell B cells (i), classification on the sorted B cell bulk RNA-seq samples (j), and predicted posterior on each validation sample (k). **l**. Markers of B cell populations SC-B3 and SC-MB4 identified in scRNA-seq data are strongly differentially expressed in bulk.



Supplemental Fig. 19. Cell density quantification. **a**. Correlation between cell density (cell counts per 200x field) from 10 histology samples and flow cytometric cell yields on B cells. **b**. Correlation between cell density (cell counts per 200x field) from 10 histology samples and flow cytometric cell yields on T cells.