Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Orange DE, Yao V, Sawicka K, et al. RNA identification of PRIME cells predicting rheumatoid arthritis flares. N Engl J Med 2020;383:218-28. DOI: 10.1056/NEJMoa2004114

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Table of Contents

Supplemental Methods.
Supplemental Figure 1. RNA quality and quantity by volume of fixative.
Supplemental Figure 2. RNA quality and quantity by time at room temperature.
Supplemental Figure 3. RNA quality and quantity of fresh and mailed samples.
Supplemental Figure 4. RNA quality and quantity by volume of extraction and washes.
Supplemental Figure 5. RNA quality and quantity with and without TriZol reagent extraction step.
Supplemental Figure 6. Cycle Times for HbgA2, 18S RNA, and TNF alpha after GlobinZero depletion.
Supplemental Figure 7. RNASeq QC metrics of RNA with various quality scores prepared with Illumina TruSeq or Kapa Hyper Prep Kits.
Supplemental Figure 8. Comparison of patient reported (RAPID3) and clinical (DAS28) disease activity scores of 4 patients.
Supplemental Figure 9. Clinical features of baseline, flare and on steroid treatment.
Supplemental Figure 10. Differentially expressed flare genes are reproducibly altered in repeated flares.
Supplemental Figure 11. Deconvolution of blood cell types over time to flare.
Supplemental Figure 12. Distinct analysis approaches identify activation of naïve B cells in blood 2 weeks prior to flare.
Supplemental Figure 13. Mean standardized gene expression of genes common to synovial sublining fibroblasts (CD34+, DKK+, and HLA-DR+ fibroblasts) and AC3 in blood over time to flare.
Supplemental Figure 14. Gating strategy for quantification of PRIME cells in blood samples.
Supplemental Figure 15. PRIME cells are nucleated.
Supplemental Figure 16. Gating strategy for FACS sorting PRIME cells.
Supplemental Figure 17. Sorted PRIME cells express synovial fibroblast genes.
Supplemental Figure 18. Sorted PRIME cells express classic synovial fibroblast genes.
Supplemental Table 1. Clinical Characteristics of 4 RA patients in longitudinal study.
Supplemental Table 2. QC stats of sorted cell RNAseq

Supplemental Methods.

Trial Oversight

This study was approved by the Rockefeller University IRB (#DOR0833 and DOR0722) as an investigator initiated clinical research study. All participating patients signed informed consents.

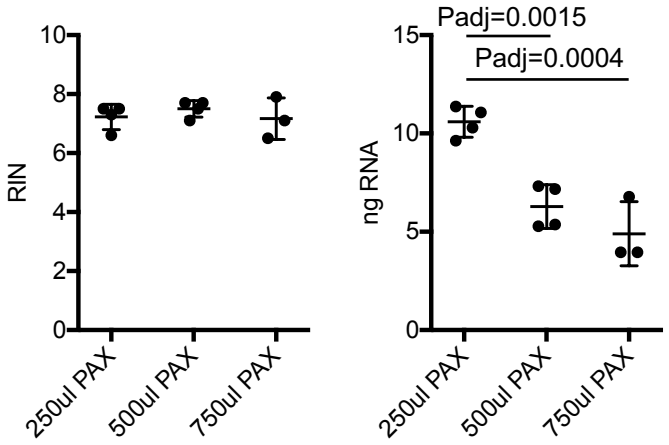
Time series analysis

ImpulseDE2 (v1.8.0, Fischer DS et al, 2018) was used to fit an impulse model to the longitudinal expression data for all non-steroid samples in the index patient from 8 weeks prior to flare up to 4 weeks after flare, followed by differential expression analysis based on a log-likelihood ratio test. For each sample, the time of flare was considered time point 0, and each sample's time point corresponded to how many weeks before or after flare the sample was collected. Collapsing the data to weeks was necessary to ensure replicates for each of the 13 weeks under consideration.

The impulse function used in ImpulseDE2 is as follows:

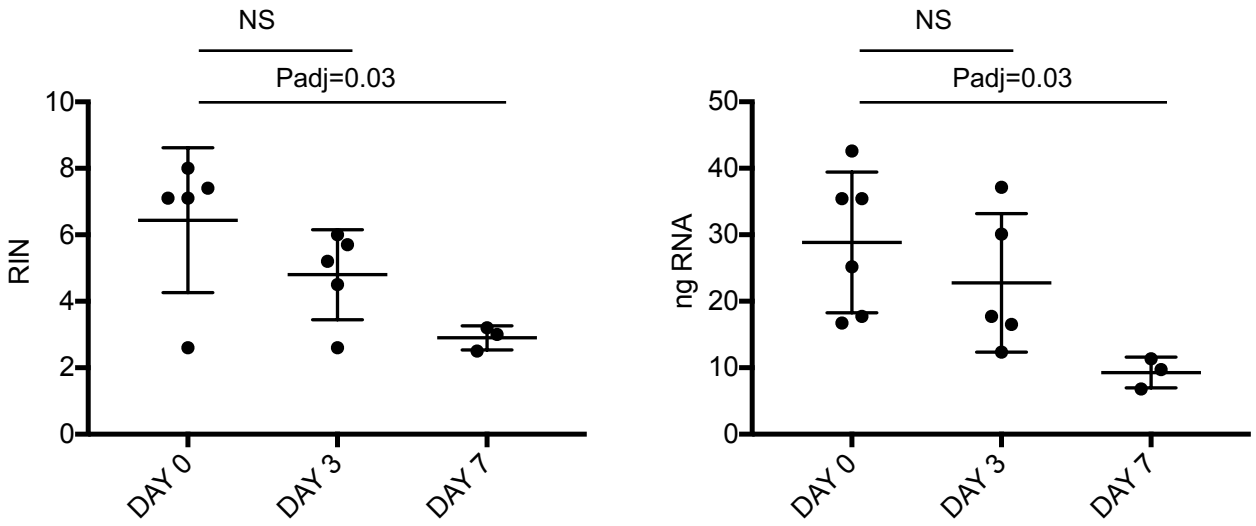
$f_{Impulse}(t) = \frac{1}{h_1}\left(h_0 + (h_1 - h_0)\frac{1}{1+e^{-\beta(t-t_1)}}\right)\left(h_2 + (h_1 - h_2)\frac{1}{1+e^{\beta(t-t_2)}}\right)$,

which essentially models the transitions between three expression states (initial, peak, steady state) using two sigmoid functions (initial -> peak, peak -> steady state). The amplitude parameters are $h_0 = f_{Impulse}(t \to -\infty), h_2 = f_{Impulse}(t \to -\infty)$, $h_1$models the intermediate expression, t_1 and t_2 are the week before / after flare at which the state transitions happen, and β is the slope parameter of both sigmoid functions.



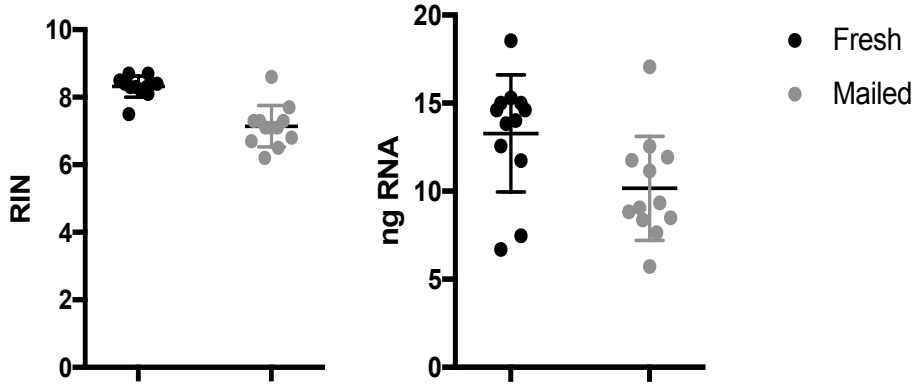
Supplemental Figure 1. RNA quality and quantity by volume of fixative.

3 drops of blood harvested with a 21 gauge lancet were added to a microtainer tube prefilled with either 250, 500 or 750ul of PAX gene fixative. Samples were stored at room temperature for 3 days and RNA was extracted using the PAX gene RNA kit. RIN scores and quantity of RNA was assessed using the Agilent 2100 Bioanalyzer picochip. Padj= ANOVA, followed by Dunnett's multiple comparisons test, using 250ul as the reference group.



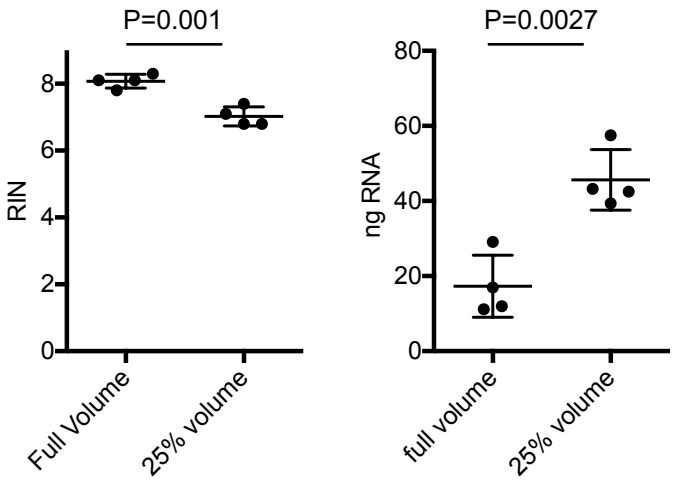
Supplemental Figure 2. RNA quality and quantity by time at room temperature.

100ul of whole blood was added to a microtainer tube prefilled with 250ul PAX gene fixative and frozen after 2 hours, 3 days, or 7 days incubation at room temperature. RNA was extracted using the PAX gene RNA kit with scaled down washes and elutions. RIN scores and quantity of RNA was assessed using the Agilent 2100 BioAnalyzer RNA picochip. Padj= ANOVA, followed by Dunnett's multiple comparisons test, using day 0 as the reference group.



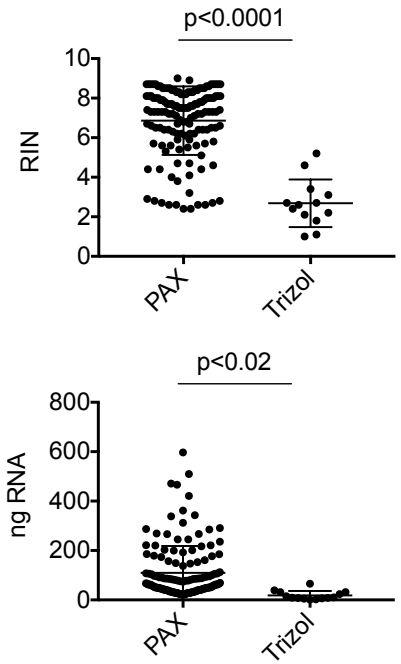
Supplemental Figure 3. RNA quality and quantity of fresh and mailed samples.

100ul of whole blood was added to a microtainer tube prefilled with 250ul PAX gene fixative and frozen after 2-hour incubation at room temperature or mailed. RNA was extracted using the PAX gene RNA kit and RIN scores and quantity of RNA was assessed using the Agilent 2100 BioAnalyzer RNA picochip.



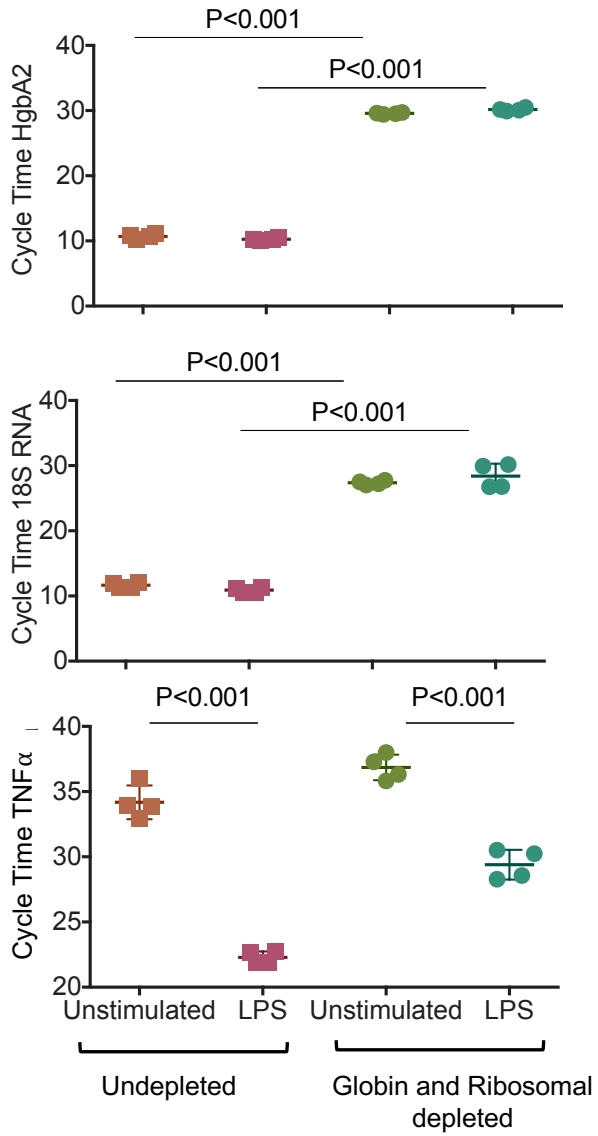
Supplemental Figure 4. RNA quality and quantity by volume of extraction and washes.

3 drops of blood harvested with a 21 gauge lancet were added to a microtainer tube prefilled with 250ul of PAX gene fixative. Samples were stored at room temperature for 3 days and RNA was extracted using the PAXgeneRNA kit according to manufacturer's directions or with a scaled down version of the PAX protocol, using 25% of the recommended volumes for all washes and elutions. RIN scores and quantity of RNA was assessed using the Agilent 2100 BioAnalyzerRNA picochip. P= unpaired two-sided t test.

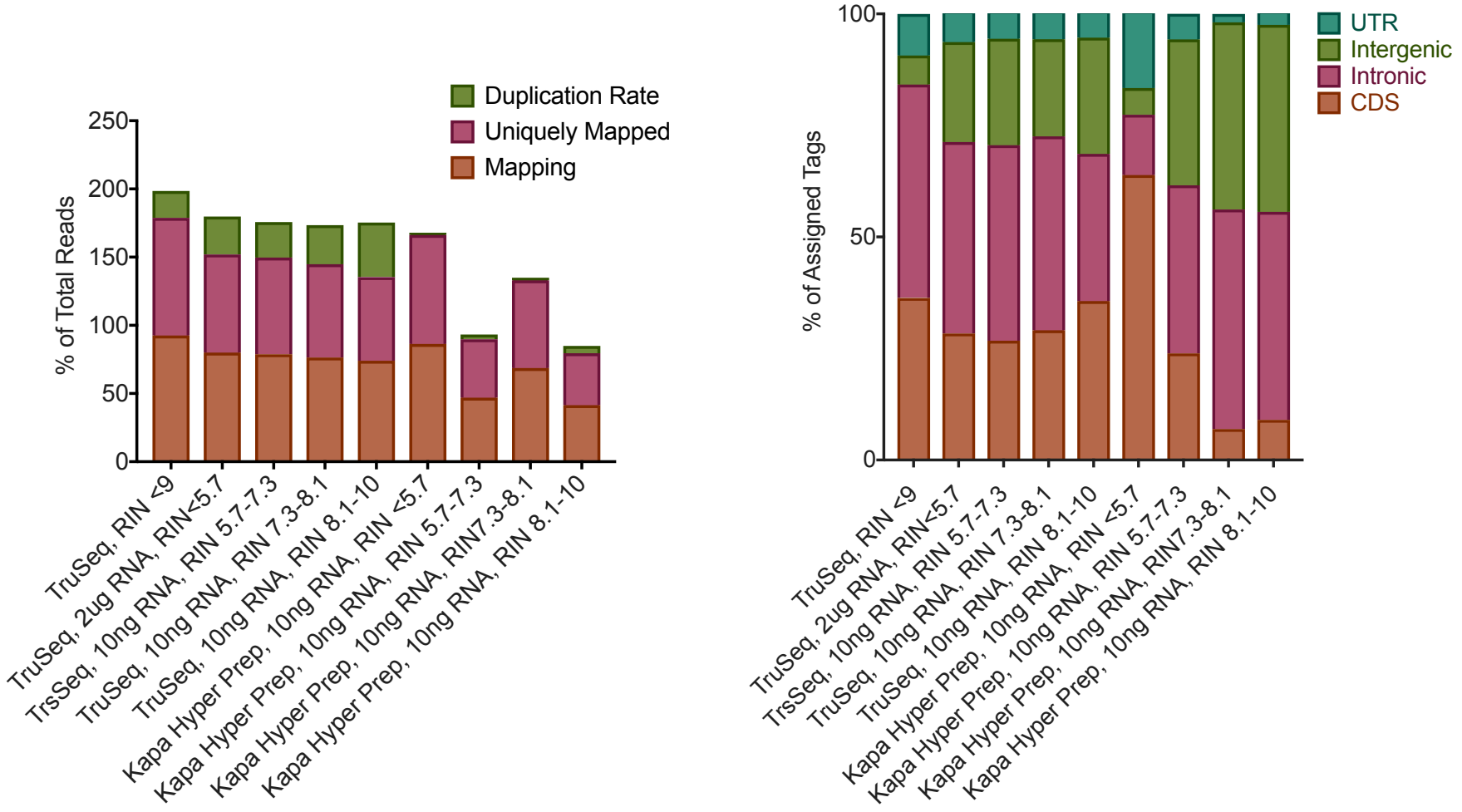


Supplemental Figure 5. RNA quality and quantity with and without TriZol reagent extraction step.

Mailed patient finger stick samples were stored in PAXgeneRNA buffer at -80°C. RNA was extracted from 142 samples using PAXgeneRNA kit with low volume washes. RNA was extracted from 13 other samples by mixing with 700ul TriZol-LS and 200ul chloroform, precipitating with isopropanol and glycogen, and washing with 80% cold ethanol. The dried pellet was resuspended in 100ul PBS and purified using the Roche High Pure Isolation kit. P values represent significance of unpaired T tests.



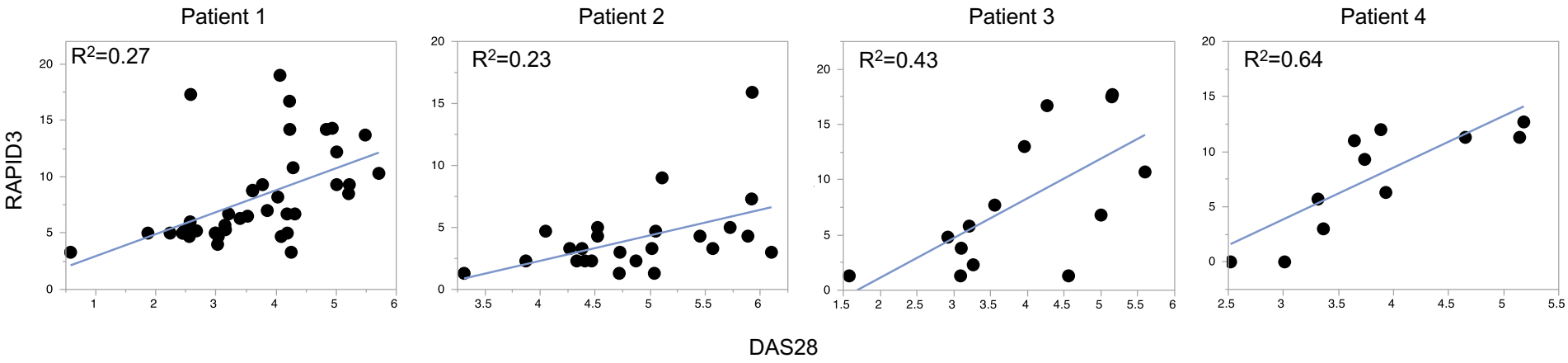
Supplemental Figure 6. Cycle times for HbgA2, 18S RNA, and TNFα after GlobinZero depletion.
Since ribosomal and hemoglobin RNA represent approximately 98% and 70% of the RNA in whole blood, respectively, we tested standard commercial kits for removing these RNAs prior to RNAseq. 4ml heparinized blood treated with 1ug/ml LPS for one hour at 37˚C. 250ul blood sample was mixed with 250ul PAXgene fixative into replicate microtainer tubes. After RNA extraction, samples were either left undepleted or treated with the globin zero depletion kit and quantitative PCR was performed to test for hemoglobin A2, 18S RNA, or TNFα mRNA expression. GlobinZero kits depleted both hemoglobin A2 and 18S ribosomal RNA (increased mean cycle time from 11 to 28 and 10 to 30, respectively) with relative preservation of TNFα mRNA. P values represent results of ordinary one-way ANOVA with Tukey's multiple comparisons test.



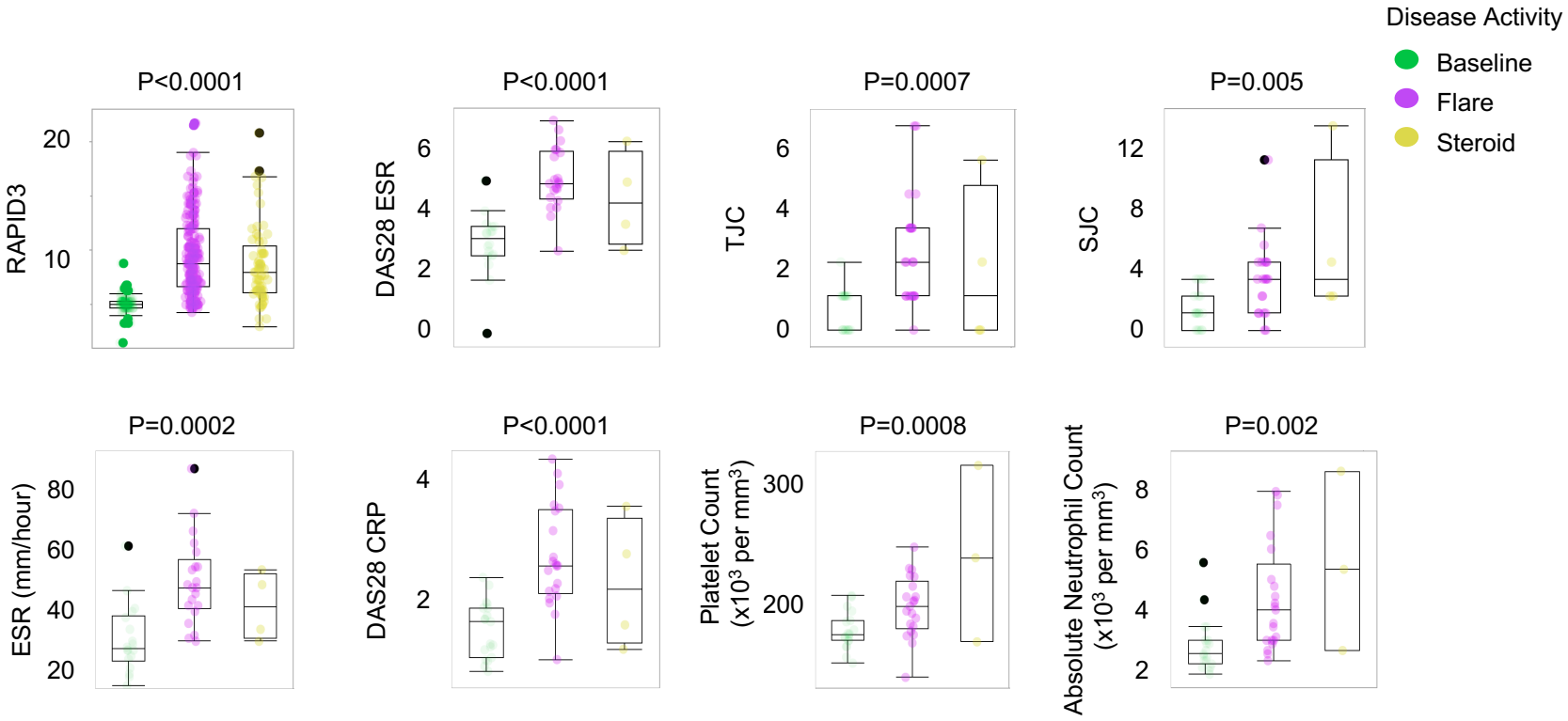
Supplemental Figure 7. RNASeq QC metrics of RNA with various quality scores prepared with Illumina TruSeq or Kapa Hyper Prep Kits.

Left Panel: Distribution of mapping, uniquely mapping, and duplicate reads.

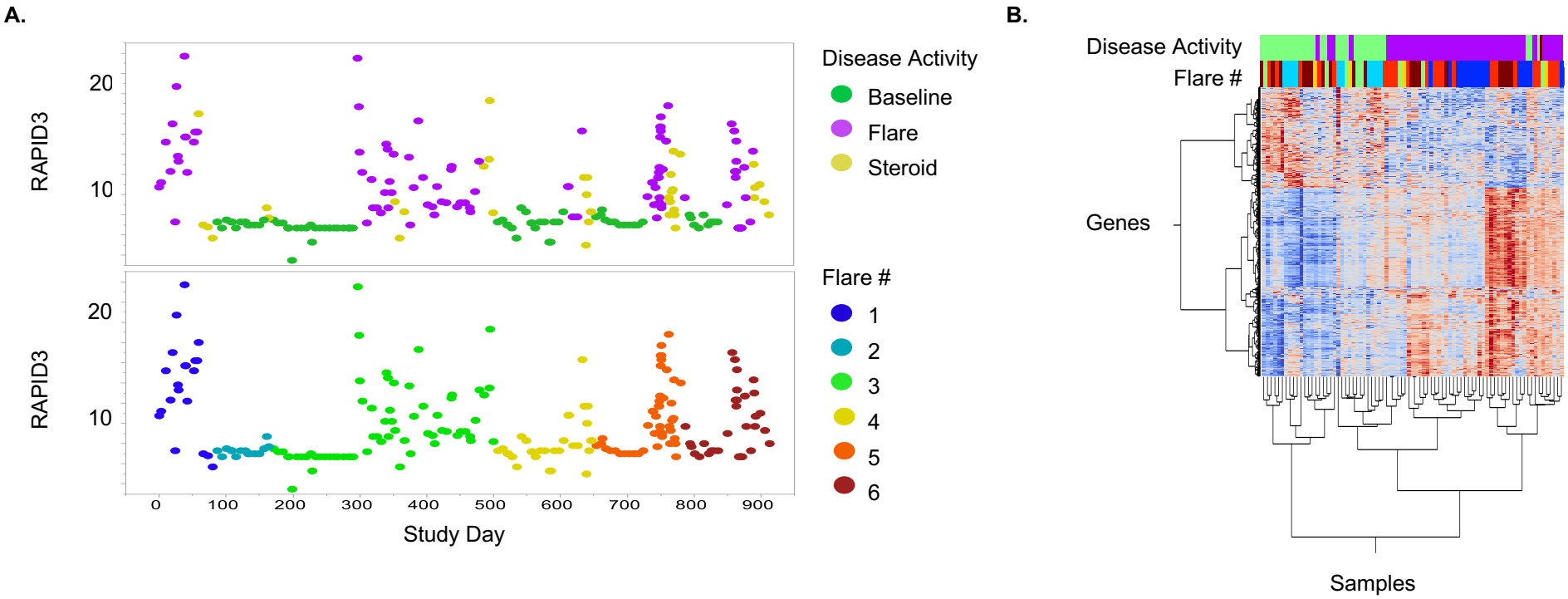
Right Panel: Distribution of tags assigned to UTR (untranslated region), intergenic, intronic, and CDS (coding sequence) of whole blood RNA samples prepared with Illumina TruSeq or Kapa Hyper Prep Kits with various input RNA quality and quantity. The Illumina TruSeq library Prep demonstrated increased mapping to coding sequence and fewer intergenic reads and was ultimately used for downstream experiments.



Supplemental Figure 8. Comparison of patient reported (RAPID3) and clinical (DAS28) disease activity scores of 4 patients.
Paired RAPID3 scores and DAS28-CRP scores were collected from 91 clinic visits of 4 RA patients. The patient reported RAPID3 questionnaire was positively correlated with clinician generated DAS28 score in all 4 patients.

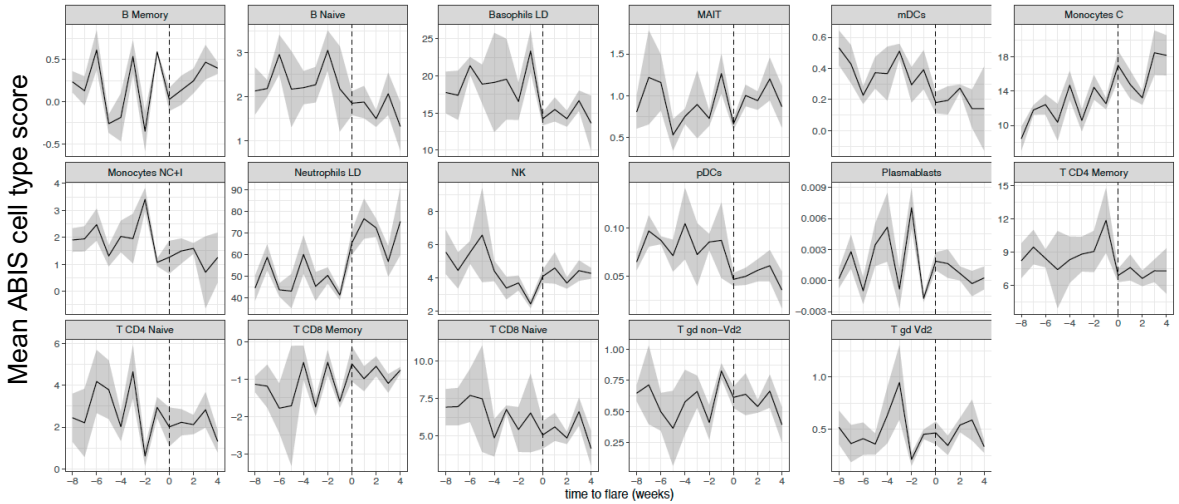


Supplemental Figure 9. Clinical features of baseline, flare and on steroid treatment.
RAPID3 questionnaire responses from 360 time points and DAS28 ESR, TJC, SJC, ESR, DAS28 CRP, platelet counts and absolute neutrophil counts from 43 clinic visits for one patient over four years. Time points are colored according to disease activity category: baseline, flare, steroid. Steroid treatment was defined as any time point when the patient took any dose of steroid that day, or if the calculated dose, using washout kinetics, was greater than 0.01mg/ml. Samples acquired between two time points that met criteria for flare were categorized as flare up until treatment with steroid. TJC indicates tender joint count. SJC indicates swollen joint count. P values represent ANOVA across three disease categories. Flare was associated with significantly increased RAPID3, DAS28 ESR, TJC, SJC, ESR, DAS28CRP, platelets and neutrophils.

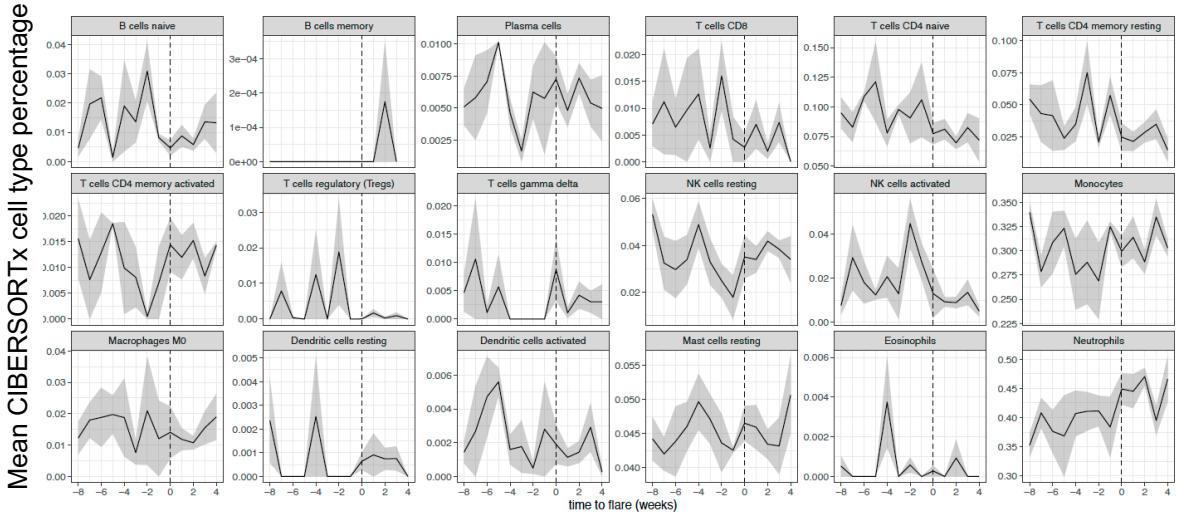


Supplemental Figure 10. Differentially expressed flare genes are reproducibly altered in repeated flares.
A. Index patient disease activity (RAPID3) over time. Top panel dots are colored by disease activity assignment. Bottom panel dots are colored according to clinical flare event number. B. Unsupervised hierarchical clustering of genes differentially expressed between baseline and flare. Top bar indicates samples colored according to disease activity assignment. Lower bar indicates samples colored according to clinical flare event number. Data shows differentially expressed flare genes are represented by multiple clinical events.

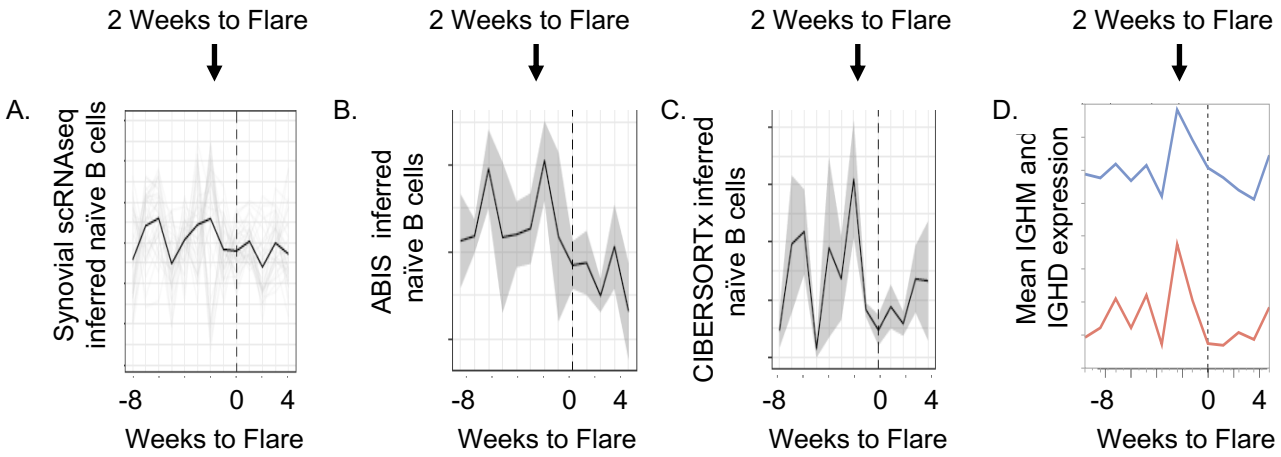
A. ABIS inferred cell types over time to flare



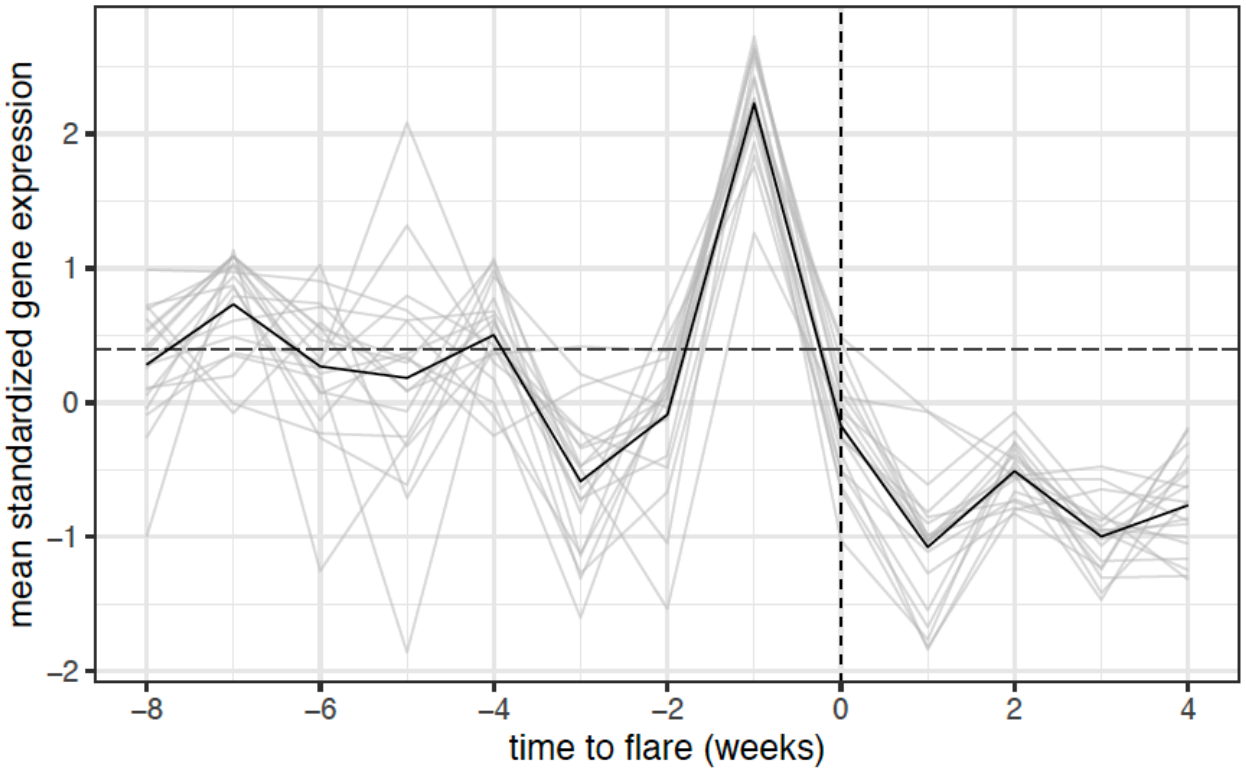
B. CIBERSORTx inferred cell types over time to flare.



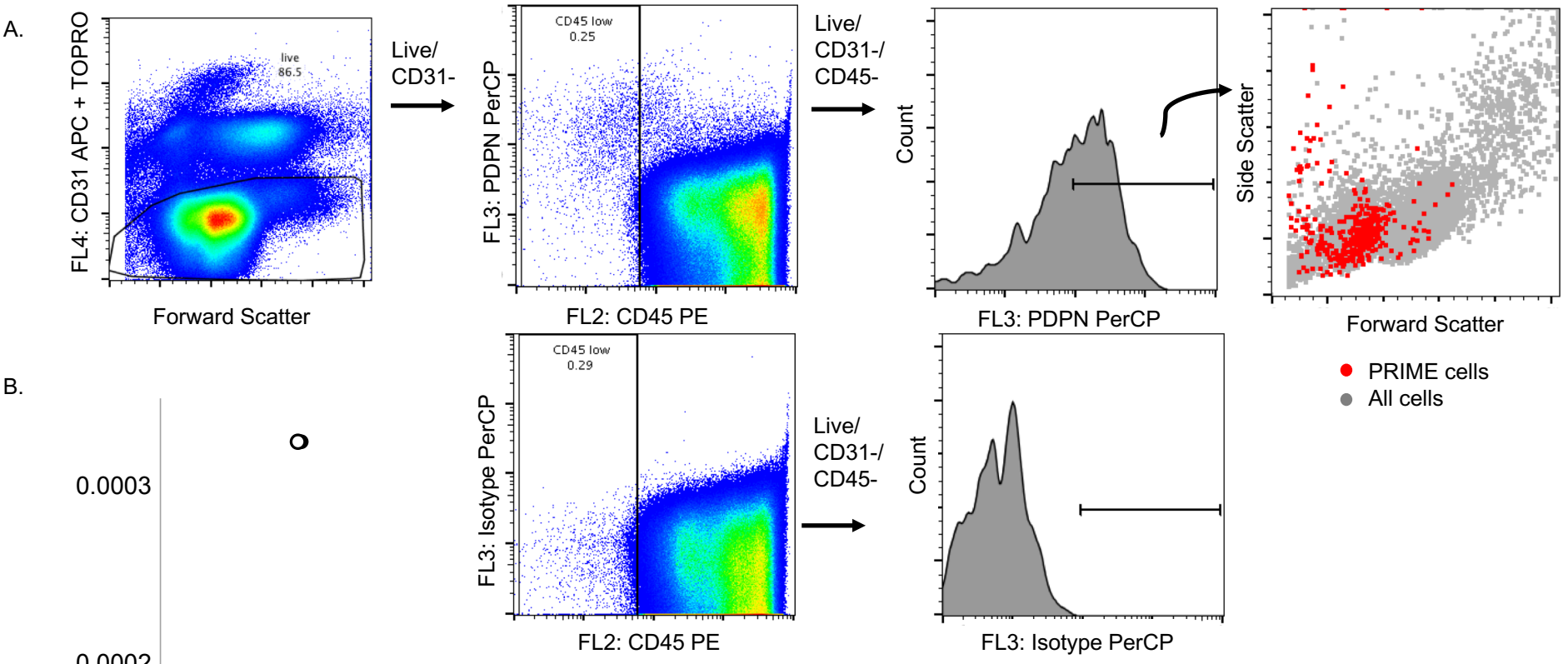
Supplemental Figure 11. Deconvolution of blood cell types over time to flare.
Mean cell type trajectories of A. ABIS inferred cell types, and B. CIBERSORTx inferred cell types over time to flare are plotted (excluding those that were 0 all throughout) showing the mean score with standard error of the mean as a ribbon. These data independently confirm data in main Figure 4A identifying activation of B cells in AC2.

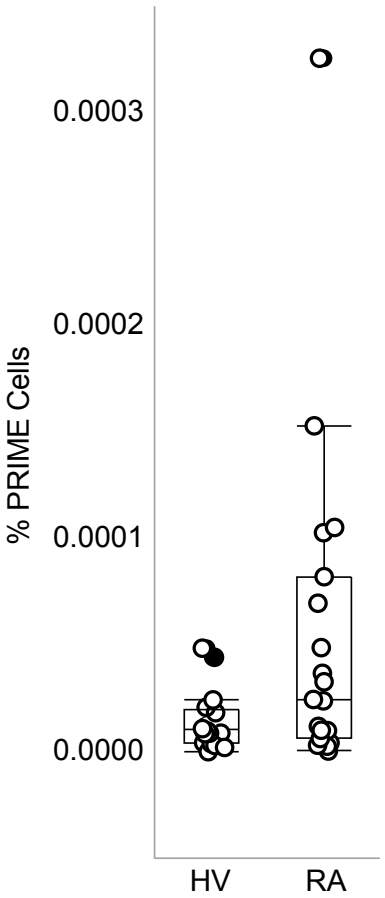


Supplemental Figure 12. Distinct analysis approaches identify activation of naïve B cells in blood 2 weeks prior to flare.
A. Mean expression of 200 synovial single cell RNAseq naïve B cell marker genes by week to flare. B. ABIS inferred naïve B cells by week to flare. C. CIBERSORTx inferred naïve B cells by week to flare. D. IGHM (blue) and IGHD (red) gene expression by week to flare. Dashed line indicates first day of symptoms of RA flare, red arrows indicate peak in B cell signature 2 weeks prior to flare.



Supplemental Figure 13. Mean standardized gene expression of genes common to synovial sublining fibroblasts (CD34+, DKK+, and HLA-DR+ fibroblasts) and AC3 in blood over time to flare.
Light gray lines represent the expression of individual genes over time to flare in patient 1 over all flares.

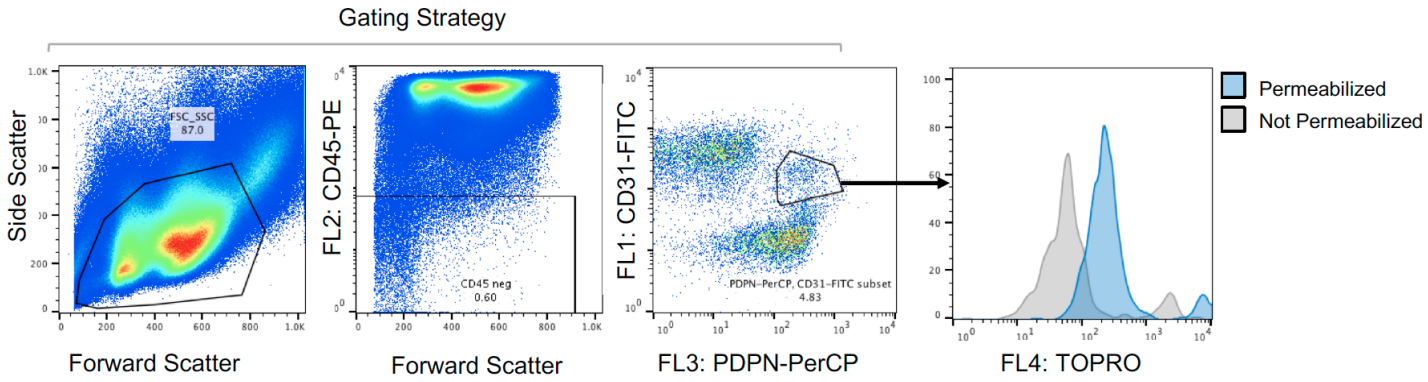




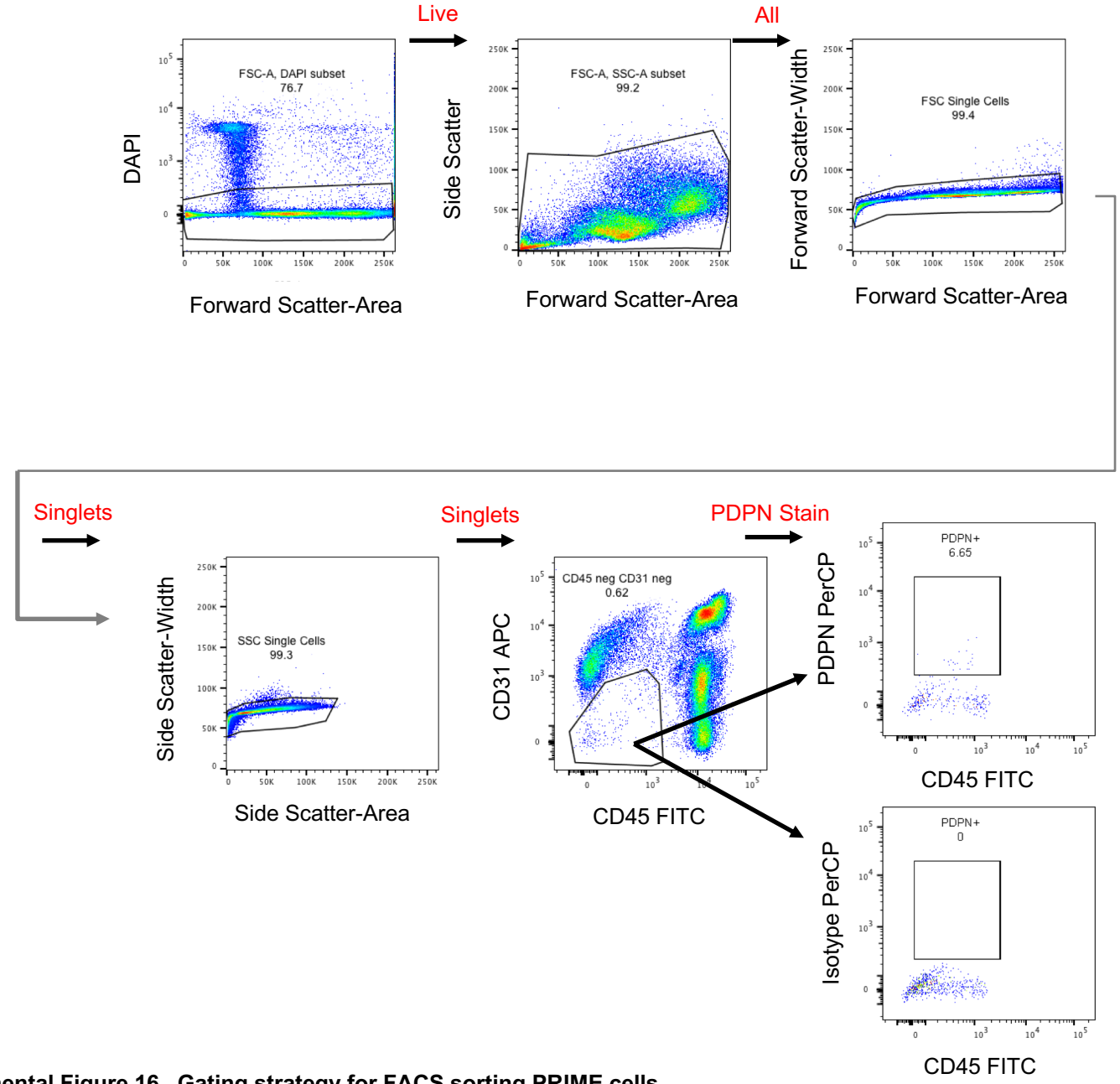
Supplemental Figure 14. Gating strategy and quantification of PRIME cells in blood samples.

A. 20-100 Million cells from CD14 depleted products of leukaphoresis were stained with CD31-APC, (WM59), PDPN-PerCP, (NZ1.3), CD45-FITC, (HI30), and DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride). Mouse IgG1-FITC, (MOPC-21), Mouse IgG1-APC, (MOPC-21), Rat IgG2a-PerCP, (eBR2a) were used as isotype matched controls. Cells were sorted on a BD FACSAria II.

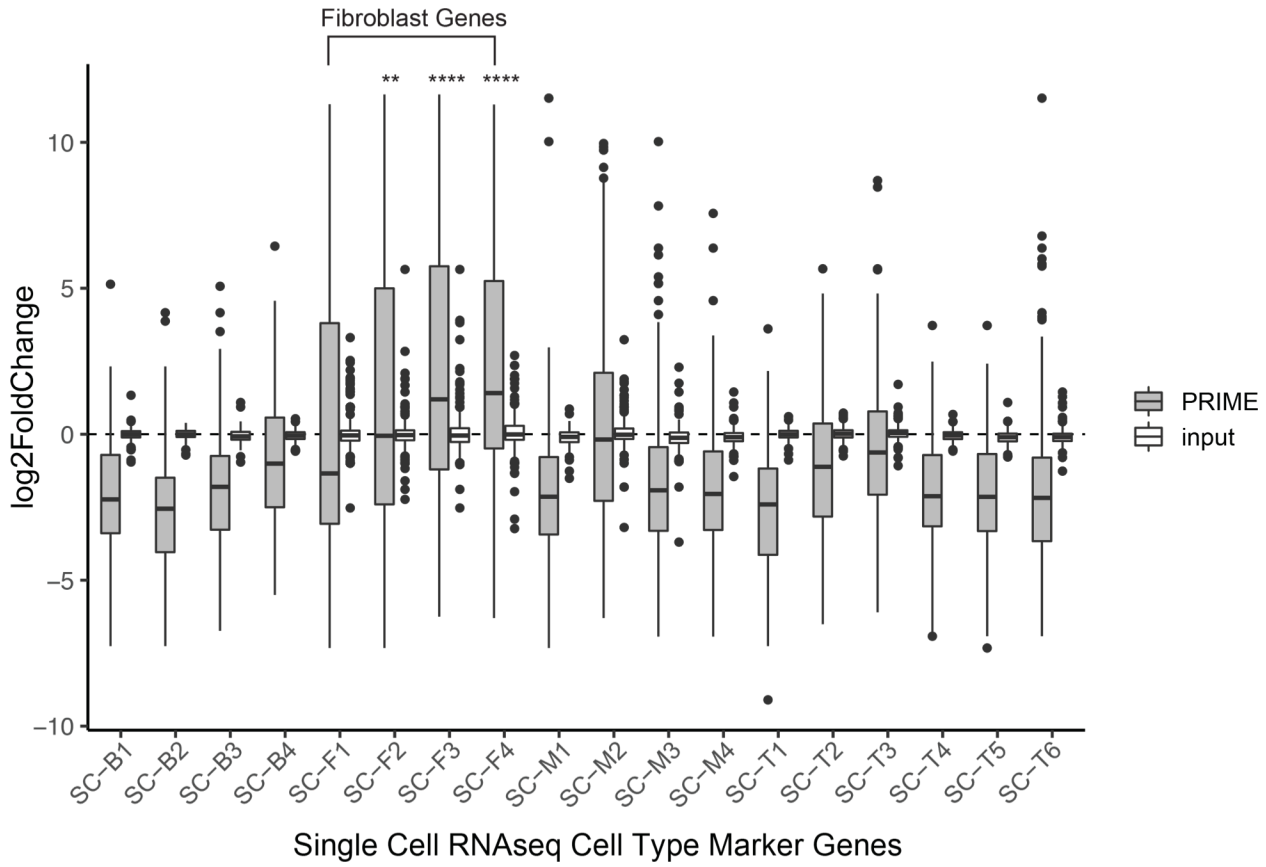
B. The percentages of PRIME cells (CD45-CD31-PDPN+) that were TO-PRO-3- (i.e., live) among TO-PRO-3-CD31- cells are shown; The box-and-whisker plots indicate the median, interquartile range, and 1.5 times the interquartile range.



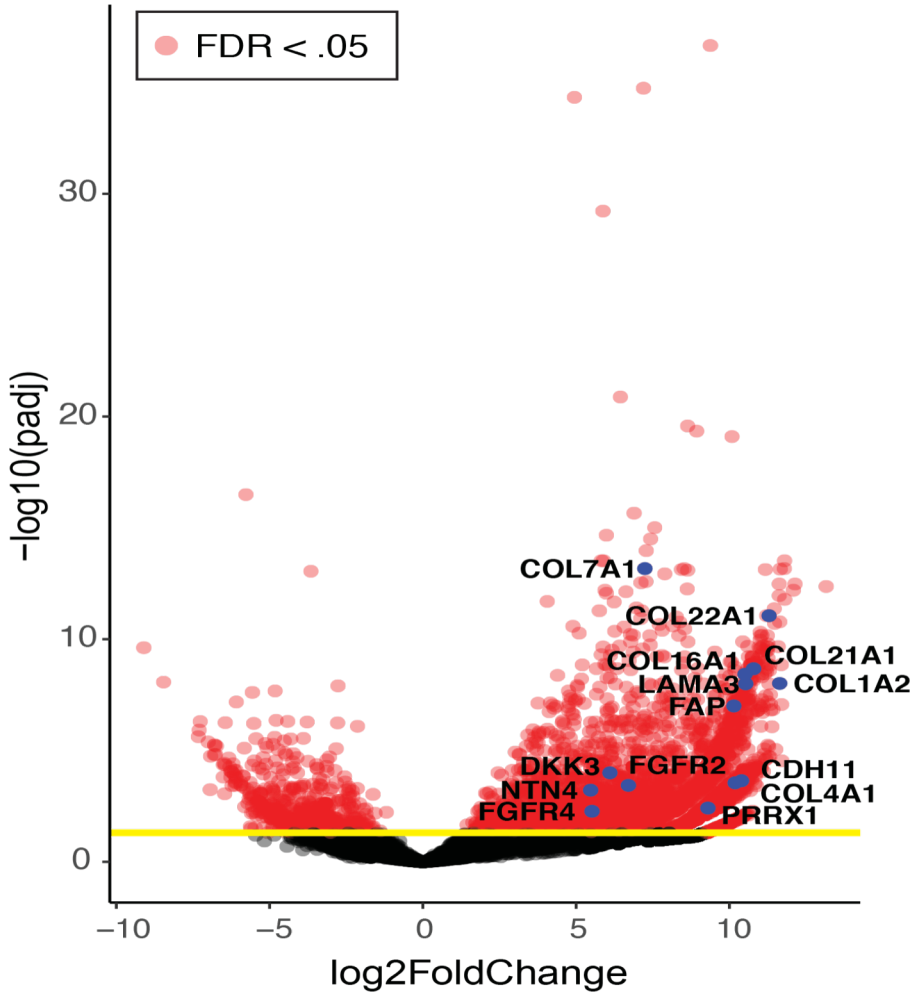
Supplemental Figure 15. PRIME cells are nucleated.
PBMC of RA donor was stained with CD45/CD31/PDPN and either TOPRO (Not Permeabilized) or permeabilization and TOPRO (Permeabilized) and assessed by flow cytometry. PRIME cells were gated as CD45-/CD31-/PDPN+ cells and TOPRO staining of permeabilized and not permeabilized cells is presented. The increased fluorescence of TOPRO in the permeabilized PRIME cells indicates the presence of double stranded nucleic acid.



Supplemental Figure 16. Gating strategy for FACS sorting PRIME cells.
20-100 Million cells from CD14 depleted leukapheresates were stained with CD31-APC, (WM59), Mouse IgG1-APC, (MOPC-21), PDPN-PerCP, (NZ1.3), Rat IgG2a-PerCP, (eBR2a), CD45-FITC, (HI30), Mouse IgG1-FITC, (MOPC-21), and DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride). Cells were sorted on a BD FACSAria II.



Supplemental Figure 17. Sorted PRIME cells express synovial fibroblast genes.
Log2 fold change of various synovial single cell RNAseq marker genes in PRIME cells (flow sorted DAPI-/CD45-/CD31-/PDPN+ cells) versus hematopoietic cells (flow sorted DAPI-/CD45+) and Log2 fold change of input cells (stained PBMC but not flow sorted) versus hematopoietic cells (flow sorted CD45+) as technical control for stress of flow sorting. These data show that single cell marker genes of fibroblasts (SC-F1, SC-F2, SC-F3, SC-F4) but not B cells (SC-B1-4), macrophages (SC-M1-4), or T cells (SC-T1-6) are enriched in sorted PRIME cells. Fibroblast genes (as marked) were the only set of synovial cell marker genes enriched in PRIME cells.



Supplemental Figure 18. Sorted PRIME cells express classic synovial fibroblast genes.
Volcano plot of log10(-padj) vs Log2 fold change of PRIME cells (flow sorted DAPI-/CD45-/CD31-/PDPN+ cells) versus hematopoietic cells (flow sorted DAPI-/CD45+). Classic fibroblast genes are significantly increased in PRIME cells relative to hematopoietic cells.

Patient ID	Age	Sex	Disease Duration	Antibody Status	Extra-Articular Manifestations	Smoking History	RA Medications	Flare #	Flare treatment
1	64	F	20	CCP>250	Rheumatoid nodules, anemia of chronic disease, Sicca	Quit, 12 pack year history	Abatacept, Leflunomide	1	Prednisone
							Abatacept, Leflunomide	2	Prednisone
							Abatacept, Leflunomide	3	Prednisone
							Abatacept, Leflunomide	4	IA Methylprednisolone
							Abatacept, Leflunomide	5	Prednisone
							Abatacept, Leflunomide	6	Prednisone
							Ritxuximab, Leflunomide	7	Prednisone
							Ritxuximab, Leflunomide	8	Prednisone
2	62	F	20	CCP>250	None	Quit, 15 pack year history	Certolizumab pegol, Methotrexate	1	Prednisone
							Certolizumab pegol, Methotrexate	2	IA Methylprednisolone
							Certolizumab pegol, Methotrexate	3	Prednisone
3	55	F	10	CCP>250	Sicca	Quit, 17 pack year history	Hydroxychloroquine	1	Prednisone
							Hydroxychloroquine	2	Prednisone
4	39	F	1	CCP>250	None	Quit, 2 pack year history	Etanercept	1	Methylprednisolone
							Adalimumab	2	Methylprednisolone

Supplemental Table 1: Clinical characteristics of 4 RA patients in longitudinal study.
Age and Disease Duration are presented in years, F=female, IA=intra-articular, all other flare treatments were oral medications

Sample ID	Cell Count	Reads	Mapped	Exonic	Intronic	Intergenic	Duplication
A_PRIME	184	2,709,266	1237994	44574	607129	562811	0.611
B_PRIME	92	2,535,751	1218029	46265	603634	545322	0.8743
C_PRIME	125	2,138,741	955919	34445	468136	430310	0.4586
A_input	5.0E+06	2,986,610	2504851	1679934	658698	82262	0.129
B_input	5.0E+06	2,972,297	2458010	1610378	688286	74450	0.1223
C_input	5.0E+06	3,466,604	2815031	1899402	651067	161399	0.1416
A_PBMC	7.2E+06	3,406,725	2815101	1905934	716195	89682	0.1418
B_PBMC	6.7E+06	2,830,425	2293824	1637695	502623	61643	0.135
C_PBMC	5.4E+06	2,783,444	2283444	1563003	557186	81809	0.1196

Supplemental Table 2: Mapping statistics of sorted-cell RNAseq. All values presented indicate number of reads. Reads indicates the number of sequenced reads, Mapped indicates reads mapped to the hg19 genome (using gencode gene annotations). Exonic, Intronic and Intergenic correspond to reads that map to corresponding genomic regions. Duplication indicates percent of reads that are duplicated.