**TITLE: Mixed Effects Association of Single Cells Identifies an Expanded Th1-Skewed Cytotoxic Effector CD4+ T Cell Subset in Rheumatoid Arthritis**

**AUTHORS:** Chamith Y. Fonseka[1,2,3,5†], Deepak A. Rao[2†], Nikola C. Teslovich[2], Ilya Korsunsky[1], Susan K. Hannes[2], Kamil Slowikowsi[1,2,3,4], Michael F. Gurish[2], Laura T. Donlin[6,7], Michael E. Weinblatt[2], Elena M. Massarotti[2], Jonathan S. Coblyn[2], Simon M. Helfgott[2], Derrick J. Todd[2], Vivian P. Bykerk[8,9], Elizabeth W. Karlson[2], Joerg Ermann[2], Yvonne C. Lee[2], Michael B. Brenner[2], and Soumya Raychaudhuri[1,2,3,4,10] *

[†]Co-first authors

*Corresponding author

**AFFILIATIONS:**

1) Center for Data Sciences, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115 USA.
2) Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115 USA.
3) Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115 USA.
4) Program in Medical and Population Genetics, Broad Institute of Massachusetts Technical Institute and Harvard University, Cambridge, MA 02138, USA.
5) Department of Biomedical Informatics, Harvard University, Cambridge, MA 02138, USA.
6) Arthritis and Tissue Degeneration Program, Hospital for Special Surgery, New York, New York 10021, USA.
7) David Z. Rosensweig Genomics Research Center, Hospital for Special Surgery, New York, New York 10021, USA.
8) Division of Rheumatology, Hospital for Special Surgery, 535 E 70th Street, New York, NY 10021 USA.
9) Department of Medicine, Weill Cornell Medical College, Cornell University, New York, NY 10021, USA.
10) Institute of Inflammation and Repair, University of Manchester, Manchester, UK.

Corresponding Author:

Soumya Raychaudhuri
77 Avenue Louis Pasteur
Harvard New Research Building, Suite 250D
Boston, MA 02446, USA.
soumya@broadinstitute.org; 617-525-4484 (tel); 617-525-4488 (fax)

**ONE SENTENCE SUMMARY:** By accounting for inter-individual and technical variation, mixed-effects regression of single cells enables identification of disease-associated cell populations and reveals a Th1-skewed effector memory CD4+ T cell population expanded in rheumatoid arthritis.

**ABSTRACT:**

High dimensional single-cell analyses have dramatically improved the ability to resolve complex mixtures of cells from human disease samples; however, identifying disease-associated cell types or cell states in patient samples remains challenging due to technical and inter-individual variation. Here we present Mixed effects modeling of Associations of Single Cells (MASC), a novel reverse single cell association strategy 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. Applying MASC to mass cytometry analyses of CD4+ T cells from blood of rheumatoid arthritis (RA) patients and controls revealed a significantly expanded population of CD4+ T cells, identified as CD27- HLA-DR+ effector memory cells, in RA patients (OR = 1.7; p = $1.1 \times 10^{-3}$). The frequency of CD27- HLA-DR+ cells was similarly elevated in blood samples from a second RA patient cohort, and CD27- HLA-DR+ cell frequency decreased in RA patients who respond to immunosuppressive therapy. Compared to peripheral blood, synovial fluid and synovial tissue samples from RA patients contained ~5-fold higher frequencies of CD27- HLA-DR+ cells, which comprised ~10% of synovial CD4+ T cells. CD27- HLA-DR+ cells express a distinctive effector memory transcriptomic program with Th1- and cytotoxicity-associated features, and produce abundant IFN-γ and granzyme A protein upon stimulation. Thus MASC identified the expansion of a unique effector memory CD4+ T cell population in RA. We propose that MASC is a broadly applicable method to identify disease-associated cell populations in high-dimensional single cell data.

## MAIN TEXT:

### INTRODUCTION

The advance of single cell technologies has enabled investigators to resolve cellular heterogeneity with unprecedented resolution. Single cell assays have been particularly useful in the study of the immune system, in which diverse cell populations often consisting of rare and transitional cell states may play an important role *(1)*. Application of single cell transcriptomic and cytometric assays in a case-control study has the potential to reveal expanded pathogenic cell populations in immune-mediated diseases.

Rheumatoid arthritis (RA) is a chronic, systemic disease affecting 0.5-1% of the adult population, making it one of the most common autoimmune disorders worldwide *(2)*. RA is triggered by environmental and genetic risk factors, leading to activation of autoreactive T cells and B cells that mediate an autoimmune response directed at the joints *(3, 4)*. CD4+ T cells have been strongly implicated in RA pathogenesis *(5, 6)*. For one, the strongest genetic association to RA is with the *HLA-DRB1* gene within the MHC; these polymorphisms affect the range of antigens that MHCII molecules can bind and present in order to activate CD4+ T cells *(3, 7, 8)*. Furthermore, many RA risk alleles outside of the MHC locus also lie in pathways important for CD4+ T cell activation, differentiation into effector ($T_{eff}$) and regulatory ($T_{reg}$) subsets, and maintenance of subset identity *(4, 8-12)*. Defining the precise CD4+ T cell subsets that are expanded or dysregulated in RA patients is critical to deciphering pathogenesis. Such cell populations may be enriched in antigen-specific T cells and may aid in discovery of dominant disease-associated autoantigens. In addition, these populations may directly carry out pathologic effectors functions that can be targeted therapeutically *(13)*.

For many autoimmune diseases, directly assaying affected tissues is difficult because samples are only available through invasive procedures. Instead, querying peripheral blood for altered immune cell populations is a rapidly scalable strategy that achieves larger sample sizes and allows for serial monitoring. Flow cytometric studies have identified alterations in specific circulating T cell subsets in RA patients, including an increased frequency of CD28- CD4+ T cells *(14-16)*; however, the expansion of CD28- T cells represents one of the relatively few T cell alterations that has been reproducibly detected by multiple groups. Limited reproducibility may be the consequence of differences in clinical cohorts, small sample sizes, methodologic variability, and use of limited, idiosyncratic combinations of phenotypic markers*(17)*.

The advent of mass cytometry now allows for relatively broad assessment of circulating immune cell populations with >30 markers *(18)*, enabling detailed, multiparametric characterization of lymphocyte subsets. This technology provides the potential to define and quantify lymphocyte subsets at high resolution using multiple markers. While rapid progress has been made in defining individual cellular populations in an unbiased fashion *(19-23)* , a key challenge that remains is establishing methods to identify cell populations associated with a disease. In particular, inter-individual variation and technical variation influence cell population frequencies and need to be accounted for in an association framework.

Here we describe a robust statistical method to test for disease associations with single cell data called MASC (Mixed effect modeling of Associations of Single Cells), which tests at the single cell level the association between population clusters and disease status. It is a 'reverse' association strategy where the case-control status is an independent variable, rather than the dependent variable. We have applied MASC to identify T cell subsets associated with RA in a mass cytometry case-control immunophenotyping dataset that we generated focused on CD4+ T cells. This high-dimensional analysis enabled us to identify disease-specific changes in canonical as well as non-canonical CD4+ T cell populations using a panel of 32 markers to

reveal cell lineage, activation, and function *(24, 25)*. Using MASC, we identify a population of memory CD4+ T cells, characterized as CD27- HLA-DR+, which is expanded in the circulation of RA patients. Further, we find that CD27- HLA-DR+ T cells are enriched within inflamed RA joints, rapidly produce IFN-γand cytolytic factors, and contract with successful treatment of RA.

## RESULTS

### *STATISTICAL AND COMPUTATIONAL STRATEGY*

Single cell association studies, with either mass cytometry or single-cell RNA-seq, require statistical strategies robust to inter-individual donor variability and technical effects that can skew cell subset estimates. For example, mass cytometry studies need to control for variability in machine sensitivity, reagent staining, and sample handling that can lead to batch effects. Inter-individual differences can lead to real shifts in cell population frequencies at baseline, while technical effects can lead to apparent shifts in cell population frequencies. MASC test for case-control differences while accounting for technical and inter-individual variation. We acquired single cell data from individual case-control samples (**Figure 1**), and then applied MASC after stringent quality controls (see **Methods**).

First, we objectively define unbiased subsets of cells using an equal number of random cells from each sample so that they contributed equally to the subsequent analyses. Here we used DensVM *(26)* to cluster mass cytometry data, but recognize the availability of alternative methodologies. Next, we used MASC, a reverse association strategy, using single cell logistic mixed-effect modeling to test individual subsets for association. Here, we predict the subset membership of each cell based upon fixed effects (e.g. sex) and random effects (e.g. batch, donor) in a null model. We then measure the improvement in model fit when a fixed effect term for the case-control status of the sample is included with a likelihood ratio test. This framework

allowed us to evaluate the significance and effect size of the case-control association for each subset while controlling for inter-individual and technical variability.

To ensure that MASC was robust to statistical inflation, we ran it 10,000 times after permuting case-control labels using the dataset described below to obviate almost any case-control association that might be present. We then recorded the reported p-values for each cluster. This approach demonstrated that MASC has only a modestly inflated type I error rate for the 19 clusters in the dataset with 6.5% of trials obtaining $p<0.05$ (**Figure S1A**). Importantly, we found that including both donor and batch random effects was critical; eliminating random effects in the model led to highly inflated p-values with 66.1% of trials obtaining $p<0.05$ (**Figure S1B**). As an alternative and frequently used strategy, we also tested a simple binomial test for case-control association. This approach is limited in our view as it fails to model donor specific and technical effects. Unsurprisingly, we found that this commonly used approach produced highly inflated results in comparison to MASC with 65.7% of trials obtaining $p<0.05$ (**Figure S1C**).

### *EXPERIMENTAL STRATEGY*

Here, we applied this method to mass cytometric analysis of CD4+ T cells obtained from peripheral blood of patients with established RA (**Table 1**). We purified memory CD4+ T cells from RA patients (cases) and non-inflammatory controls and either (1) rested the cells for 24 hours or (2) stimulated the cells with anti-CD3/anti-CD28 beads for 24 hours. We then analyzed the cells using a 32-marker mass cytometry panel that included 22 markers of lineage, activation, and function (**Table S1**). This design allowed us to interrogate immune states across cases and controls and also capture stimulation-dependent changes. After stringent quality control measures, we analyzed a total of 26 RA cases and 26 controls. We randomly sampled

1000 cells from each of the 52 samples so that each sample contributed equally to the analysis, preventing samples that happened to have more cells captured by the mass cytometry assay from being overrepresented. We projected resting and stimulated data separately using the Barnes-Hut SNE algorithm so that all cells from all samples were projected into the same two dimensions using all markers in the panel with the exception of CD4 and CD45RO, which were used to gate CD4+ memory T cells for analysis.

### *CLUSTERING APPROACH*

Projecting the data into SNE space revealed areas of local density that consisted of predominantly cells from RA or OA samples (**Figure S2**). We wanted to identify CD4+ memory T cell populations in an unbiased way. Currently, clustering single cell data is an active area of research, and there is no consensus on the best clustering strategy. Hence we objectively considered multiple clustering algorithms options. We evaluated DensVM *(26)*, FlowSOM *(27)*, and Phenograph *(23)*, which were identified as among the best performing in a recent benchmarking comparison of clustering methods for high dimensional cytometry data *(28)*. The DensVM method uses an SNE projection of the dataset to first estimate the number of clusters by searching for local densities on the projection with varying bandwidths before classifying cells based on the similarity of expression. Phenograph works by creating a graph representing phenotypic similarities between cells and identifying clusters using Louvain community detection. The FlowSOM algorithm builds a self-organizing map with a minimum spanning tree to detect populations, then classifies cells in a meta-clustering step using consensus hierarchical clustering. We clustered cells with all three methods using the same selection of markers that was used to create SNE projections.

After running all three algorithms on the dataset, we identified 19, 19, and 21 clusters for DensVM, FlowSOM, and Phenograph respectively. An ideal clustering algorithm would define clusters that are distinct from each other in terms of marker intensities. However, we would ideally not want those cluster intensity differences to not be the cause of batch effects; that is that clusters should not be disproportionately constituted by cells from any individual batch.

In order to quantify the extent to which clustering approaches were defining clusters with distinct marker intensities, we utilized an information-theory based metric (**Methods**). All three methods generated clusters that were similar in their ability to generate clusters with distinct marker intensities (**Figure S3A**). Then, to determining whether those intensity differences were dependent on batch differences, we used a similar information-theory based metric to assess whether clusters were disproportionately represented by individual batches (**Methods**). Here, we observed that Phenograph and FlowSOM were much more sensitive to batch effects than clusters identified by DensVM ($p < 2 \times 10^{-3}$, Wilcoxon rank sum test, **Figure S3B**). Consistent with this quantitative assessment, we observed that in our data FlowSOM and Phenograph produced clusters that were constituted exclusively or dominated by cells from a specific batch. Thus, we chose to analyze clusters identified by the DensVM algorithm going forward.

### *Mass Cytometry Reveals The Landscape of CD4+ Memory T Cell Subsets*

We observed substantial diversity among resting CD4+ memory T cells, consistent with previous reports demonstrating the breadth of phenotypes in CD8+ T cells and CD4+ T cells *(29-31)*. We identified 19 distinct subsets in resting (*R*) memory CD4+ T cells (**Figure 2A, S4A, S15**). Central memory T cells ($T_{CM}$) segregated from effector memory T cells ($T_{EM}$) by the expression of CD62L (**Figure 2B**). Five subsets (subsets 1 – 5) of central memory T ($T_{CM}$) cells, all expressing CD62L, varied in the expression of CD27 and CD38, highlighting the

heterogeneity within the $T_{CM}$ compartment (**Figure 2E**). We identified two $T_h1$ subsets (subsets 8 and 12) as well as two $T_{reg}$ subsets (subsets 7 and 11). While both $T_{reg}$ subsets expressed high levels of CD25 and FoxP3, subset 11 also expressed HLA-DR, reflecting a known diversity among $T_{reg}$ populations in humans *(32)*.

When applied to the stimulated CD4+ T cell data in a separate analysis, DensVM identified 21 subsets (**Figure 2C, S4B, S15**). As expected, certain activation markers, like CD25 and CD40L, were broadly expressed across most subsets after stimulation (**Figure 2D**). Mass cytometry robustly detected cytokine production from stimulated CD4+ memory T cells (**Figure 2D, 2F**). Activated effector cells identified by the production of IL-2 were separated into three groups (subsets 1 – 3) by relative expression of TNF. Cytokine expression after stimulation also improved the ability to resolve certain CD4+ T effector subsets, such as $T_h17$ cells (subset 15) and $T_h1$ cells (subset 12). However, we were unable to resolve the $T_{reg}$ subtypes that were observed in the resting T cells; after stimulation, all CD25+ FoxP3+ cells were grouped together (cluster 21). The set of cells that did not activate after stimulation (subsets 16, 17, and 19) were easily identified by the lack of expression of activation markers such as CD25, CD40L and cytokines and the retention of high CD3 expression.

### *IDENTIFYING POPULATIONS THAT ARE ENRICHED OR DEPLETED IN RA SAMPLES*

We next sought to identify subsets that are significantly overrepresented or underrepresented by cells from cases. The percentage of RA cells in each subset ranged from 36.7% to 63.6% in resting state, and 38.9% to 65.7% in the stimulated state (**Table S2**, **Figure S5A**). Visualizing the density of the SNE projections revealed that related cells clustered into dense groups both at rest and after stimulation (**Figure 2B, 2D**), while plotting SNE projections of cells from cases and controls separately and coloring by density clearly suggested differential

abundance of RA cells among clusters (**Figure S2**). When we formally tested whether differences in cell populations were statistically significant accounting for subject-specific and batch-specific random effects with MASC (**Methods**), we observed three populations with significantly altered proportions of cells from cases in the resting T cell data. Most notably, we observed enrichment in subset 18 (p = $5.9 \times 10^{-4}$, **Table 2**, **Figure 3A**); this subset consisted 3.1% of total cells from cases compared to 1.7% of cells from controls and achieved an odds ratio of 1.9 (95% CI = 1.3 – 2.7). Conversely, subset 7 (p = $8.8 \times 10^{-4}$, OR = 0.6, 95% CI = 0.5 – 0.8) and subset 12 (p = $2.0 \times 10^{-3}$, OR = 0.5, 95% CI = 0.4 – 0.8) were underrepresented by RA cells.

To confirm the robustness of this analytical model, we conducted a stringent permutation test that was robust to potential batch effects. For this test, we reassigned case-control labels randomly to each sample; to stringently control for batch effects, retaining the same number of cases and controls in each batch. For each subset we recorded the number of case cells after randomization to define a null distribution. In the real data, we observed that 753 out of the 1184 cells in subset 18 were from case samples. In contrast, when we permuted the data we observed a mean of 550 cells and a standard deviation of 63 cells from case samples. In only 93 out of 100,000 permutations did we observed more than 753 cells from case samples in the randomized data (permutation p = $9.4\times10^{-4}$, **Table S2**). We applied permutation testing to every subset and found that the *p*-values produced by permutation were similar to *p*-values derived from the mixed-effect model framework (Spearman's r = 0.86, **Figure S5B, S5C**). When we considered the subsets identified as significant by MASC, both subsets 18 (permutation p = $9.4\times10^{-4}$) and 7 (permutation p = $1.8\times10^{-4}$) retained significance while subset 12 demonstrated nominal evidence of case-control association (permutation p = $1.3\times10^{-2}$).

Next, for each resting subset, we identified corresponding subsets in the stimulated dataset using a cluster centroid alignment strategy to calculate the distance between subsets across datasets (**Methods**). Subset 18 in the resting dataset was most similar to subset 18 in the stimulated dataset, while subset 7 in the resting dataset was most similar to subset 21 in the stimulated dataset (**Figure S5D, S5E**). Applying MASC, we observed case-control association for subset 18 in the stimulated data (p = $1.2\times10^{-3}$, OR = 1.7, 95% CI = 1.2 – 2.2), while subset 21 (p = 0.55) was not significant in the stimulated data **(Table 2, Figure 3B**). We identified one additional population subset that was significant in the stimulated dataset: subset 20 (p = $1.7\times10^{-3}$, OR = 1.7, 95% CI = 1.2 – 2.3) (**Table S3**).

We wanted to ensure that the results we observed were not an artifact of using DensVM to cluster the cytometry data. To this end, when we used Phenograph *(23)* and FlowSOM *(27)* to cluster the same dataset, we observed a CD27-, HLA-DR+ cluster with either method (**Figure S6A, S6C**), with association to RA by MASC (**Figure S6B, S6D**). We also assessed how analysis by MASC compares to Citrus, an algorithm that uses hierarchical clustering to define cellular subsets and then builds a set of models using the clusters to stratify cases and controls. When applied to the same case-control resting dataset, Citrus was unable to produce models with an acceptable cross-validation error rate, regardless of the method used (**Figure S7**).

### *CD27- HLA-DR+ CD4+ Effector Memory T Cells Are Expanded in RA Patients*

After noting that subset 18 demonstrated robust association to RA, we sought to understand the key features of this population. First, the lack of CD62L indicates that this subset is an effector memory T cell population (**Figure 2E, 2F**). To define the markers that best differentiated this subset from other cells, we calculated the Kullback-Liebler divergence for the

expression of each marker between the subset and the full dataset; we applied this approach to both the resting and stimulated datasets (**Methods**, **Figure 4A**). The expression of HLA-DR and perforin were notably increased in subset 18 compared to all other cells, while the expression of CD27 was decreased in this subset (**Figure 4B**). We observed that gating on CD27 and HLA-DR largely recapitulated subset 18 in the resting cells and subset 18 in the stimulated cells (**Figure 4C and 4D**), with an F measure of 0.8 and 0.7 respectively **(Methods**). Although HLA-DR is a known to be expressed on T cells in response to activation, it takes several days to induce strong HLA-DR expression *(33)*. Thus, it is likely that cells in subset 18 expressed HLA-DR prior to stimulation, such that analyses of both resting and stimulated cells identified the same HLA-DR+ CD27- effector memory T cell population.

In order to confirm the expansion of the CD27- HLA-DR+ T cell population in RA patients that we observed by mass cytometry, we evaluated the frequency of CD27- HLA-DR+ T cells in an independent cohort of 39 seropositive RA patients and 27 controls using conventional flow cytometry (**Table 1**). We determined the percentage of memory CD4+ T cells with a CD27- HLA-DR+ phenotype by gating individual samples from each group (**Figure 3C, Figure S8A, Figure S9**). Consistent with the mass cytometry analysis, CD27- HLA-DR+ cells were significantly expanded in the RA patient samples ($p = 0.011$, one-tailed *t* test, **Figure 3D**). The frequency of this subset was 0.8% in controls and 1.7% in RA samples, which is similar to the two-fold enrichment we observed in mass cytometry data.

To assess the effect of RA treatment on CD27- HLA-DR+ cell frequency, we quantified CD27- HLA-DR+ cell frequencies in 23 RA patients before and 3 months after initiation of a new medication for RA. We dichotomized patients as those who experienced a clinical response, defined as a reduction in CDAI scores (**Δ**CDAI-), versus those that did not, defined as an increase in CDAI scores (**Δ**CDAI+). We observed that changes in CD27- HLA-DR+ cell frequency tracks with their clinical response to treatment initiation ($p = 3.49 \times 10^{-4}$, Wilcoxon

signed-rank test, **Figure 5A**). Specifically, in the 18 **Δ**CDAI- patients, the cell frequency significantly reduced by 0.7 fold ($p = 0.006$, Wilcoxon signed-rank test, **Figure S10A**); in contrast, the 5 **Δ**CDAI+ patients in this same trial had 1.8 fold increase in CD27- HLA-DR+ cells. We observed a significant difference in the CD27- HLA-DR+ frequency fold-change between patients experiencing a CDAI reduction versus not ($p = 0.02$, Wilcoxon rank sum test, **Figure S10B**). The decrease in frequency of CD27- HLA-DR+ cells in patients responding to treatment escalation was accompanied by a slight increase in the frequency of CD27+ HLA-DR- cells: CD27+ HLA-DR- cells represented an average of 86.2% of CD4+ memory T cells before treatment, and an average of 87.9% of CD4+ memory T cells afterwards ($p = 0.009$, Wilcoxon signed-rank test, **Figure S10C**). These results confirm that CD27- HLA-DR+ CD4+ T cells are expanded in the circulation of RA patients and decrease with effective disease treatment.

To determine whether CD27- HLA-DR+ T cells are further enriched at the sites of inflammation in seropositive RA patients, we evaluated T cells in inflamed synovial tissue samples obtained at the time of arthroplasty (**Table 1**) and in inflammatory synovial fluid samples from RA patients. In a set of 9 synovial tissue samples with lymphocytic infiltrates observed by histology, the frequency of CD27- HLA-DR+ cells was significantly increased 5-fold ($p < 0.01$, Wilcoxon rank-sum, median 10.5% of memory CD4+ T cells) compared to blood (**Figure 5B**). Notably, in 2 of the tissue samples, >20% of the memory CD4+ T cells displayed this phenotype. CD27- HLA-DR+ cells were similarly expanded in synovial fluid samples from seropositive RA patients ($p < 0.001$, Wilcoxon rank-sum, median 8.9% of memory CD4+ T cells, n=8) Thus, CD27- HLA-DR+ T cells are enriched at the primary sites of inflammation in RA patients.

### *CD27- HLA-DR+ CD4+ Effector Memory T Cells Are Th1-skewed And Express Cytotoxic Factors*

To evaluate the potential function of the CD27- HLA-DR+ cell subset, we performed RNA-seq on CD27- HLA-DR+ CD4+ T cells with other effector populations to identify transcriptomic signatures for this subset (**Figure S8B**). We sorted and sequenced the following CD4+ T cell populations: naïve CD4+ T cells, central memory CD4+ T cells, regulatory CD4+ T cells, and all four CD27-/+ HLA-DR-/+ subsets from PBMCs from 7 RA cases and 6 OA controls. In total we generated 1.1 billion reads for 90 samples. We aligned reads with Kallisto*(34)* and applied stringent quality control.

We applied principal components analysis (PCA) on the expression data (**Figure 6A**). Strikingly, the first PC, capturing 4% of total variation, separated cell types along a naïve to effector axis, with the CD27- HLA-DR+ subset representing the extreme case with the highest PC1 values, and naïve T cells with the lowest PC1 values. We examined the genes with the highest PC1 loadings, and found that PC1 was associated negatively with CCR7 and positively with CXCR3 and CCR5, which we had previously observed as being elevated among CD27- HLA-DR+ cells. We note that CXCR3 and CCR5 are both chemokine receptors associated with a Th1 phenotype. In addition, PRF1 (perforin), a cytotoxic factor, was also strongly associated with PC1. The extreme position of CD27- HLA-DR+ cells along the continuum of CD4+ T cells suggested a possible late or terminal effector memory phenotype. To further explore this naïve to effector gradient, we used the gene loadings along PC1 to perform gene set enrichment analysis (GSEA) and identify the pathways that were most associated. Intriguingly, the naïve to CD27- HLA-DR+ axis was strongly correlated with naïve vs effector and natural killer (NK) vs CD4+ T cell gene signatures (**Figure 6B**).

The association of CXCR3 and CCR5 with PC1 prompted us to examine the expression of Th1-associated genes in these cells. CD27- HLA-DR+ cells expressed increased levels of typical Th1-associated genes including the Th1 signature cytokine IFN-γ and the transcription factor TBX21 (Tbet), in addition to CXCR3 and CCR5 (**Figure 6C**). In contrast, cytokines

characteristic of other polarized Th subsets (IL17A, IL4, IL21, TGFB1) and transcription factors associated with other Th subsets (RORC, GATA3, BCL6, FOXP3) were not elevated in CD27- HLA-DR+ cells compared to other effector populations (**Figure S11A**). Interestingly, we noted that the transcription factor CIITA was also increased in CD27- HLA-DR+ cells (**Figure 6C**). CIITA is well-known for its role as a regulator of MHC class II; indeed, we observed that targets of CIITA like HLA genes were significantly overexpressed in the CD27- HLA-DR+ population compared to other populations (**Figure S11B**).

Given the association of PRF1 with PC1, with CD27- HLA-DR+ cells at the extreme, we assessed the expression of cytotoxic gene programs more broadly by analyzing the enrichment of genes differentially expressed in CD27- HLA-DR+ cells in a set of genes highly differentially expressed in natural killer (NK) cells as compared to CD4+ T cells. Indeed, genes that were highly expressed in NK cells similarly showed high expression in CD27- HLA-DR+ cells, while genes with low expression in NK cells showed similarly low expression in CD27- HLADR+ cells (**Table S4**). Consistent with the enrichment results, a set of genes characteristically associated with cytolytic cells, including PRF1, GZMB, GZMA, GNLY, and NKG7 were all increased in expression in CD27- HLADR+ cells compared to most other T cell populations analyzed (**Figure 6D**). These results indicate that CD27- HLA-DR+ cells express a transcriptomic signature characteristic of cytotoxic cells.

We also evaluated the production and expression of cytokines and cytolytic factors at the protein level by intracellular flow cytometry. We assessed production of effector molecules by cells after *in vitro* stimulation with PMA + ionomycin, a stimulation method that readily reveals T cell capacity for cytokine production. Consistent with RNA-seq analyses, CD27- HLA-DR+ cells also more frequently expressed perforin ($p = 0.031$, one-tailed Wilcoxon signed-rank test) and granzyme A ($p = 0.015$, one-tailed Wilcoxon signed-rank test), than did CD27+ HLA-DR- cells, which constitute the majority of memory CD4+ T cells. In addition, CD27- HLA-DR+ cells

produced IFN-γ at a much higher frequency ($p = 0.009$, one-tailed Wilcoxon signed-rank test, **Figure 6E**). Percent positivity for each marker was determined by selecting an expression level threshold to determine positive staining (**Figure S12**). Taken together, broad analyses of broad gene expression and targeted measures of effector molecule production at the protein level indicate that CD27 HLA-DR+ T cells are a Th1-skewed T cell population capable of producing cytotoxic molecules.

## DISCUSSION

Mass cytometry has now been successfully applied to decipher heterogeneity of human T cells in multiple settings, including identification of disease-specific changes to circulating immune cell populations and mapping of developmental pathways *(29, 35-38)*. It, and other emerging single cell strategies, offer a promising avenue to characterize the T cell features, and other immunological features, of a wide-range of diseases in humans. However, successful application of mass cytometry on a larger scale (>20 samples) to conduct association studies on clinical phenotypes in humans has been limited thus far *(39-42)*, in part due to limited availability of effective association testing strategies.

Although there are many approaches to cluster single cell data *(19-23, 26, 27)*, extension of these methods to perform case-control association testing is not straightforward. The simplest strategy would be to define subsets of interest by clustering the cells and then comparing frequencies of these subsets in cases and controls with a univariate test such as a t-test or a non-parametric Mann-Whitney test. While commonly used in flow cytometry analysis, this approach is dramatically underpowered, as it relies upon reducing single-cell data to per-sample subset frequencies.

In contrast, our methodology takes full advantage of single cell measurements by using a “reverse-association” framework to test whether case-control status influences the membership of a given cell in a population. That is – each single cell is treated as a single event. But importantly – these events are not independent. Failing to account for dependencies, for example by using a binomial test or not accounting for random effects, results in dramatically inflated statistics and irreproducible associations (**Figure S1**). Using a mixed-effect logistic regression model allows us to account for covariance in single cell data induced by technical and biological factors that could confound association signals (**Figure 1, Figure S1B**), without inflated association tests. MASC also allows users to utilize technical covariates that might be

relevant to a single cell measurement (e.g. read depth per cell for single cell RNA-seq or signal quality for mass cytometry) that might influence cluster membership for a single cell.

Here we demonstrate that MASC identified a CD4+ T cell population that is significantly expanded in the circulation of RA patients in two different mass cytometry datasets. We determined a simple marker pair (CD27 and HLA-DR) that allowed for direct, manual gating of the T cell population of interest. We interrogated and reproduced the CD27- HLA-DR+ T cell population revealed by MASC using more targeted flow cytometric analysis in blood samples from an independent RA cohort. We note that Citrus, an association strategy to automatically highlight statistical differences between experimental groups and identify predictive populations in mass cytometry data *(43)*, was unable to identify the expanded CD27- HLA-DR+ population. We believe that our methodology compares favorably to Citrus because it incorporates both technical covariates (e.g. batch) and clinical covariates when modeling associations, a key feature when analyzing high-dimensional datasets of large disease cohorts. Additionally, the agglomerative hierarchical clustering framework in Citrus dramatically increases the testing burden and limits power.

Although we found that clustering with DensVM outperformed FlowSOM and Phenograph on our data (**Figure S3**), we want to emphasize that many clustering methods have emerged to analyze both cytometry and single cell RNA-seq data; clustering single cell data remains an active field of research *(44, 45)*. There continues to be controversy as to the best strategies, most appropriate metrics, and the optimal parameter choices*(44, 46-48)*. Available clustering approaches utilize a variety of methods, such as graph-based community detection, self-organizing maps *(27)*, and density-based clustering *(20, 23, 49)*. As neural networks (NNs) and as a subclass, autoencoders, have been proven to be powerful general nonlinear models, we implemented an autoencoder-based clustering approach that we tested on our datasets. We found this clustering method also had potential, and should be further investigated in the future

(**Figure S13**). As different clustering strategies may be more appropriate for different datasets, we have implemented MASC specifically to allow for different clustering options based on user preference.

The CD27- HLA-DR+ T cells identified by MASC in blood samples from RA patients were further enriched in both synovial tissue and synovial fluid of RA patients. The accumulation of these cells in chronically inflamed RA joints, combined with lack of CD27, suggests that these cells have been chronically activated. Loss of CD27 is characteristic of T cells that have been repeatedly activated, for example T cells that recognize common restimulation antigens *(50-52)*. Importantly, the broader CD27- CD4+ T cell population does not itself differ in frequency between RA patients and controls, in contrast to the expanded population of CD27- HLA-DR+ cells identified by MASC. We hypothesize the CD27- HLA-DR+ T cell population in RA patients may be enriched in RA antigen-specific T cells, offering a potential tool to identify relevant antigens in RA.

Expression of HLA-DR suggests recent or continued activation of these cells *in vivo*. The well-known HLA class II association to RA may also suggest an important role for disease susceptibility for this subset. Despite the suspected chronic activation of these cells, CD27- HLA-DR+ cells do not appear functionally exhausted. CD27- HLA-DR+ cells rapidly produce multiple effector cytokines upon stimulation *in vitro*, with a dramatically increased predisposition to IFN-γ production. These cells also show increased expression of cytolytic molecules like granzyme A and perforin on both the transcriptomic and proteomic level. CD4+ cytotoxic T cells expressing granzyme A and perforin, also with a CD27- phenotype, are reported to be expanded in patients with chronic viral infections *(53)*. Of note, CD4+ cells that are perforin+ and granzyme A+ have been observed in RA synovial samples *(54, 55)* and other chronic inflammatory conditions *(56)*. Interestingly, cytolytic CD4+ T cells lack the capacity to provide B cell help, suggesting that this is a distinct population from the expanded T peripheral helper cell

population in seropositive RA *(13, 55)*. Taken together, these findings nominate CD27- HLA-DR+ T cells as a potential pathogenic T cell population that may participate in the chronic autoimmune response in RA *(57)*.

Our data suggests that RA treatment response in a cohort of patients initiating a new RA therapy is associated with a reduction in the frequency of CD27- HLA-DR+ T cells. However, before this cellular population can be used effectively as a biomarker, additional longitudinal data on non-responders and healthy controls will be necessary to understand how variable population frequencies are both within and across individuals. We hypothesize that effective RA treatments inhibit the strong activation or expansion of effector CD4+ T cells, thereby reducing the number of cells that acquire a CD27- HLA-DR+ phenotype. Although we were able to detect significant associations between the frequency of CD27- HLA-DR+ cells and clinical response (**Figure 5A**), we recognize the need for a larger study to detect other subtler subphenotypic associations, such as association with specific therapeutics or disease activity (**Figure S14**). To this end, larger prospective cohorts with deep immunophenotyping of CD4+ T cells in blood and tissue will be critical. Currently, the Accelerating Medical Partnerships (AMP) Rheumatoid Arthritis/Systematic Lupus Erythematosus network is now recruiting and immunophenotyping prospective cohorts, and we anticipate that finding associations of this T cell population to other key clinical parameters will soon become possible.

In summary, the MASC single-cell association modeling framework identified a Th1-skewed cytotoxic effector memory CD4+ T cell population expanded in RA using a case-control mass cytometry dataset. The MASC method is extendible to any case-control experiment in which single cell data is available, including flow cytometry, mass cytometry, and single cell RNA-seq datasets. Although current single cell RNA-seq studies are not yet large scale, ongoing projects like the AMP RA/SLE network or the Human Cell Atlas aim to provide these

datasets and may benefit from using the MASC framework in case-control testing, or testing for other clinical subphenotypes such as specific treatment response or disease progression.

## MATERIALS AND METHODS

### *EXPERIMENTAL DESIGN*

Human subjects research was performed in accordance with the Institutional Review Boards at Partners HealthCare and Hospital for Special Surgery via approved protocols with appropriate informed consent as required. Patients with RA fulfilled the ACR 2010 Rheumatoid Arthritis classification criteria, and electronic medical records were used to ascertain patients' rheumatoid factor and anti-CCP antibody status, C-reactive protein level, and medication use. Synovial tissue samples for mass and flow cytometry were collected from seropositive RA patients undergoing arthroplasty at the Hospital for Special Surgery, New York or at Brigham and Women's Hospital, Boston. Samples with lymphocytic infiltrates on histology were selected for analysis. Sample inclusion criteria were established prospectively, with the exception of samples that were excluded from the study due to poor acquisition by the mass cytometer.

Synovial fluid samples were obtained as excess material from a separate cohort of patients undergoing diagnostic or therapeutic arthrocentesis of an inflammatory knee effusion as directed by the treating rheumatologist. These samples were de-identified; therefore, additional clinical information was not available.

Blood samples for clinical phenotyping were obtained from consented patients seen at Brigham and Women's Hospital. We performed medical record review for ACPA (anti-citrullinated protein antibody) positivity according to CCP2 (cyclic citrullinated protein) assays, C-reactive protein (CRP), and RA-specific medications including methotrexate and biologic DMARDS. For blood cell analyses in the cross-sectional cohort, the treating physician measured the clinical disease activity index (CDAI) on the day of sample acquisition. For RA patients followed longitudinally, a new disease-modifying antirheumatic drug (DMARD) was initiated at the discretion of the treating rheumatologist, and CDAIs were determined at each visit by trained research study

staff. Blood samples were acquired before initiation of a new biologic DMARD or within 1 week of starting methotrexate and 3 months after initiating DMARD therapy*(58)*. Concurrent prednisone at doses ≤10mg/day were permitted. All synovial fluid and blood samples were subjected to density centrifugation using Ficoll-Hypaque to isolate mononuclear cells, which were cryopreserved for batched analyses.

### SAMPLE PREPARATION FOR MASS CYTOMETRY

We rapidly thawed cryopreserved peripheral blood mononuclear cells (PBMCs) and isolated total CD4+ Memory T cells by negative selection using MACS magnetic bead separation technology (Miltenyi #130-091-893). Subsequently, we rested the CD4+ Memory T cells for 24 hours in complete RPMI (Gibco #11875093) sterile-filtered and supplemented with 15% FBS, 1% Pen/Strep (Gibco #15140), 0.5% Essential and Non-Essential Amino Acids (Gibco #11130051 and Gibco #11140050), 1%Sodium Pyruvate (Gibco #11360070), 1% HEPES (Gibco #15630080), and 55μM 2-mercaptoethanol (Gibco #21985023). We activated the cells using Human T-Activator CD3/CD28 Dynabeads (ThermoFisher #11131D) at a density of 1 bead:2 cells. At 6 hours prior to harvesting (t=18 hours of stimulation), we added Monensin and Brefeldin A 1:1000 (BD GolgiPlug #555029 and BD GolgiStop #554724). After 24 hours of stimulation, we incubated the cells with a rhodium metallointercalator (Fluidigm #201103A) in culture at a final dilution of 1:500 for 15 minutes as a viability measure. We then harvested cells into FACS tubes and washed with CyTOF Staining Buffer (CSB) composed of PBS with 0.5% BSA (Sigma-Aldrich #A3059), 0.02% sodium azide (Sigma-Aldrich #71289), and 2μM EDTA (Ambion #AM9260G). We spun the cells at 500xg for 7 minutes at room temperature. We incubated the resulting cell pellets in 10ul Fc Receptor Binding Inhibitor Polyclonal Antibody (eBiosciene #14-9161-71) and 40ul of CSB for 10 minutes at 4C. The samples were then incubated for 30 minutes at 4C on a shaker rack with 1ul of the following

eighteen CyTOF surface antibodies in a cocktail brought to a volume of 50ul/sample in CSB: Anti-Human CD49D (9F10)-141Pr (Fluidigm #3141004B), Anti-Human CCR5 (CD195)(NP-6G4) - 144Nd (Fluidigm #3144007A), Anti-Human CD4 (RPA-T4) -145Nd (Fluidigm #3145001B), Anti-Human CD8a (RPA-T8) -146Nd (Fluidigm #3146001B), Anti-Human CD45RO-147Sm (BWH CyTOF Core), Anti-Human CD28-148Nd (BWH CyTOF Core), Anti-Human CD25 (IL-2R) (2A3) - 149Sm (Fluidigm #3149010B), Anti-Human PD1-151Eu (BWH CyTOF Core), Anti-Human CD62L (DREG-56)-153Eu (Fluidigm #3153004B), Anti-Human CD3 (UCHT1) -154Sm (Fluidigm #3154003B), Anti-Human CD27 (L128) -155Gd (Fluidigm #3155001B), Anti-Human CD183[CXCR3](G025H7)-156Gd (Fluidigm #3156004B), Anti-Human CCR7-170Er (BWH CyTOF Core), Anti-Human ICOS-160Gd (BWH CyTOF Core), Anti-Human CD38 (HIT2)-167Er (Fluidigm #3167001B), Anti-Human CD154 (CD40L) (24-31)-168Er (Fluidigm #3168006B), Anti-Human CXCR5[CD185](51505)-171Yb (Fluidigm #3171014B), and Anti-Human HLA-DR (L243) -174Yb (Fluidigm #3174001B). We washed the cells with 1 ml of CSB and spun at 700xg for 5 minutes at RT. Post spin, we aspirated the buffer from pellet and added 1 ml 1:4 ratio of concentrate to diluent of a Foxp3 / Transcription Factor Staining Buffer Set (eBioscience #00-5523-00) supplemented with formaldehyde solution (Sigma-Aldrich #F1268) to a final concentration of 1.6%. We incubated the cells at room temperature on a gentle shaker in the dark for 45 minutes. We washed the cells with two ml of CSB + 0.3% saponin (CSB-S), and spun at 800xg for 5 minutes. We incubated the cell pellet with 1ul of the following fourteen intracellular antibodies and 10ul of a solution of Iridium (1:25) in CSB-S brought to a total volume of 100ul for 35 minutes at r.t. on a gentle shaker: Anti-Human IL-4 (MP4-25D2)-142 (Fluidigm #3142002B), Anti-Mouse/Human IL-5 (TRFK5) -143Nd (Fluidigm #3143003B), Anti-Human IL-22 (22URTI) -150Nd (Fluidigm #3150007B), Anti-Human TNFα (Mab11) -152Sm (Fluidigm #3152002B), Anti-Human IL-2 (MQ1-17H12)-158Gd (Fluidigm #3158007B), Anti-Human IL21-159Tb (BWH CyTOF Core), Anti-Human IFNg (B27) -165Ho (Fluidigm #3165002B), Anti-Human GATA3-166Er (BWH CyTOF Core), Anti-Human IL9-172Yb (BWH CyTOF Core), Anti-Human Perforin (B-

D48)-175Lu (Fluidigm #3175004B), Anti-Human IL10-176Yb (BWH CyTOF Core), Anti-Human IL17A-169Tm (BWH CyTOF Core), Anti-Human Foxp3 (PCH101)-162Dy (Fluidigm #3162011A), and Anti-Human Tbet-164Dy (BWH CyTOF Core). Post-incubation, we washed the cells in PBS and spun them at 800xg for 5 minutes. We resuspended the pellet in 1 ml of 4% formaldehyde prepared in CSB and incubated the cells for 10 minutes at r.t. on a gentle shaker, followed by another PBS wash and spin at 800xg for 5 minutes. We washed the pellet in 1ml of MilliQ deionized water, spun at 800xg for 6 minutes, and subsequently resuspended the resulting pellet in deionized water at a concentration of 700,000 cells per ml for analysis via the CyTOF 2. We transferred the suspensions to new FACS tubes through a 70μm cell strainer and added MaxPar EQ Four Element Calibration Beads (Fluidigm #201078) at a ratio of 1:10 by volume prior to acquisition.

### MASS CYTOMETRY PANEL DESIGN

We designed an antibody panel for mass cytometry with the goal of both accurately identifying CD4+ effector memory T cell populations and measuring cellular heterogeneity within these populations. We chose markers that fell into one of five categories to generate a broadly informative panel: chemokine receptors, transcription factors, lineage markers, effector molecules, and markers of cellular activation and exhaustion (**Table S1**). Antibodies were obtained from DVS or from the Longwood Medical Area CyTOF Antibody Resource Core (Boston, MA).

### MASS CYTOMETRY DATA ACQUISITION

We analyzed samples at a concentration of 700,000 cells/ml on a Fluidigm-DVS CyTOF 2 mass cytometer. We added Max Par 4-Element EQ calibration beads to every sample that was run on the CyTOF 2, which allowed us to normalize variability in detector sensitivity for samples run in

different batches using previously described methods*(59)*. We used staining for iridium and rhodium metallointercalators to identify viable singlet events. We excluded samples where acquisition failed or only yielded a fraction of the input cells (< 5%). The criteria for sample exclusion were not set prospectively but were maintained during all data collection runs.

Since samples were processed and analyzed on different dates, we ran equal numbers of cases and controls each time to guard against batch effects. However, as CyTOF data are very sensitive to day-to-day variability, we took extra steps to pre-process and normalize data across the entire study.

### FLOW CYTOMETRY SAMPLE PREPARATION

For flow cytometry analysis of the validation and longitudinal blood cohorts and synovial samples, cryopreserved cells were thawed into warm RPMI/10% FBS, washed once in cold PBS, and stained in PBS/1% BSA with the following antibodies for 45 minutes: anti-CD27-FITC (TB01), anti-CXCR3-PE (CEW33D), anti-CD4-PE-Cy7 (RPA-T4), anti-ICOS-PerCP-Cy5.5 (ISA-3), anti-CXCR5-BV421 (J252D4), anti-CD45RA-BV510 (HI100), anti-HLA-DR-BV605 (G46-6), anti-CD49d-BV711 (9F10), anti-PD-1-APC (EH12.2H7), anti-CD3-AlexaFluor700 (HIT3A), anti-CD29-APC-Cy7 (TS2/16), propidium iodide. Cells were washed in cold PBS, passed through a 70-micron filter, and data acquired on a BD FACSAria Fusion or BD Fortessa using FACSDiva software. Samples were analyzed in uniformly processed batches containing both cases and controls.

### FLOW CYTOMETRY INTRACELLULAR CYTOKINE STAINING

Effector memory $CD4^+$ T cells were purified from cryopreserved PBMCs by magnetic negative selection (Miltenyi) and rested overnight in RPMI/10%FBS media. The following day, cells were stimulated with PMA (50ng/mL) and ionomycin (1µg/mL) for 6 hours. Brefeldin A and monensin (both 1:1000, eBioscience) were added for the last 5 hours. Cells were washed twice in cold PBS, incubated for 30 minutes with Fixable Viability Dye eFluor 780 (eBioscience), washed in PBS/1%BSA, and stained with anti-CD4-BV650 (RPA-T4), anti-CD27-BV510 (TB01), anti-HLA-DR-BV605 (G46-6), anti-CD20-APC-Cy7 (2H7), and anti-CD14-APC-Cy7 (M5E2). Cells were then washed and fixed and permeabilized using the eBioscience Transcription Factor Fix/Perm Buffer. Cells were then washed in PBS/1%BSA/0.3% saponin and incubated with anti-IFN-γ-FITC (B27), anti-TNF-PerCp/Cy5.5 (mAb11), anti-IL-10-PE (JES3-9D7) and anti-IL-2-PE/Cy7 (MQ1-17H12) or anti-granzyme A-AF647 (CB9) and anti-perforin-PE/Cy7 (B-D48) for 30 minutes, washed once, filtered, and data acquired on a BD Fortessa analyzer. Gates were drawn to identify singlet T cells by FSC/SSC characteristics, and dead cells and any contaminating monocytes and B cells were excluded by gating out eFluor 780-positive, CD20+, and CD14+ events.

### SYNOVIAL TISSUE PROCESSING

Synovial samples were acquired after removal as part of standard of care during arthroplasty surgery. Synovial tissue was isolated by careful dissection, minced, and digested with 100µg/mL LiberaseTL and 100µg/mL DNaseI (both Roche) in RPMI (Life Technologies) for 15 minutes, inverting every 5 minutes. Cells were passed through a 70µm cell strainer, washed, subjected to red blood cell lysis, and cryopreserved in Cryostor CS10 (BioLife Solutions) for batched analyses.

**RNA LIBRARY PREPARATION AND SEQUENCING**

Seven rheumatoid arthritis (RA) case samples and six osteoarthritis (OA) control samples were flow sorted into the following seven subsets for low-input RNA-Sequencing: Central Memory T cells (TCM), Naïve T cells (TN), Regulatory T cells (TReg), DR+27+ Effector T cells (DR+27+), DR+27- Effector T Cells (DR+27-), DR-27+ Effector T Cells (DR-27+), and DR-27- Effector T Cells (DR-27-). We rapidly thawed the case and control samples of cryopreserved peripheral blood mononuclear cells, and MACS enriched the samples for CD4+ T cells (Miltenyi 130-096-533). We rested the samples overnight in complete RPMI/10% FBS. Following the rest, we prepared the samples for fluorescence activated cell sorting (FACS). We washed the cells once in cold PBS, and incubated them with eBioscience human FC Receptor Binding Inhibitor (Thermo #14-9161-73). Subsequently, we stained the samples in PBS/5% FBS for 45 minutes with the following antibodies: FITC CD27 (Biolegend #302806), PE CD25 (Biolegend #356104), Pe/Cy7 CD127(IL-7Ra) (Biolegend #351320), Brilliant Violet 510 HLA-DR (Biolegend #307646), Brilliant Violet 605 CD45RA (Biolegend #304134), APC CD62L (Biolegend #304810), Alexa Fluor 700 CD4 (Biolegend #300526), APC/Cy7 CD14 (Biolegend #367108), and APC/Cy7 CD19 (Biolegend #302217). Post-stain, we washed the cells in cold PBS, passed them through a 70-micron filter, and acquired the samples on a BD FACSAria Fusion cytometer. 1000 cells from each subset were sorted and collected into 5ul of TCL Lysis Buffer (Qiagen #1031576) with 1% b-me. We processed and collected the samples uniformly in batches containing both cases and controls, and randomized the samples within the plate. We prepared sequencing libraries using the Smart-Seq2 protocol. Sequenced libraries were pooled and sequenced with the Illumina HiSeq 2500 using 25bp paired-end reads. We removed one outlier with low read depth. The remaining libraries were sequenced to a depth of 6-19M reads.

### GENE EXPRESSION QUANTIFICATION

åWe quantified cDNAs on canonical chromosomes in Ensembl release 83 with Kallisto v0.43.1 in transcripts per million (TPM) and summed to get gene-level expression values. We filtered transcripts that did not have at least 5 counts in at least 10 samples. For differential expression analysis and principal components analysis (PCA), we used log (base 2) transformed TPM values.

### PRINCIPAL COMPONENT ANALYSIS

We selected 3087 well expressed, variable genes for PCA, by excluding genes with lower than 1 mean or 0.75 standard deviation expression. We used the removeBatchEffect function in the R Limma package to regress out donor specific contributions to gene expression. We normalized the resulting data by scaling and centering each gene and then scaling and centering each sample. Finally, we used prcomp in R to perform PCA on the filtered, regressed, and normalized data.

### CORRELATION ANALYSIS

For individual genes, we computed association of their expression with the proposed ordering of cell types, along the naïve to effector gradient. In this association, we modeled expression as a linear function of cell type, encoded as an ordinal variable, and donor, one-hot encoded as a categorical variable. The statistical significance of the association was estimated with the lm function in R, followed by adjustment using the Benjamini Hochberg procedure.

### GENE SET ENRICHMENT ANALYSIS

We performed gene set enrichment analysis using the R package gage directly on orderings defined by previous analyses. To enrich pathways from the PCA results, we used gene loadings for each principle component. For differential expression, we used the estimated *t* statistics. In

order to produce barcode plots for select pathways, we reanalyzed these pathways using the R *liger* package.

## ***STATISTICAL ANALYSIS***

### MIXED EFFECTS MODELING OF ASSOCIATIONS OF SINGLE CELLS (MASC)

Here, we present a flexible method of finding significant associations between subset abundance and case-control status that we have named MASC (Mixed effects modeling of Associations of Single Cells). The MASC framework has three steps: (1) stringent quality control, (2) definition of population clusters, and (3) association testing. Here we assume that we have single cell assays each quantifying *M* possible markers where markers can be genes (RNA-seq) or proteins (cytometry).

***Quality control.*** To mitigate the influence of batch effects and spurious clusters, we first remove poorly recorded events and low quality markers before further analysis. We remove those markers (1) that have little expression, as these markers are not informative, and (2) with significant batch variability. First, we concatenated samples by batch and measured the fraction of cells negative and positive for each marker. We then calculated the ratio of between-batch variance to total variance for each marker's negative and positive populations, allowing us to rank and retain 20 markers that were the least variable between batches. We also removed markers that were either uniformly negative or positive across batches, as this indicated that the antibody for that marker was not binding specifically to its target. For single-cell transcriptomic data, an analogous step would involve removing genes with low numbers of supporting reads or genes whose expression varies widely between batches.

Once low quality markers have been identified and removed, we remove events that are likely to be artifacts. We first remove events that have extremely high signal for a single marker: events that have recorded expression values at or above the 99.9$^{th}$ percentile for that marker are removed. These events are unlikely to be intact, viable cells. Next, a composite "information content" score (eq. 1) for each event *i* is created in the following manner: the expression *x* for each marker *M* is rescaled from 0 to 1 across the entire dataset to create normalized expression values $y_i$ for each event *i*. The sum of these normalized expression values is used to create the event's information content score.

1) $INFO_i = \sum_{m=1}^{m=M} y_{i,m}$

The information content score reflects that events with little to no expression in every channel are less informative than events that have more recorded expression. Events with low scores ($INFO_i < 0.05$) and are unlikely to be informative in downstream analysis are removed. In addition, events that derive more than half of their information content score from expression in a single channel are also removed (eq. 2):

2) $INFO_i * 0.5 < \max_{m \in M}(y_{i,m})$

Potential explanations for these events include poorly stained cells or artifacts caused by the clumping of antibodies with DNA fragments. A final filtering step retains events that were recorded as having detectable expression in at least $M_{min}$ markers, where $M_{min}$ may vary from experiment to experiment based on the panel design and expected level of co-expression between channels. The quality control steps described here are specific for mass cytometry analysis and will need to be optimized for use with transcriptomic data.

***Clustering.*** After applying quality control measures to each sample, we combine data from cases and controls into a single dataset. It is critical to ensure that each sample contributes an equal numbers of cells to this dataset, as otherwise the largest samples will dominate the analysis and confound association testing. We suggest sampling an equal number of cells from each sample. After sampling the data, we partition these cells into populations using DensVM, which performs unsupervised clustering based on marker expression. We note that partitioning the data can be accomplished with different clustering approaches – such as SPADE or PhenoGraph for mass cytometry data – or even by using traditional bivariate gating, as MASC is not dependent on any particular method of clustering (**Figure S5**)

***Association testing.*** Once all cells have been assigned to a given cluster, the relationship between single cells and clusters is modeled using mixed-effect logistic regression to account for donor or technical variation (eq. 3). We model the age and sex of sample *k as* fixed effect covariates, while the donor and batch that cell *i* belongs to are modeled as random effects. The random effects variance-covariance matrix treats each sample and batch as independent gaussians. Each cluster is individually modeled. Note that this baseline model does not explicitly include any single cell expression measures.

3) $\log\left[\frac{Y_{i,j}}{1-Y_{i,j}}\right] = \theta_j + \beta_{clinical}X_{i,k} + (\phi_i|k) + (\kappa_i|m)$

Where $Y_{i,j}$ is the odds of cell *i* belonging to cluster *j*, $\theta_j$ is the intercept for cluster *j*, $\beta_{clinical}$ is a vector of clinical covariates for the k$^{th}$ sample, $(\phi_i|k)$ is the random effect for cell *i* from k$^{th}$ sample, $(\kappa_i|m)$ is the random effect for cell *i* from batch *m*.

To determine if any clusters are associated with case-control status, we include an additional covariate that indicates whether the k$^{th}$ sample is a case or control (eq. 4)

4) $\log\left[\frac{Y_{i,j}}{1-Y_{i,j}}\right] = \theta_j + \beta_{clinical}X_{i,k} + (\phi_i|k) + (\kappa_i|m) + \beta_{case}X_{i,k}$

Here, $Y_{i,j}$ is the odds of cell *i* belonging to cluster *j*, $\theta_j$ is the intercept for cluster *j*, $\beta_{clinical}$ is a vector of clinical covariates for the k$^{th}$ sample, $(\phi_i|k)$ is the random effect for cell *i* from k$^{th}$ sample, $(\kappa_i|m)$ is the random effect for cell *i* from batch *m*, $\beta_{case}$ indicates the effect of k$^{th}$ sample's case-control status.

5) $D = -2 * \ln\left(\frac{\text{likelihood for null model}}{\text{likelihood for full model}}\right)$

6) $p = 1 - \left(\frac{t^{(v-2)/2}e^{-t/2}}{2^{v/2}\Gamma(\frac{v}{2})}\right)$

We compare the two models using a likelihood ratio test (eq. 5) to find the test statistic *D*, which is the ratio of the likelihoods for the baseline and full models. The term *D* is distributed under the null by a $\chi^2$ distribution with 1 degree of freedom, as there is only one additional parameter in the full model compared to the null (case-control status). We derive a *p*-value by comparing test statistic *D* of the likelihood ratio test to the value of the $\chi^2$ distribution with 1 degrees of freedom (eq. 6), allowing us to find clusters in which case-control status significantly improves model fit. A significant result indicates that cluster membership for a single cell is influenced by case-status after accounting for technical and clinical covariates. The effect size of the case-control association can be estimated by calculating the odds ratio from $\beta_{case}$. If a dataset includes multiple groups, then we can test for association between *g* groups using *g*-1 indicator variables. This approach allows us to capture inter-individual differences between donors, as well as model the influence of technical and clinical covariates that might influence a cell to be included as a member of one cluster versus another.

## MASS CYTOMETRY DATA ANALYSIS

We analyzed 50 samples (26 cases, 24 controls) in the resting condition and 52 samples (26 cases, 26 controls) in the stimulated condition by CyTOF. Two samples were only analyzed after stimulation due to low numbers of available PBMCs. We ran aliquots of a standard PBMC sample alongside cases and controls with each CyTOF run to allow us to measure batch variability directly, as these aliquots should not be biologically dissimilar. We then used these data to find markers that had stained poorly or varied significantly between batches and removed them analysis. After acquisition, each sample was gated to a $CD4^+$, $CD45RO^+$ population using FlowJo 10.1 (TreeStar, USA) and combined into a single dataset before analyzing the data using MASC as previously described. We performed the data filtration steps requiring cells to demonstrate expression in at least 5 markers ($M_{min}$ = 5). This removed 3.0-6.0% of all events captured in each sample. We first removed the initial noise factor applied to all zero expression values in mass cytometry by subtracting 1 from expression values and setting any negative values to 0, then applied the inverse hyperbolic sine transform with a cofactor of 5 to the raw expression data, using the following equation: $y = \sinh^{-1} \frac{\max(x-1,0)}{5}$

To partition the data, we first randomly selected 1000 cells from each sample and applied the t-Distributed Stochastic Neighbor Embedding (tSNE) algorithm (Barnes-Hut implementation*(60)*) to the reduced dataset with the following parameters: perplexity = 30 and theta = 0.5. We did not include channels for CD4 or CD45RO in the SNE clustering as these markers were only used in gating samples to confirm the purity of CD4 memory T cell selection. We performed separate SNE projections for resting and stimulated cells. To identify high-dimensional populations, we used a modified version of DensVM*(26)*. 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 SNE projection.

To create elbow plots, we ran DensVM using 25 bandwidth settings evenly spaced along the interval [0.61, 5] for the resting data and [0.66, 5] for the stimulated data. We normalized marker expression to mean 0 variance 1 in each dataset before calculating the fraction of total variance explained by between-cluster variance.

Association testing for each cluster was performed using mixed effect logistic regression according to the MASC method. Donor and batch were included as random-effect covariates, donor sex and age was included as a fixed-effect covariate, and donors were labeled as either cases (RA) or controls (OA). To confirm the associations found by MASC, we conducted exact permutation testing in which we permuted the association between case-control status and samples within batches 10,000 times, measuring the fraction of cells from RA samples that contributed to each cluster in each permutation. This allowed us to build an empirical null for the case-control skew of each cluster, and we could then determine for each cluster how often a skew equal or greater to the observed skew occurred. We adjusted the p-values for the number of tests we performed (the number of clusters analyzed in each condition) using the Bonferroni correction.

Cluster Alignment

We aligned subsets between experiments using the following strategy: In each experiment, expression data was first scaled to mean zero, variance one to account for differences in sensitivity. We used the mean expression value for marker column to define a centroid for each cluster in both datasets. For a given cluster in the first dataset (query dataset), we calculated the Euclidean distance (eq. 7) between that cluster and cluster centroids in the second dataset (target dataset) across all shared markers. The cluster that is most similar to the cluster in query dataset is the cluster with the lowest distance in the target dataset, relative to all other clusters in that dataset.

7) $\mathrm{dist}(q,t) = \sum_{k=1}^{K} \sqrt{(q_k - t_k)^2}$

Here, $q$ refers to the query cluster and $t$ to the target cluster, while $q_k$ and $t_k$ indicate the normalized mean expression of marker $k$ (out of $K$ total markers) in clusters $q$ and $t$ respectively.

Marker Specificity

We wanted to determine which markers best separated a given population from the rest of the data in a quantitative manner, as finding a set of population-specific markers is crucial for isolating the population *in vivo*. In order to do this, we examined the distribution of expression for each marker individually in the entire dataset (*Q*) and the population of interest (*P*). We then grouped expression values for *P* and *Q* into 100 bins, normalized the binned vector to 1, and calculated the Kullback-Leibler divergence from *Q* to *P* for that marker with the following equation:

$$D_{KL}(P \parallel Q) = \sum_{i=1}^{i=100} P_i \log \frac{P_i}{Q_i}$$

The divergence score can be interpreted as a measure of how much the distribution of expression for a given marker in the entire dataset resembles the distribution of expression for that marker in the population of interest. Higher scores represent lower similarity of the marker's expression distributions for *P and Q*, indicating that the expression profile of that marker is more specific for that population. By calculating this score for every marker, we can rank and identify markers that best differentiate the population of interest from the dataset.

Biaxial Gating and Cluster Overlap

To determine the concordance between biaxial gating of CD27- HLA-DR+ cells and cluster 18, we first selected cells that had normalized expression values of CD27 < 1 and HLA-DR >= 1.

We then calculated an F-measure statistic between the cells selected using the CD27 and HLA-DR gates and each cluster identified in the resting and stimulated datasets (eq. 8). Here, precision is defined as the number of cells in each cluster tested that fall into the CD27- HLA-DR+ gate, while recall is defined as the number of cells gated as CD27- HLA-DR+ that are in each cluster.

8) $F_{measure} = 2\times\frac{precision \times recall}{precision + recall}$

Clustering Method Comparison

We compared clustering sets that contained either 19 (FlowSOM and DensVM) or 21 (Phenograph) clusters. We independently clustered the resting dataset with Phenograph and FlowSOM using the same cells and markers used to cluster the data with DensVM. We set *k* to 19 for FlowSOM clustering to match the number of clusters found by DensVM; for Phenograph, we used the default setting of k = 30.

To evaluate quantify the ability of different clustering algorithms to define clusters that was explaining marker fluctuations, we defined an information theory based metric to evaluate the relative information content captured by each set of clusters in terms of marker intensity. We selected this approach since it is separate from the objective functions that the clustering algorithms were attempting to optimize.

First, for each cell, we normalized marker intensities so that they summed to one. Then we defined a null $Q_i$ representing the average normalized intensity for marker *i* across all cells. We also defined $P_{i,j}$ which is the mean intensity of marker *i* of cells from cluster *j*. Then for each cluster *j* we calculate their KL divergence for each of the *M* markers (eq. 9).

9) $D_{KL,j}\left(P_j \parallel Q\right) = \sum_i^M P_{i,j} \ln\frac{P_{i,j}}{Q_i}$

A cluster with low divergence from the average expression of markers across the entire dataset has captured less marker intensity information than one with a high divergence, as biologically valid clusters will have unique marker profiles that differ greatly from one another and from the average marker expression profile.

We defined a similar metric to quantify the extent to which individual batches were accounting for differences in cluster composition. In this instance we calculated $P_{i,j}$ which is the proportion of cells from cluster *j* that batch *i* contributed. We also calculate $Q_i$ which is the proportion of cells that batch *i* contributes overall to the dataset. With this definition we calculate the KL divergence for each of the *M* batches (eq. 10).

$$10)\ D_{KL,j}\left(P_{\cdot j} \parallel Q\right) = \sum_i^M P_{i,j} \ln \frac{P_{i,j}}{Q_i}$$

A cluster that contains cells with low divergence from the null distribution of cells across batches is affected less by batch effects than one with a high divergence score, and a cluster completely free of batch effects should have a K-L divergence of zero.

Autoencoder Clustering

We performed clustering analysis using a deep autoencoder and a Gaussian Mixture Model (GMM). The autoencoder was designed with an architecture of 3 hidden layers, with depths of 8, 2, and 8 nodes, respectively. We trained the model with no regularization and 300 epochs. The two modes from the middle hidden layer were then used as features to learn a GMM. The number of clusters was chosen a priori to match the number discovered in the DensVM analysis. All parameters not specific in this section were set to default values.

All analyses were performed using custom scripts for R 3.4.0. We used the following packages: *flowCore* to read and process FCS files for further analysis, *lme4* to apply mixed effect logistic

regression, *liger*, *ggplot2,* and *pheatmap* for data visualization, and *cytofkit* for the FlowSOM and Phenograph clustering algorithms. The *h2o* package and *mclust* packages were used to implement the autoencoder clustering method.

### FLOW CYTOMETRY DATA ANALYSIS

Flow cytometry data were analyzed using FlowJo 10.0.7, with serial gates drawn to identify singlet lymphocytes by FSC/SSC characteristics. Viable memory CD4+ T cells were identified as propidium iodide-negative CD3+ CD4+ CD45RA- cells. We then calculated the frequency of HLA-DR- CD27+ cells among memory CD4+ T cells.

## LIST OF SUPPLEMENTARY MATERIALS

Figure S1. MASC Type 1 Error.

Figure S2. SNE Projection Density.

Figure S3. Kullback-Leibler Divergence Analysis of Clustering Approaches.

Figure S4. DensVM Clustering Elbow Plots.

Figure S5. Association Permutation Testing and Cluster Alignment.

Figure S6. Phenograph and FlowSOM Clustering.

Figure S7. Association Testing with Citrus.

Figure S8. Flow Cytometry and RNA-seq Gating Strategies.

Figure S9. CD27 and HLA-DR Expression in Flow Cytometry Cohort.

Figure S10. CD4+ T Cell Populations in a Clinical Response Cohort.

Figure S11. RNA-seq Analysis of CD4+ T Cell Subsets.

Figure S12. Flow Cytometry Expression Quantification.

Figure S13. Auto-encoder Clustering Method.

Figure S14. CD27- HLA-DR+ Frequency and Clinical Characteristics.

Figure S15. Marker Expression Distribution Plots for DensVM Clusters.

Table S1. Panel design for mass cytometry experiments.

Table S2. MASC analysis of the 19 clusters identified in the resting dataset.

Table S3. MASC analysis of the 21 clusters identified in the stimulated dataset.

Table S4. Gene set enrichment analysis of genes differentially expressed in CD27- HLA-DR+ cells.

## REFERENCES

1. K. E. Neu, Q. Tang, P. C. Wilson, A. A. Khan, Single-Cell Genomics: Approaches and Utility in Immunology, *Trends Immunol.* **38**, 140–149 (2017).

2. M. Cross, E. Smith, D. Hoy, L. Carmona, F. Wolfe, T. Vos, B. Williams, S. Gabriel, M. Lassere, N. Johns, R. Buchbinder, A. Woolf, L. March, The global burden of rheumatoid arthritis: estimates from the global burden of disease 2010 study, *Annals of the Rheumatic Diseases* **73**, 1316–1322 (2014).

3. S. Viatte, D. Plant, S. Raychaudhuri, Genetics and epigenetics of rheumatoid arthritis, *Nat Rev Rheumatol* **9**, 141–153 (2013).

4. J. Kurkó, T. Besenyei, J. Laki, T. T. Glant, K. Mikecz, Z. Szekanecz, Genetics of Rheumatoid Arthritis — A Comprehensive Review, *Clinic Rev Allerg Immunol* **45**, 170–179 (2013).

5. A. M. Gizinski, D. A. Fox, T cell subsets and their role in the pathogenesis of rheumatic disease, *Current Opinion in Rheumatology* **26**, 204–210 (2014).

6. W. Wang, S. Shao, Z. Jiao, M. Guo, H. Xu, S. Wang, The Th17/Treg imbalance and cytokine environment in peripheral blood of patients with rheumatoid arthritis, *Rheumatol. Int.* **32**, 887–893 (2012).

7. S. Raychaudhuri, C. Sandor, E. A. Stahl, J. Freudenberg, H.-S. Lee, X. Jia, L. Alfredsson, L. Padyukov, L. Klareskog, J. Worthington, K. A. Siminovitch, S.-C. Bae, R. M. Plenge, P. K. Gregersen, P. I. W. de Bakker, Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis, *Nature Publishing Group* **44**, 291–296 (2012).

8. C. Perricone, F. Ceccarelli, G. Valesini, An overview on the genetic of rheumatoid arthritis: A never-ending story, *Autoimmunity Reviews* **10**, 599–608 (2011).

9. D. Diogo, Y. Okada, R. M. Plenge, Genome-wide association studies to advance our understanding of critical cell types and pathways in rheumatoid arthritis: recent findings and challenges, *Current Opinion in Rheumatology* **26**, 85–92 (2014).

10. S. Eyre, J. Bowes, D. Diogo, A. Lee, A. Barton, P. Martin, A. Zhernakova, E. Stahl, S. Viatte, K. McAllister, C. I. Amos, L. Padyukov, R. E. M. Toes, T. W. J. Huizinga, C. Wijmenga, G. Trynka, L. Franke, H.-J. Westra, L. Alfredsson, X. Hu, C. Sandor, P. I. W. de Bakker, S. Davila, C. C. Khor, K. K. Heng, R. Andrews, S. Edkins, S. E. Hunt, C. Langford, D. Symmons, P. Concannon, S. Onengut-Gumuscu, S. S. Rich, P. Deloukas, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, L. Ärlsetig, J. Martin, S. Rantapää-Dahlqvist, R. M. Plenge, S. Raychaudhuri, L. Klareskog, P. K. Gregersen, J. Worthington, High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis, *Nat Genet* **44**, 1336–1340 (2012).

11. Y. Okada, D. Wu, G. Trynka, T. Raj, C. Terao, K. Ikari, Y. Kochi, K. Ohmura, A. Suzuki, S. Yoshida, R. R. Graham, A. Manoharan, W. Ortmann, T. Bhangale, J. C. Denny, R. J. Carroll, A. E. Eyler, J. D. Greenberg, J. M. Kremer, D. A. Pappas, L. Jiang, J. Yin, L. Ye, D.-F. Su, J. Yang, G. Xie, E. Keystone, H.-J. Westra, T. Esko, A. Metspalu, X. Zhou, N. Gupta, D. Mirel, E. A. Stahl, D. Diogo, J. Cui, K. Liao, M. H. Guo, K. Myouzen, T. Kawaguchi, M. J. H. Coenen, P. L.

C. M. van Riel, M. A. F. J. van de Laar, H.-J. Guchelaar, T. W. J. Huizinga, P. Dieudé, X. Mariette, S. L. Bridges, A. Zhernakova, R. E. M. Toes, P. P. Tak, C. Miceli-Richard, S.-Y. Bang, H.-S. Lee, J. Martin, M. A. Gonzalez-Gay, L. Rodriguez-Rodriguez, S. Rantapää-Dahlqvist, L. Ärlestig, H. K. Choi, Y. Kamatani, P. Galan, M. Lathrop, RACI consortium, GARNET consortium, S. Eyre, J. Bowes, A. Barton, N. de Vries, L. W. Moreland, L. A. Criswell, E. W. Karlson, A. Taniguchi, R. Yamada, M. Kubo, J. S. Liu, S.-C. Bae, J. Worthington, L. Padyukov, L. Klareskog, P. K. Gregersen, S. Raychaudhuri, B. E. Stranger, P. L. De Jager, L. Franke, P. M. Visscher, M. A. Brown, H. Yamanaka, T. Mimori, A. Takahashi, H. Xu, T. W. Behrens, K. A. Siminovitch, S. Momohara, F. Matsuda, K. Yamamoto, R. M. Plenge, Genetics of rheumatoid arthritis contributes to biology and drug discovery, *Nature* **506**, 376–381 (2014).

12. X. Hu, H. Kim, E. Stahl, R. Plenge, M. Daly, S. Raychaudhuri, Integrating Autoimmune Risk Loci with Gene-Expression Data Identifies Specific Pathogenic Immune Cell Subsets, *The American Journal of Human Genetics* **89**, 496–506 (2011).

13. D. A. Rao, M. F. Gurish, J. L. Marshall, K. Slowikowski, C. Y. Fonseka, Y. Liu, L. T. Donlin, L. A. Henderson, K. Wei, F. Mizoguchi, N. C. Teslovich, M. E. Weinblatt, E. M. Massarotti, J. S. Coblyn, S. M. Helfgott, Y. C. Lee, D. J. Todd, V. P. Bykerk, S. M. Goodman, A. B. Pernis, L. B. Ivashkiv, E. W. Karlson, P. A. Nigrovic, A. Filer, C. D. Buckley, J. A. Lederer, S. Raychaudhuri, M. B. Brenner, Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis, *Nature* **542**, 110–114 (2017).

14. P. B. Martens, J. J. Goronzy, D. Schaid, C. M. Weyand, Expansion of unusual CD4+ T cells in severe rheumatoid arthritis, *Arthritis Rheum* **40**, 1106–1114 (1997).

15. M. Scarsi, T. Ziglioli, P. Airo, Decreased circulating CD28-negative T cells in patients with rheumatoid arthritis treated with abatacept are correlated with clinical response, *J. Rheumatol.* **37**, 911–916 (2010).

16. K. J. Warrington, S. Takemura, J. J. Goronzy, C. M. Weyand, CD4+,CD28- T cells in rheumatoid arthritis patients combine features of the innate and adaptive immune systems, *Arthritis Rheum* **44**, 13–20 (2001).

17. C. Y. Fonseka, D. A. Rao, S. Raychaudhuri, Leveraging blood and tissue CD4+ T cell heterogeneity at the single cell level to identify mechanisms of disease in rheumatoid arthritis, *Current Opinion in Immunology* **49**, 27–36 (2017).

18. S. C. Bendall, E. F. Simonds, P. Qiu, E.-A. D. Amir, P. O. Krutzik, R. Finck, R. V. Bruggner, R. Melamed, A. Trejo, O. I. Ornatsky, R. S. Balderas, S. K. Plevritis, K. Sachs, D. Pe'er, S. D. Tanner, G. P. Nolan, Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum, *Science* **332**, 687–696 (2011).

19. T. Sörensen, S. Baumgart, P. Durek, A. Grützkau, T. Häupl, immunoClust-An automated analysis pipeline for the identification of immunophenotypic signatures in high-dimensional cytometric datasets, *Cytometry*, n/a–n/a (2015).

20. N. Samusik, Z. Good, M. H. Spitzer, K. L. Davis, G. P. Nolan, Automated mapping of phenotype space with single-cell data, *Nat Meth* **13**, 493–496 (2016).

21. P. Lin, M. Troup, J. W. K. Ho, CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data, *Genome Biol.* **18**, 59 (2017).

22. E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, S. A. McCarroll, Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets, *Cell* **161**, 1202–1214 (2015).

23. J. H. Levine, E. F. Simonds, S. C. Bendall, K. L. Davis, E.-A. D. Amir, M. D. Tadmor, O. Litvin, H. G. Fienberg, A. Jager, E. R. Zunder, R. Finck, A. L. Gedman, I. Radtke, J. R. Downing, D. Pe'er, G. P. Nolan, Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis, *Cell* **162**, 184–197 (2015).

24. O. Ornatsky, D. Bandura, V. Baranov, M. Nitz, M. A. Winnik, S. Tanner, Highly multiparametric analysis by mass cytometry, *Journal of Immunological Methods* **361**, 1–20 (2010).

25. J. Ermann, D. A. Rao, N. C. Teslovich, M. B. Brenner, S. Raychaudhuri, Immune cell profiling to guide therapeutic decisions in rheumatic diseases, *Nat Rev Rheumatol* **11**, 541–551 (2015).

26. B. Becher, A. Schlitzer, J. Chen, F. Mair, H. R. Sumatoh, K. W. W. Teng, D. Low, C. Ruedl, P. Riccardi-Castagnoli, M. Poidinger, M. Greter, F. Ginhoux, E. W. Newell, High-dimensional analysis of the murine myeloid cell system, *Nat. Immunol.* **15**, 1181–1189 (2014).

27. S. Van Gassen, B. Callebaut, M. J. Van Helden, B. N. Lambrecht, P. Demeester, T. Dhaene, Y. Saeys, FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data, *Cytometry A* **87**, 636–645 (2015).

28. L. M. Weber, M. D. Robinson, Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data, *Cytometry A* **89**, 1084–1096 (2016).

29. M. T. Wong, D. E. H. Ong, F. S. H. Lim, K. W. W. Teng, N. McGovern, S. Narayanan, W. Q. Ho, D. Cerny, H. K. K. Tan, R. Anicete, B. K. Tan, T. K. H. Lim, C. Y. Chan, P. C. Cheow, S. Y. Lee, A. Takano, E.-H. Tan, J. K. C. Tam, E. Y. Tan, J. K. Y. Chan, K. Fink, A. Bertoletti, F. Ginhoux, M. A. Curotto de Lafaille, E. W. Newell, A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures, *Immunity* **45**, 442–456 (2016).

30. Y. Cheng, M. T. Wong, L. Van Der Maaten, E. W. Newell, Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding, *J. Immunol.* **196**, 924–932 (2016).

31. M. T. Wong, J. Chen, S. Narayanan, W. Lin, R. Anicete, H. T. K. Kiaang, M. A. C. De Lafaille, M. Poidinger, E. W. Newell, Mapping the Diversity of Follicular Helper T Cells in Human Blood and Tonsils Using High-Dimensional Mass Cytometry Analysis, *Cell Rep* **11**, 1822–1833 (2015).

32. C. Baecher-Allan, E. Wolf, D. A. Hafler, MHC class II expression identifies functionally distinct human regulatory T cells, *The Journal of Immunology* **176**, 4622–4631 (2006).

33. S. Oshima, D. D. Eckels, Selective expression of class II MHC isotypes by MLC-activated human T lymphocytes, *Hum. Immunol.* **27**, 208–219 (1990).

34. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification, *Nat Biotechnol* **34**, 525–527 (2016).

35. G. M. Mason, K. Lowe, R. Melchiotti, R. Ellis, E. de Rinaldis, M. Peakman, S. Heck, G. Lombardi, T. I. M. Tree, Phenotypic Complexity of the Human Regulatory T Cell Compartment Revealed by Mass Cytometry, *J. Immunol.* **195**, 2030–2037 (2015).

36. S. C. Bendall, K. L. Davis, E.-A. D. Amir, M. D. Tadmor, E. F. Simonds, T. J. Chen, D. K. Shenfeld, G. P. Nolan, D. Pe'er, Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development, *Cell* **157**, 714–725 (2014).

37. P. B. Ferrell, K. E. Diggins, H. G. Polikowsky, S. R. Mohan, A. C. Seegmiller, J. M. Irish, High-Dimensional Analysis of Acute Myeloid Leukemia Reveals Phenotypic Changes in Persistent Cells during Induction Therapy, *PLoS ONE* **11** (2016), doi:10.1371/journal.pone.0153207.

38. M. Setty, M. D. Tadmor, S. Reich-Zeliger, O. Angel, Wishbone identifies bifurcating developmental trajectories from single-cell data, *Nature* (2016), doi:10.1038/nbt.3569.

39. M. Mingueneau, S. Boudaoud, S. Haskett, T. L. Reynolds, G. Nocturne, E. Norton, X. Zhang, M. Constant, D. Park, W. Wang, T. Lazure, C. Le Pajolec, A. Ergun, X. Mariette, Cytometry by time-of-flight immunophenotyping identifies a blood Sjögren's signature correlating with disease activity and glandular inflammation, *J. Allergy Clin. Immunol.* (2016), doi:10.1016/j.jaci.2016.01.024.

40. L. Hansmann, L. Blum, C.-H. Ju, M. Liedtke, W. H. Robinson, M. M. Davis, Mass cytometry analysis shows that a novel memory phenotype B cell is expanded in multiple myeloma, *Cancer Immunol Res* **3**, 650–660 (2015).

41. D. M. Strauss-Albee, J. Fukuyama, E. C. Liang, Y. Yao, J. A. Jarrell, A. L. Drake, J. Kinuthia, R. R. Montgomery, G. John-Stewart, S. Holmes, C. A. Blish, Human NK cell repertoire diversity reflects immune experience and correlates with viral susceptibility, *Sci Transl Med* **7**, 297ra115–297ra115 (2015).

42. B. Gaudillière, G. K. Fragiadakis, R. V. Bruggner, M. Nicolau, R. Finck, M. Tingle, J. Silva, E. A. Ganio, C. G. Yeh, W. J. Maloney, J. I. Huddleston, S. B. Goodman, M. M. Davis, S. C. Bendall, W. J. Fantl, M. S. Angst, G. P. Nolan, Clinical recovery from surgery correlates with single-cell immune signatures, *Sci Transl Med* **6**, 255ra131 (2014).

43. R. V. Bruggner, B. Bodenmiller, D. L. Dill, R. J. Tibshirani, G. P. Nolan, Automated identification of stratifying signatures in cellular subpopulations, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E2770–7 (2014).

44. Y. Saeys, S. V. Gassen, B. N. Lambrecht, Computational flow cytometry: helping to make sense of high-dimensional immunology data, *Nature Reviews Immunology* **16**, 449–462 (2016).

45. A. Wagner, A. Regev, N. Yosef, Revealing the vectors of cellular identity with single-cell genomics, *Nat Biotechnol* **34**, 1145–1160 (2016).

46. C. Chester, H. T. Maecker, Algorithmic Tools for Mining High-Dimensional Cytometry Data, *J. Immunol.* **195**, 773–779 (2015).

47. D. Y. Orlova, L. A. Herzenberg, G. Walther, Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets, *Nature Reviews Immunology* **18**, 77–77 (2017).

48. Y. Saeys, S. Van Gassen, B. Lambrecht, Response to Orlova et al. "Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets", *Nature Reviews Immunology* **18**, 78–78 (2017).

49. X. Qiu, A. Hill, J. Packer, D. Lin, Y.-A. Ma, C. Trapnell, Single-cell mRNA quantification and differential analysis with Census, *Nat Meth* **14**, 309–315 (2017).

50. A. Larbi, T. Fulop, From "truly naïve" to “exhausted senescent” T cells: when markers predict functionality, *Cytometry A* **85**, 25–35 (2014).

51. R. De Jong, M. Brouwer, B. Hooibrink, T. Van der Pouw-Kraan, F. Miedema, R. A. Van Lier, The CD27- subset of peripheral blood memory CD4+ lymphocytes contains functionally differentiated T lymphocytes that develop by persistent antigenic stimulation in vivo, *Eur. J. Immunol.* **22**, 993–999 (1992).

52. A. Schiott, M. Lindstedt, B. Johansson-Lindbom, E. Roggen, C. A. K. Borrebaeck, CD27-CD4+ memory T cells define a differentiated memory population at both the functional and transcriptional levels, *Immunology* **113**, 363–370 (2004).

53. V. Appay, J. J. Zaunders, L. Papagno, J. Sutton, A. Jaramillo, A. Waters, P. Easterbrook, P. Grey, D. Smith, A. J. McMichael, D. A. Cooper, S. L. Rowland-Jones, A. D. Kelleher, Characterization of CD4(+) CTLs ex vivo, *The Journal of Immunology* **168**, 5954–5958 (2002).

54. G. M. Griffiths, S. Alpert, E. Lambert, J. McGuire, I. L. Weissman, Perforin and granzyme A expression identifying cytolytic lymphocytes in rheumatoid arthritis, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 549–553 (1992).

55. T. Namekawa, U. G. Wagner, J. J. Goronzy, C. M. Weyand, Functional subsets of CD4 T cells in rheumatoid synovitis, *Arthritis Rheum* **41**, 2108–2116 (1998).

56. H. Mattoo, V. S. Mahajan, T. Maehara, V. Deshpande, E. Della-Torre, Z. S. Wallace, M. Kulikova, J. M. Drijvers, J. Daccache, M. N. Carruthers, F. V. Castelino, J. R. Stone, J. H. Stone, S. Pillai, Clonal expansion of CD4(+) cytotoxic T lymphocytes in patients with IgG4-related disease, *J. Allergy Clin. Immunol.* **138**, 825–838 (2016).

57. C. L. Kohem, R. I. Brezinschek, H. Wisbey, C. Tortorella, P. E. Lipsky, N. Oppenheimer-Marks, Enrichment of differentiated CD45RBdim,CD27- memory T cells in the peripheral blood, synovial fluid, and synovial tissue of patients with rheumatoid arthritis, *Arthritis Rheum* **39**, 844–854 (1996).

58. Y. C. Lee, C. O. Bingham, R. R. Edwards, W. Marder, P. Kristine, M. B. Bolster, D. J. Clauw, L. W. Moreland, B. Lu, A. Wohlfahrt, Z. Zhang, T. Neogi, Pain Sensitization is Associated with Disease Activity in Rheumatoid Arthritis Patients: A Cross-Sectional Study, *Arthritis Care Res (Hoboken)* (2017), doi:10.1002/acr.23266.

59. R. Finck, E. F. Simonds, A. Jager, S. Krishnaswamy, K. Sachs, W. Fantl, D. Pe'er, G. P. Nolan, S. C. Bendall, Normalization of mass cytometry data with bead standards, *Cytometry*

*Part A* **83A**, 483–494 (2013).

60. L. Van Der Maaten, Accelerating t-SNE using tree-based algorithms, *The Journal of Machine Learning Research* **15**, 3221–3245 (2014).

**ACKNOWLEDGEMENTS:**

This work was supported in part by funding from the National Institutes of Health (UH2AR067677, U19AI111224, and 1R01AR063759 (to S.R.)), the Doris Duke Charitable Foundation Grant #2013097, and funding from the Smith Fund at Brigham and Women's Hospital, T32 AR007530-31 and the William Docken Inflammatory Autoimmune Disease Fund (to M.B.B), Rheumatology Research Foundation Tobe and Stephen Malawista, MD Endowment in Academic Rheumatology (to D.A.R), R01 AR064850-03 (to Y.C.L).

**AUTHOR CONTRIBUTIONS**

C.Y.F., S.R. conceived the MASC association testing method. C.Y.F., S.R., and K.F.S. conducted all statistical analyses. D.A.R., N.C.T., S.K.H., M.F.G., J.E., M.B.B., and S.R. designed and conducted all mass cytometry, flow cytometry, and functional immunology assays. D.A.R., L.T.D., M.E.W., E.M.M., J.S.C., S.M.H., D.J.T., V.P.B., E.W.K., and Y.C.L. recruited patients and obtained samples for this study. C.Y.F, D.A.R. and S.R. wrote the initial manuscript. All authors edited and revised the final manuscript.

**DATA AND SOFTWARE AVAILABILITY**

The data that support the findings of this study are available from the corresponding author upon reasonable request. MASC and all other custom scripts used in this analysis are available from the corresponding author upon request.

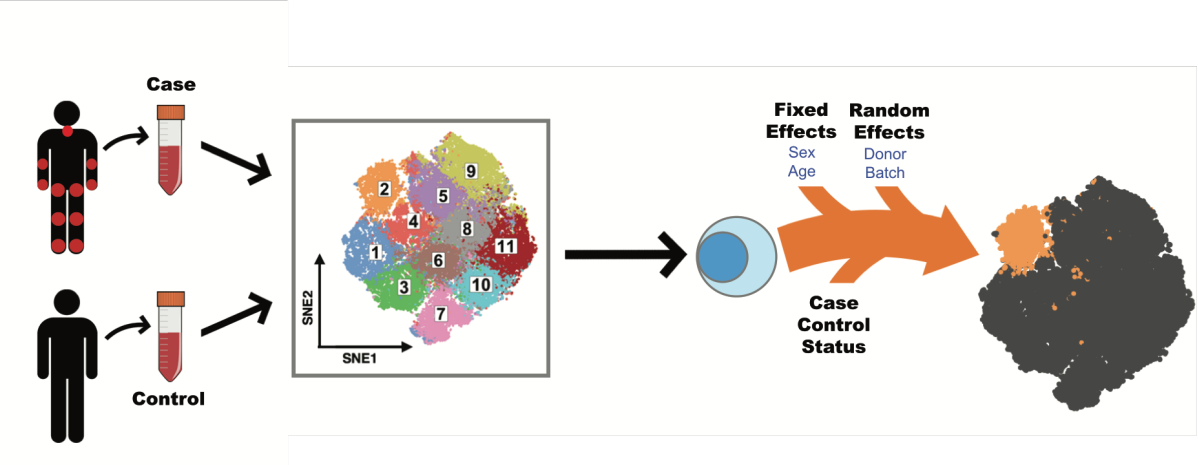


**Figure 1**: **Mixed effect modeling of Associations of Single Cells (MASC) Overview**

Single cell transcriptomics or proteomics are used to assay samples from cases and controls, such as immunoprofiling of peripheral blood. The data is then clustered to define populations of similar cells. Mixed-effect logistic regression is used to predict individual cell membership in previously defined populations. The addition of a case-control term to the regression model allows the user to identify populations for which case-control status is significantly associated.

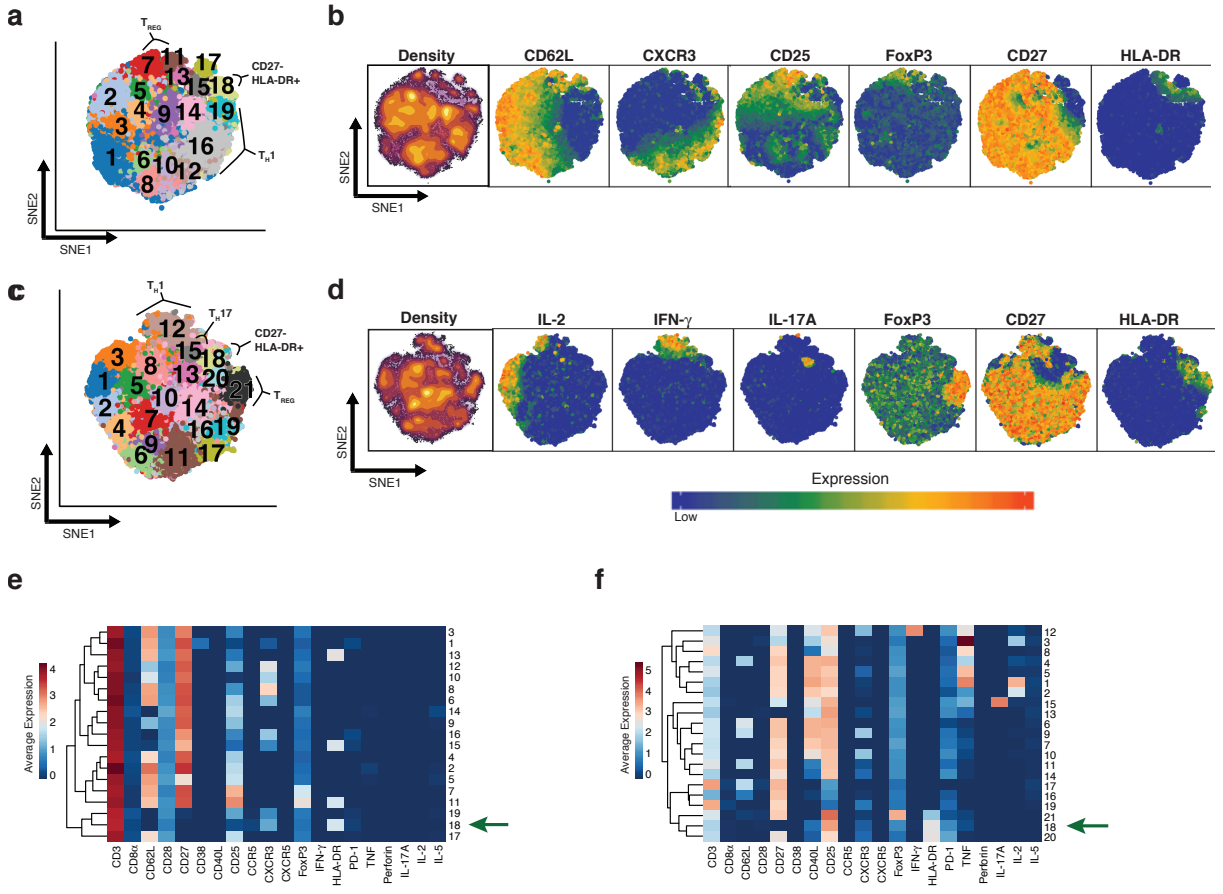


**Figure 2: Diversity of CD4$^+$ memory T cells before and after stimulation**

(a) SNE projection of 50,000 resting CD4+ memory T cells sampled equally from RA atients (n=24) and controls (n=26). DensVM identified 19 populations in this dataset. (b) Same SNE projections as in (a) colored by the density of cells on the SNE plot or the expression of the markers labeled above each panel. (c) SNE projection of 52,000 CD4$^+$ stimulated memory T cells sampled equally from RA patients (n=26) and controls (n=26). Cells were stimulated for 24 hours with anti-CD3/anti-CD28 beads. (d) Same SNE projections as in (c) colored by the density of cells on the SNE plot or the expression of the markers labeled above each panel. (e) Heatmap showing mean expression of indicated markers across the 19 populations found in resting cells. (f) Heatmap showing mean expression of indicated markers across the 21 populations found after stimulation. Protein expression data are shown after arcsinh transformation. All markers but CD4 and CD45RO were used to create SNE projections and perform clustering.

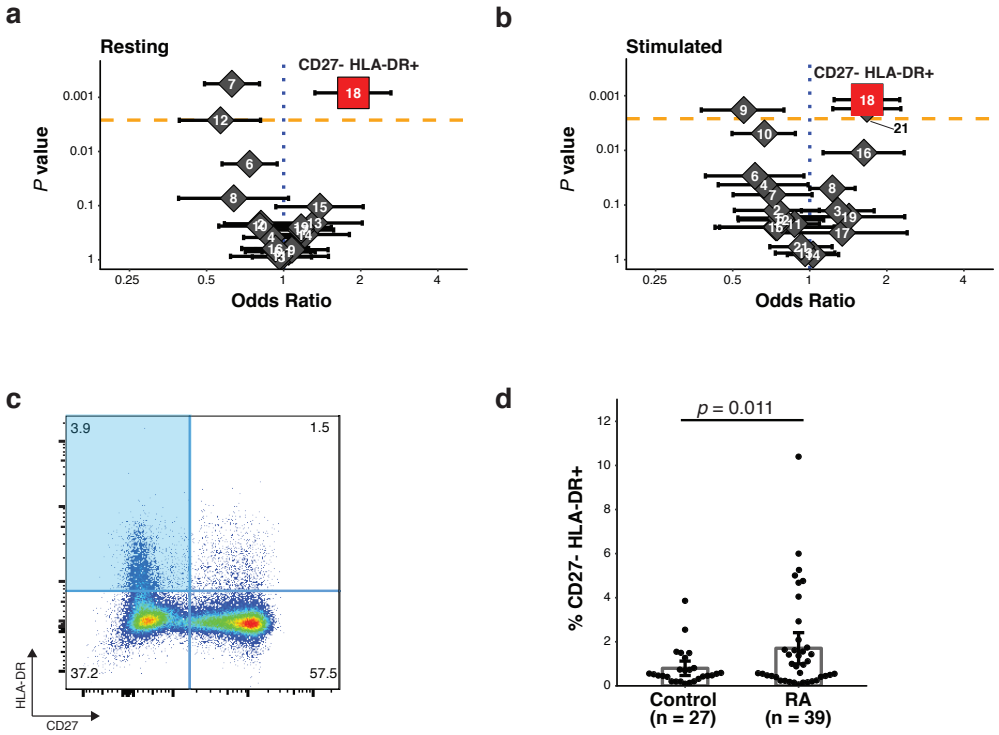


**Figure 3: MASC identifies a population that is expanded in RA**

(a,b) Odds ratios and association p-values were calculated by MASC for each population identified the resting (a) and stimulated (b) datasets. The yellow line indicates the significance threshold after applying the Bonferroni correction for multiple testing. (c) Flow cytometry dot plot of gated memory CD4+ T cells from a single RA donor shows the gates used to identify $CD27^-$ $HLA\text{-}DR^+$ memory CD4+ T cells (blue quadrant). (d) Flow cytometric quantification of the percentage of $CD27^-$, $HLA\text{-}DR^+$ cells among blood memory CD4+ T cells in an independent cohort of seropositive RA patients (n = 39) and controls (n = 27). Statistical significance was assessed using a one-tailed *t*-test.

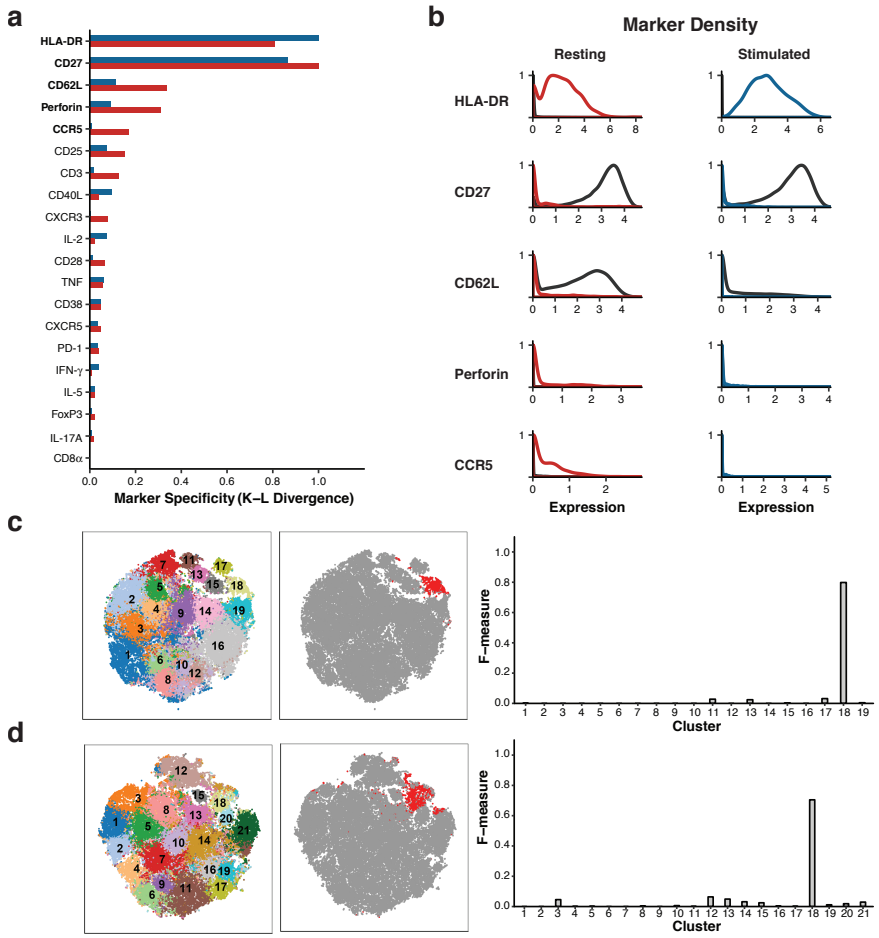


**Figure 4: CD27 and HLA-DR expression specifically mark the expanded population**

(a) Plot of the Kullback-Liebler divergence for each marker comparing cluster 18 to all other cells in both the resting dataset (red) and the stimulated dataset (blue). (b) Density plots showing expression of the five markers most different between cluster 18 cells (resting = red, stimulated = blue) and all other cells in the same dataset (black line). (c) Left: SNE projection of clusters identified in resting dataset; Middle: Same SNE projection, with cells gated as CD27- HLA-DR+ colored in red; Right: F-measure scores were calculated for the overlap between gated cells and each cluster in the resting dataset. (d) Left: SNE projection of clusters identified in stimulated dataset; Middle: Same SNE projection, with cells gated as CD27- HLA-DR+ colored in red; Right: F-measure scores were calculated for the overlap between gated cells and each cluster in the stimulated dataset.

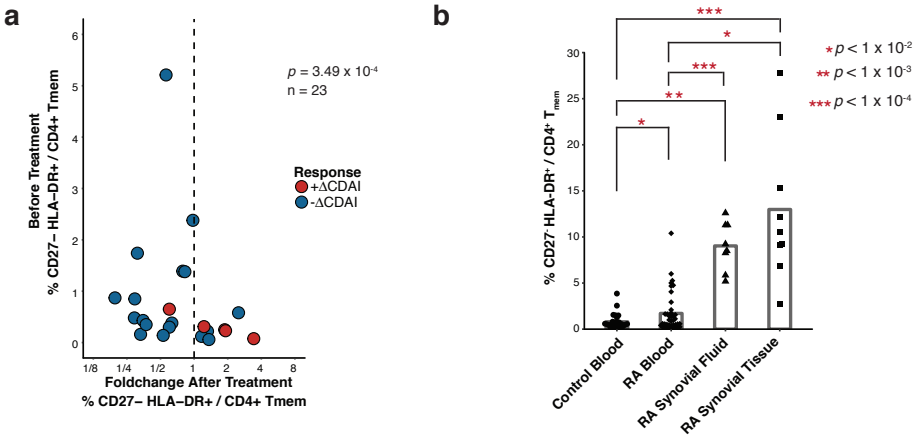


**Figure 5: CD27$^-$ HLA-DR$^+$ memory CD4+ T cells are expanded in the blood and joints of patients with active RA.**

(a) Flow cytometric quantification of the frequency of CD27- HLA-DR+ memory CD4+ T cells in 18 RA patients prior to starting a new medication, plotted against change in cell frequency after 3 months of new therapy. Treatment significantly reduced CD27- HLA-DR+ cell frequency as determined by a Wilcoxon signed-rank test. b) Flow cytometric quantification of the percentage of memory CD4+ T cells with a CD27- HLA-DR+ phenotype in cells from seropositive RA synovial fluid (n=8) and synovial tissue (n=9), compared to blood samples from RA patients and controls. Blood sample data are the same as shown in panel 3d. Significance was assessed using one-tailed *t*-test and applying a Bonferroni correction for multiple testing.

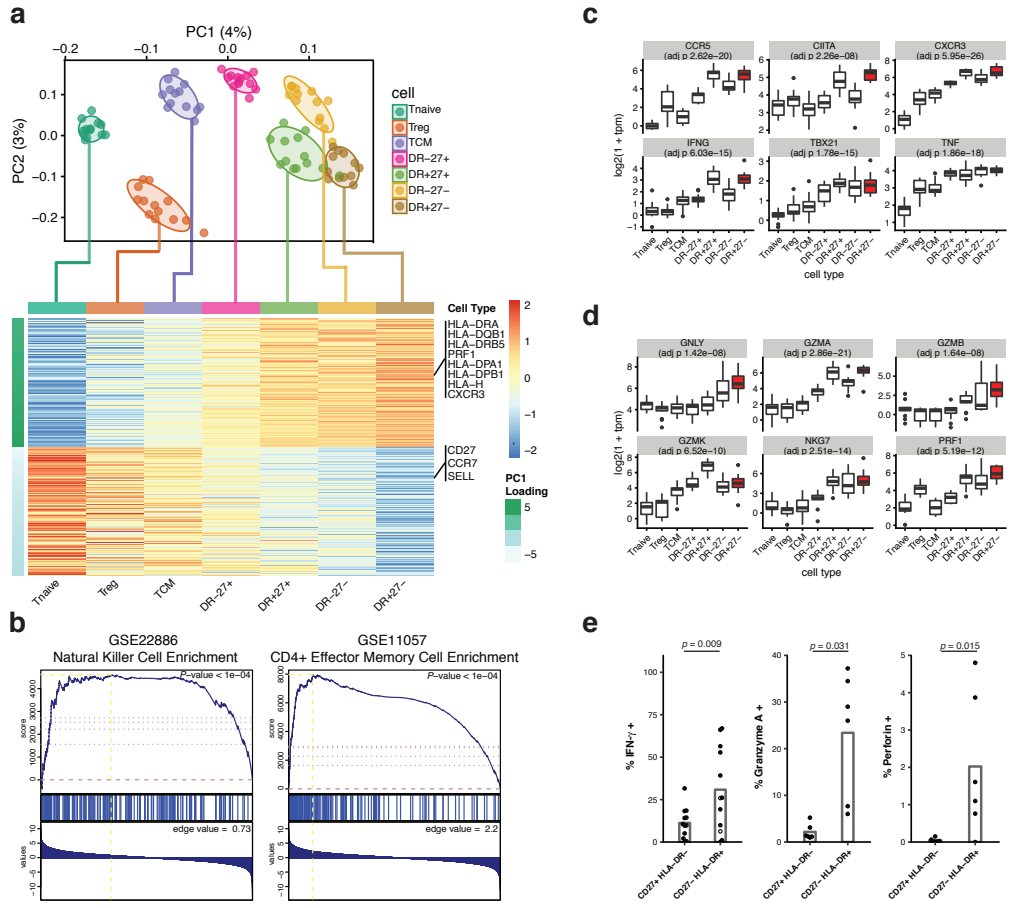


**Figure 6: Transcriptomic characterization of CD27- HLA-DR+ memory CD4+ T cells identifies a Th1-skewed cytotoxic phenotype.**

RNA-seq characterization of CD27- HLA-DR+ (DR+27-) cells and 6 related CD4+ T cell populations: naive T cells (Tnaive), regulatory T cells (Treg), central memory t cells (TCM), and three populations of effector memory T cells, CD27+ HLA-DR- (DR-27+), CD27+ HLA-DR+ (DR+27+), and CD27- HLA-DR- (DR-27-). (a) PCA plot (top) and PC1 gene loadings (bottom) of 90 samples from the 7 CD4+ T cell populations. Cells were colored on the PCA plot according to known cell type. Normal confidence ellipses at 1 standard deviation were plotted for each cell type. The 300 most positive and 300 most negative PC1 gene loadings for each cell type were averaged and plotted in the heatmap the PCA plot. Rows were mean centered and scaled to variance 1 to highlight trends among genes. Relevant genes to the CD27- HLA-DR+ population were labeled. The ordering of cell types along the x-axis of the PCA plot suggests that PC1 is associated with a naive to effector gradient. (b) Gene set enrichment analysis was performed on all genes, ranked on their PC1 loadings. Two significantly enriched genesets: NK signature (GSE22886 NAIVE CD4 T CELL VS NK CELL DN) and effector memory t cell signature (GSE11057 NAIVE VS EFF MEMORY CD4 T CELL) are shown, indicating that genes involved in each signature had statistically significantly high PC1 loadings and were therefore associated

with the effector phenotype gradient. (c) Distribution of log-scaled gene expression of six canonical Th1 genes: CCR5, CIITA, CXCR3, IFNG, TBX21 (Tbet), and TNF. Populations are ordered by PC1 loading values, with CD27- HLA-DR+ population highlighted in red. (d) Distribution of log-scaled gene expression of six canonical cytotoxic genes: GNLY, GZMA, GZMB, GMZK, NKG7, and PRF1. Populations are ordered by PC1 loading values, with CD27- HLA-DR+ population highlighted in red. Reported p-values in (c) and (d) correspond to a linear model of gene expression against ordered cell type (as an ordinal variable), with p-values adjusted for multiple testing by the Benjamini Hochberg procedure. (e) Cytokine expression determined by intracellular cytokine staining of peripheral effector memory CD4+ T cells after *in vitro* stimulation with PMA/ionomycin. The percentage of cells positive for each stain is plotted for CD27+ HLA-DR- and CD27- HLA-DR+ subsets. Each dot represents a separate donor (n = 12; 6 RA patients and 6 controls, except for the quantification of Granyzme A and perforin where n = 6; 3 RA patients and 3 controls).

| | | Mass cytometry cohort | | Flow cytometry cohort | |
|---|---|---|---|---|---|
| | | **Controls** | **Cases** | **Controls** | **Cases** |
| Blood Cross-sectional cohorts | **Number** | 26 | 26 | 35 | 42 |
| | **Age** | 57 ± 15 | 66 ± 9 | 61 ± 13 | 58 ± 14 |
| | **Female** | 19 (73) | 20 (77) | 22 (63) | 33 (78) |
| | **ACPA- or RF-positive** | N/A | 22 (85) | N/A | 42 (100) |
| | **CRP (mg/L)** | N/A | 8.6 ± 16.9 | N/A | 9.3 ± 17.4 |
| | **CDAI** | N/A | 9.3 ± 4.4 | N/A | 13.7 ± 8.1 |
| | **Methotrexate** | 0 | 18 (69) | 0 | 19 (45) |
| | **Anti-TNF** | 0 | 10 (38) | 0 | 16 (38) |
| | **Other biologics** | 0 | 5 (19) | 0 | 10 (24) |

| | | Longitudinal cohort |
|---|---|---|
| Blood Longitudinal cohort | **Number** | 18 |
| | **Age** | 49 ± 17 |
| | **Female** | 17 (94) |
| | **ACPA- or RF-positive** | 18 (100) |
| | **CDAI Before** | 17.6 ± 9.3 |
| | **CDAI After** | 6.3 ± 4.2 |
| | **Started methotrexate** | 7 |
| | **Started anti-TNF** | 4 |
| | **Started other biologic** | 7 |

| | **Patient** | **#1** | **#2** | **#3** | **#4** | **#5** | **#6** | **#7** | **#8** | **#9** |
|---|---|---|---|---|---|---|---|---|---|---|
| Synovial Tissue Donors | **Age** | 57 | 54 | 76 | 46 | 46 | 79 | 62 | 63 | 52 |
| | **Sex** | F | F | F | F | F | F | M | M | F |
| | **CRP (mg/L)** | 25 | 8 | 8 | 11 | 17 | 19 | 13 | 66 | 76 |
| | **CDAI** | 14 | 9 | 17 | 15 | 21 | 25 | 5 | 9 | N/A |
| | **Methotrexate** | No | Yes | No | No | No | No | No | Yes | No |
| | **Biologic therapy** | Yes | Yes | Yes | Yes | Yes | No | No | No | No |

**Table 1**: Clinical characteristics of patient samples used. Average ± SD is shown. Parentheses indicate percentages. CRP: C-reactive protein. CDAI: clinical disease activity index. Other biologics include rituximab, tofacitinib, and abatacept.

| Condition | Description | Subset | RA Proportion | *P* value | Odds Ratio | Test |
|---|---|---|---|---|---|---|
| Resting | HLA-DR+,CD27- | 18 | 0.636 | $\mathbf{5.9 \times 10^{-4}}$ | 1.9 (1.3 – 2.7) | MASC |
| Stimulated | HLA-DR+, CD27- | 18 | 0.619 | $\mathbf{1.3 \times 10^{-3}}$ | 1.7 (1.2 – 2.2) | MASC |
| Flow cytometry replication | Gated HLA-DR+, CD27- | NA | NA | $\mathbf{1.1 \times 10^{-2}}$ | NA | One-tailed *t* test |

**Table 2:** Overview of the subsets found to be significantly expanded in RA. RA proportion reflects the fraction of cells in the subset that were from RA donors. The 95% confidence interval is shown next to odds ratio.