REVERSE ASSOCIATION OF SINGLE CELLS TO RHEUMATOID ARTHRITIS ACCOUNTING FOR MIXED EFFECTS IDENTIFIES AN EXPANDED CD27- HLA-DR+ EFFECTOR MEMORY CD4+ T CELL POPULATION

Chamith Y. Fonseka[1,2,3,5†], Deepak A. Rao[2†], Nikola C. Teslovich[2], 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

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

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

Correspondence to:

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)

ABSTRACT

Defining disease associated cellular subsets from single cell assays can be confounded by technical variation and inter-individual differences in subset frequencies. We present Mixed effects modeling of Associations of Single Cells (MASC), a novel reverse single cell association strategy that tests whether case-control status influences the membership of single cells in any of multiple cellular subsets. MASC is able to account for random technical artifacts that confound clustering and person-to-person biological variation. After demonstrating appropriate type I error, we applied this method to mass cytometry data collected from 52 cases and controls to identify CD4+ memory T cells altered in the circulation of rheumatoid arthritis (RA) patients. MASC identified a population of effector memory T cells expanded in RA patients compared to non-inflammatory controls (OR = 1.7; p = 1.1×10^{-3}). This population was characterized by absence of CD27 and expression of HLA-DR and constituted 1.7% of the total resting CD4+ memory T cells in controls and 3.1% in cases. We replicated the expansion of CD27- HLA-DR+ T cells in an independent cohort of 39 RA patients (p = 0.011), and find that effective RA treatment reduces the frequency of these cells in circulation. Further, we find that CD27- HLA-DR+ cells infiltrate the joints of RA patients and produce abundant effector cytokines upon stimulation. We propose that MASC is a broadly applicable method to identify disease-associated cell populations in high-dimensional single cell data.

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 where 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 potentially pathogenic immune cell populations.

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 mediating 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 subpopulations 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.

The advent of mass cytometry now allows for relatively broad assessment of circulating immune cell populations with >30 markers[17], enabling detailed, multiparametric characterization of lymphocyte subpopulations. This technology provides the potential to define and quantify lymphocyte subpopulations at high resolution using multiple markers. While rapid progress has been made in defining individual cellular populations in an unbiased fashion[18-22] , a key challenge that remains is defining methods to identify association of single cell populations with a disease. The challenge is that the random effects of inter-individual variation and technical variation influence real or perceived cellular population subset frequencies, and need to be accounted 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[23,24]. Using MASC, we identify a population of memory CD4+ T cells, characterized as CD27- HLA-DR+, that 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-γ, 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 numbers of random cells from each sample so that they contributed equally to the subsequent analyses. Here we used DensVM[25] 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 S2A**). 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 S2B**). 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 approach produced highly inflated results in comparison to MASC with 65.7% of trials obtaining $p<0.05$ (**Figure S2C**).

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 (**Supplementary Table 1**). 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 cases and 26 controls. We randomly sampled 1000 cells from each of the 52 samples so that each sample contributed equally to the analysis. We then used Barnes-Hut SNE algorithm to project the data altogether

so that all cells from all samples were projected into the same two dimensions, and then applied DensVM to assign cells to subsets by their similarity of marker expression[25].

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[26-28]. We identified 19 distinct subsets in resting (*R*) memory CD4+ T cells (**Figure 2A**). 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 were able to identify 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[29].

When applied to the stimulated CD4+ T cell data, DensVM identified 21 subsets (**Figure 2C**). As expected, certain activation markers, like CD25 and CD40L, were broadly expressed across all subsets after stimulation, while other markers like CD38 and HLA-DR were restricted to specific subsets (**Figure 2D**). Mass cytometry robustly detected cytokine production from stimulated CD4+ memory T cells (**Figure 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 (**Figure S1A,B**).

When we formally tested whether they 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 = 5.9×10^{-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 = 8.8×10^{-4}, OR = 0.6, 95% CI = 0.5 – 0.8) and subset 12 (p = 2.0×10^{-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. Only in 93 out of 100,000 permutations did we observed more than 753 cells from case samples in the

randomized data (permutation p = 9.4×10^{-4}, **Supplementary Table 2**). 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 S1C, D**). When we considered the subsets identified as significant by MASC, both subsets 18 (permutation p = 9.4×10^{-4}) and 7 (permutation p = 1.8×10^{-4}) retained significance while subset 12 demonstrated nominal evidence of case-control association (permutation p = 1.3×10^{-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 S1E, F**). Applying MASC, we observed case-control association for subset 18 in the stimulated data (p = 1.2×10^{-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 = 1.7×10^{-3}, OR = 1.7, 95% CI = 1.2 – 2.3) (**Supplementary Table 3**).

We wanted to ensure that the results we observed were not an artifact of using DensVM to cluster the cytometry data. To this end, we used Phenograph[22] and FlowSOM[30] to cluster the same dataset. We observed a CD27-, HLA-DR+ cluster with either method (**Figure S3A,C**), which we found to be significantly associated in RA by MASC (**Figure S3B,D**). 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 S4**).

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,F**). In order 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 4**). The expression of HLA-DR and perforin are notably increased in subset 18 compared to all other cells, while the expression of CD27 is decreased in this subset. Although HLA-DR is a known to be expressed on T cells in response to activation, it normally takes several days to induce strong HLA-DR expression[31]. 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+ cells in RA patients, 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 in each group (**Figure 3C**). 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). 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 18 RA patients before and 3 months after initiation of a new medication for RA. In these patients, all of whom experienced a reduction in disease activity after treatment escalation, RA drug therapy significantly reduced the frequency of CD27- HLADR+ cells (**Figure 5A**, $p = 0.003$, Wilcoxon test). 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 increased 5-fold (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 (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.

To evaluate the potential function of the CD27- HLA-DR+ cell subset, we assessed cytokine production by these cells after *in vitro* stimulation with PMA + ionomycin, a stimulation method that readily reveals T cell capacity for cytokine production. We find that CD27- HLA-DR+ cells more frequently produce IFN-γ and IL-17A than do CD27+ HLA-DR- cells, which constitute the majority of memory CD4+ T cells (**Figure 5C**). A higher frequency of cells in both populations produce IL-2 and TNF. Taken together, observations suggest that the CD27- HLA-DR+ T cell population expanded in RA patients may be chronically activated effector T cells in RA.

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[26,32-35] 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[36-39], in part due to limited availability of effective association testing strategies.

Although there are many approaches to cluster single cell data[18-22,25,30], straightforward extension of these methods to perform case-control association testing is not as simple as it might seem at first glance. The simplest strategy would be to define subsets of interest by clustering and 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 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. 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. 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.

Applying MASC to mass cytometry analysis of CD4+ T cells identified a T cell population that is significantly expanded in the circulation of RA patients, detected 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 could then interrogate and reproduce 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 also demonstrated enrichment of CD27- HLA-DR+ cells in both synovial tissue and synovial fluid of RA patients.

The accumulation of these cells in chronically inflamed RA synovial tissue, combined with lack of CD62L and robust production of IFN-γ and TNF suggest that CD27- HLA-DR- T cells may be terminally differentiated effector cells. Expression of HLA-DR suggests recent or continued activation of these cells *in vivo*. Lack of CD27 suggests chronic or repeated activation, as is seen in T cells that respond to common re-stimulation antigens[40-42]. Together, these results nominate CD27- HLA-DR+ T cells as a potential pathogenic T cell population that may participate in the chronic autoimmune response in RA patients[43]. Notably, 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. The ability to identify potentially pathogenic T cell populations such as this one in the circulation, amidst all of the other uninvolved T cells, has multiple potential values. These cells may represent a key population with the most prominent effector functions driving disease in individual patients. Further quantification of this population itself or its functional subphenotypes might be a clinically predictive peripheral blood biomarker. This cell population is also likely to be enriched in antigen-specific T cells, such that assessment of their T cell receptors may help identify disease-associated autoantigens.

We note that Citrus is an alternative association strategy to automatically highlight statistical differences between experimental groups and identify predictive populations in mass cytometry data[44]. In the setting of our rheumatoid arthritis mass cytometry data, Citrus was unable to identify the expanded CD27- HLA-DR+ population that MASC. We believe our methodology compares favorably to this method since 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.

The single-cell association modeling framework we have developed, MASC, is extendible to any case-control experiment in which single cell data is available, including both flow and mass cytometry data, as well as to single cell RNA-seq data. 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 disease progression or activity.

MATERIALS AND METHODS

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 subpopulation 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 (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 S3**)

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*, θ_j is the intercept for cluster *j*, $\beta_{clinical}$ is a vector of clinical covariates for the kth sample, $(\phi_i|k)$ is the random effect for cell *i* from kth 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 kth 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*, θ_j is the intercept for cluster *j*, $\beta_{clinical}$ is a vector of clinical covariates for the kth sample, $(\phi_i|k)$ is the random effect for cell *i* from kth

sample, $(\kappa_i|m)$ is the random effect for cell *i* from batch *m*, β_{case} indicates the effect of kth sample's case-control status.

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

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

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 χ^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 χ^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 β_{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.

SAMPLE ACQUISITION

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.

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[45]. 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.

Synovial Tissue Processing

Synovial samples were acquired from discarded arthroplasty tissue. 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.

Sample Preparation for mass cytometry

We rapidly thawed cryopreserved peripheral blood mononuclear cells (PBMCs) at 37° C and isolated total CD4+ memory T cells by negative selection with magnetic bead separation technology. Subsequently, we cultured the CD4+ memory T cells for 24 hours in RPMI 1640, sterile-filtered and supplemented with 15% FBS, 1% Pen/Strep, 0.5% Essential and Non-Essential Amino Acids, 0.5mM Sodium Pyruvate, 30mM HEPES, and 55µM 2-mercaptoethanol. We activated the cells with CD3/CD28 T cell activation beads at a density of 1 bead: 2 cells, and at 6 hours prior to harvesting (t=18 hours), added Monensin and Brefeldin A at a dilution of 1:1000. At 24 hours, we incubated the cells with a rhodium metallointercalator 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, 0.02% sodium azide, and 2µM EDTA. We spun the cells at 500xg for 7 minutes at room temperature. We incubated the resulting cell pellets with 10ul Fc block and 40ul of CSB for 10 minutes at 4C, and moved to ice until the surface antibody cocktail had been prepared. We then added the CyTOF surface antibodies to the cells and incubated for 30 minutes at 4C on a shaker rack. We

washed the cells with 2 ml of CSB and spun at 700xg for 5 minutes at 4C. Post spin, we aspirated the buffer from pellet and added 1 ml of FOXP3 fixative/permeabilization solution supplemented with 1.6% paraformaldehyde. We incubated the cells at room temperature on a gentle shaker in the dark for 45 minutes. We washed the cells with CSB + 0.3% saponin, and spun at 800xg for 5 minutes. To the resulting pellet, we incubated the intracellular antibodies and Iridium in a total volume of 100ul of CSB containing saponin for 35 minutes at r.t. on a gentle shaker. Post-incubation, we washed the cells in PBS and spun at 800xg for 5 minutes. We resuspended the cells in 1 ml of 4% formaldehyde prepared in CyTOF PBS and incubated for 10 minutes at room temperature on a gentle shaker, followed by another PBS wash and spin at 800xg for 5 minutes. We washed the pellet in deionized water, spun at 700xg for 6 minutes, and subsequently resuspended the resulting pellet in deionized water at 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 4-element MaxPar beads 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 (**Supplemental Table 1**). 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[46]. 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.

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 with the following settings: s = 0.05 and N = 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[47]) 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 used for gating samples. To identify subpopulations, we used a modified version of DensVM[25]. DensVM performs kernel density estimation across the dimensionally reduced SNE map to build a training set, then assigns cells to clusters by their expression of all markers using an SVM classifier. We modified the DensVM code to increase the range of potential bandwidths searched during the density estimation step and to return the SVM model generated from the SNE projection.

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.

We aligned subpopulations 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. We then aligned clusters by calculating the Euclidean distance between cluster centroids in the query and target datasets.

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.

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.

All analyses were performed using custom scripts for R 3.4.0[47]. We used the following packages: *flowCore*[48] to read and process FCS files for further analysis, *lme4*[49] to apply mixed effect logistic regression, *ggplot2*[50] and *pheatmap*[51] for data visualization, and *cytofkit*[52] to implement FlowSOM and Phenograph.

Flow Cytometry Sample Preparation and Data Acquisition

For analysis of the validation and longitudinal blood cohorts and synovial samples analyzed by flow cytometry, 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. 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.

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-17A-eFluor 450 (eBio64DEC17), anti-IL-10-PE (JES3-9D7), anti-IL-21-APC (3A3-N2), anti-IL-2-PE/Cy7 (MQ1-17H12), and anti-CXCL13-AlexaFluor700 (53610, R&D Systems) 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.

CODE AVAILABILITY

MASC and all other custom scripts used in this analysis are available from the corresponding author upon request.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

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.

REFERENCES

1. Neu, K. E., Tang, Q., Wilson, P. C. & Khan, A. A. Single-Cell Genomics: Approaches and Utility in Immunology. *Trends Immunol.* **38,** 140–149 (2017).
2. Cross, M. *et al.* 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. Viatte, S., Plant, D. & Raychaudhuri, S. Genetics and epigenetics of rheumatoid arthritis. *Nat Rev Rheumatol* **9,** 141–153 (2013).
4. Kurkó, J. *et al.* Genetics of Rheumatoid Arthritis — A Comprehensive Review. *Clinic Rev Allerg Immunol* **45,** 170–179 (2013).
5. Gizinski, A. M. & Fox, D. A. T cell subsets and their role in the pathogenesis of rheumatic disease. *Current Opinion in Rheumatology* **26,** 204–210 (2014).
6. Wang, W. *et al.* The Th17/Treg imbalance and cytokine environment in peripheral blood of patients with rheumatoid arthritis. *Rheumatol. Int.* **32,** 887–893 (2012).
7. Raychaudhuri, S. *et al.* 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. Perricone, C., Ceccarelli, F. & Valesini, G. An overview on the genetic of rheumatoid arthritis: A never-ending story. *Autoimmunity Reviews* **10,** 599–608 (2011).
9. Diogo, D., Okada, Y. & Plenge, R. M. 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. Eyre, S. *et al.* High-density genetic mapping identifies new susceptibility loci for rheumatoid arthritis. *Nat Genet* **44,** 1336–1340 (2012).
11. Okada, Y. *et al.* Genetics of rheumatoid arthritis contributes to biology and drug discovery. *Nature* **506,** 376–381 (2014).
12. Hu, X. *et al.* 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. Rao, D. A. *et al.* Pathologically expanded peripheral T helper cell subset drives B cells in rheumatoid arthritis. *Nature* **542,** 110–114 (2017).
14. Martens, P. B., Goronzy, J. J., Schaid, D. & Weyand, C. M. Expansion of unusual CD4+ T cells in severe rheumatoid arthritis. *Arthritis Rheum* **40,** 1106–1114 (1997).
15. Scarsi, M., Ziglioli, T. & Airo, P. 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. Warrington, K. J., Takemura, S., Goronzy, J. J. & Weyand, C. M. CD4+,CD28- T cells in rheumatoid arthritis patients combine features of the innate and adaptive immune systems. *Arthritis Rheum* **44,** 13–20 (2001).
17. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332,** 687–696 (2011).
18. Sörensen, T., Baumgart, S., Durek, P., Grützkau, A. & Häupl, T. immunoClust-An automated analysis pipeline for the identification of immunophenotypic signatures in high-dimensional cytometric datasets. *Cytometry* n/a–n/a (2015). doi:10.1002/cyto.a.22626
19. Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L. & Nolan, G. P. Automated mapping of phenotype space with single-cell data. *Nat Meth* **13,** 493–496 (2016).
20. Lin, P., Troup, M. & Ho, J. W. K. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biol.* **18,** 59 (2017).
21. Macosko, E. Z. *et al.* Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* **161,** 1202–1214 (2015).

22. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162,** 184–197 (2015).
23. Ornatsky, O. *et al.* Highly multiparametric analysis by mass cytometry. *Journal of Immunological Methods* **361,** 1–20 (2010).
24. Ermann, J., Rao, D. A., Teslovich, N. C., Brenner, M. B. & Raychaudhuri, S. Immune cell profiling to guide therapeutic decisions in rheumatic diseases. *Nat Rev Rheumatol* **11,** 541–551 (2015).
25. Becher, B. *et al.* High-dimensional analysis of the murine myeloid cell system. *Nat. Immunol.* **15,** 1181–1189 (2014).
26. Wong, M. T. *et al.* A High-Dimensional Atlas of Human T Cell Diversity Reveals Tissue-Specific Trafficking and Cytokine Signatures. *Immunity* **45,** 442–456 (2016).
27. Cheng, Y., Wong, M. T., Van Der Maaten, L. & Newell, E. W. Categorical Analysis of Human T Cell Heterogeneity with One-Dimensional Soli-Expression by Nonlinear Stochastic Embedding. *J. Immunol.* **196,** 924–932 (2016).
28. Wong, M. T. *et al.* 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).
29. Baecher-Allan, C., Wolf, E. & Hafler, D. A. MHC class II expression identifies functionally distinct human regulatory T cells. *The Journal of Immunology* **176,** 4622–4631 (2006).
30. Van Gassen, S. *et al.* FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A* **87,** 636–645 (2015).
31. Oshima, S. & Eckels, D. D. Selective expression of class II MHC isotypes by MLC-activated human T lymphocytes. *Hum. Immunol.* **27,** 208–219 (1990).
32. Mason, G. M. *et al.* Phenotypic Complexity of the Human Regulatory T Cell Compartment Revealed by Mass Cytometry. *J. Immunol.* **195,** 2030–2037 (2015).
33. Bendall, S. C. *et al.* Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. *Cell* **157,** 714–725 (2014).
34. Ferrell, P. B. *et al.* High-Dimensional Analysis of Acute Myeloid Leukemia Reveals Phenotypic Changes in Persistent Cells during Induction Therapy. *PLoS ONE* **11,** (2016).
35. Setty, M., Tadmor, M. D., Reich-Zeliger, S. & Angel, O. Wishbone identifies bifurcating developmental trajectories from single-cell data. *Nature* (2016). doi:10.1038/nbt.3569
36. Mingueneau, M. *et al.* 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
37. Hansmann, L. *et al.* Mass cytometry analysis shows that a novel memory phenotype B cell is expanded in multiple myeloma. *Cancer Immunol Res* **3,** 650–660 (2015).
38. Strauss-Albee, D. M. *et al.* Human NK cell repertoire diversity reflects immune experience and correlates with viral susceptibility. *Sci Transl Med* **7,** 297ra115–297ra115 (2015).
39. Gaudillière, B. *et al.* Clinical recovery from surgery correlates with single-cell immune signatures. *Sci Transl Med* **6,** 255ra131 (2014).
40. Larbi, A. & Fulop, T. From "truly naïve" to ‘exhausted senescent’ T cells: when markers predict functionality. *Cytometry A* **85,** 25–35 (2014).
41. De Jong, R. *et al.* 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).
42. Schiott, A., Lindstedt, M., Johansson-Lindbom, B., Roggen, E. & Borrebaeck, C. A. K. CD27- CD4+ memory T cells define a differentiated memory population at both the functional and transcriptional levels. *Immunology* **113,** 363–370 (2004).
43. Kohem, C. L. *et al.* 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).

44. Bruggner, R. V., Bodenmiller, B., Dill, D. L., Tibshirani, R. J. & Nolan, G. P. Automated identification of stratifying signatures in cellular subpopulations. *Proc. Natl. Acad. Sci. U.S.A.* **111,** E2770–7 (2014).
45. Lee, Y. C. *et al.* 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
46. Finck, R. *et al.* Normalization of mass cytometry data with bead standards. *Cytometry Part A* **83A,** 483–494 (2013).
47. Van Der Maaten, L. Accelerating t-SNE using tree-based algorithms. *The Journal of Machine Learning Research* **15,** 3221–3245 (2014).

		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
Resting 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.

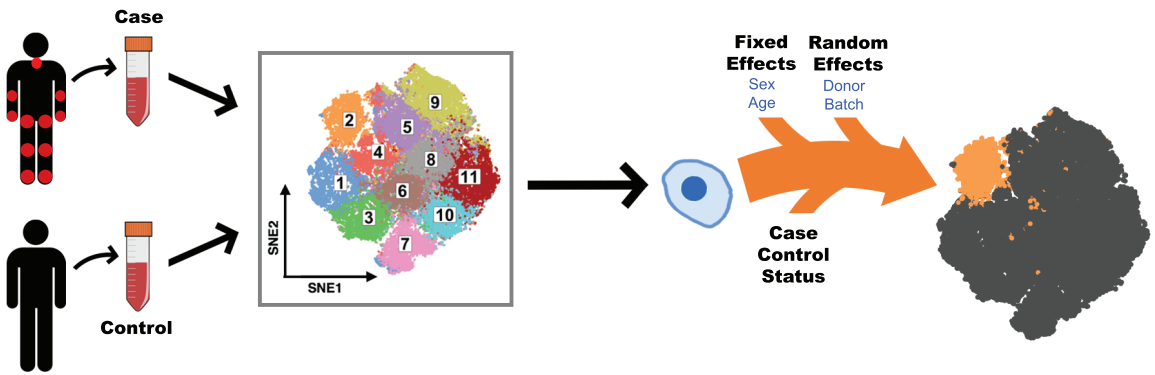

Figure 1: **Mixed effect modeling of Associations of Single Cells (MASC) Overview**

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.

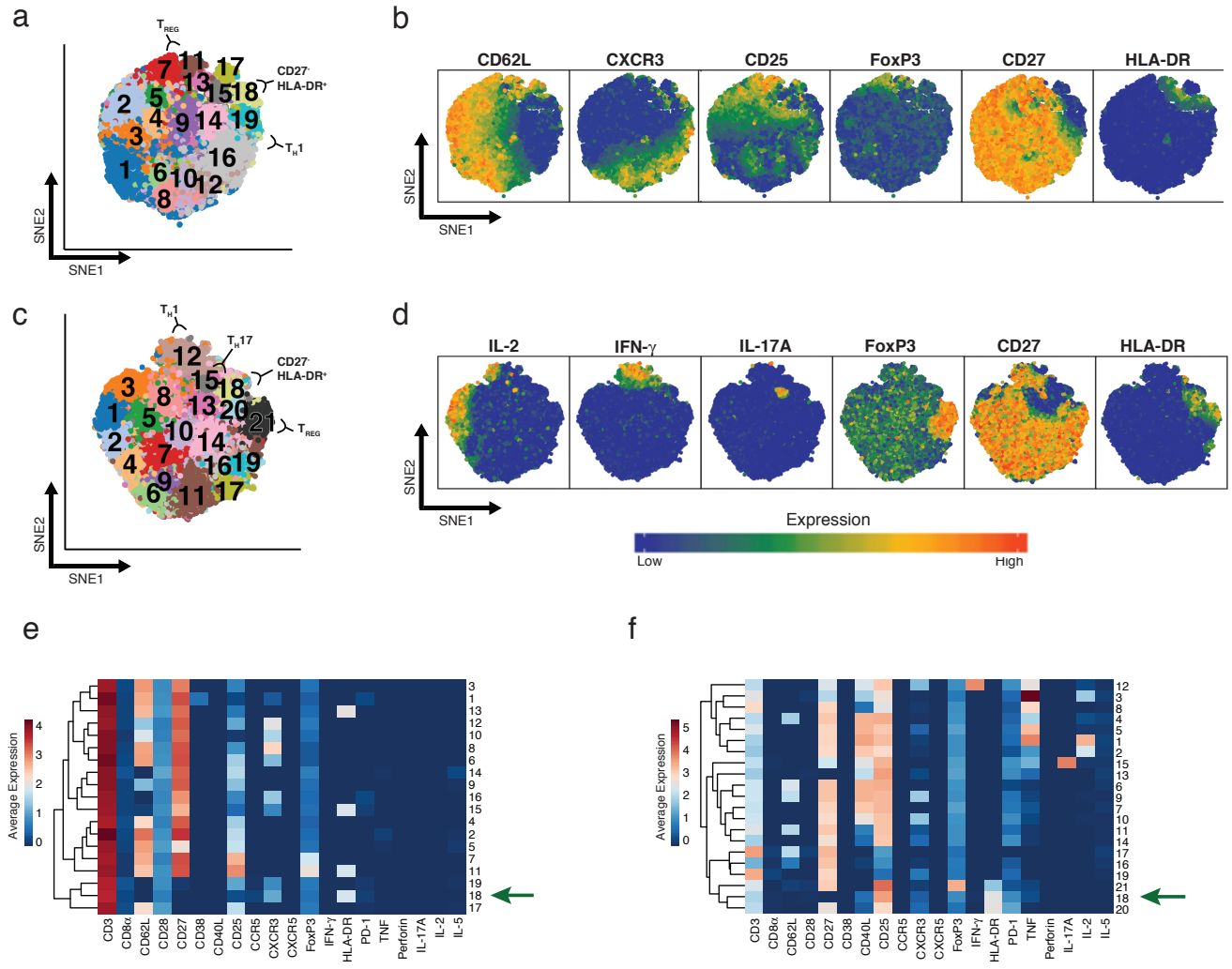

Figure 2: Diversity of CD4$^+$ memory T cells before and after stimulation

(a) SNE projection of 50,000 resting CD4+ memory T cells sampled equally from RA patients (n=24) and controls (n=26). DensVM identified 19 populations in this dataset. (b) Same SNE projections as in (a) colored by the expression of the marker labeled above

each plot. (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 expression of the marker labeled above each plot. (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.

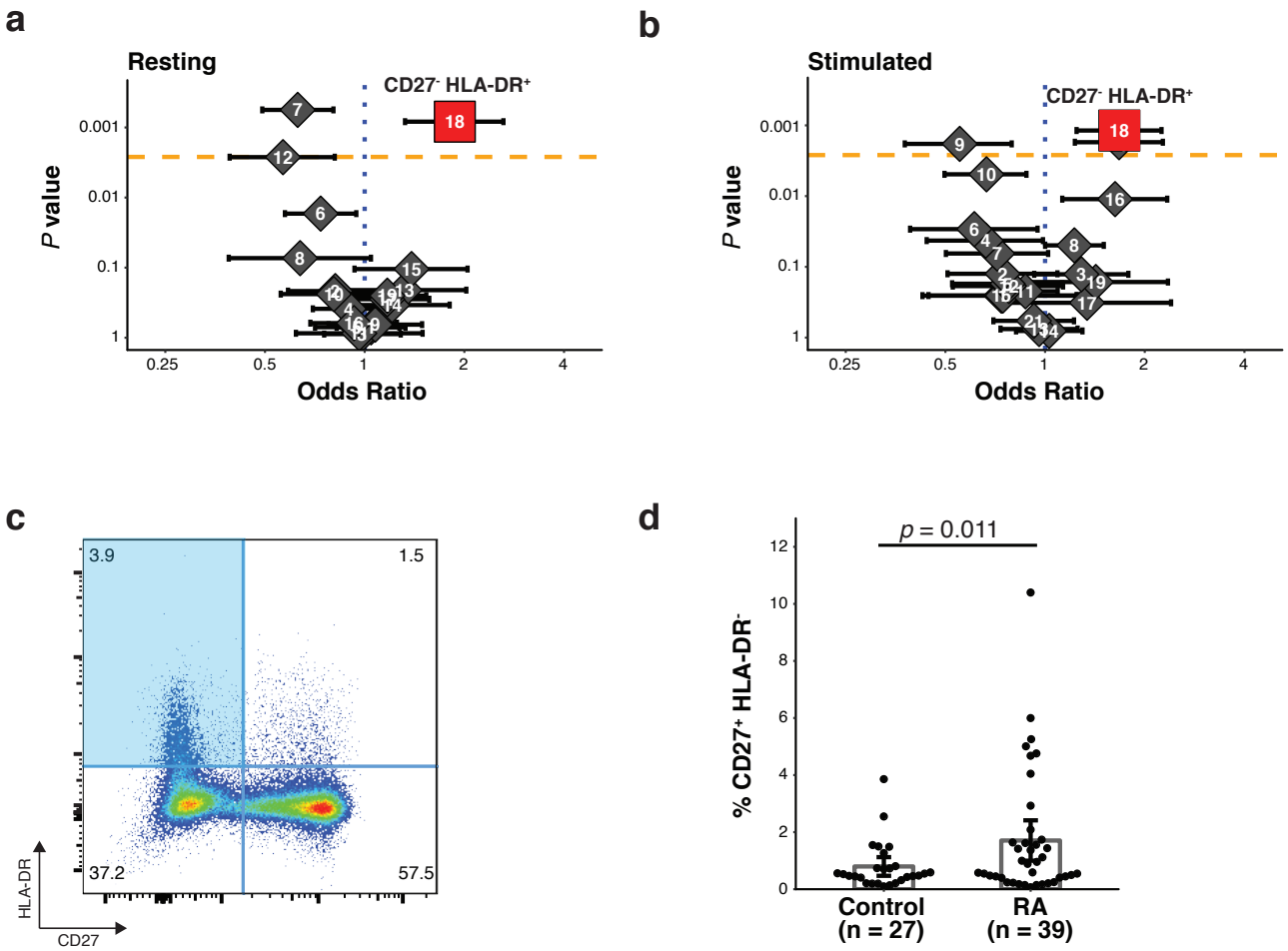

Figure 3: MASC identifies a population that is expanded in RA

(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.

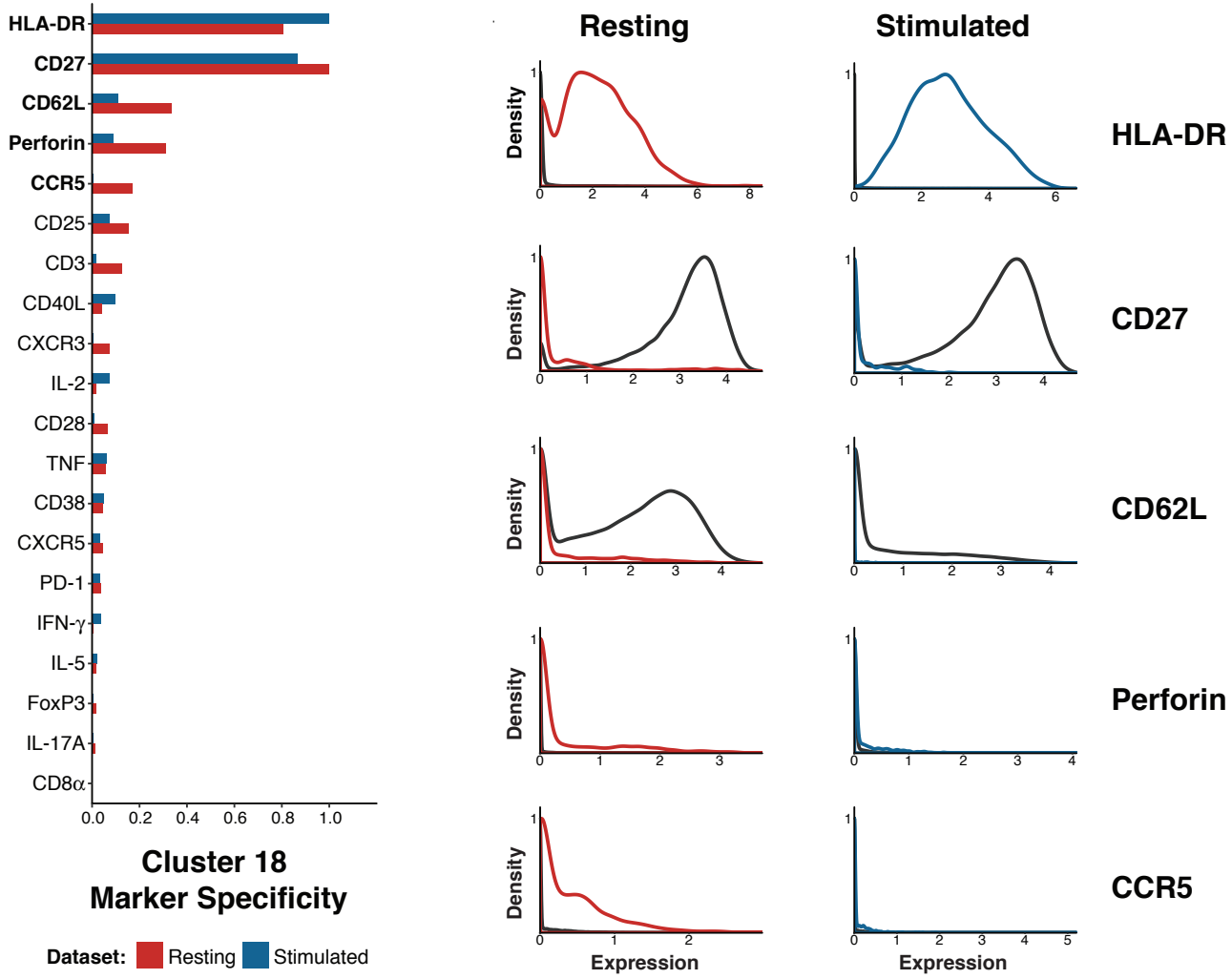

Figure 4: CD27 and HLA-DR expression specifically mark the expanded population

(a) Plot of the Kullback-Liebler divergence for each marker comparing cluster 18 to all other cells in both the resting cell dataset (red) and the stimulated cell 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).

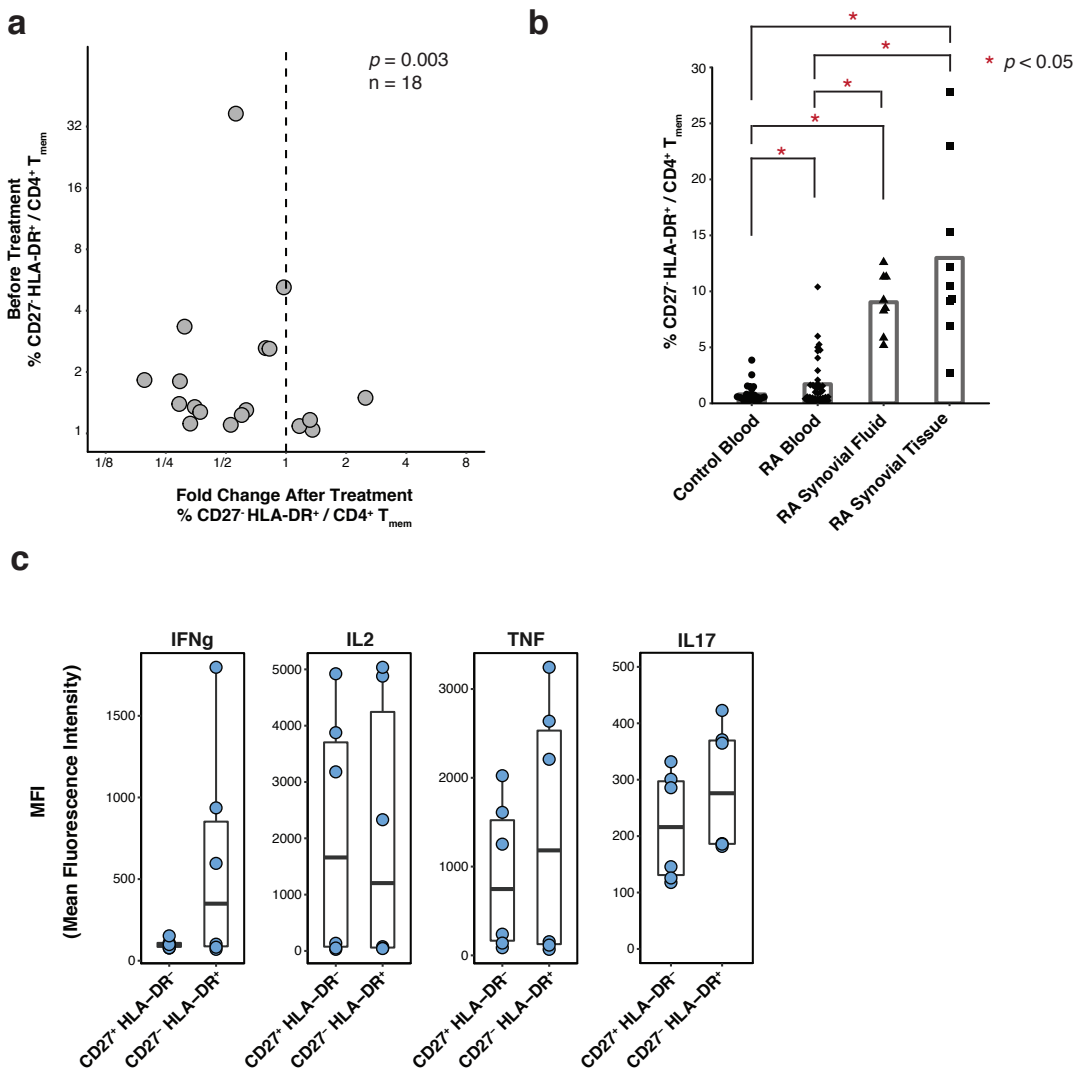

Figure 5: CD27$^-$ HLA-DR$^+$ memory CD4+ T cells are expanded in the blood and joints of patients with active RA.

(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 (p=0.003 by 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. (c) Cytokine expression determined by intracellular cytokine staining of blood effector memory CD4+ T cells after *in vitro* stimulation with PMA/ionomycin. Mean fluorescence intensity (MFI) plotted. Each dot represents a different T cell donor (n = 6; 3 RA patients and 3 controls).