Mixed-effects association of single cells identifies an expanded effector CD4+ T cell subset in rheumatoid arthritis

Chamith Y. Fonseka1,2,3,4,5,*, Deepak A. Rao2,*, Nikola C. Teslovich2, Ilya Korsunsky1, Susan K. Hannes2, Kamil Slowikowski1,2,3,4,5, Michael F. Gurish2, Laura T. Donlin6,7,8, James A. Lederer1, Michael E. Weinblatt2, Elena M. Massarotti2, Jonathan S. Coblyn2, Simon M. Helfgott2, Derrick J. Todd9, Vivian P. Bykerk8,9, Elizabeth W. Karlson2, Joerg Ermann2, Yvonne C. Lee5,10, Michael B. Brenner2, Soumya Raychaudhuri1,2,3,5,11†, Michael E. Weinblatt2, Elena M. Massarotti2, Jonathan S. Coblyn2, Simon M. Helfgott2, Derrick J. Todd9, Vivian P. Bykerk8,9, Elizabeth W. Karlson2, Joerg Ermann2, Yvonne C. Lee5,10, Michael B. Brenner2, Soumya Raychaudhuri1,2,3,5,11†

High-dimensional single-cell analyses have 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 because of technical and interindividual variation. Here, we present mixed-effects modeling of associations of single cells (MASC), a 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 confounders and biological variation. Applying MASC to mass cytometry analyses of CD4+ T cells from the 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 (odds ratio, 1.7; P = 1.1 × 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 responded to immunosuppressive therapy. Mass cytometry and flow cytometry analyses indicated that CD27− HLA-DR+ cells were associated with RA (meta-analysis P = 2.3 × 10−4). Compared to peripheral blood, synovial fluid and synovial tissue samples from RA patients contained about fivefold higher frequencies of CD27− HLA-DR+ cells, which comprised ~10% of synovial CD4+ T cells. CD27− HLA-DR+ cells expressed a distinctive effector memory transcriptomic program with T helper 1 (Th1)− and cytotoxicity-associated features and produced abundant interferon-γ (IFN-γ) and granzyme A protein 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, 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 to 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 major histocompatibility complex (MHC); these polymorphisms affect the range of antigens that MHC II molecules can bind and present 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 (Teff) and regulatory (Treg) 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 the discovery of dominant disease-associated autoantigens. In addition, these populations may directly carry out pathologic effector 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 have 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. Although the discovery of novel cellular populations has been enabled by rapid progress in developing sensitive clustering methods (19–23), a key challenge that remains is establishing methods to identify cell populations associated with a disease. In particular, interindividual variation and technical variation can influence cell population frequencies and need to be accounted for in an association framework. Single-cell association studies, with either mass cytometry or single-cell RNA sequencing (RNA-seq), require statistical strategies robust to interindividual 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. Interindividual differences can lead to real shifts in cell population frequencies at baseline, whereas technical effects can lead to apparent shifts in cell population frequencies.

Here, we describe a robust statistical method to test for disease associations with single-cell data called MASC (mixed-effects 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 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 noncanonical CD4+ T cell populations using a panel of 32 markers to reveal cell lineage, activation, and function (24, 25). Using MASC, we identified a population of memory CD4+ T cells, characterized as CD27− human leukocyte antigen (HLA)-DR+ which was expanded in the circulation of RA patients. Further, we found that CD27− HLA-DR+ T cells were enriched within inflamed RA joints, rapidly produced interferon-γ (IFN-γ) and cytolytic factors, and contracted with successful treatment of RA.

RESULTS
Statistical and computational strategy
We acquired single-cell mass cytometry data from RA case and osteoarthritis (OA) control peripheral blood samples (Fig. 1) and then applied MASC after stringent quality controls to remove technical artifacts and poorly stained cells that are typically observed in mass cytometry data (Materials and Methods). First, we objectively defined subsets of cells using an equal number of random cells from each sample so that they contributed equally to the subsequent analyses. We then used DensVM (26) to cluster the mass cytometry data. Finally, we applied MASC to identify differentially abundant cellular populations associated with disease.

MASC is a reverse association strategy that uses single-cell logistic mixed-effect modeling to test individual cellular populations for association by predicting the subset membership of each cell based on fixed effects (e.g., sex) and random effects (e.g., batch and donor). It assumes a null model where the subset membership of each single cell is estimated by fixed and random effects without considering the case-control status of the samples. Thus, under this null framework, we assume that variation in cluster frequencies is not associated with case-control status. We then measured the improvement in model fit when a fixed-effect term for the case-control status of the sample was 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 interindividual and technical variability.

To ensure that MASC had appropriately calibrated type 1 error, we ran the analysis 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. Under ideal circumstances, 5% of these 10,000 trials would achieve a P value of <0.05 purely by chance in this random dataset; conversely, an inflated method would have much greater than 5% of trials obtaining a P value of <0.05. This approach demonstrated that MASC has only a modestly inflated type 1 error rate for the 19 clusters in the dataset, with 6.5% of trials obtaining P < 0.05 (fig. S1A). Under ideal circumstances, 5% of these 10,000 trials would achieve a P value of <0.05 purely by chance in this random dataset; conversely, an inflated method would have much greater than 5% of trials obtaining a P value of <0.05. This approach demonstrated that MASC has only a modestly inflated type 1 error rate for the 19 clusters in the dataset, with 6.5% of trials obtaining P < 0.05 (fig. S1A). 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 (fig. 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 because 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 (fig. S1C).

Experimental strategy
We applied MASC to mass cytometric analysis of memory CD4+ T cells that were magnetically isolated from peripheral blood mononuclear cells (PBMCs) from patients with established RA (cases) and non-inflammatory OA controls (Table 1). We either (i) rested the cells for 24 hours or (ii) stimulated them with anti-CD3/anti-CD28 beads for 24 hours before analyzing cells with a 32-marker mass cytometry panel that included 22 markers of lineage, activation, and function (table S1). This experimental 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 T cell data separately using the Barnes-Hut modification to the t-SNE (t-distributed stochastic neighbor embedding) algorithm (27) 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 we used to gate CD4+ memory T cells for analysis.

### Clustering approach

Projecting the data into t-SNE space revealed areas of local density that consisted predominantly of cells from RA or OA samples (fig. S2). We wanted to identify CD4+ memory T cell populations in an unsupervised manner. Currently, clustering high-dimensional single-cell data (such as mass cytometry or single-cell RNA-seq data) is an active area of research, and there is no consensus on the best clustering strategy. Hence, we objectively considered multiple clustering algorithm options. We evaluated DensVM (26), FlowSOM (28), and PhenoGraph (23), which

| Table 1. Clinical characteristics of patient samples used. Mean ± SD is shown. Parentheses indicate percentages. “Other biologics” includes rituximab, tofacitinib, and abatacept. ACPA, anticitrullinated protein antibody; CDAL, clinical disease activity index; CRP, C-reactive protein; NA, not applicable; TNF, tumor necrosis factor. |
|---|---|---|---|---|
| **Blood cross-sectional cohorts** | **Controls** | **Cases** | **Controls** | **Cases** |
| **Number** | 26 | 26 | 27 | 39 |
| **Age** | 57 ± 15 | 66 ± 9 | 61 ± 14 | 58 ± 14 |
| **Female** | 19 (73) | 20 (77) | 18 (64) | 30 (77) |
| **ACPA- or RF-positive** | NA | 22 (85) | NA | 39 (100) |
| **CRP (mg/liter)** | NA | 8.6 ± 16.9 | NA | 9.8 ± 17.9 |
| **CDAI** | NA | 9.3 ± 4.4 | NA | 13.7 ± 7.4 |
| **Methotrexate** | 0 | 18 (69) | 0 | 18 (46) |
| **Other biologics** | 0 | 5 (19) | 0 | 9 (23) |

| **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 |

| **Synovial tissue donors** | **Patient** | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 |
| **Age** | 57 | 54 | 76 | 46 | 46 | 79 | 62 | 63 | 52 |
| **Sex** | F | F | F | F | F | F | M | M | F |
| **CRP (mg/liter)** | 25 | 8 | 8 | 11 | 17 | 19 | 13 | 66 | 76 |
| **CDAI** | 14 | 9 | 17 | 15 | 21 | 25 | 5 | NA |
| **Methotrexate** | No | Yes | No | No | No | No | No | Yes | No |
| **Biologic therapy** | Yes | Yes | Yes | Yes | Yes | No | No | No | No |
were identified as among the best performing in a recent benchmarking comparison of clustering methods for high-dimensional cytometry data (29). The DensVM method uses a t-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 and 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 t-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 similar in their ability to generate clusters with distinct marker intensities (fig. S3A). Then, to determine whether those intensity differences were dependent on batch differences, we used a clustering informativeness metric (CIM)–based metric to assess whether clusters were disproportionately represented by individual batches (Materials and Methods). Here, we observed that PhenoGraph and FlowSOM were much more sensitive to batch effects than clusters identified by DensVM (P < 2 × 10⁻³, Wilcoxon rank-sum test; fig. S3B). Consistent with this quantitative assessment, we observed that, in our data, FlowSOM and PhenoGraph produced clusters that were constituted exclusively of or dominated by cells from a specific batch. Thus, we chose to analyze clusters identified by the DensVM algorithm going forward.

**Landscape of CD4⁺ memory T cell subsets**

We observed substantial diversity among resting CD4⁺ memory T cells in both cases and controls, consistent with previous reports demonstrating...
a breadth of phenotypes in CD8− T cells and CD4+ T cells (30–32). We identified 19 distinct subsets in resting (R) memory CD4+ T cells (Fig. 2A and figs. S4A and S5). Central memory T cells (T*CM) segregated from effector memory T cells (T*EM) by the expression of CD62L (Fig. 2B). Five subsets (subsets 1 to 5) of T*CM all expressing CD62L, varied in expression of CD27 and CD38, highlighting the heterogeneity within the T*CM compartment (Fig. 2E). We identified two T helper 1 (T*H1) cell subsets (subsets 8 and 12) and two Treg subsets (subsets 7 and 11). Both Treg subsets expressed high levels of CD25 and FoxP3, and subset 11 also expressed HLA-DR, reflecting a known diversity among Treg populations in humans (33).

When applied to the stimulated CD4+ T cell data in a separate analysis, DensVM identified 21 subsets among all case and control samples (Fig. 2C and figs. S4B and S5). As expected, certain activation markers, such as CD25 and CD40L, were broadly expressed across most subsets after stimulation (Fig. 2D). Mass cytometry robustly detected cytokine production from stimulated CD4+ memory T cells (Fig. 2, D and F). Activated effector cells identified by the production of interleukin-2 (IL-2) were separated into three groups (subsets 1 to 3) according to the relative expression of TNF. Cytokine expression after stimulation also improved the ability to resolve certain CD4+ T eff subsets, such as T*H1.7 cells (subset 15) and T*H1 cells (subset 12). However, we were unable to resolve the Treg subtypes that were observed in the resting T cells; after stimulation, all CD25+FoxP3+ cells were grouped together (subset 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 were significantly overrepresented or underrepresented in patient cells. The frequency of RA cells in each subset ranged from 36.7 to 63.6% in the resting state and from 38.9 to 65.7% in the stimulated state (table S2 and fig. S6A). Visualizing the density of the t-SNE projections revealed that related cells clustered into dense groups both at rest and after stimulation (fig. 2, B and D), and plotting t-SNE projections of cells from cases and controls separately while coloring by density suggested differential abundance of RA cells among clusters (fig. S2). Accounting for subject- and batch-specific random effects with MASC (Materials and Methods), we observed three populations with significantly altered proportions of cells from cases in the resting T cell data. We observed enrichment in subset 18 (P = 5.9 × 10−4; Table 2 and Fig. 3A); this subset consisted of 3.1% of total cells from cases compared to 1.7% of cells from controls and achieved an odds ratio (OR) of 1.9 [95% confidence interval (CI), 1.3 to 2.7]. Conversely, subset 7 (P = 8.8 × 10−2; OR, 0.6; 95% CI, 0.5 to 0.8) and subset 12 (P = 2.0 × 10−3; OR, 0.5; 95% CI, 0.4 to 0.8) were underrepresented for RA cells.

To confirm the robustness of this analytical model, we conducted a stringent permutation test that was robust to potential batch effects. To control for batch, we reassigned case-control labels randomly to each sample, retaining the same number of cases and controls in each batch. For each T cell subset, we recorded the number of case cells after randomization to define a null distribution. In the real data, we observed that 753 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 an SD of 63 cells from case samples. In only 93 of 100,000 permutations did we observe more than 753 cells from case samples in the randomized data (permutation P = 9.4 × 10−4; table S2). We applied permutation testing to every subset and found that the P values produced by permutation were similar to those derived from the mixed-effects model framework (Spearman’s ρ = 0.86; fig. S6, B and C). 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, whereas subset 12 demonstrated only nominal evidence of case-control association (permutation P = 1.3 × 10−2).

Next, for each resting T cell subset, we identified corresponding subsets in the stimulated T cell dataset using a cluster centroid alignment strategy to calculate the distance between subsets across datasets (Materials and Methods). Subset 18 in the resting dataset was most similar to subset 18 in the stimulated dataset, whereas subset 7 in the resting dataset was most similar to subset 21 in the stimulated dataset (fig. S6, D and E). 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 to 2.2), whereas subset 21 (P = 0.55) was not significant in the stimulated data (Table 2 and Fig. 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 to 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. When we used PhenoGraph (23) and FlowSOM (28) to cluster the same mass cytometry dataset, we observed a CD27+, HLA-DR+ cluster with either method (fig. S7, A and C) with association to RA by MASC (fig. S7, B and D). We also assessed how analysis by MASC compared to Citrus (34), 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 rested dataset, Citrus was unable to produce models with an acceptable cross-validation error rate, regardless of the method used (fig. S8).

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% CI is shown next to the OR.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Description</th>
<th>Subset</th>
<th>RA proportion</th>
<th>P</th>
<th>OR</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resting</td>
<td>HLA-DR+, CD27−</td>
<td>18</td>
<td>0.636</td>
<td>5.9 × 10−4</td>
<td>1.9 (1.3–2.7)</td>
<td>MASC</td>
</tr>
<tr>
<td>Stimulated</td>
<td>HLA-DR+, CD27−</td>
<td>18</td>
<td>0.619</td>
<td>1.3 × 10−3</td>
<td>1.7 (1.2–2.2)</td>
<td>MASC</td>
</tr>
<tr>
<td>Flow cytometry replication</td>
<td>Gated</td>
<td>NA</td>
<td>NA</td>
<td>4.4 × 10−4</td>
<td>NA</td>
<td>One-tailed t test</td>
</tr>
<tr>
<td>Meta-analysis</td>
<td>HLA-DR+, CD27−</td>
<td>NA</td>
<td>NA</td>
<td>2.3 × 10−3</td>
<td>NA</td>
<td>Stouffer’s Z-score method</td>
</tr>
</tbody>
</table>

CD27⁻ HLA-DR⁺ CD4⁺ TEM expansion in RA

After noting that subset 18 demonstrated robust association with RA, we interrogated the key features of this population. The lack of CD62L expression in subset 18 indicated that this subset was a TEM population. To define the markers that best differentiated this subset from other cells, we calculated the MIM for the expression of each marker in the subset for both the resting and stimulated datasets (Materials and Methods and Fig. 4A). The expression of HLA-DR and perforin was notably increased in subset 18 compared to all other cells, whereas the expression of CD27 was decreased in this subset (Fig. 4B). We observed that gating on CD27 and HLA-DR largely recapitulated subset 18 in both the resting cells and the stimulated cells (Fig. 4, C and D), with an F measure of 0.8 and 0.7, respectively (Materials and Methods). Although HLA-DR is known to be expressed on T cells in response to activation, it takes several days to induce strong HLA-DR expression (25). Thus, it is likely that cells in subset 18 expressed HLA-DR before stimulation such that analyses of both resting and stimulated cells identified the same HLA-DR⁺ CD27⁻ TEM population.

To confirm the expansion of the CD27⁻ HLA-DR⁺ TEM population in RA patients that we observed by mass cytometry, we evaluated the frequency of CD27⁻ HLA-DR⁺ TEM cells in an independent cohort of 39 seropositive RA patients and 27 noninflammatory controls with OA 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 (Fig. 3C and figs. S9A and S10). Consistent with the mass cytometry analysis, CD27⁻ HLA-DR⁺ cells were significantly expanded in the RA patient samples (P = 0.044, one-tailed t test; P > 0.52, Shapiro-Wilk normality test; Fig. 3D). The frequency of this subset was 0.8% in controls and 1.7% in RA samples, which was similar to the twofold enrichment we observed in the mass cytometry data. We then considered that the mass cytometry and flow cytometry association results together in a meta-analysis, confirming that CD27⁻ HLA-DR⁺ cells significantly associated with RA (P = 2.3 × 10⁻⁴, Stouffer’s Z-score method; Table 2).

To assess the effect of RA treatment on CD27⁻ HLA-DR⁺ TEM frequency, we quantified CD27⁻ HLA-DR⁺ TEM 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 (36 scores (ΔCDAI−)), versus those who did not, defined as an increase in CDAI scores (ΔCDAI+). We observed that changes in CD27⁻ HLA-DR⁺ TEM frequency tracked with clinical response to treatment initiation (P = 3.49 × 10⁻³, Wilcoxon signed-rank test; Fig. 5A). Specifically, in the 18 ΔCDAI− patients, the frequency of CD27⁻ HLA-DR⁺ TEM cells significantly reduced by 0.7-fold (P = 0.006, Wilcoxon signed-rank test; fig. S11B); in contrast, the 5 ΔCDAI+ patients in this same trial had a 1.8-fold increase in CD27⁻ HLA-DR⁺ TEM cells, although the increase was not statistically significant. We did observe a significant difference in the CD27⁻ HLA-DR⁺ TEM frequency fold change between patients experiencing a CDAI reduction versus not (P = 0.02, Wilcoxon rank-sum test; fig. S11A). The decrease in frequency of CD27⁻ HLA-DR⁺ TEM cells in patients responding to treatment escalation was accompanied by a slight increase in the frequency of CD27⁺ HLA-DR⁺ TEM cells: CD27⁺ HLA-DR⁺ TEM 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 afterward (P = 0.009, Wilcoxon signed-rank test; fig. S11C). These results confirm that CD27⁻ HLA-DR⁺ TEM cells were expanded in the circulation of RA patients and decreased with effective disease treatment. We also examined the relationship between the frequency of CD27⁻ HLA-DR⁺ CD4⁺ T cells and disease activity or therapeutic use, but we found no significant associations (fig. S12).

To determine whether CD27⁻ HLA-DR⁺ T cells were 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 nine synovial tissue samples with lymphocytic infiltrates observed by histology, the frequency of CD27⁻ HLA-DR⁺ TEM cells was significantly increased fivefold (P < 0.01, Wilcoxon rank-sum test; median, 10.5% of memory CD4⁺ T cells) compared to blood (Fig. 5B). In two of the tissue samples, >20% of the memory CD4⁺ T cells displayed this phenotype. CD27⁻ HLA-DR⁺ TEM cells were evaluated by microarray analysis, and we found that the expression of HLA-DR was significantly increased in these cells (P = 0.03, Wilcoxon signed-rank test; fig. S11D).
similarly expanded in synovial fluid samples from seropositive RA patients ($P < 0.001$, Wilcoxon rank-sum test; median, 8.9% of memory CD4$^+$ T cells; $n = 8$). Thus, CD27$^-$ HLA-DR$^+$ T cells were enriched at the primary sites of inflammation in RA patients.

**Functional and transcriptional features of CD27$^-$ HLA-DR$^+$ CD4$^+$ TEM**

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 (fig. S9B to D). We sorted and sequenced the following CD4$^+$ T cell populations: naïve CD4$^+$ T cells (T$_{N}$), central memory CD4$^+$ T cells (T$_{CM}$), regulatory CD4$^+$ T cells (T$_{reg}$), and all four CD27$^-/-$ HLA-DR$^-/-$ subsets—defined as DR$^+$ T$_{eff}$ (DR$^+$), DR$^+$ T$_{eff}$ (DR$^+$), DR$^-$ T$_{eff}$ (DR$^-$), and DR$^-$ T$_{eff}$ (DR$^-$)—from PBMCs in seven RA cases and six OA controls. In total, we generated 1.1 billion reads for 90 samples. We aligned reads with Kallisto (37) and applied stringent quality controls to remove genes that lacked sufficient expression (Materials and Methods), ultimately resulting in a set of 15,234 genes for analysis.

---

**Fig. 4. CD27 and HLA-DR expression specifically mark the expanded population.** (A) Plot of the Kullback-Liebler (KL) divergence for each marker comparing cluster 18 to all other cells (gray) 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: t-SNE projection of clusters identified in resting dataset. Middle: Same t-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: t-SNE projection of clusters identified in stimulated dataset. Middle: Same t-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.
We performed principal components analysis (PCA) on the expression data (Fig. 6A). The first PC, capturing 4% of the 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. This finding was consistent with the elevated expression of CXCR3 and CCR5 we observed in the mass cytometry analyses of CD27− HLA-DR+ cells. We note that CXCR3 and CCR5 are both chemokine receptors associated with a Tfh1 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 naive to effector gradient, we used the gene loadings along PC1 to perform GSEA and identify the pathways that were most associated. Intriguingly, the naïve to CD27− HLA-DR+ axis was strongly correlated with naïve versus effector and natural killer (NK) versus CD4+ T cell gene signatures (Fig. 6B).

The association of CXCR3 and CCR5 with PC1 prompted us to examine the expression of T1H1-associated genes in these cells. The effector memory cell populations showed increased expression of multiple T1H1-associated genes, including IFNG (IFN-γ) and TBX21 (Tbet), compared to TΝ, Tem, and TCM. In addition, expression of these T1H1-associated genes was higher in CD27− HLA-DR+ compared to CD27+ HLA-DR+ effector memory cells, which constituted most of the Tem pool (Fig. 6C). In contrast, cytokines characteristic of other polarized T1H subsets (IL17A, IL4, IL21, and TGFβ1) and transcription factors associated with other T1H subsets (RORC, GATA3, BCL6, and FOXP3) were not elevated in CD27− HLA-DR+ cells compared to other effector populations (fig. S13A). We noted that the transcription factor CIITA was also increased in CD27− HLA-DR+ cells (Fig. 6C). CIITA is well known for its role as a regulator of MHC class II; we observed that targets of CIITA such as HLA genes were significantly overexpressed in the CD27− HLA-DR+ population compared to other populations (fig. S13B).

As the expression of PRF1 was positively associated with PC1, we wanted to assess the relationship between the expression of cytotoxic gene programs and the naïve to effector PC1 gradient. We used GSEA to query whether NK cell–associated genes were up-regulated in the CD27− HLA-DR+ population. Specifically, we ranked genes by their degree of differential expression, comparing expression in CD27− HLA-DR+ cells versus that seen in all other cell types. For NK cell–associated genes, we used the previously defined geneset from MSigDB that defines genes up-regulated in NK cells versus CD4+ T cells. Genes that were highly expressed in NK cells similarly showed high expression in CD27− HLA-DR+ cells, whereas genes with low expression in NK cells showed similarly low expression in CD27− HLA-DR+ cells (table S4). Consistent with the enrichment results, a set of genes characteristicaly associated with cytolytic cells, including PRF1, GZMB, GZMA, GNLY, and NKG7, were all increased in expression in CD27− HLA-DR+ cells compared to most other T cell populations analyzed (Fig. 6D).

These results indicate that the CD27− HLA-DR+ cells expressed 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 phorbol 12-myristate 13-acetate (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) before treatment.

Fig. 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 before 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 Fig. 3D. Significance was assessed using one-tailed t test after determining normality with a Shapiro-Wilk test (P > 0.52) and applying a Bonferroni correction for multiple testing.
signed-rank test) than did CD27+ HLA-DR+ cells, which constituted most of the memory CD4+ T cells. In addition, CD27+ HLA-DR+ cells produced IFN-γ at a much higher frequency \((P = 0.009, \text{one-tailed Wilcoxon signed-rank test; Fig. 6E})\). Percent positivity for each marker was determined by selecting an expression level threshold to determine positive staining (fig. S14). Together, broad analyses of gene expression and targeted measures of effector molecule production at the protein level indicated that CD27+ HLA-DR+ T cells are a T111-skewed T cell population capable of producing cytotoxic molecules.

**DISCUSSION**

Mass cytometry has been successfully applied to decipher the heterogeneity of human T cells in multiple settings, including the identification of...
of disease-specific changes to circulating immune cell populations and mapping of developmental pathways (30, 38–41). It and other emerging single-cell strategies offer a promising avenue to characterize the T cell 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 (42–45), in part due to the limited availability of effective association testing strategies.

Although there are many approaches to cluster single-cell data (19–23, 26, 28), 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 nonparametric Mann-Whitney test. Although commonly used in flow cytometry analysis, this approach is markedly underpowered because it relies upon reducing single-cell data to potentially inaccurate 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 was treated as a single event. However, these cells are not entirely statistically independent. Failing to account for dependencies, for example, by using a binomial test or not accounting for random effects, can result in considerably inflated statistics and irreproducible associations (fig. S1). Using a mixed-effects logistic regression model allowed us to account for covariance in single-cell data induced by technical and biological factors that could confound association signals (Fig. 1 and fig. S1B), without inflated association tests. MASC also allows users to use technical covariates relevant to a single-cell measurement (for example, read depth per cell for single-cell RNA-seq or signal quality for mass cytometry) that might influence cluster membership for a single cell.

We note that Citrus, an association strategy to automatically highlight statistical differences between experimental groups and identify predictive populations in mass cytometry data (34), was unable to identify the expanded CD27− HLA-DR+ population. We believe that our methodology compared favorably to Citrus because it incorporated both technical covariates (e.g., batch) and clinical covariates when modeling associations, a key feature when analyzing high-dimensional datasets of large disease cohorts. In addition, the agglomerative hierarchical clustering framework in Citrus substantially increased the testing burden and limited power.

Although we found that clustering with DensVM outperformed FlowSOM and PhenoGraph on our data (fig. 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 (46, 47). There continues to be controversy as to the best strategies, most appropriate metrics, and the optimal parameter choices (46, 48–50). Available clustering approaches use a variety of methods, such as graph-based community detection, self-organizing maps (28), and density-based clustering (20, 23, 51). Because neural networks and 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 that this clustering method also had potential and should be further investigated in the future (fig. S15). Because 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 the 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 (52–54). The broader CD27− CD4+ T cell population did 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 that 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 in disease susceptibility for this subset. Despite the suspected chronic activation of these cells, CD27− HLA-DR+ cells did not appear functionally exhausted. CD27− HLA-DR+ cells rapidly produced multiple effector cytokines upon stimulation in vitro, with an increased predisposition to IFN-γ production. These cells also showed increased expression of cytolytic molecules such as 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 (55). Of note, CD4+ cells that are perforin+ and granzyme A+ have been observed in RA synovial samples (56, 57) and other chronic inflammatory conditions (58, 59). 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, 57). 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.

Although we were able to detect significant associations between the frequency of CD27− HLA-DR− T cells and clinical response (Fig. 5A), we recognize the need for a larger study to detect other subtle subphenotypic associations, such as association with specific therapeutics or disease activity (fig. S12). To this end, larger prospective cohorts with deep immunophenotyping of CD4+ T cells in blood and tissue will be critical.

We note that this study has certain limitations. First, the degree of expansion of CD27− HLA-DR− T cells in the periphery is modest and varies between individuals. Although we were able to recover most of the cells identified as expanded by MASC in our mass cytometry experiments using CD27 and HLA-DR as markers, we note that the more fine-grained immunophenotyping may help to refine phenotypes that even more specifically pinpoint the disease-associated T cell population. Although we have characterized these cells as effector memory CD4+ T cells that produce molecules associated with cytotoxicity and TH1 phenotype, further characterization is needed in future studies. For example, we assigned an effector memory phenotype to the CD27− HLA-DR− cell subset based on mass cytometry data; however, the mass cytometry panel did not specifically measure the expression of CD45RA. Although this would appear to leave open the possibility that these cells might belong to a TEmRA (CD45RO− CD45RA+) subset, we note that our sorting strategy for cells before assay by mass cytometry specifically excluded CD45RA+ cells and consider this possibility unlikely. In addition, our study focused exclusively on CD45RO+ memory CD4+ T cells; thus, we have not assessed other CD4+ T cell populations that may also have cytotoxic features, including T effector memory CD45RA+ (TEmRA) cells (60). A broader cytometric assessment of T cell and other immune
cell phenotypes in RA will be of interest in subsequent studies. Also, we did observe that the mRNA and protein expression of specific markers was not always concordant. Applying more recently developed single-cell techniques that ascertain protein and mRNA expression simultaneously would better describe the dynamics of CD27^+ HLA-DR^+ T cells in response to stimulation (61, 62). We anticipate that MASC will be applied to test for case-control associated differential abundance across multiple cell types in the future.

In summary, the MASC single-cell association modeling framework identified a T_{H1}-skewed cytotoxic effector memory CD4^+ T cell population expanded in RA using a case-control mass cytometry dataset. The MASC method is adaptable to any case-control experiment in which single-cell data are 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 may benefit from using the MASC framework in case-control testing or testing for other clinical sub-phenotypes such as specific treatment response or disease progression.

MATERIALS AND METHODS

Study design

The objective of this research study was to profile immune subsets in peripheral blood samples from RA patients and controls with OA to identify disease-related alterations or changes in frequency among CD4^+ T cells. Human subject research was performed in accordance with the Institutional Review Boards at Partners HealthCare and Hospital for Special Surgery via approved protocols (Partners HealthCare protocol 2014P00255) with appropriate informed consent as required. Patients with RA fulfilled the American College of Rheumatology 2010 Rheumatoid Arthritis classification criteria, and electronic medical records were used to ascertain patients’ rheumatoid factor and anti-CCP (cytometrically activated protein) antibody status, CRP 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 (BWH), 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 because of poor acquisition by the mass cytometer. The study sample size was a result of including all samples that passed quality control and was not set prospectively.

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 consenting patients seen at BWH. We performed medical record review for ACPA positivity according to CCP2 assays, CRP, and RA-specific medications including methotrexate and biologic disease-modifying antirheumatic drugs (DMARDs). For blood cell analyses in the cross-sectional cohort, the treating physician measured the CDAI on the day of sample acquisition. For RA patients followed longitudinally, a new 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 (63). Concurrent prednisone at doses ≤10 mg/day was 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 PBMCs and isolated total CD4^+ memory T cells by negative selection using MACS (magnetic-activated cell sorting) magnetic bead separation technology (Miltenyi). Subsequently, we rested the CD4^+ memory T cells for 24 hours in complete RPMI (Gibco) sterile filtered and supplemented with 15% fetal bovine serum (FBS), 1% penicillin/streptomyacin (Gibco), 0.5% essential and nonessential amino acids (Gibco), 1% sodium pyruvate (Gibco), 1% Heps (Gibco), and 55 μM 2-mercaptoethanol (Gibco). We activated the cells using Human T-Activator CD3/CD28 Dynabeads (Thermo Fisher) at a density of 1 bead 2 cells. At 6 hours before harvesting (t = 18 hours of stimulation), we added menasin and brefnelin A (1:1000) (BD GolgiPlug and BD Golglstop). After 24 hours of stimulation, we incubated the cells with a rhodium metallointercalator (Fluidigm) in culture at a final dilution of 1:500 for 15 min as a viability measure. We then harvested cells into fluorescence-activated cell sorting (FACS) tubes and washed with CyTOF staining buffer (CSB) composed of phosphate-buffered saline (PBS) with 0.5% bovine serum albumin (BSA) (Sigma-Aldrich), 0.02% sodium azide (Sigma-Aldrich), and 2 μM EDTA (Ambion). We spun the cells at 500 g for 7 min at room temperature. We incubated the resulting cell pellets in 10 μl of Fc receptor binding inhibitor polyclonal antibody (ebioscience) and 40 μl of CSB for 10 min at 4°C. The samples were then incubated for 30 min at 4°C on a shaker rack with 1 μl of the following 18 CyTOF surface antibodies in a cocktail brought to a volume of 50 μl per sample in CSB: anti-human CD49D (9F10)—141Pr (Fluidigm), anti-human CCR5 (CD195)—F-6G4—144Nd (Fluidigm), anti-human CD4 (RPA-TA)—145Nd (Fluidigm), anti-human CDSa (RPA-TA)—146Nd (Fluidigm), anti-human CD45RO—147Sm (BHW CyTOF Core), anti-human CD28—148Nd (BHW CyTOF Core), anti-human CD25—IL-2R—(2A3)—149Sm (Fluidigm), anti-human PD1—151Eu (BHW CyTOF Core), anti-human CD62L—DREG56—153Eu (Fluidigm), anti-human CD3—UCHT1—154Sm (Fluidigm), anti-human CD27—L28—155Gd (Fluidigm), anti-human CD183—CXCR3—G025H7—156Gd (Fluidigm), anti-human CCR7—170Er (BHW CyTOF Core), anti-human ICOS—160Gd (BHW CyTOF Core), anti-human CD38—HIT2—167Er (Fluidigm), anti-human CD154—CD40L—(24—31)—168Er (Fluidigm), anti-human CXCR5—CD185—S150S5—171Yb (Fluidigm), and anti-human HLA-DR—L243—174Yb (Fluidigm). We washed the cells with 1 ml of CSB and spun at 700g for 5 min at room temperature. After spin, we aspirated the buffer from pellet and added 1 ml of 1:4 ratio of concentrate to diluent of a Foxp3/Transcription Factor Staining Buffer Set (ebioscience) supplemented with formaldehyde solution (F1268, Sigma-Aldrich) to a final concentration of 1.6%. We incubated the cells at room temperature on a gentle shaker in the dark for 45 min. We washed the cells with 2 ml of CSB + 0.3% saponin (CSB-S) and spun at 800g for 5 min. We incubated the cell pellet with 1 μl of the following 14 intracellular antibodies and 10 μl of a solution of iridium (1:25) in CSB-S brought to a total volume of 100 μl for 35 min at room temperature on a gentle shaker: anti-human IL-4—MP4-25D2—142 (Fluidigm), anti-mouse/human IL-5—TRFK5—143Nd (Fluidigm), anti-human IL-22—22U RTI—150Nd (Fluidigm), anti-human TNFα—Mab11—152Sm (Fluidigm), anti-human IL-2—MQ1-17H12—158Gd (Fluidigm), anti-human IL-21—159Yb (BHW CyTOF Core), anti-human IFN-γ—B27—165Ho (Fluidigm), anti-human GATA3—166Er (BHW CyTOF Core), anti-human IL-9—172Yb (BHW CyTOF Core), anti-human perforin—B-4D8—175Lu (Fluidigm), anti-human IL-10—176Yb (BHW CyTOF
Core), anti-human IL-17A–169Tm (BWH CyTOF Core), anti-human Foxp3 (PCH101)–162Dy (Fluidigm), and anti-human Tbet-164Dy (BWH CyTOF Core). After incubation, we washed the cells in PBS and spun them at 800g for 5 min. We resuspended the pellet in 1 ml of 4% formaldehyde prepared in CSB and incubated the cells for 10 min at room temperature on a gentle shaker, followed by another PBS wash and spin at 800g for 5 min. We washed the pellet in 1 ml of Milli-Q deionized water, spun at 800g for 6 min, and subsequently resuspended the resulting pellet in deionized water at a concentration of 700,000 cells/ml for analysis via the CyTOF 2. We transferred the suspensions to new FACS tubes through a 70-μm strainer and added MaxPar EQ Four Element Calibration Beads (Fluidigm) at a ratio of 1:10 by volume before acquisition.

Mass cytometry panel design
We designed an antibody panel for mass cytometry with the goal of both accurately identifying CD4+ T EM 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).

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 (64). 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.

Because samples were processed and analyzed on different dates, we ran equal numbers of cases and controls each time to guard against day variability, we took extra steps to preprocess 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 min: anti-CD27–fluorescein isothiocyanate (FITC) (TB01), anti-CXCR3–phycocerythrin (PE) (CE3W3D), anti-CD4–PE-Cy7 (RPA-T4), anti-ICOS–peridinin chlorophyll protein (PerCP)–Cy5.5 (ISA-3), anti-CXCR5–BV421 (J52D4), anti-CD45RA–BV510 (H100), anti–HLA-DR–BV605 (G46-6), anti–CD49d–BV711 (9F10), anti–PD-1–allophycocyanin (APC) (E1H2.27H), anti–CD3–Alexa Fluor 700 (HIT3A), anti–CD29–APC-Cy7 (TS2/16), and propidium iodide. Cells were washed in cold PBS and passed through a 70-μm filter, and data were 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 medium. The following day, cells were stimulated with PMA (50 ng/ml) and ionomycin (1 μg/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 min with Fixable Viability Dye eFlour 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, 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–APC647 (CB9) and anti–perforin–PE/Cy7 (B-D48) for 30 min, washed once, and filtered, and data were acquired on a BD Fortessa analyzer. Gates were drawn to identify singlet T cells by forward scatter (FSC)/side scatter (SSC) characteristics, and dead cells and any contaminating monocytes and B cells were excluded by gating out eFlour 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 Liberase TL (100 μg/ml) and deoxyribonuclease I (100 μg/ml) (both Roche) in RPMI (Life Technologies) for 15 min, inverting every 5 min. Cells were passed through a 70-μm 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 RA case samples and six OA control samples were flow sorted into the following seven cellular subsets for low-input RNA-seq: TCM, TEM, T reg, DR+27+, DR+27−, and DR−27-. We rapidly thawed the case and control samples of cryopreserved PBMCs, and MACS enriched the samples for CD4+ T cells (Miltenyi). We rested the samples overnight in complete RPMI/10% FBS. After the rest, we prepared the samples for FACS. We washed the cells once in cold PBS and incubated them with eBioscience Human Fc Receptor Binding Inhibitor (Thermo Fisher). Subsequently, we stained the samples in PBS/5% FBS for 45 min with the following antibodies: FITC CD27 (BioLegend), PE CD25 (BioLegend), PE/Cy7 CD127(IL-7Ra) (BioLegend), Brilliant Violet 510 HLA-DR (BioLegend), Brilliant Violet 605 CD45RA (BioLegend), APC CD62L (BioLegend), Alexa Fluor 700 CD4 (BioLegend), APC/Cy7 CD14 (BioLegend), and APC/Cy7 CD19 (BioLegend). After staining, we washed the cells in cold PBS, passed them through a 70-μm filter, and acquired the samples on a BD FACSaria Fusion cytometer. One thousand cells from each subset were sorted and collected into 5 μl of TCL lysis buffer (Qiagen) 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 25–base pair paired-end reads. We removed one outlier with low read depth. The remaining libraries were sequenced to a depth of 6 million to 19 million reads.

Flow cytometry data analysis
Flow cytometry data were analyzed using FlowJo 10.0.7 (TreeStar Inc.), 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 various populations from the pool of memory CD4+ T cells.

Gene expression quantification
We quantified complementary DNAs on canonical chromosomes (autosomal, X, Y, and mitochondrial) in Ensembl release 83 with Kallisto v0.43.1 in TPM. This analysis quantified inferred counts and length-normalized expression (TPM) of transcripts. To quantify gene expression, we collapsed transcripts mapping to the same HGNC gene symbol by summing the TPM values over. We filtered transcripts that were not sufficiently well expressed, omitting those that did not have at least five counts in at least 10 samples. This resulted in 15,234 well-expressed genes out of 27,717 total genes. For differential expression analysis and PCA, we used log (base 2)–transformed TPM values.

Principal components analysis
Using the gene-level log2 TPM expression metric described above, we selected the top 500 most variably expressed genes for PCA, excluding genes with lower than 1 mean or 0.75 SD expression. We used the removeBatchEffect function in the R limma package, with default parameters, to regress out donor-specific contributions to gene expression. To prevent the absolute range of a strongly expressed gene from dominating the signal in the PCA, we scaled gene expression using the base R scale function on the rows of the expression matrix. This function centers and scales a vector by subtracting the mean and dividing by SD. We then renormalized the samples by centering and scaling the columns, with the same R function. Finally, we used prcomp in R to perform PCA on the resulting gene expression matrix.

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 GSEA 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 principal component. For differential expression, we used the estimated t statistics. To produce barcode plots for select pathways, we reanalyzed these pathways using the R package liger.

Mixed-effects modeling of associations of single cells
Here, we present a flexible method of finding significant associations between subset abundance and case-control status that we have named MASC. The MASC framework has three steps: (i) stringent quality control, (ii) definition of population clusters, and (iii) 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 removed poorly recorded events and low-quality markers before further analysis. We removed those markers (i) that have little expression, because these markers are not informative, and (ii) 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, because 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 were identified and removed, we removed events that were likely to be artifacts. We first removed events that had extremely high signal for a single marker: Events that have recorded expression values at or above the 99.9th percentile for that marker are removed. These events were considered unlikely to be intact, viable cells. Next, a composite “information content” score (Eq. 1) for each event $i$ was 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 $y_{i,m}$ for each event $i$. The sum of these normalized expression values was used to create the event’s information content score.

$$\text{INFO}_i = \sum_{m=1}^{M} y_{i,m}$$  \hspace{1cm} (1)

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 < 0.05) were considered unlikely to be informative in downstream analysis and were removed. In addition, events that derived more than half of their information content score from expression in a single channel were also removed (Eq. 2):

$$\text{INFO}_i \leq 0.5 \leq \max_{m=1}^{M}(y_{i,m})$$  \hspace{1cm} (2)

Potential explanations for these events include poorly stained cells or artifacts caused by the clumping of antibodies with DNA fragments. A final filtering step retained events that were recorded as having detectable expression in at least $M_{\text{min}}$ markers, where $M_{\text{min}}$ may vary from experiment to experiment based on the panel design and expected level of coexpression 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 combined data from cases and controls into a single dataset. It was critical to ensure that each sample contributed equal numbers of cells to this dataset, because otherwise the largest samples would dominate the analysis and confound association testing. After sampling an equal number of cells from each sample, we partitioned these cells into populations using DensVM (26), 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, because MASC is not dependent on any particular method of clustering (fig. S5).

Association testing
Once all cells were assigned to a given cluster, the relationship between single cells and clusters was modeled using mixed-effects logistic regression to account for donor or technical variation (Eq. 3). We modeled the
age and sex of sample $k$ as fixed-effect covariates, whereas the donor and batch that cell $i$ belongs to were modeled as random effects. The random effects variance-covariance matrix treated each sample and batch as independent Gaussians. Each cluster was individually modeled. Note that this baseline model did not explicitly include any single-cell expression measures.

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

(3)

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

To determine whether any clusters were associated with case-control status, we included an additional covariate that indicated whether the $k$th sample is a case or control (Eq. 4):

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

(4)

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

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

(5)

$$p = 1 - \left( \frac{t^{(y_2/2)} e^{-t/2}}{2^{y/2} \Gamma(y/2)} \right)^2$$

(6)

We compared 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 df, because there is only one additional parameter in the full model compared to the null (case-control status). We derived a $P$ value by comparing test statistic $D$ of the likelihood ratio test to the value of the $\chi^2$ distribution with 1 df (Eq. 6), allowing us to find clusters in which case-control status significantly improves model fit. A significant result ($P < 0.05$ after multiple testing correction) indicated that cluster membership for a single cell is influenced by case-control status after accounting for technical and clinical covariates. The effect size of the case-control association can be estimated by calculating the OR from $\beta_{\text{case}}$. If a dataset includes multiple groups, then we can test for association between $g$ groups using $g - 1$ indicator variables. This approach allowed us to capture interindividual differences between donors and 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 magnetically sorted peripheral blood samples (26 cases and 24 controls) under the resting condition and 52 samples (26 cases and 26 controls) in the stimulated condition by CyTOF. Here, we stimulated samples by incubating them with Human T-Activator CD3/CD28 Dynabeads (Thermo Fisher) at a density of 1 bead2 cells for 24 hours. 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, because 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 from analysis. After acquisition, each sample was gated to a CD4+, CD45RO+ population using FlowJo 10.1 (TreeStar) 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 measurable expression (arcsinh-transformed expression > 0) in at least five markers ($M_{\min} = 5$). This removed 3.0 to 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 and then applied the inverse hyperbolic sine transform with a cofactor of 5 to the raw expression data, using the following equation: $y = \sinh^{-1}(x - 1.0)$.

To partition the data, we first randomly selected 1000 cells from each sample and applied the t-SNE algorithm [Barnes-Hut implementation (27)] 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 t-SNE clustering because these markers were only used in gating samples to confirm the purity of CD4 memory T cell selection. We performed separate t-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 t-SNE map to build a training set and then assigns cells to clusters by their expression of all markers using a support vector machine (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 t-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-effects logistic regression according to the MASC method. Donor and batch were included as random-effect covariates, donor sex and age were 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 were first scaled to mean 0, variance 1 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.

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

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

**Marker informativeness metric**

We wanted to determine which markers best separated a given population from the rest of the data in a quantitative manner, because finding a set of population-specific markers is crucial for isolating the population in vivo. 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 KL divergence from $Q$ to $P$ for that marker with the following equation:

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

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 $\geq 1$. 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, whereas recall is defined as the number of cells gated as CD27$^{-}$ HLA-DR$^{+}$ that are in each cluster.

$$F_{\text{measure}} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$  \hspace{1cm} (8)

**Clustering informativeness metric**

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 and 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 because 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 $P_{ij}$ 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).

$$D_{\text{KL}}(P_j || Q) = \sum_{i=1}^{M} P_{ij} \log \frac{P_{ij}}{Q_i}$$  \hspace{1cm} (9)

A cluster with low divergence from the average expression of markers across the entire dataset will capture less marker intensity information than one with a high divergence, because 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 $P_{ij}$, 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).

$$D_{\text{KL}}(P_j || Q) = \sum_{i=1}^{M} P_{ij} \log \frac{P_{ij}}{Q_i}$$  \hspace{1cm} (10)

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 KL 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 three 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.

**Meta-analysis**

We used Stouffer’s Z-score method to define a meta-analysis $P$ value for the significant expansion of CD27$^{-}$ HLA-DR$^{+}$ cells in RA. We converted $P$ values from the resting mass cytometry and flow cytometry analyses to Z scores and found a meta-analysis Z score by taking the sum of these scores divided by the square root of the number of scores—which was two, in our case. We then derived a meta-analysis $P$ value from the Z score using the standard normal distribution.

All analyses were performed using custom scripts for R 3.4.0. We used the following packages: flowCore (65) to read and process FCS files for further analysis, lim4e (66) to apply mixed-effects logistic regression, ggplot2 (67), pheatmap (68) for data visualization, and cytofkit (69) for the implementation of FlowSOM and PhenoGraph clustering algorithms. RNA-seq analyses were conducted using Kallisto (37) to align reads and count reads and gage (70) and liger (71) to perform GSEA. The h2o (72) and mclust (73) packages were used to implement the autoencoder clustering method.
REFERENCES AND NOTES


48. V. P. Atli, A. Magdal, B. J. Schmiedel, J. Clarke, P. O.


Funding: This work was supported, in part, by funding from the NIH (U54CA189677 to S.R., U19AI111224 to S.R., 1R01AR063759 to S.R., and R01 AR064850-03 to T.L.), the Doris Duke Charitable Foundation (grant 2013097 to S.R.), the William D. Dockery Inflammatory Autoimmune Disease Fund (to M.B.B. and S.R.), the Ruth L. Kirschstein National Research Service Award (F31-AR707582 to K.S.), and the Rheumatology Research Foundation Towb and Stephen Malwista, MD Endowment in Academic Rheumatology (to D.A.R.). Author contributions: C.Y.F. and S.R. conceived the study. C.Y.F., D.A.R., and S.R. designed and conducted all experiments. S.R. wrote the initial manuscript. All authors contributed to the final manuscript. Competing interests: The authors declare that they have no competing interests. I.K. has been a paid bioinformatics consultant for Outlet Bio LLC since November 2017. Data and materials availability: All data associated with this study can be found in the paper or the Supplementary Materials. RNA-seq data from this study have been deposited in the Gene Expression Omnibus database (GSE118209). Masc and all other custom scripts used in this analysis are available at www.github.com/immunogenomics.

Submitted 22 September 2017 Resubmitted 9 May 2018 Accepted 25 September 2018 Published 17 October 2018 10.1126/scitranslated.aaq0305
Mixed-effects association of single cells identifies an expanded effector CD4⁺ T cell subset in rheumatoid arthritis


Sci Transl Med 10, eaaq0305.
DOI: 10.1126/scitranslmed.eaaq0305

Pinpointing culprits in single-cell data

New techniques analyzing single cells are being applied to clinical samples. These techniques generate large datasets, and it can be challenging to identify rare cell types associated with disease. Fonseka et al. developed a simple statistical method to do just that while simultaneously controlling for confounding biological and technical variation. Using their method on single-cell mass and flow cytometry data revealed an expanded CD4⁺ T cell subset in the blood of rheumatoid arthritis patients that was later seen to contract upon treatment. The authors' method performed better than the current gold standard and could be a potentially widely applied tool for the analysis of high-dimensional single-cell data.