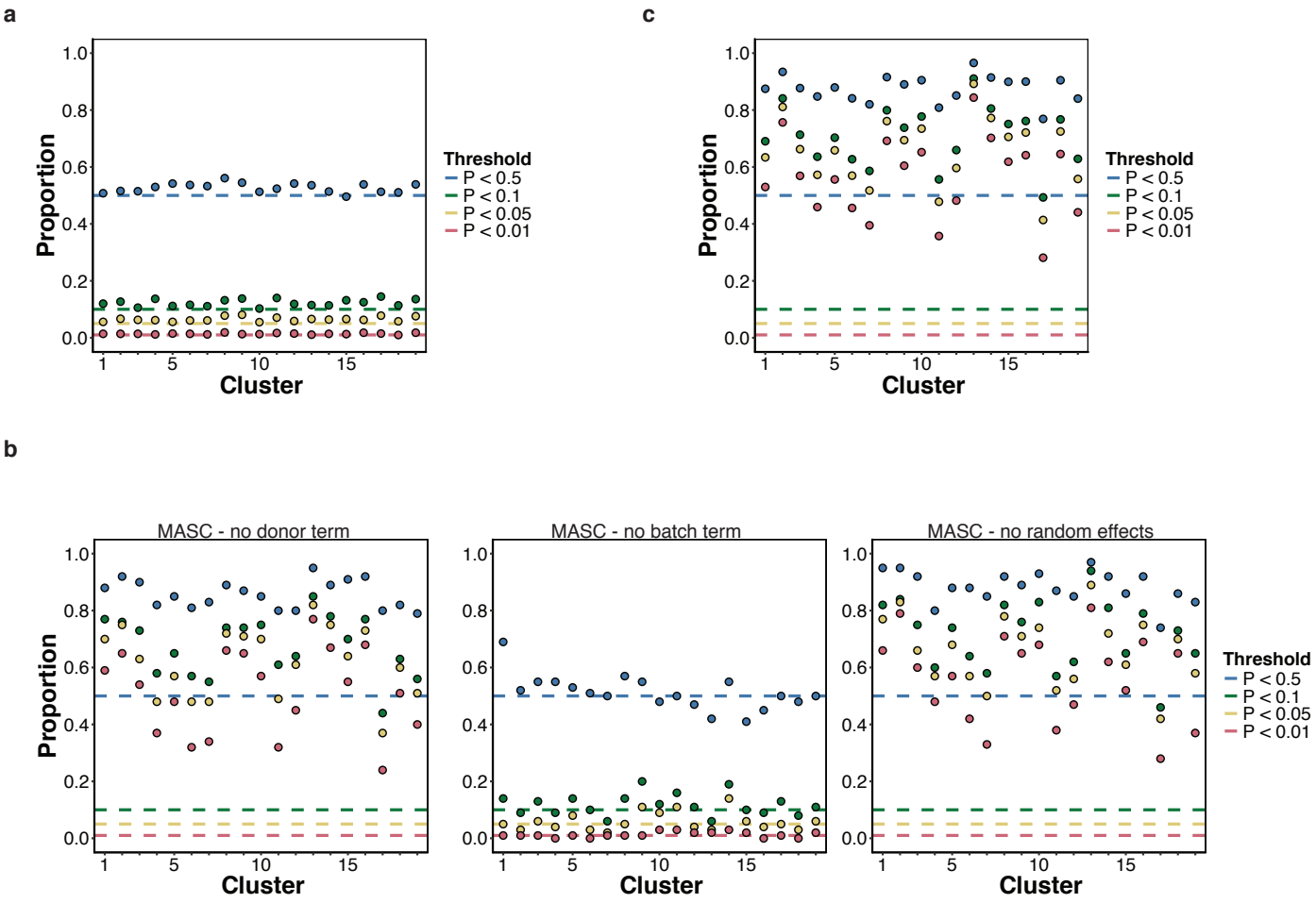
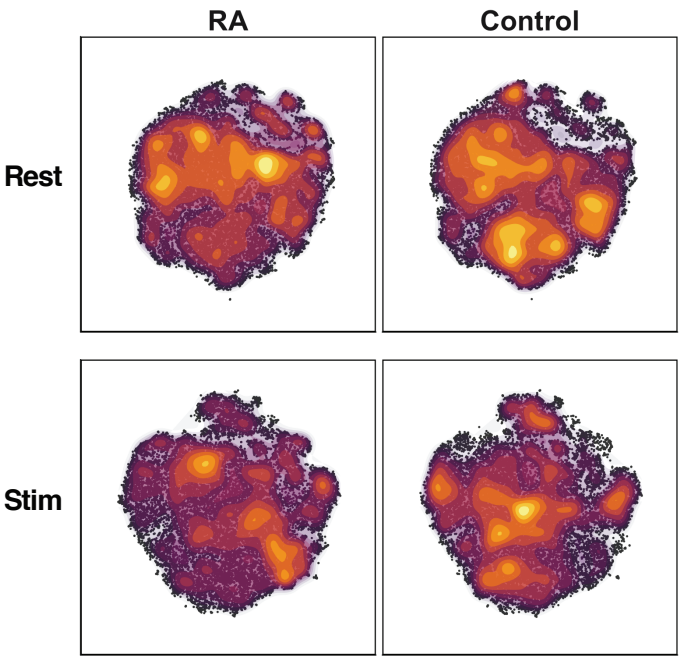


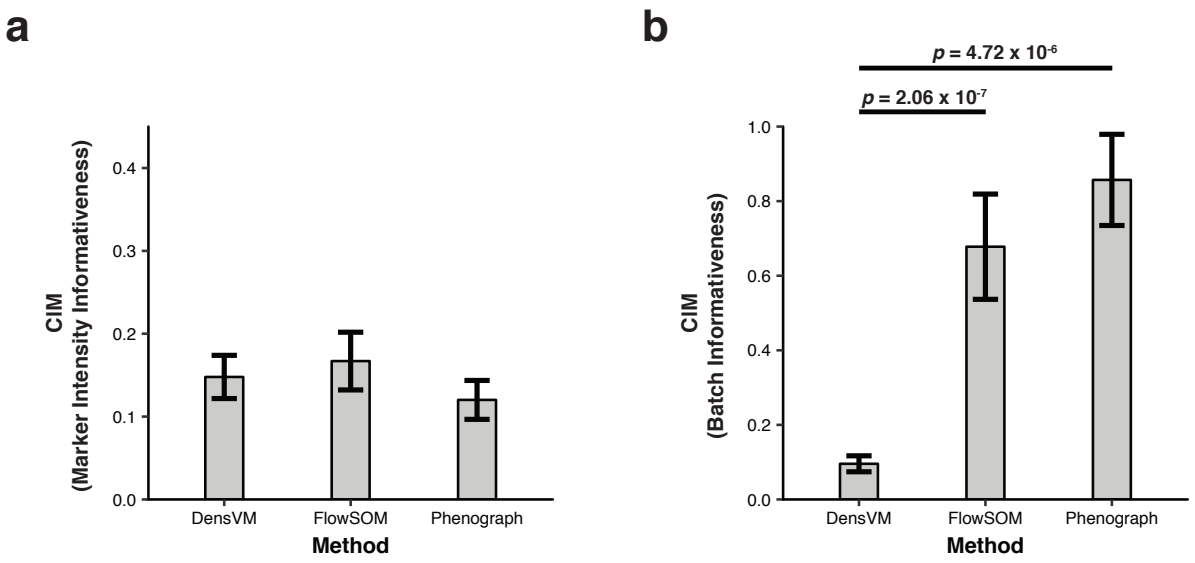
Supplementary Figure 1 – MASC Type 1 Error

(a) MASC demonstrates well-controlled type 1 error rates. MASC was run on the resting dataset after randomizing case-control labels 10000 times to eliminate any case-control associations. The proportion of p-values at different thresholds are plotted for each cluster. (b) MASC p-values obtained in the same manner as previous, but without donor or batch specific random effect terms. (c) P-values obtained in the same manner for binomial association tests on clusters found in the resting dataset.



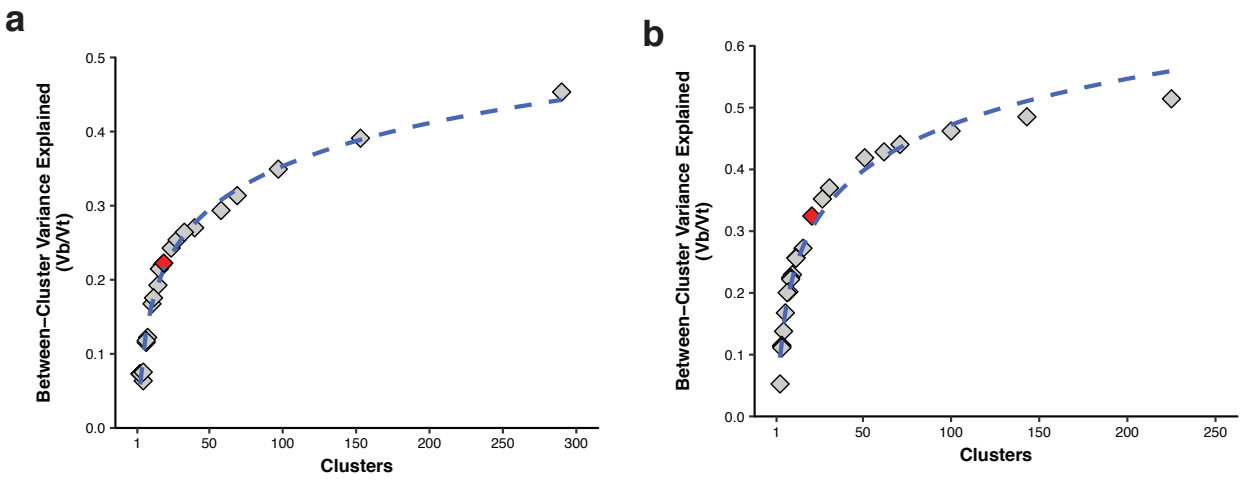
Supplementary Figure 2 – SNE Projection Density

(a) SNE projections of datasets before (top) and after (bottom) stimulation, split by case-control status. Coloring the SNE projections by density identifies regions that are differentially abundant between RA and control samples.



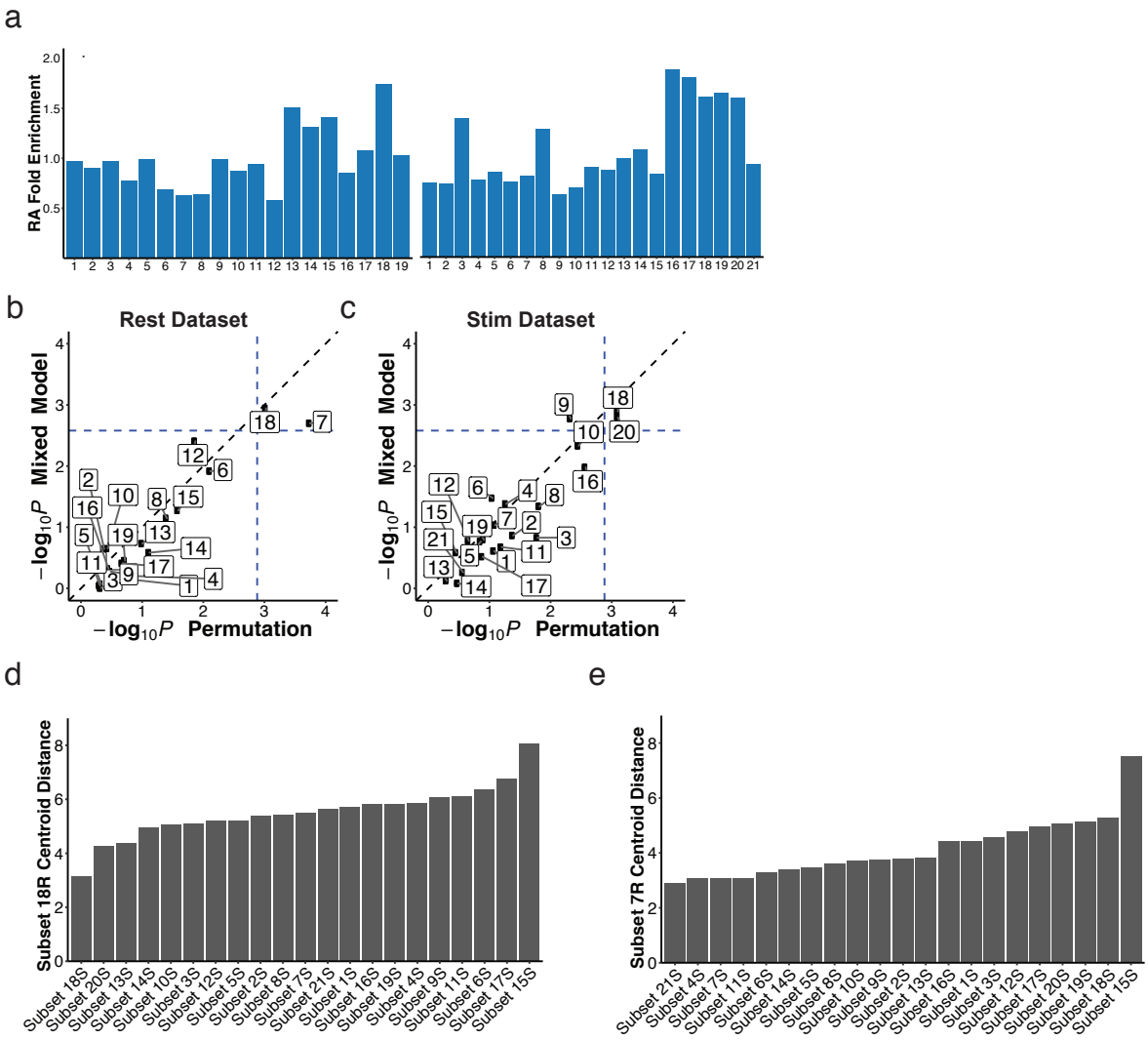
Supplementary Figure 3 – Cluster Informativeness Metric Analysis of Clustering Approaches

We clustered the same dataset using three different clustering algorithms, DensVM, Phenograph, and FlowSOM. These algorithms identified 19 (DensVM and FlowSOM) or 21 (Phenograph) clusters. (a) Clusters found by DensVM, Phenograph, and FlowSOM had similar average CIM scores when considering marker expression, indicating that the clusters found by these algorithms were similarly informative. That is, marker intensities were different from the average marker expression profile across clusters to the same extent. (b) Clusters found by Phenograph and FlowSOM had a significantly higher CIM score when considering batch than those found by DensVM, indicating that the Phenograph and FlowSOM clusters were more affected by batch effects. We assessed significance using a Wilcoxon rank sum test and p-values were Bonferroni adjusted to control for multiple testing.



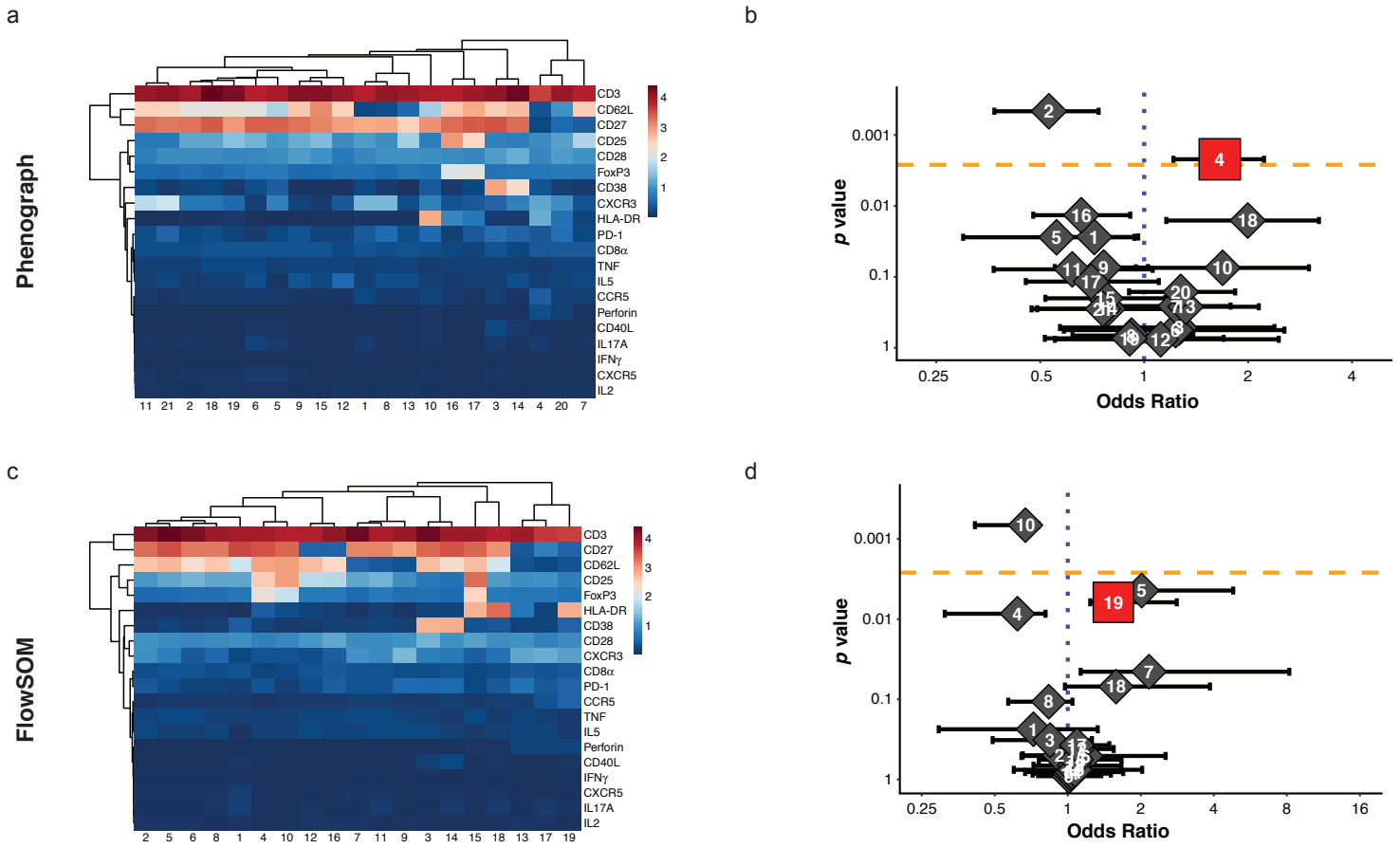
Supplementary Figure 4 – DensVM Clustering Elbow Plots

In order to define the optimal number of clusters, we use the the elbow strategy. We clustered data with DensVM across a range of bandwidth values, yielding different numbers of clusters at each of the 25 bandwidth values chosen. We then took the ratio of between-cluster variance to total variance to measure the amount of variance explained by each set of clusters. The set of clusters used in our analyses is marked in red and an exponential fit to the points shown is plotted as a dashed blue line. (a) DensVM clustering of the resting dataset produced 3-290 clusters across different bandwidths. The bandwidth producing 19 clusters (red) is at an inflection point for the amount of between-cluster variance explained. (b) DensVM clustering of the stimulated dataset produced 3-225 clusters across different bandwidths. The bandwidth producing 21 clusters (red) is at an inflection point for the amount of between-cluster variance explained.



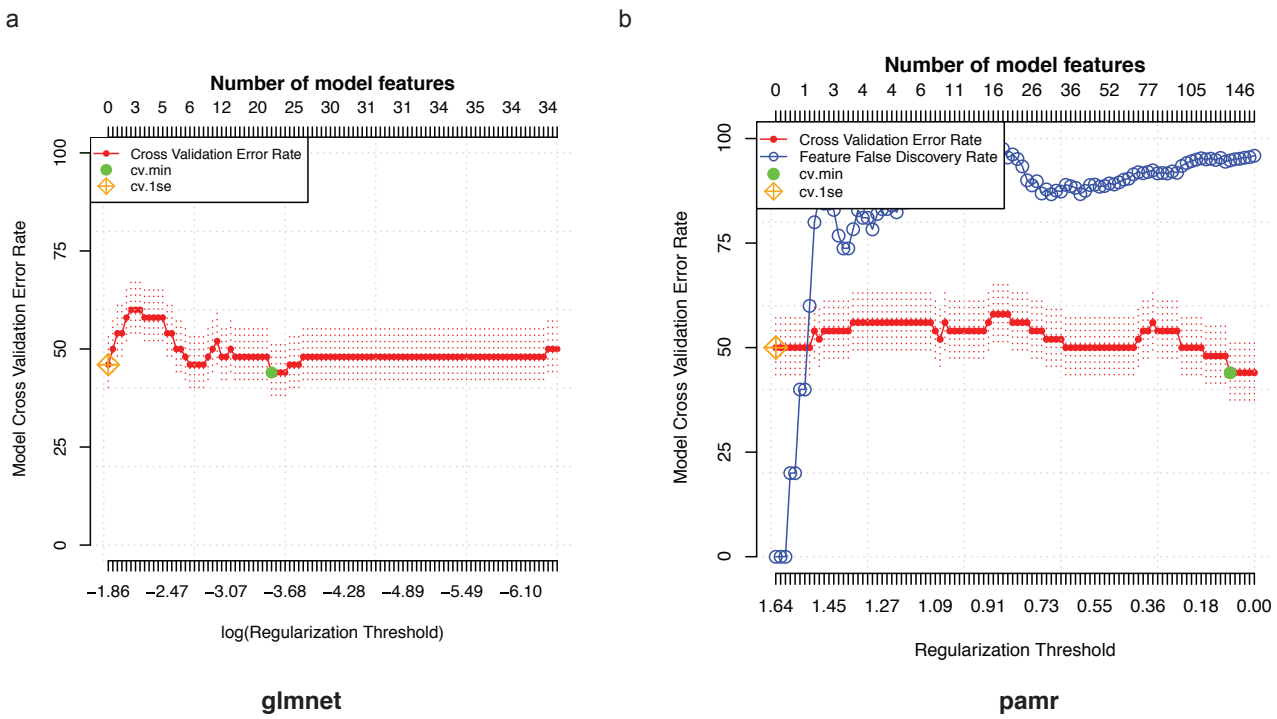
Supplementary Figure 5 – Association Permutation Testing and Cluster Alignment

(a) The enrichment or depletion of RA cells relative to the overall proportion is shown for all clusters identified in the resting (left) and stimulated (right) datasets. (b, c) Association p-values as calculated by MASC (y-axis) and by explicit permutation (x-axis) correlate in both resting and stimulated datasets. Spearman's correlation coefficients for (b) and (c) were $r_s = 0.82$ and $r_s = 0.86$, respectively. (d) Clusters in the stimulated dataset ranked by their overall distance from cluster 18. After normalizing marker expression in each cluster, cluster centroids were created and Euclidean distances were calculated between all clusters in the stimulated dataset and cluster 18 in the resting dataset. (e) Same as (d), but for distance from cluster 7 in the resting dataset.



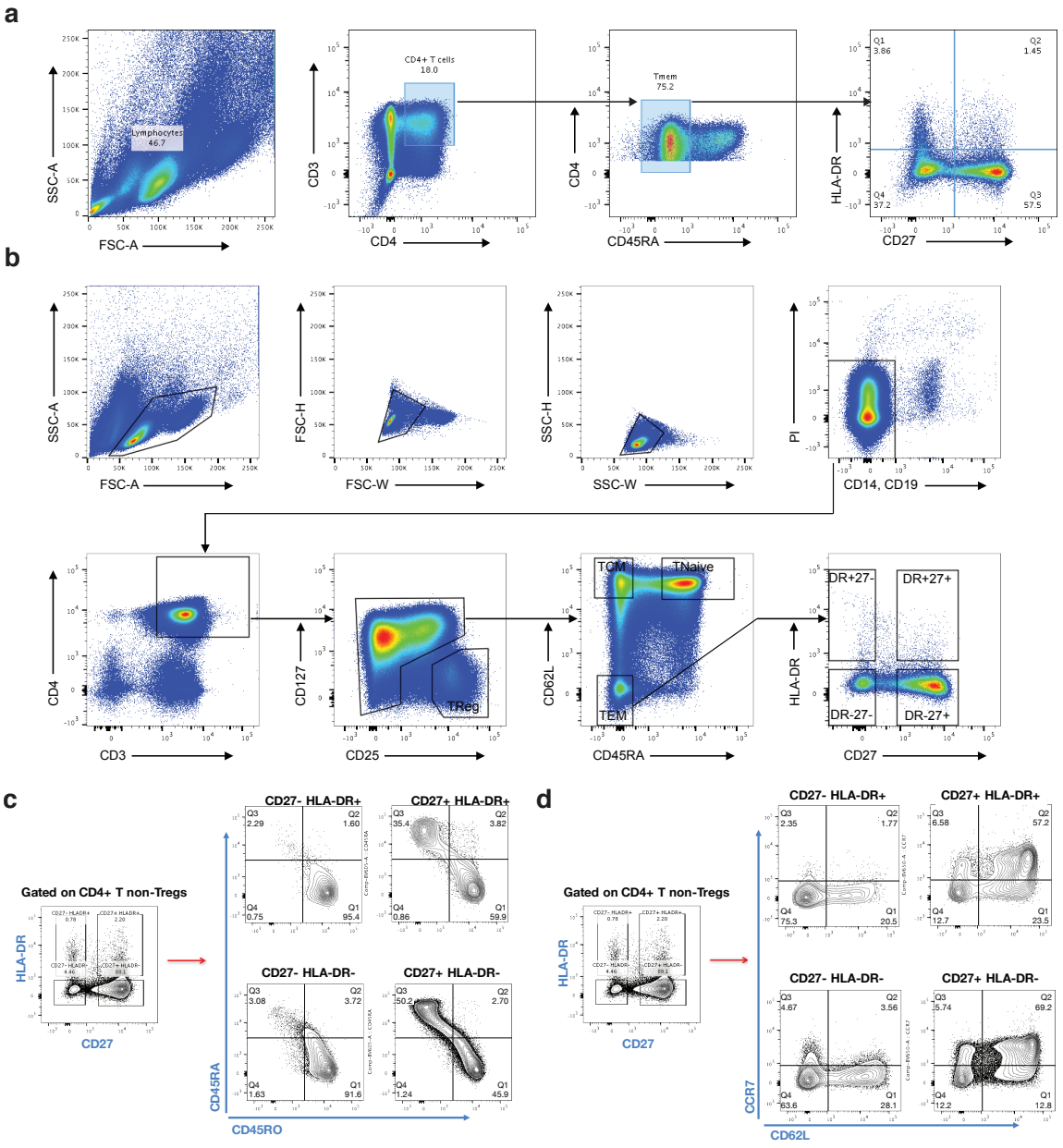
Supplementary Figure 6 – Phenograph and FlowSOM Clustering.

(a) Phenograph identified 21 clusters in the resting dataset, including an CD27- HLA-DR+ T_{EM} population (cluster 4) that is significantly expanded in RA. (b) Odds ratios and association p-values were calculated by MASC for each cluster identified by Phenograph. The yellow line indicates the significance threshold after applying the Bonferroni correction for multiple testing. (a) FlowSOM identified 19 clusters in the resting dataset, including an CD27- HLA-DR+ T_{EM} population (cluster 19) that is nominally expanded in RA. (b) Odds ratios and association p-values were calculated by MASC for each cluster identified by FlowSOM. The yellow line indicates the significance threshold after applying the Bonferroni correction for multiple testing.



Supplementary Figure 7 – Association Testing with Citrus

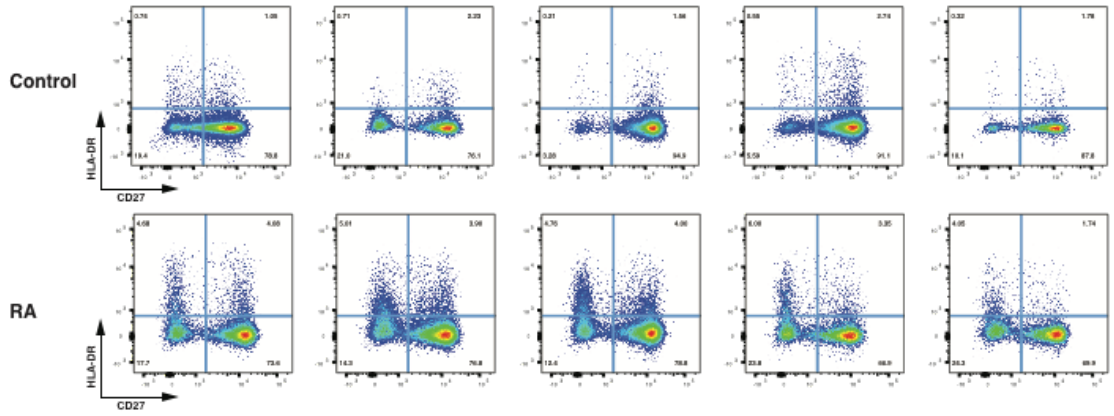
Citrus was run on the resting dataset but failed to produce models with acceptable error rates using either L1-penalized regression (a) or nearest shrunken centroid (b) methods. Model features found to be associated with case-control status by either method are unlikely to be meaningful given the extremely high cross-validation error.



Supplementary Figure 8 – Flow Cytometry and RNA-seq Gating Strategies

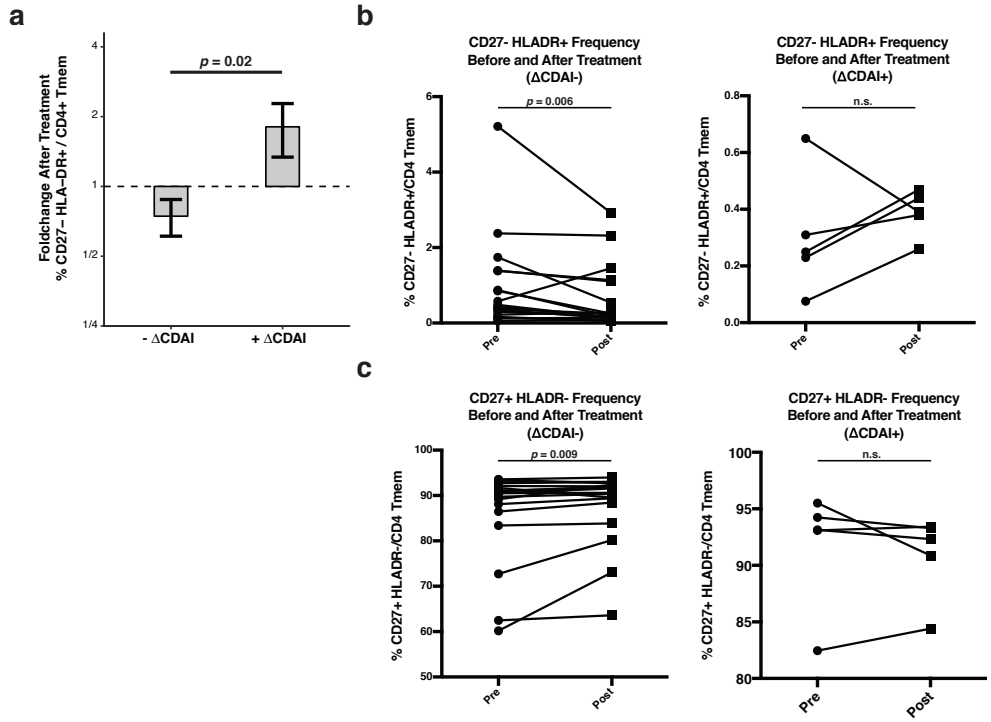
(a) Gating strategy used to isolate CD27- HLA-DR+ cells for flow cytometry quantification. Cells were first gated to lymphocytes using forward and side scatter parameters, then to CD4+ memory T cells before being split into four populations based upon the expression of CD27 and HLA-DR. (b) Gating strategy used to isolate populations for RNA sequencing. Cells were gated to lymphocytes using forward and side scatter parameters, then gated as CD14- CD19- to remove any non T cell lymphocytes. Cells were then gated to CD4+ T cells before isolating the following populations: regulatory T cells (CD25+ CD127-), central memory T cells (CD62L+ CD45RA-), naïve T cells (CD62L+ CD45RA+) and effector memory T cells (CD62L- CD45RA-). Effector memory T cells were then split into four populations based upon the expression of

CD27 and HLA-DR. (c) Expression of CD45RO and CD45RA is shown for all four effector memory populations analyzed by RNA-seq. CD27- HLA-DR+ cells are uniformly CD45RA-CD45RO+. (d) Same as (c), but the expression of CD62L and CCR7 are shown.



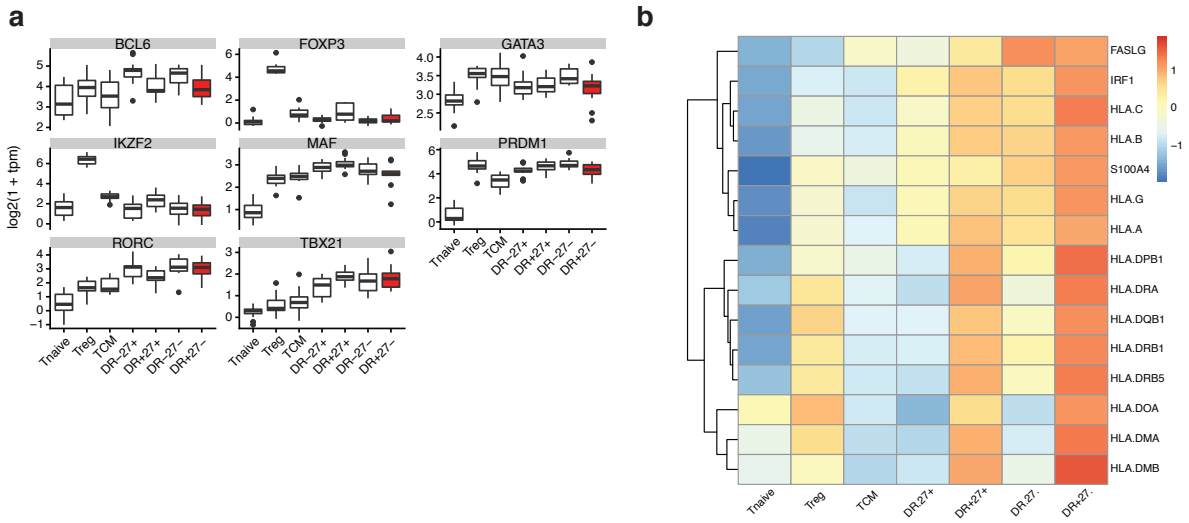
Supplementary Figure 9 – CD27 and HLA-DR Expression in Flow Cytometry Cohort

The expansion of the CD27- HLA-DR+ T cell population in RA patients was validated in an independent cohort of 39 seropositive RA patients and 27 controls using flow cytometry. The frequency of CD27 and HLA-DR cells among CD4+ memory T cells is shown for 10 representative donors, 5 cases and 5 controls.



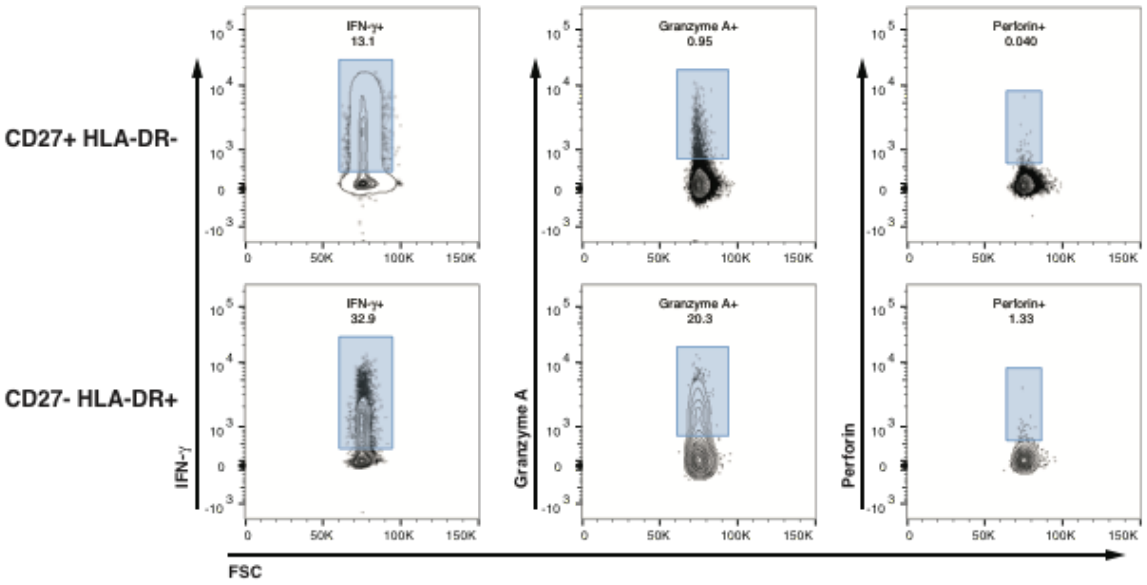
Supplementary Figure 10 – CD4+ Effector Memory T Cell Populations in a Clinical Response Cohort

We quantified the frequency of CD27- HLA-DR+ T cells in 23 RA patients before and 3 months after initiation of a new medication for RA. Patients were separated into those who experienced a clinical response (n = 18) versus those that did not (n = 5), defined as a reduction (**-ΔCDAI**) or an increase in CDAI scores (**+ΔCDAI)**. (a) The fold-change in CD27- HLA-DR+ frequency was significantly different between the two groups of patients (p=0.02, Wilcoxon rank sum test). (b) We quantified CD27- HLA-DR+ cell frequencies in patients who experienced a reduction in disease activity after initiation of a new medication for RA and those who did not. The frequency of the CD27- HLA-DR+ subset significantly decreased in **-ΔCDAI** individuals (p=0.006, Wilcoxon rank sum test), but did not significantly change among **+ΔCDAI** individuals. (c) Same as (b), except that the frequency of CD27+ HLA-DR- was quantified in **ΔCDAI** and **+ΔCDAI** individuals. We calculated frequencies of CD27- HLA-DR+ and CD27+ HLA-DR- cells from all CD4+ memory T cells, and assessed significance with Wilcoxon signed-rank tests.



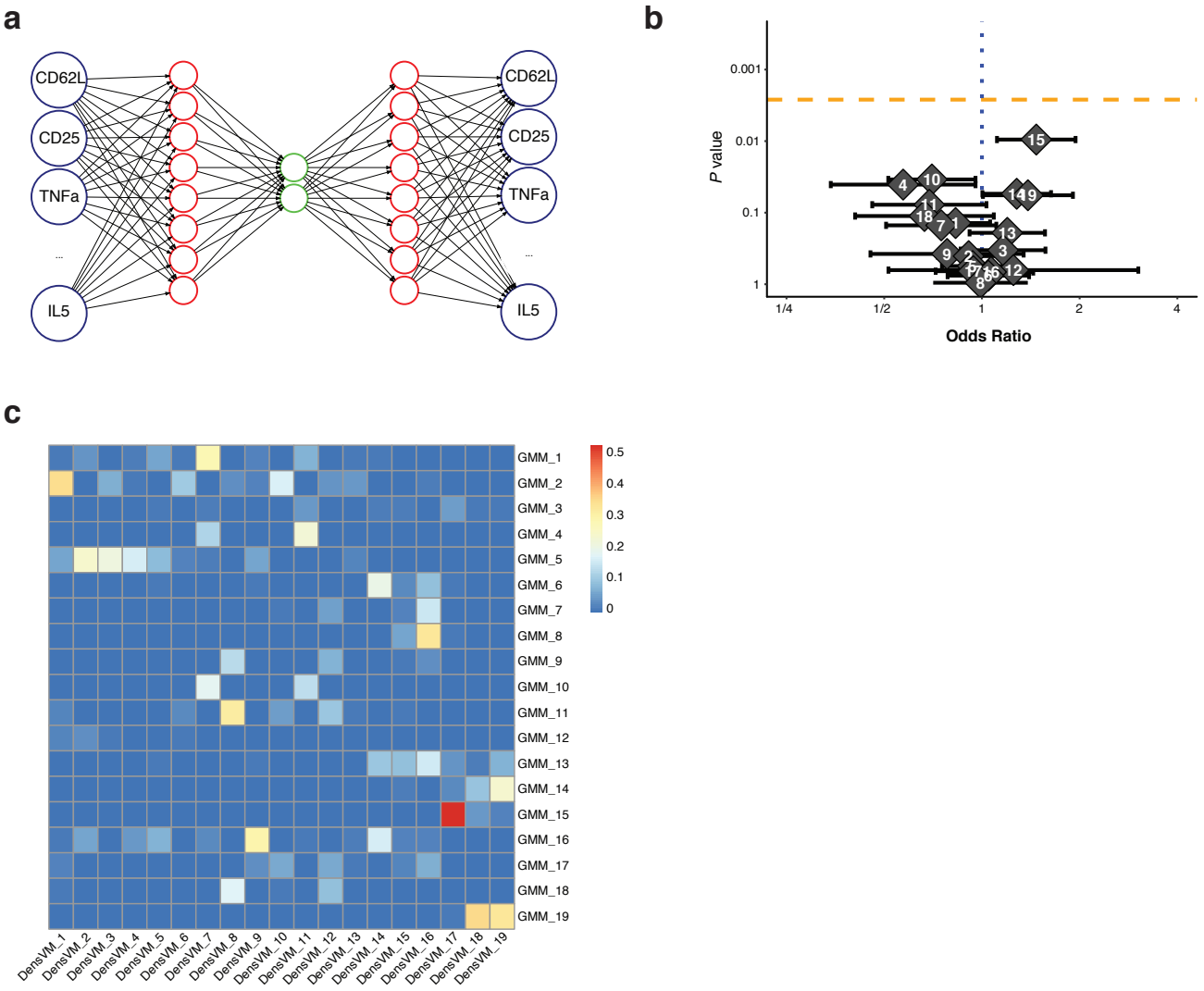
Supplementary Figure 11 – RNA-seq Analysis of CD4+ T Cell Subsets

(a) Expression of lineage-defining transcription factors for CD4+ T helper subsets shown for each population analyzed by RNA-seq. Populations are ordered by principal component 1 loadings, from naïve to effectors. (c) The expression of selected targets of transcription factor CIITA is shown for each sequenced T cell population.



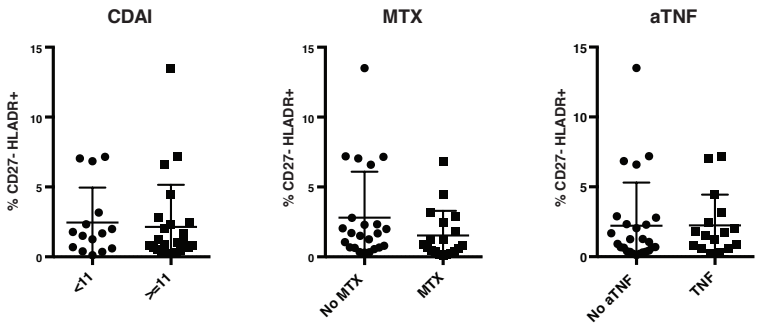
Supplementary Figure 12 – Flow Cytometry Expression Quantification

The expression of markers granzyme A, perforin, and IFN-γ across all samples are displayed, with the gating used to to define percent positivity for those markers. The same gates were used to analyze CD27+ HLA-DR- and CD27- HLA-DR+ populations. The expression plots of granzyme A and perforin show the concatenation of six samples (3 RA, 3 OA), while the expression plot of IFN-γ shows concatenation of 12 samples (6 RA, 6 OA).



Supplementary Figure 13 – Using a Neural-Net Auto-encoder to Cluster Mass Cytometry Data

(a) Schematic of the deep auto-encoder that was trained upon the rest data using 3 hidden layers. Clustering was then performed on the middle layer (2 node) projection (highlighted in green). (b) Clusters identified by the auto-encoder were tested for case-control associations using MASC. None of the clusters reached significance after correcting for multiple hypothesis testing. (c) The Jaccard index was calculated between each cluster identified by DensVM (x-axis) and the autoencoder (y-axis). The most significant disease-associated auto-encoder clusters share significant overlap with the original DensVM clusters.



Supplementary Figure 14 – CD27- HLA-DR+ Frequency and Clinical Characteristics

The frequency of CD27- HLA-DR+ cells was quantified as the percentage of memory CD4+ T cells in an independent cohort of 39 seropositive RA patients and 27 controls using conventional flow cytometry. RA patients were then dichotomized by clinical disease activity index (CDAI) scores, methotrexate use (MTX) or anti-TNF therapy use (aTNF). The frequency of CD27- HLA-DR+ cells was not significantly different between groups in any comparison.

Supplementary Figure 15 – Marker Expression Distribution Plots for DensVM Clusters

For each cluster in the resting (n = 19) and stimulated (n = 21) datasets, we plotted the distribution of marker expression for cells in the cluster against the expression distribution for that marker for cells across the entire dataset. Expression specific to the cluster is colored in red, while dataset expression is colored in dark grey. Mass cytometry expression values are shown after applying a standard arcsinh transformation.

Isotope	Marker	Clone
Nd143Di	IL-5	TRFK5
Nd144Di	CCR5	NP-6G4
Nd145Di	CD4*	RPA-T4
Nd146Di	CD8α	RPA-T8
Sm147Di	CD45RO*	UCHL1
Nd148Di	CD28	CD28.2
Sm149Di	CD25	2A3
Eu151Di	PD-1	EH12.2H7
Sm152Di	TNF	MAb11
Eu153Di	CD62L	DREG-56
Sm154Di	CD3	UCHT1
Gd155Di	CD27	L128
Gd156Di	CXCR3	G025H7
Gd158Di	IL-2	MQ1-17H12
Dy162Di	FoxP3	PCH101
Ho165Di	IFN-γ	B27
Er167Di	CD38	HIT2
Er168Di	CD40L	24-31
Tm169Di	IL-17A	BL168
Yb171Di	CXCR5	51505
Yb174Di	HLA-DR	L243
Lu175Di	Perforin	B-D48

Supplementary Table 1: Panel design for mass cytometry experiments. Markers that are starred were only used for gating purposes to confirm the purity of CD4 memory T cell isolation and were not including in clustering or downstream analyses.

Cluster	Cell Number	RA Proportion	Permutation *p* value	MASC *p* value	Odds Ratio	Odds Ratio, 2.5% CI	Odds Ratio, 97.5% CI
1	6000	0.493	2.43E-01	7.12E-01	1.0	0.8	1.3
2	4506	0.475	4.18E-01	2.36E-01	0.8	0.6	1.1
3	3994	0.492	4.71E-01	9.60E-01	1.0	0.8	1.3
4	2776	0.437	1.24E-01	6.40E-01	0.9	0.7	1.2
5	2205	0.498	4.69E-01	8.78E-01	1.0	0.7	1.3
6	1578	0.406	7.68E-03	9.31E-03	0.7	0.5	0.9
7	2100	0.385	1.80E-04	8.78E-04	0.6	0.5	0.8
8	2138	0.389	3.94E-02	6.20E-02	0.6	0.3	1.0
9	3856	0.497	3.27E-01	5.39E-01	1.1	0.8	1.6
10	2317	0.467	3.64E-01	1.81E-01	0.8	0.5	1.1
11	790	0.486	4.89E-01	8.96E-01	1.0	0.6	1.5
12	1421	0.367	1.34E-02	2.03E-03	0.5	0.4	0.8
13	965	0.602	9.90E-02	1.54E-01	1.4	0.9	2.1
14	2781	0.568	7.59E-02	2.56E-01	1.3	0.8	1.9
15	1048	0.586	2.57E-02	5.29E-02	1.5	1.0	2.2
16	7507	0.461	3.54E-01	4.89E-01	0.9	0.7	1.2
17	1020	0.520	2.09E-01	3.30E-01	1.2	0.9	1.5
18	1184	0.636	9.40E-04	5.59E-04	1.9	1.3	2.7
19	1814	0.507	1.91E-01	2.17E-01	1.2	0.9	1.6

Supplementary Table 2: MASC analysis of the 19 clusters identified in the resting dataset.

Cluster	Cell Number	RA Proportion	Permutation *p* value	MASC *p* value	Odds Ratio	Odds Ratio, 2.5% CI	Odds Ratio, 97.5% CI
1	2518	0.430	8.35E-02	2.45E-01	0.7	0.4	1.2
2	1757	0.429	4.10E-02	1.37E-01	0.7	0.5	1.1
3	2981	0.586	1.64E-02	1.48E-01	1.3	0.9	1.8
4	2086	0.440	5.32E-02	4.15E-02	0.7	0.4	1.0
5	3202	0.463	1.38E-01	1.71E-01	0.8	0.5	1.1
6	2031	0.435	8.83E-02	3.35E-02	0.6	0.4	1.0
7	3687	0.452	8.10E-02	9.26E-02	0.7	0.5	1.1
8	3834	0.565	1.52E-02	4.56E-02	1.2	1.0	1.5
9	1160	0.389	4.66E-03	1.68E-03	0.5	0.4	0.8
10	2908	0.416	3.47E-03	4.67E-03	0.7	0.5	0.9
11	4742	0.478	6.32E-02	2.11E-01	0.9	0.7	1.1
12	3934	0.469	2.17E-01	1.67E-01	0.8	0.5	1.1
13	2436	0.501	4.97E-01	7.52E-01	1.0	0.7	1.3
14	4462	0.522	3.26E-01	8.27E-01	1.0	0.8	1.3
15	755	0.458	3.45E-01	2.61E-01	0.7	0.4	1.3
16	1459	0.657	2.66E-03	1.05E-02	1.6	1.1	2.3
17	1701	0.646	1.29E-01	3.04E-01	1.4	0.8	2.4
18	1182	0.619	7.90E-04	1.28E-03	1.7	1.2	2.2
19	1483	0.626	1.23E-01	1.60E-01	1.4	0.9	2.4
20	911	0.618	7.90E-04	1.61E-03	1.7	1.2	2.3
21	2771	0.487	2.68E-01	5.48E-01	0.9	0.7	1.2

Supplementary Table 3: MASC analysis of the 21 clusters identified in the stimulated dataset.

Gene Set	Pathway	p value	q value	Size	Enrichment
GSE22886	NAÏVE CD4 T CELL VS NK CELL	8.45E-13	1.01E-11	159	NK CELL
GSE3982	CENT MEMORY CD4 T CELL VS NK CELL	3.53E-08	1.74E-07	159	NK CELL
GSE27786	CD4 T CELL VS NK CELL	1.59E-04	6.38E-04	170	NK CELL
GSE3039	CD4 T CELL VS NKT CELL	5.07E-03	1.22E-02	166	NKT CELL
GSE3982	EFF MEMORY CD4 T CELL VS NK CELL	1.67E-01	3.35E-01	149	NK CELL

Supplementary Table 4: Gene set enrichment analysis of genes differentially expressed in CD27- HLA-DR+ cells. Q values represent FDR corrected p-values, using an FDR of 5%. The number of genes in each set is listed as size. Enrichment indicates which cell type gene signature was enriched in genes specific for CD27- HLA-DR+ cells.