

Antigen Tracking Figures

Ryan Sheridan

2020-08-16

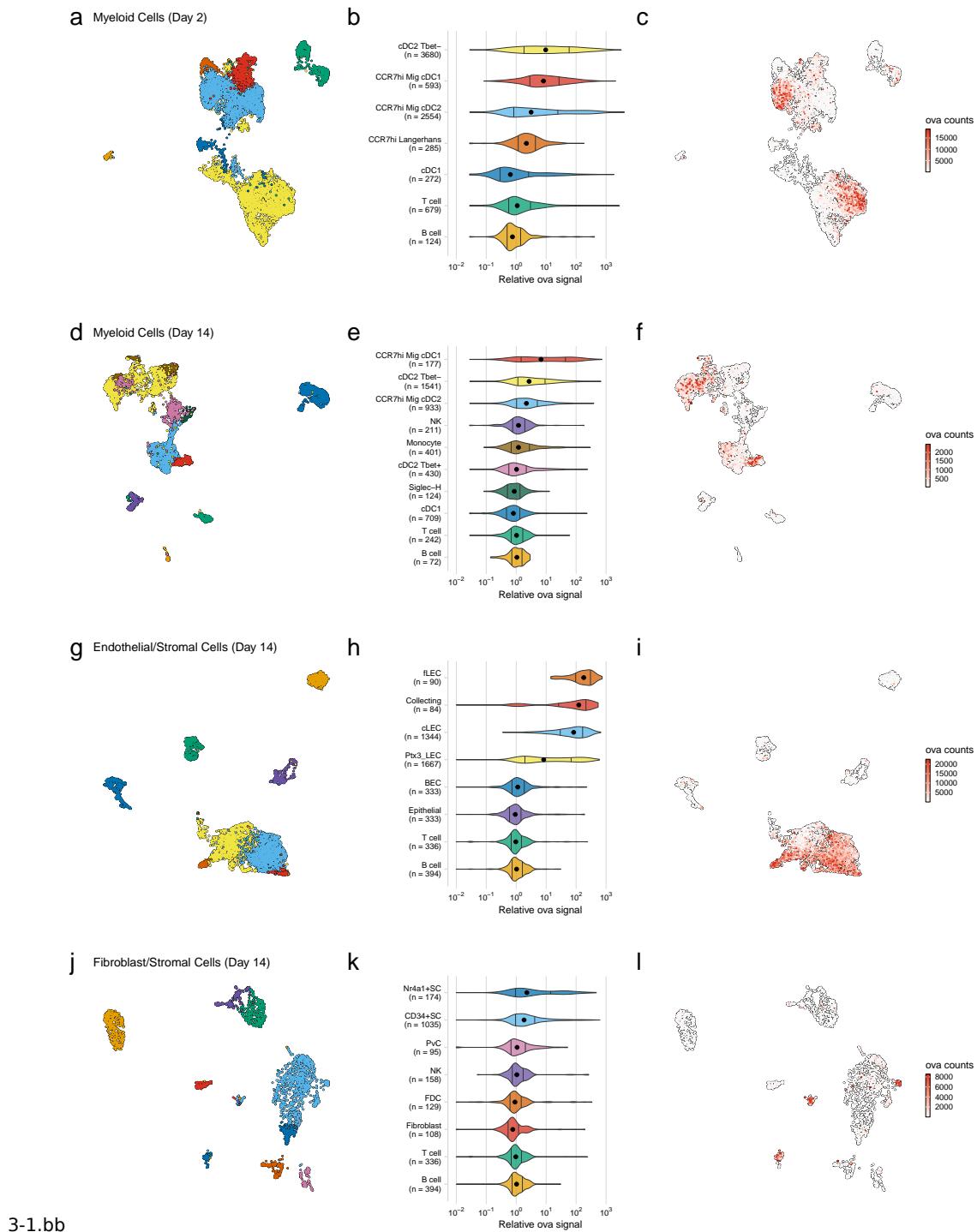
Methods

Fastq files from the gene expression and antigen tracking libraries were processed using the feature barcode version of the cellranger count pipeline (v3.1.0). Reads were aligned to the mm10 and Vaccinia virus (NC_006998) reference genomes. Analysis of gene expression and antigen tracking data was performed using the Seurat R package (v3.2). Antigen tracking and gene expression data were combined into the same Seurat object for each sample (CD45-/day2, CD45+/day2, CD45-/day14, CD45+/day14). Cells were filtered based on the number of detected genes (>250 and <5000) and the percent of mitochondrial reads (<15%). Gene expression counts were log normalized (NormalizeData), and relative ova signal was calculated by dividing ova-psDNA counts by the median ova-psDNA counts for all T and B cells present in the sample. Gene expression data were scaled and centered (ScaleData). 2000 variable features (FindVariableFeatures) were used for PCA (RunPCA) and the first 40 principal components were used to find clusters (FindNeighbors, FindClusters) and calculate uniform manifold approximation and projection (UMAP) (RunUMAP). Cell types were annotated using the R package clustifyr (<https://rnabi-oco.github.io/clustifyr>) along with reference bulk RNA-seq data from ImmGen (available for download through the clustifyrdata R package, <https://rnabioco.github.io/clustifyrdata>). The samples were then divided into separate objects for DCs, LECs, and FRCs. Cell subsets were annotated using clustifyr with reference bulk RNA-seq data for DCs, LECs, and FRCs.

Identification of ova-low and ova-high populations was accomplished using a gaussian mixture model implemented with the R package mixtools (<https://cran.r-project.org/web/packages/mix-tools/index.html>). All LECs were used when identifying ova-low and ova-high cells (Fig. 4). For DCs (Fig. S10), ova-low and -high populations were identified independently for each DC cell type. For ova-low and ova-high populations, differentially expressed genes were identified using the R package presto (wilcoxauc, <https://github.com/immunogenomics/presto>). Differentially expressed genes were filtered to include those with an adjusted p-value <0.05, log fold change >0.25, AUC >0.5, and with at least 50% of ova-high cells expressing the gene. A reproducible analysis pipeline is available at <https://github.com/rnabioco/antigen-tracking>.

Figure 3

Cell types associated with high antigen counts. (a, d, g, j) UMAP projections are shown for DCs, LECs, and FRCs. (b, e, h, k) Relative ova signal was calculated by dividing antigen counts for each cell by the median antigen counts for T and B cells. (c, f, i, l) Antigen counts are displayed on UMAP projections for each cell type.



3-1.bb

Figure 4

Genes associated with high antigen counts for day 2 and day 14 cDC2 Tbet⁻ cells. (a, d) Day 2 (a) and day 14 (d) cDC2 Tbet⁻ cells containing low and high antigen counts were identified using a gaussian mixture model. A UMAP projection is shown for ova-low and ova-high cells. Cell types not included in the comparison are shown in white (Other). (b, e) The distribution of ova antigen counts is shown for ova-low and ova-high cDC2 Tbet⁻ cells. Dotted lines indicate the mean counts for each population. (c, f) UMAP projections show the expression of top markers associated with ova-high cDC2 Tbet⁻ cells.

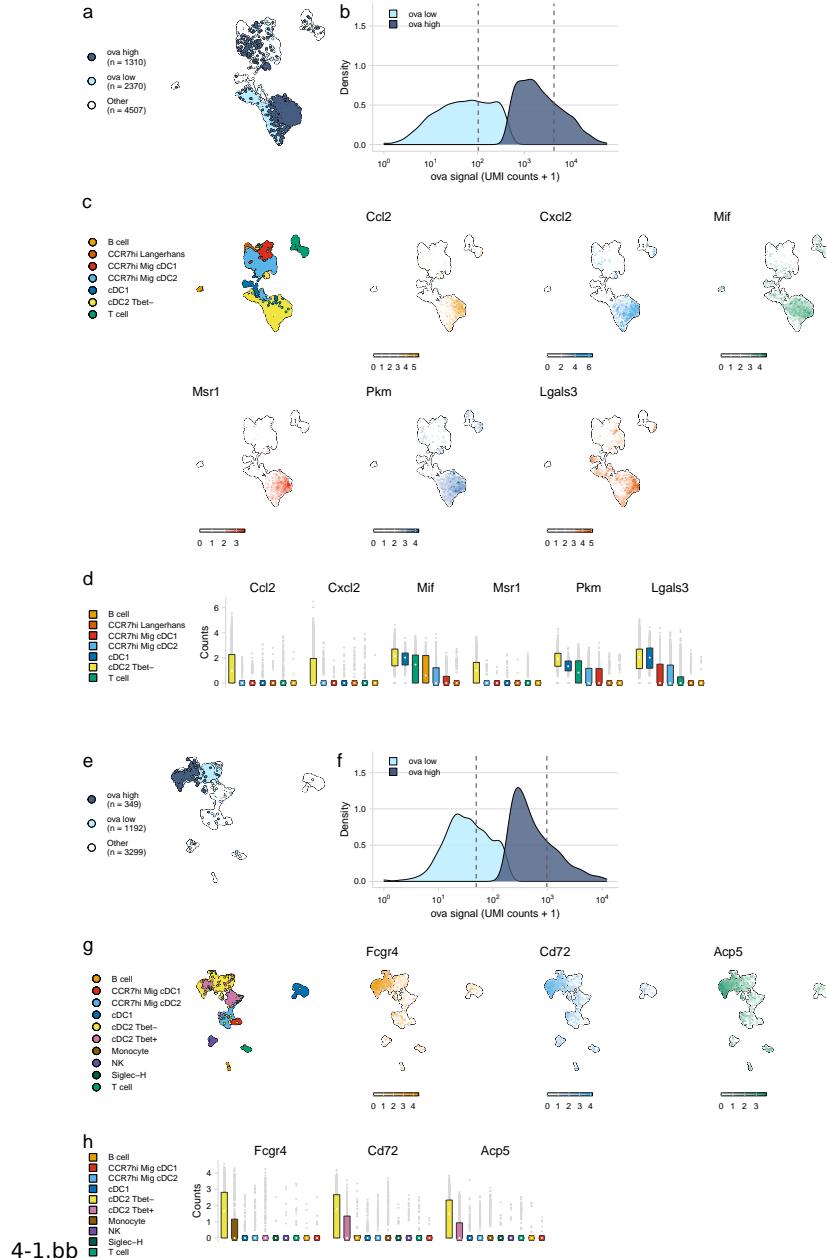


Figure 5

Genes associated with high antigen counts for day 14 LECs. (a) LECs containing low and high antigen counts were identified using a gaussian mixture model. A UMAP projection is shown for ova-low and ova-high cells. T cells, B cells, and epithelial cells are shown in white (Other). (b) The distribution of ova antigen counts is shown for ova-low and ova-high cells. Dotted lines indicate the mean counts for each population. (c) The fraction of cells belonging to each LEC cell type is shown for ova-low and ova-high populations. (d) UMAP projections show the expression of top ova-high markers. (e) The expression of ova-high markers was plotted for each cell type.

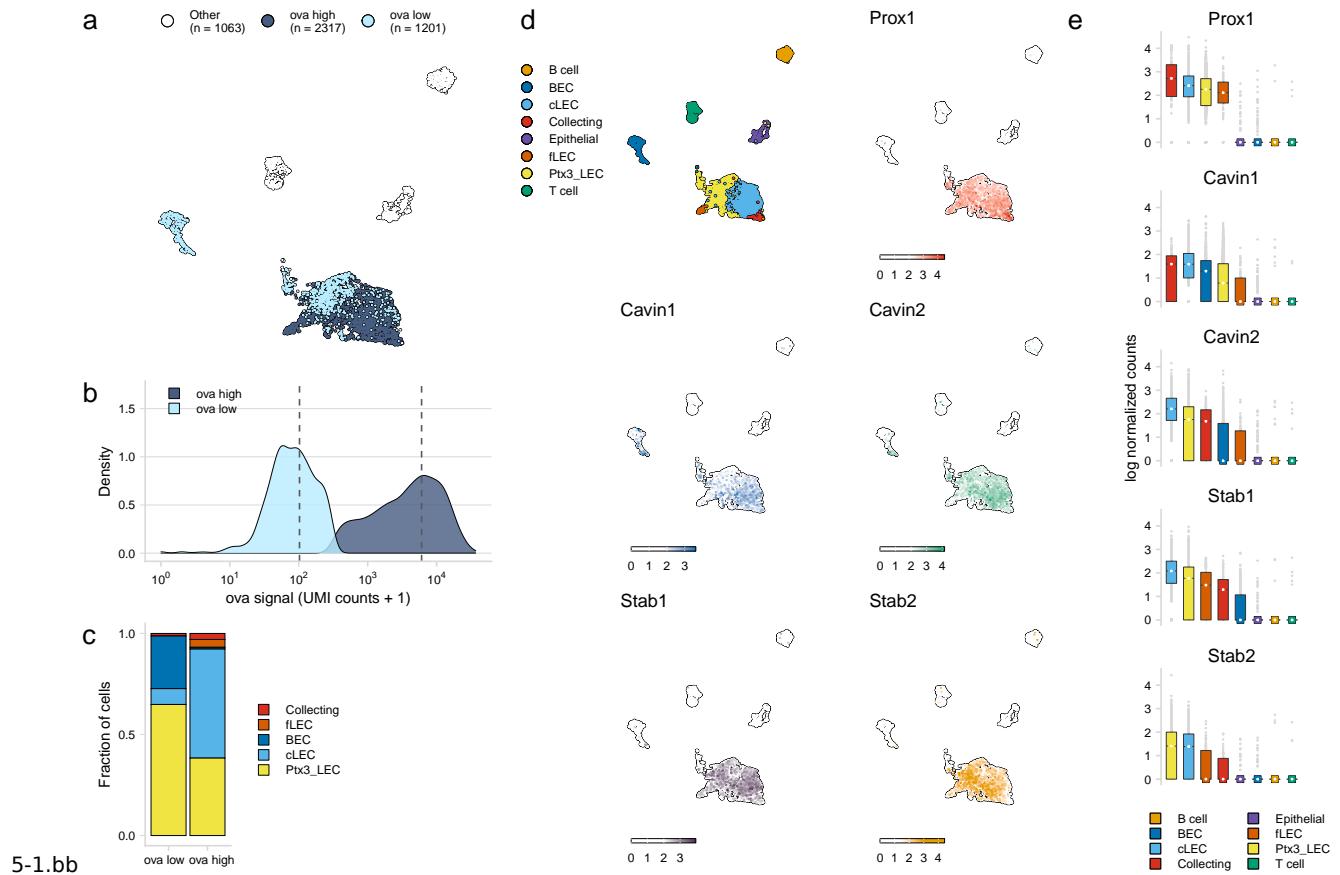


Figure S5

Comparison of ova-psDNA, psDNA, and pDNA signal for each cell type. Counts are shown for ova-psDNA, psDNA, and pDNA for DCs (a, b), LECs (c), and FRCs (d).

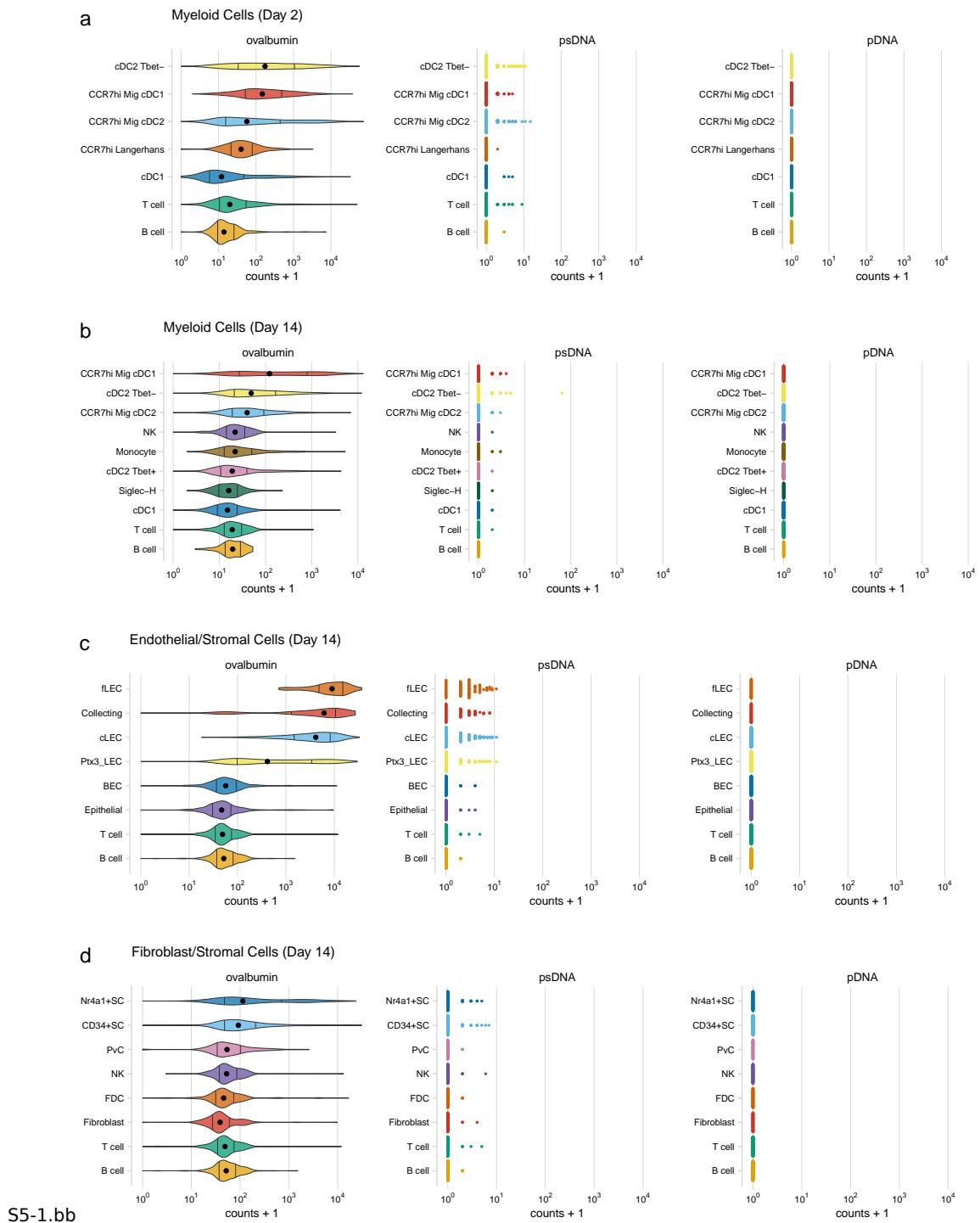
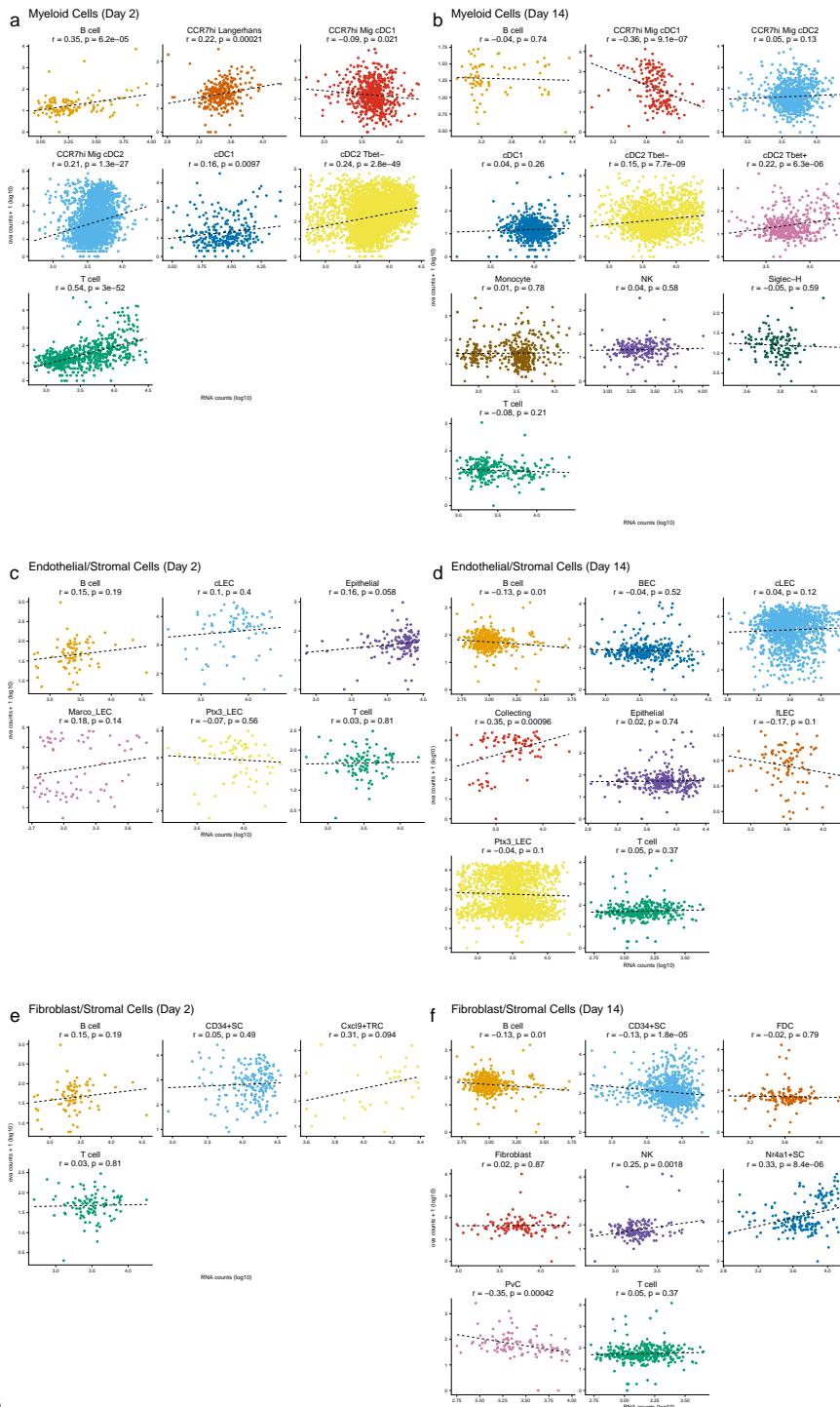


Figure S6

Antigen counts were compared with total mRNA counts for each cell for DCs (a, b), LECs (c, d), and FRCs (e, f).



S6-1.bb

Figure S7

LEC cell types associated with high antigen counts for the day 2 timepoint. (a) A UMAP projection is shown for LEC cell types, epithelial cells, B cells, and T cells identified for the day 2 timepoint. (b) Relative ova signal is shown for each cell type. Relative ova signal was calculated by dividing antigen counts for each cell by the median antigen counts for T and B cells. (c) Antigen counts are displayed on the UMAP projection shown in (a). (d, e) Correlation coefficients are shown comparing each identified LEC cell type with the reference cell types from Xiang et al.

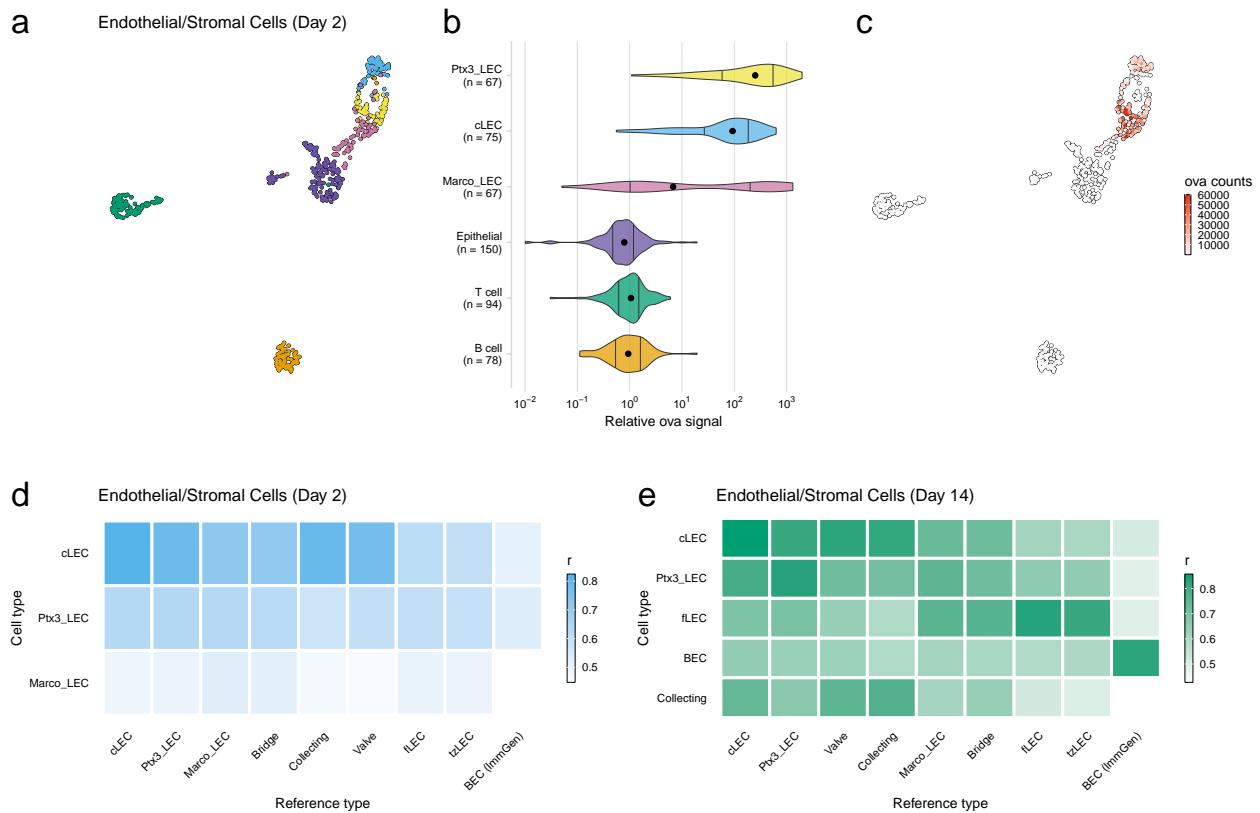


Figure S9

FRC cell types associated with high antigen counts for the day 2 timepoint. (a) A UMAP projection is shown for FRC cell types, B cells, and T cells identified for the day 2 timepoint. (b) Relative ova signal is shown for each cell type. Relative ova signal was calculated by dividing antigen counts for each cell by the median antigen counts for T and B cells. (c) Antigen counts are displayed on the UMAP projection shown in (a).

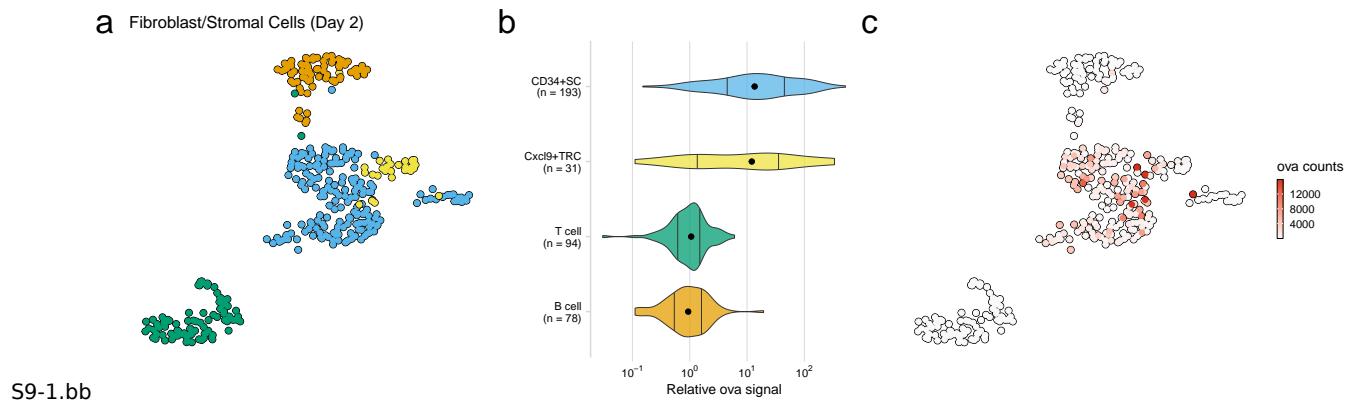


Table S2

Comparison of relative ova signal for cell types shown in Fig 3, S7b, and S9b. A two-sided Wilcoxon rank sum test was used to compare relative ova signal for each cell type. Relative ova signal was calculated by dividing antigen counts for each cell by the median antigen counts for T and B cells. The Bonferroni method was used to correct for multiple comparisons. The number of cells in each group (n cells), fraction of the total cells for the sample (frac cells), median relative ova signal, test statistic (statistic), estimation of the median difference (estimate), and confidence interval (conf.low, conf.high) are included.

Table S3

Genes associated with ova-high cells for DCs, FRCs, and LECs. Ova-high cells were independently identified for DCs, FRCs, and LECs using a gaussian mixture model implemented with the R package mixtools. Differentially expressed genes were identified using a Wilcoxon rank sum test performed using the R package presto (wilcoxauc). The Benjamini-Hochberg method was used to correct for multiple comparisons. Genes were filtered to only include those with an adjusted p-value <0.05, log fold change >0.25, auc >0.5, and with at least 50% of ova-high cells expressing the gene. The average expression, log fold change, test statistic (statistic), area under the receiver operator curve (auc), percentage of ova-high cells that express the gene (pct_in), and percentage of ova-low cells that express the gene (pct_out) are included.

Table S4

Genes associated with ova-high cells for different DC, FRC, and LEC cell types. Ova-high cells were identified independently for each cell type shown in Fig 3 using a gaussian mixture model implemented with the R package mixtools. Differentially expressed genes were identified using a Wilcoxon rank sum test performed using the R package presto (wilcoxauc). The Benjamini-Hochberg method was used to correct for multiple comparisons. Genes were filtered to only include those with an adjusted p-value <0.05, log fold change >0.25, auc >0.5, and with at least 50% of ova-high cells expressing the gene. The average expression, log fold change, test statistic (statistic), area under the receiver operator curve (auc), percentage of ova-high cells that express the gene (pct_in), and percentage of ova-low cells that express the gene (pct_out) are included.