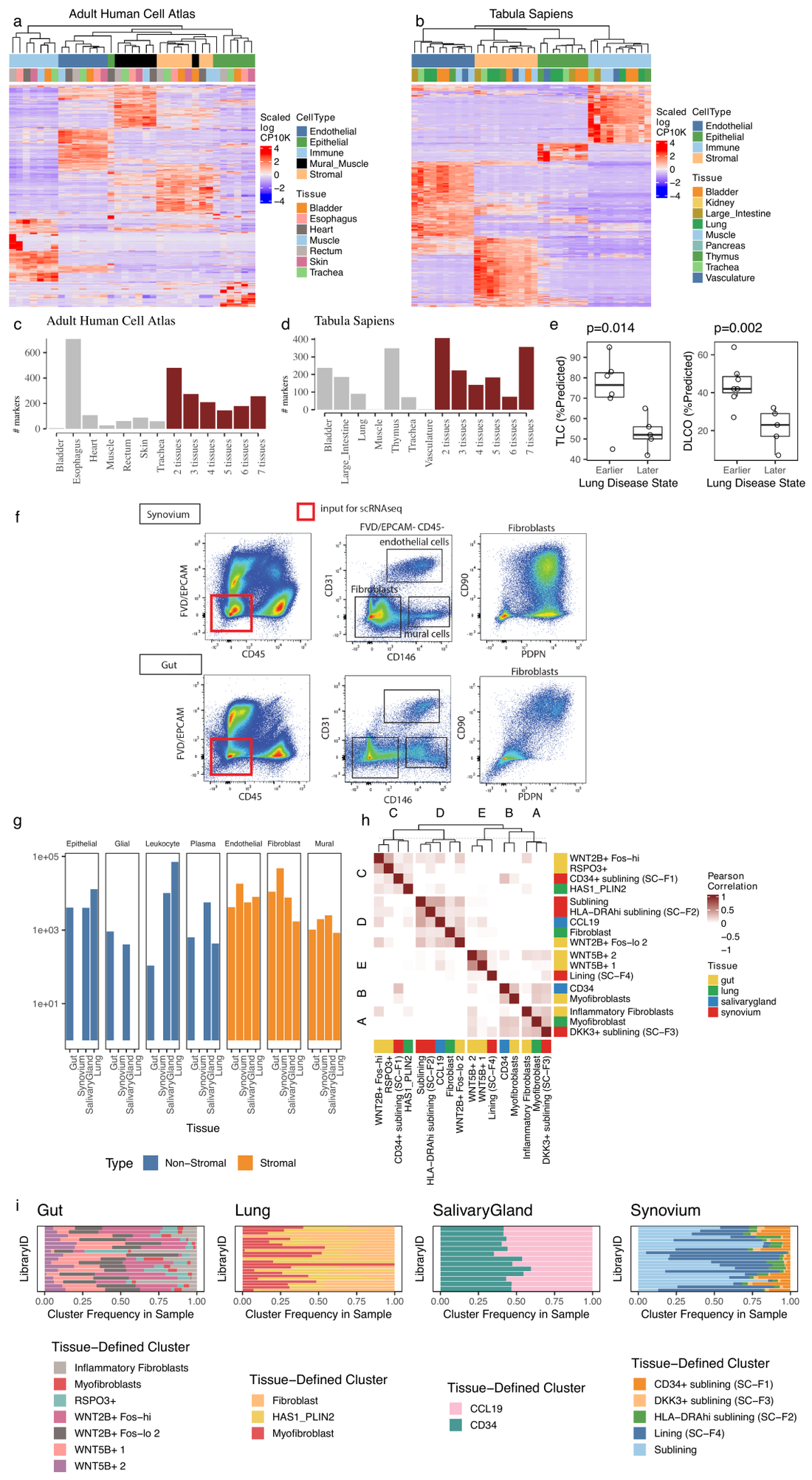

Figure S1. Cross-tissue atlas analyses and scRNAseq profiles of and fibroblast heterogeneity within intestine, lung, salivary gland, and synovium; related to Figure 1. We analyzed gene expression profiles of cells grouped into major lineages in the (a) AHCA and (b) TS datasets. Columns and rows order determined with hierarchical clustering. Columns colored by tissue and major lineage. Heatmap color denotes relative expression (scaled logCP10K). Breakdown of fibroblast genes from (c) AHCA and (d) TS datasets into tissue-specific markers (in grey) and tissue-shared markers (in red). (e) Patients with earlier-stage interstitial lung disease (ILD) had greater total lung capacity (TLC) and greater diffusing capacity for carbon monoxide (DLCO) compared to late-stage ILD. For patients with pulmonary function testing within six months prior to enrollment (n=7 of 8 earlier-stage; n=5 of 11 late-stage), TLC and DLCO are shown. P-values computed with one-tailed, unpaired t-test. (f) Flow sorting synovial and intestinal surgical samples to enrich for live (FVD-), EpCAM-CD45- stromal cells. (g) Number of stromal and non-stromal cells identified in the scRNAseq dataset in each tissue. (h) Correlation heatmap comparing relative gene expression profiles of tissue-defined clusters. Rows and columns were grouped using hierarchical clustering into five meta-cluster groups: A through E. Heatmap color denotes Pearson correlation coefficient and rows and columns colored by tissue in which cluster was defined. (i) Per-donor frequencies of tissue-defined clusters, grouped by tissue.

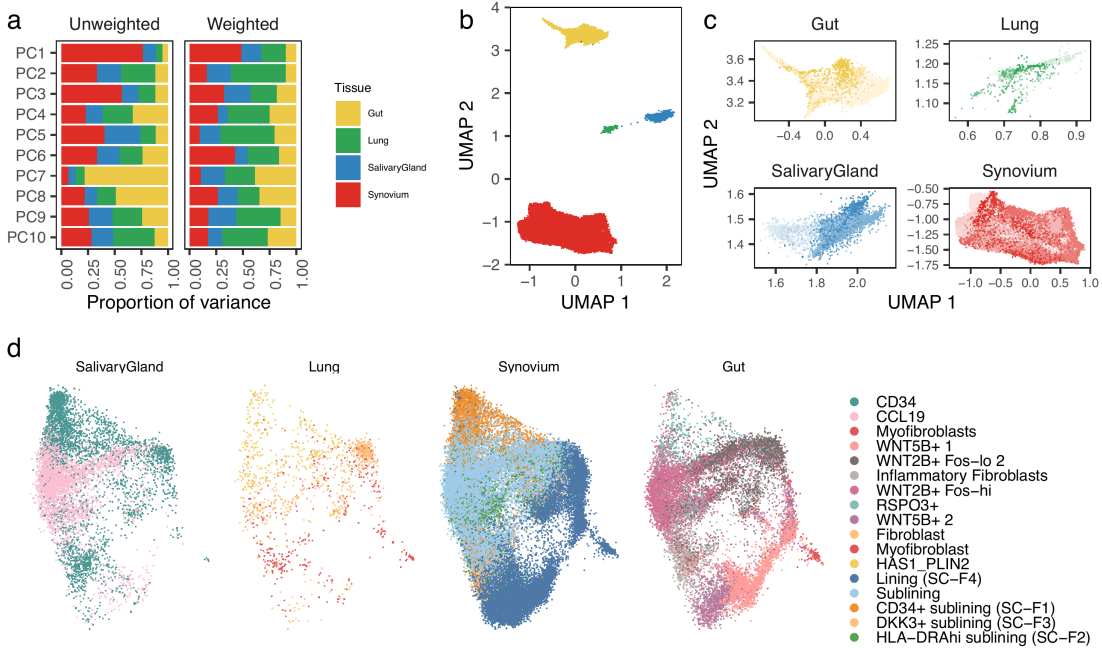

Figure S2. Integrated cross-tissue fibroblast reference atlas; related to Figure 2. (a) Breakdown of variance captured in the first 10 principle components for unweighted PCA and weighted PCA shows that weighted PCA creates a more balanced embeddings among tissues. (b) Before Harmony integration, UMAP embedding of fibroblasts separates entirely by tissue. (c) Within each tissue, there is substantial separation by donor, denoted by a different hue of the corresponding tissue's color. UMAP coordinates are the same as in (b), zoomed in to focus on each tissue separately. (d) After Harmony integration, the clusters identified in tissue-specific analyses are still separated, suggesting that the Harmony embedding preserves within tissue variation.

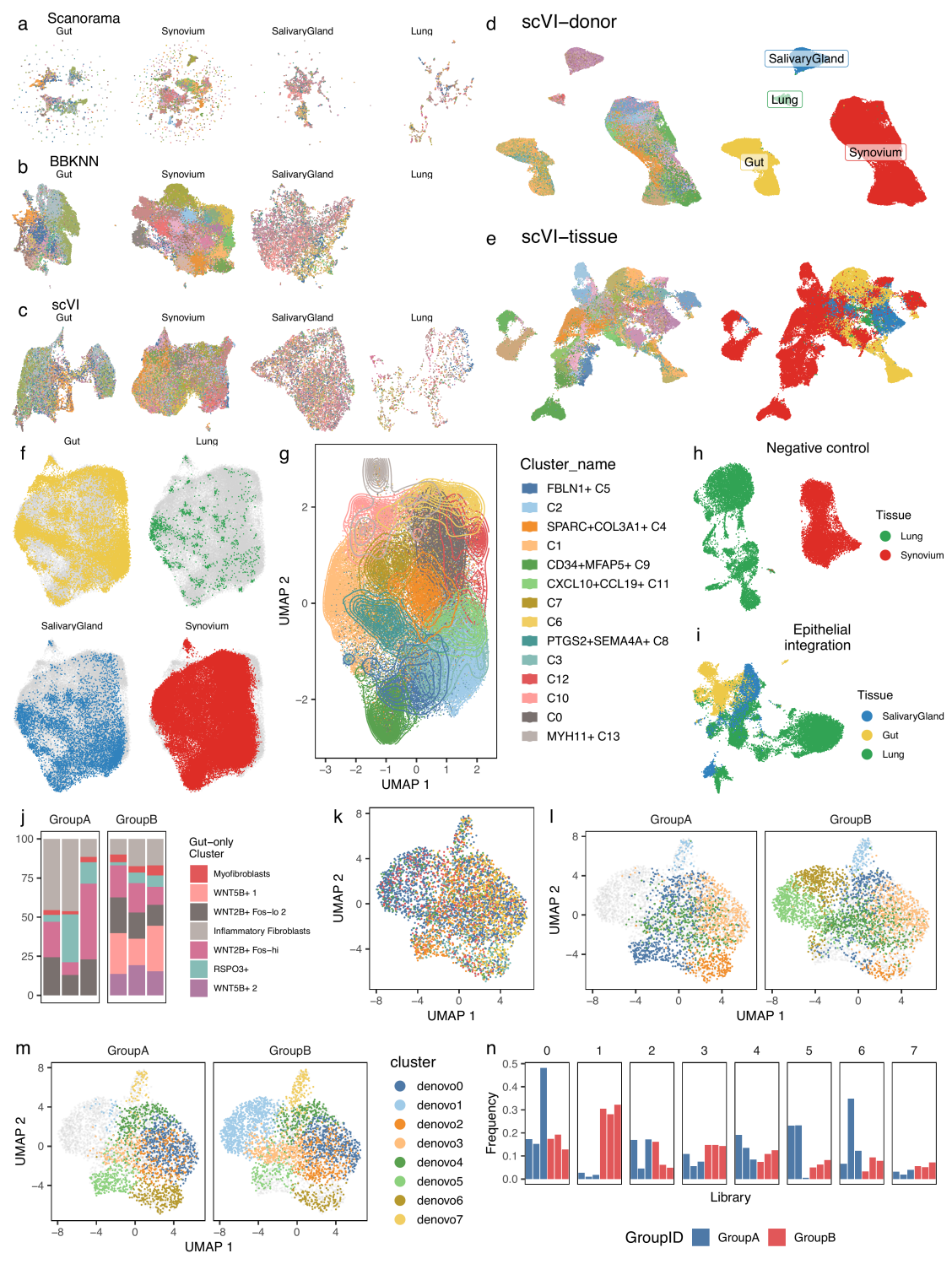

Figure S3. Testing for over-integration; related to Figure 2. Evaluation of alternative integration algorithms, (a) Scanorama, (b) BBKNN, and (c) scVI to integrate donors within each tissue. UMAP plots colored by donor identity show degree of mixing achieved by each method. Two attempts to integrate the full multi-tissue dataset with scVI, first by integrating over (d) donor identity and then by (e) tissue identity. UMAP plots colored by donor (left) and tissue (right) show degree of mixing relative to each variable. We repeated Harmony integration over donor identity alone and achieved (f) high degree of mixing among tissues and (g) separation among previously defined cross-tissue clusters. (h) Negative control experiment jointly analyzed fibroblasts from synovium with epithelial cells from lung. UMAP plot colored by tissue shows embedding of cells post Harmony integration. (i) We repeated the two-level Harmony integration pipeline on epithelial cells and plotted their UMAP projection after Harmony, colored by tissue. (j) Six donors were selected and artificially split into 2 groups. WNT5B+ fibroblasts were removed from Group A. (k) UMAP embedding of Harmony-integrated downsampled data, colored by donor. (l) Same UMAP, colored by ground-truth cell types, split by group A (left) and group B (right). (m) Downsampled data were re-clustered and re-labeled with de novo cluster labels. (n) Relative abundances of de-novo clusters in each library, colored by group.

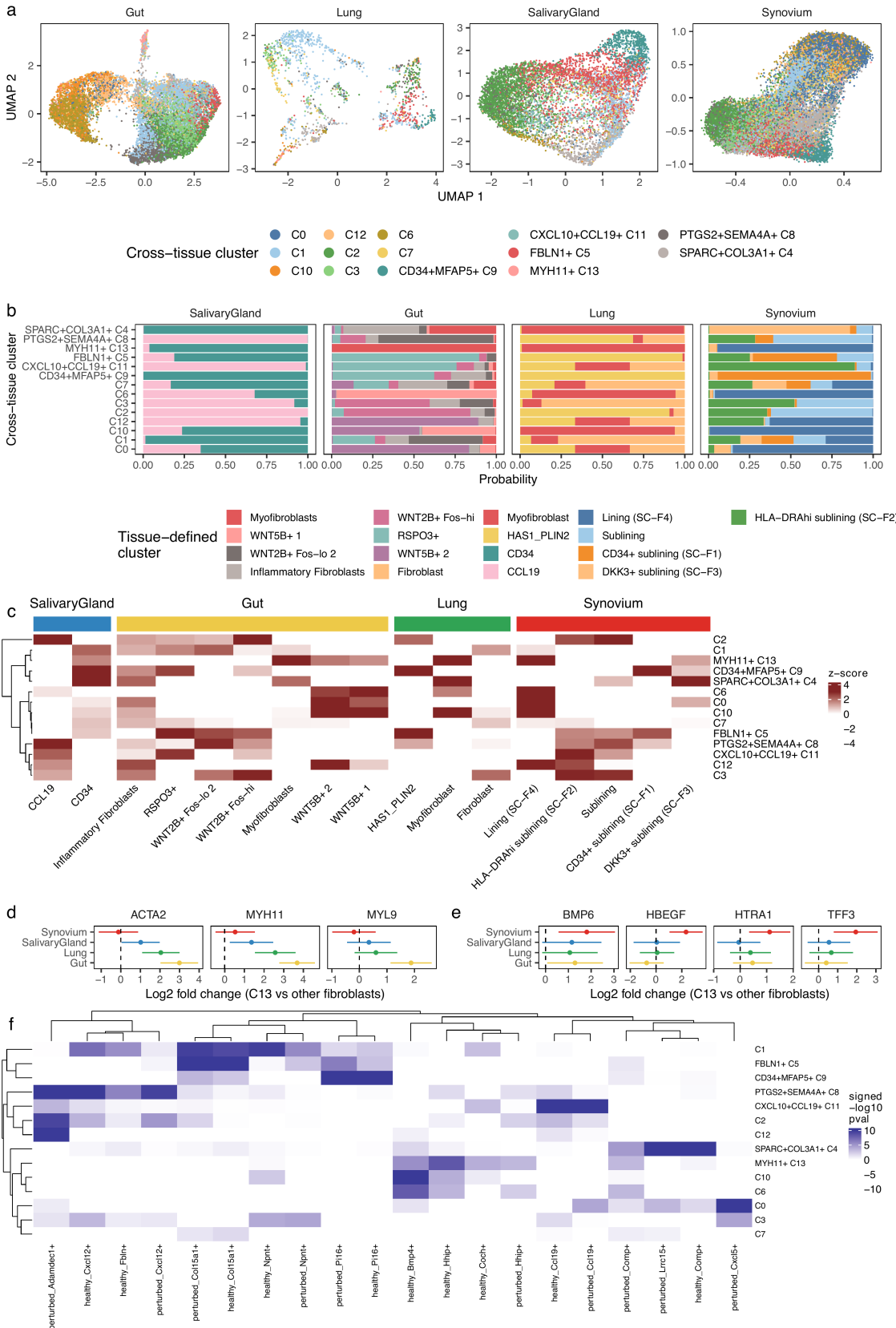

Figure S4. Correspondence analysis; related to Figures 1 and 2. (a) Cells from each tissue plotted in UMAP space from within-tissue analyses but colored by their cluster identity from cross-tissue analysis. (b) Probabilistic assignment of tissue-defined clusters to cross-tissue clusters. (c) Statistical association of tissue-defined clusters (columns) to cross-tissue clusters (rows) using logistic regression analysis. Color denotes (scaled) log odds from logistic regression. (d) Gene expression fold change of genes associated with myofibroblast lineage in cluster C13 (vs other clusters). (e) Same, for genes associated with bone and cartilage repair. (f) Enrichment of fibroblasts states from Buechler et al mouse fibroblast atlases from healthy and perturbed tissues. Each column represents a mouse cluster while the rows represent our human cross-tissue clusters. The color of the heatmap denotes degree of enrichment, as measured by the (signed) -log10 p value from gene set enrichment analysis.

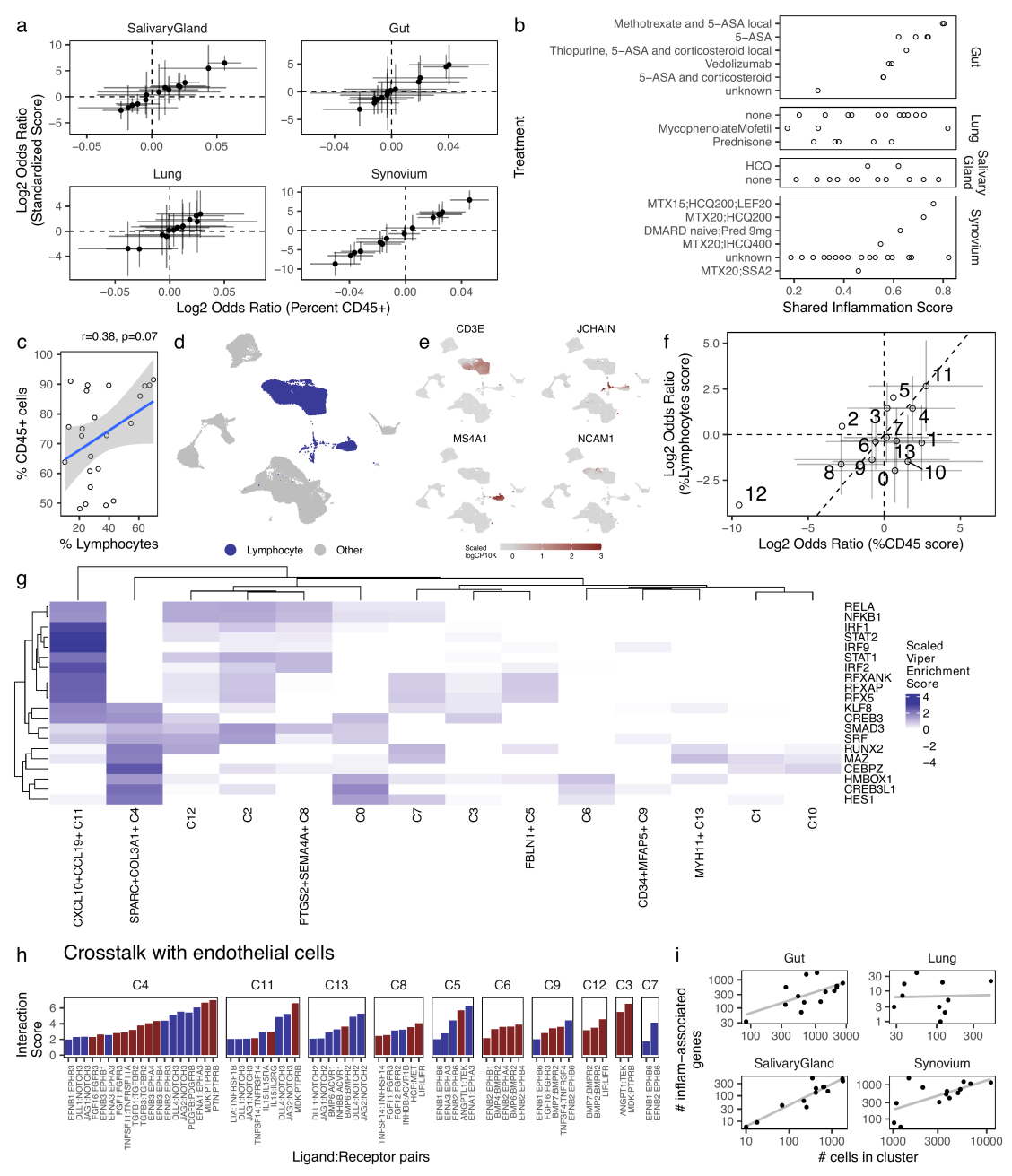

Figure S5. Enrichment of inflammation expanded clusters; related to Figure 3. (a) Comparison of differential abundance analysis using %CD45 relative abundance (x-axis) and normalized cross-tissue scores (y-axis). Error bars denote 95% confidence intervals. (b) Distribution of shared inflammation score among treatment categories for patients with clinical diagnosis (i.e. no healthy controls). (c) In lung samples only, correlation between proportion of lymphocytes (x-axis) vs proportions of CD45+ cells (y-axis). Pearson correlation r and p values shown. (d) UMAP projection of all lung cells, with lymphocytes labeled in blue, which have high expression of (e) canonical markers for lymphocyte populations: CD3E for T cells, JCHAIN for plasma cells, MS4A1 (CD20) for B cells, and NCAM1 (CD56) for NK cells. (f) Comparison of fibroblast cluster inflammation association results for lung samples only, quantified using scores based on %CD45+ cells (x-axis) and scores based on % lymphocytes (y-axis). Error bars denote 95% confidence intervals and omitted when variance was poorly estimated (for clusters C2 and C12). (g) Results of VIPER algorithm for TF enrichment. Top 10 TFs were selected for the C4 and C11 clusters and enrichment results plotted in row-scaled heatmap for all cross-tissue clusters. (h) Ligand receptor analysis of endothelial cell crosstalk with fibroblast populations. Each column is a putative ligand receptor cognate pair, faceted by fibroblast subtype. Y-axis represents the strength of the putative crosstalk, while color denotes direction of interaction: (blue) endothelial ligand to fibroblast receptor or (red) fibroblast ligand to endothelial receptor. (i) Summary of within-cluster analyses to associated gene expression with inflammation score. Each panel represents tissue-specific results relating (y) the number of nominally ($p<0.05$) inflammation-associated genes inside cluster to (x) total number of cells inside cluster. Both x and y axes are represented in log-scale. Correlation between log(x) and log(x) represented by grey linear regression line.

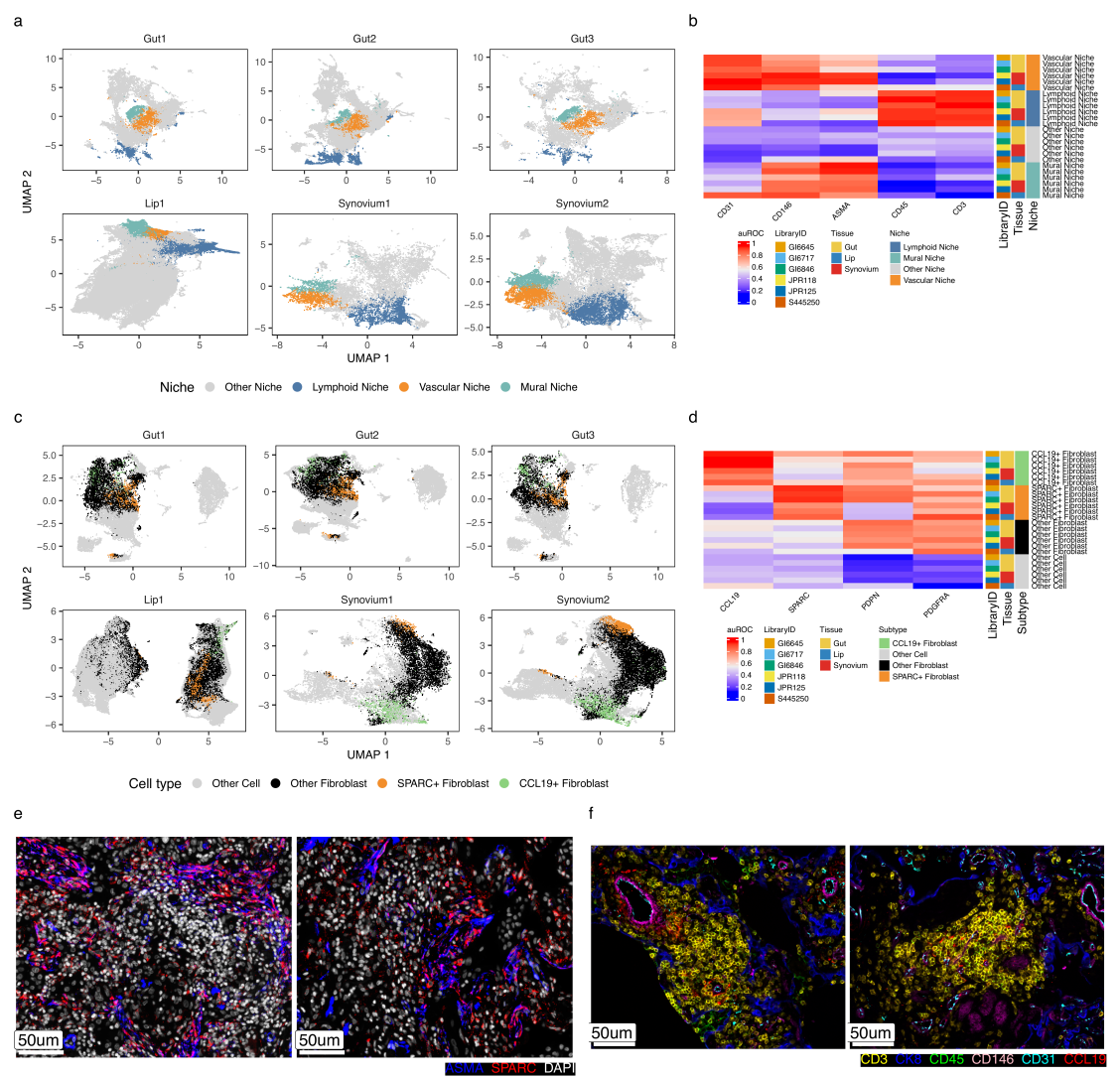

Figure S6; related to Figure 4. Co-localization of inflammation-expanded fibroblast phenotypes in vascular and lymphoid niches. (a) UMAP plots of spatial neighborhoods from each tissue sample. Each spatial neighborhood is defined by taking the mean intensity around each cell's local neighborhood. Neighborhoods are colored by their assigned anatomical niche. (b) Differential expression of relevant marker intensities for each anatomical niche. Color denote the auROC statistic. Each row denotes one niche from one tissue sample. (c) UMAP plots of individual cells from each tissue sample. Cells are colored by broad cell type (fibroblast vs other) and by type of fibroblast (CCL19+, SPARC+, or other). (d) Differential expression of relevant marker intensities for each cell type. Color denote the auROC statistic. Each row denotes one cell type from one tissue sample. (e) Manually selected regions of interest from lung immunofluorescence images highlight colocalization of ASMA and SPARC and (f) localization of CCL19 near CD3 but not on cells with CK8, CD45, CD146, or CD31.

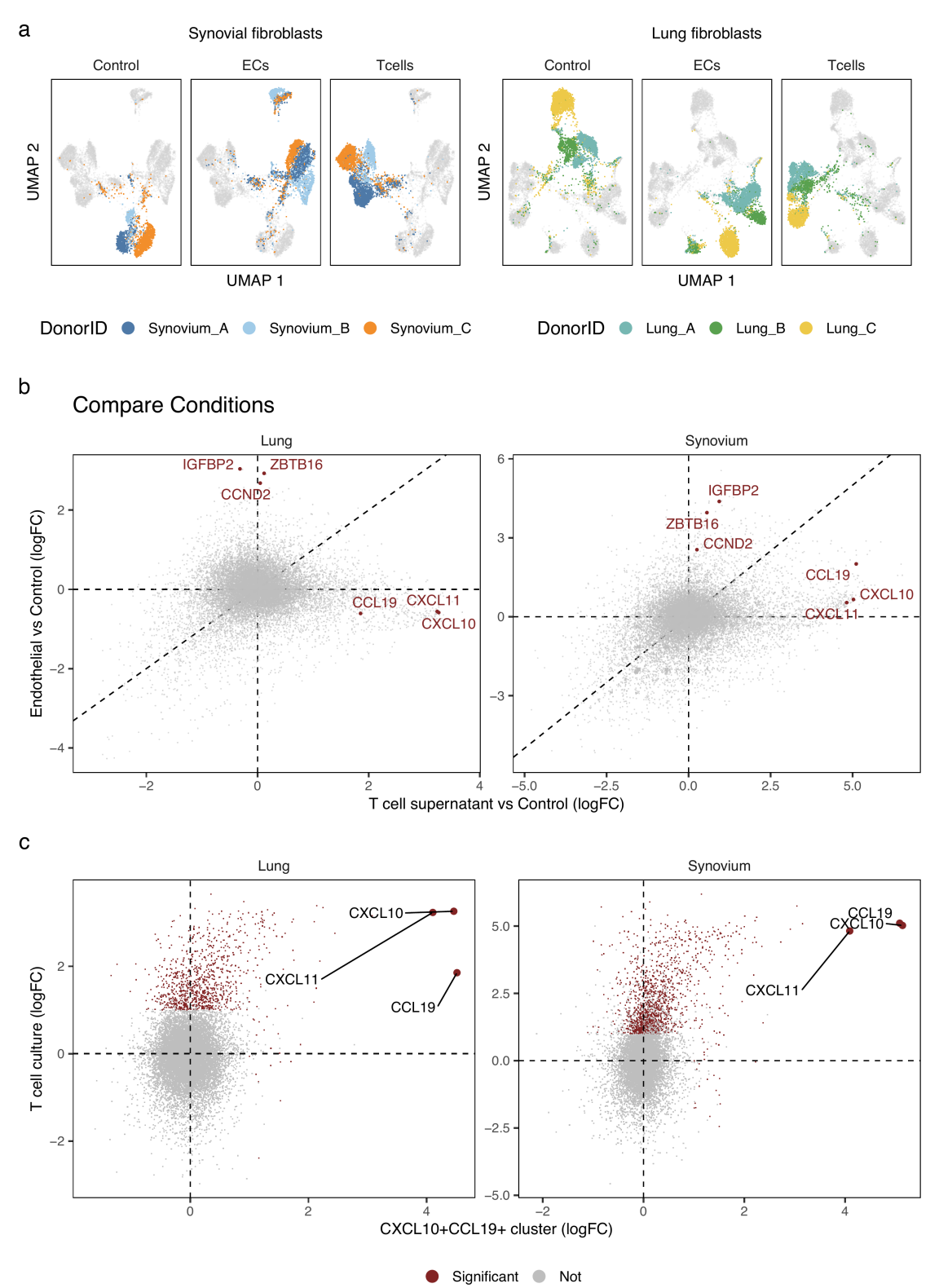

Figure S7. Convergence of fibroblasts from distinct tissues with in vitro activation; related to Figure 5. (a) UMAP analyses was performed separately, first for the synovial fibroblasts and then for the lung fibroblasts. Plots color cells by distinct cell line and facet by experimental culture condition. (b) Within each tissue, we compared the log-fold-change (vs control) of fibroblasts with T cell derived signals (x-axis) to that of fibroblasts with endothelial cell derived signals (y-axis). Representative genes from Figure 6 plotted here for comparison. (c) Zoom in on the correlation of CXCL10+CCL19+ (C11) cluster markers (x-axis) with the T cell activation signature (y-axis). Genes significantly ($p<0.01$,log_2 F-C>1) upregulated in either axis are colored red.

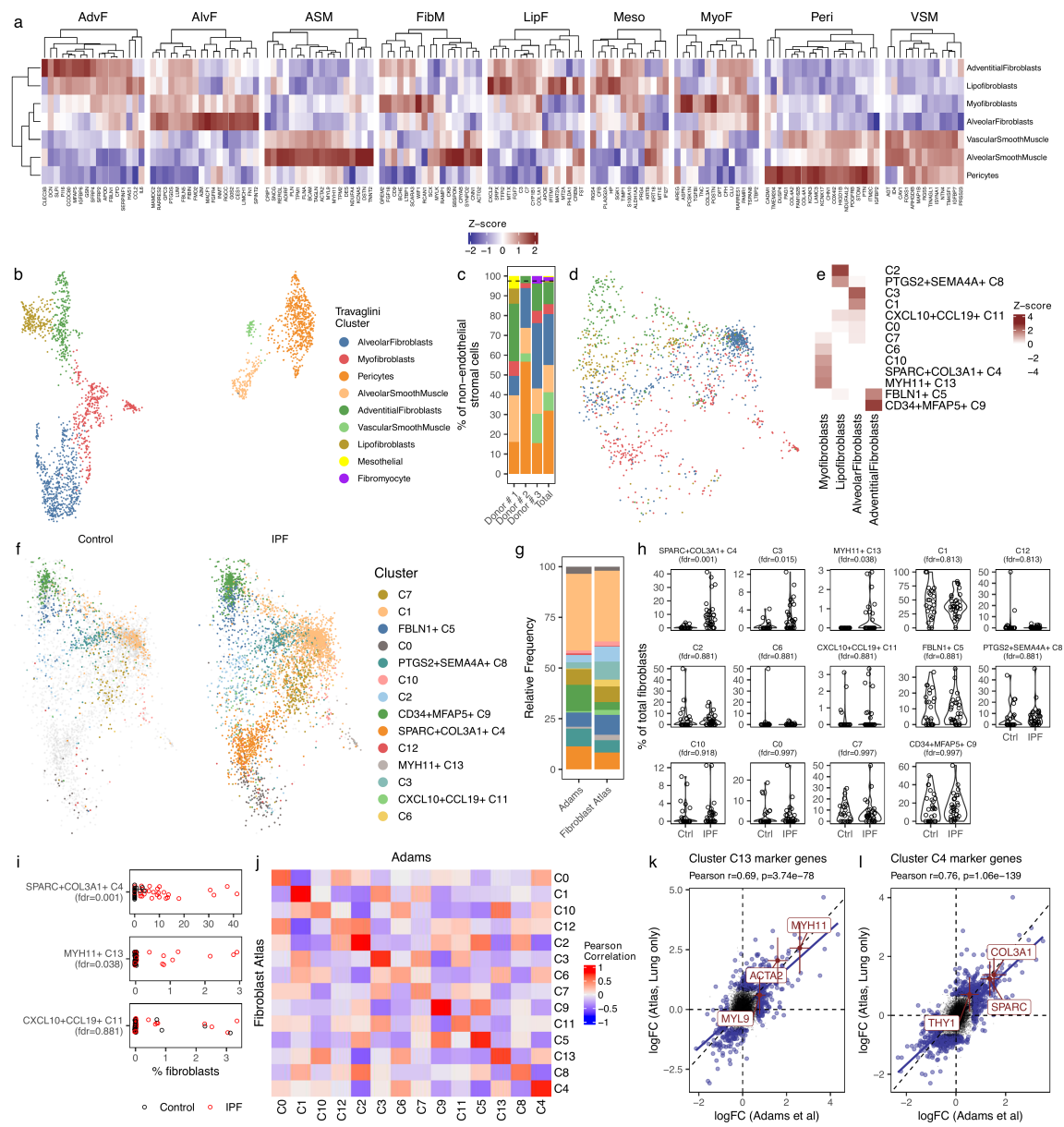

Figure S8. Validation of lung clusters with independent datasets; related to Figure 2. (a) Heatmap of canonical marker signatures defined in Travaglini et al. Each column block denotes top 20 genes per cluster, filtered for logFC≥0.5. Rows denote mean expression of cluster in the single-cell lung dataset from our study. Color is scaled logFC statistic (one-vs-all clusters). (b) We performed clustering analysis of our lung stromal cells to identify 7 non-endothelial stromal clusters defined in atlas of healthy human lung by Travaglini et al. UMAP projection of these cells is colored by these 7 clusters. (c) Relative frequencies of fibroblasts in Travaglini et al. dataset. First row denotes overall frequencies, with dotted line drawn at 97.5%. Bottom three rows denote frequency breakdowns per donor in dataset. (d) We plotted these labels in our cross-tissue fibroblast UMAP to see how they map onto our cross-tissue clusters. Color scheme shared with panel (b). (e) Confusion matrix between Travaglini clusters (x-axis) and cross-tissue clusters (y-axis), colored by number of co-occurring cells scaled by rows then columns. (f) Fibroblasts from study of healthy and IPF lung tissue projected into our cross-tissue atlas, colored by inferred cross-tissue cluster identity. Cells from healthy and IPF donors displayed separately to visualize areas of differential abundance. (g) Overall frequencies of cross-tissue clusters in our lung cohort and in Adams et al. lung cohort. (h) Differential abundance of cross-tissue clusters in Adams et al cohort. Each dot represents per-donor cluster frequency. FDR values computed with two-tailed t-test and Benjamini Hochberg adjustment. (i) Per donor frequencies of fibroblast clusters show differential abundance between IPF donors (in red) and healthy donors (in black) for C4 and C13 clusters but not for C4 cluster. FDR values computed with two-tailed t-test and Benjamini Hochberg adjustment. (j) Pairwise correlation between relative gene expression profiles in our cohort (rows) and Adams cohort (columns). Correlation computed with Pearson statistic and used genes differentially expressed (|log_2FC |≥ 1,p<0.01) in at least one cohort. Correlation of relative (one-vs-all) cluster marker expression in (k) C13 cluster and (l) C4 cluster. Differentially expressed genes (|log_2FC |≥1,p<0.01) shown in blue. Blue regression line computed on differentially expressed genes only. In red, canonical genes plotted with point estimate (dot) and 95% confidence intervals (lines).

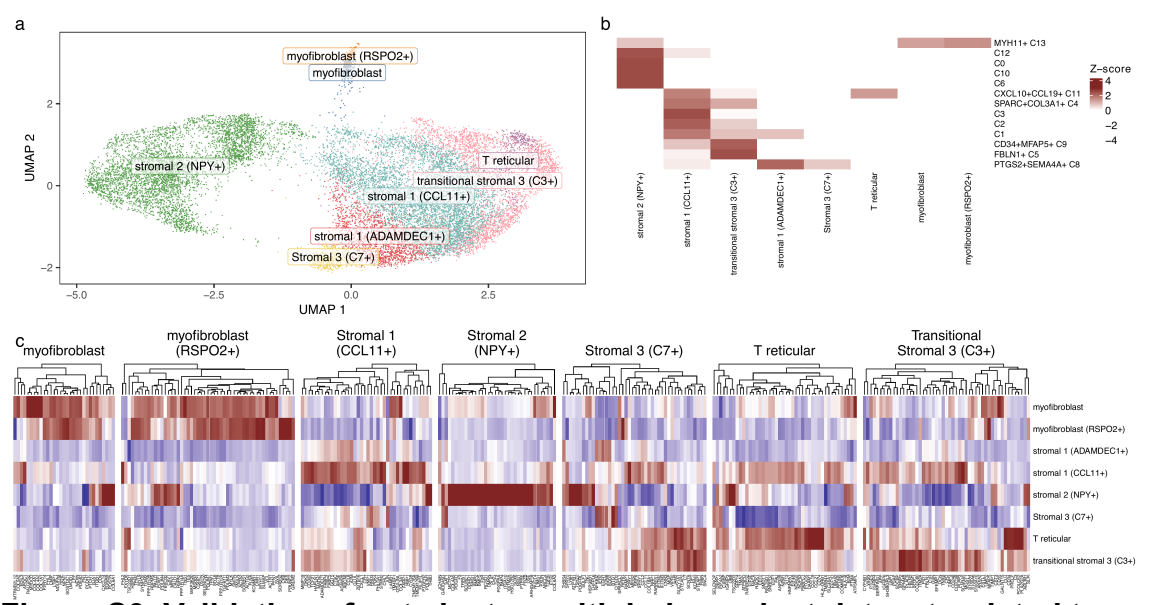

Figure S9. Validation of gut clusters with independent dataset; related to Figure 2. (a) Gut fibroblasts from healthy adult gut atlas reproduce in cross-tissue atlas. Eight phenotypes from independent gut atlas identified in our gut fibroblast dataset. (b) Heatmap of canonical marker signatures defined in healthy gut atlas. Each column block denotes top 50 genes per cluster, filtered for auROC≥0.6. Rows denote mean expression of cluster in the single-cell lung dataset from our study. Color is scaled auROC statistic (one-vs-all clusters). (c) Correspondence between atlas phenotypes and cross-tissue integrative cluster labels. Color reflects z-score of co-occurrence of two labels.

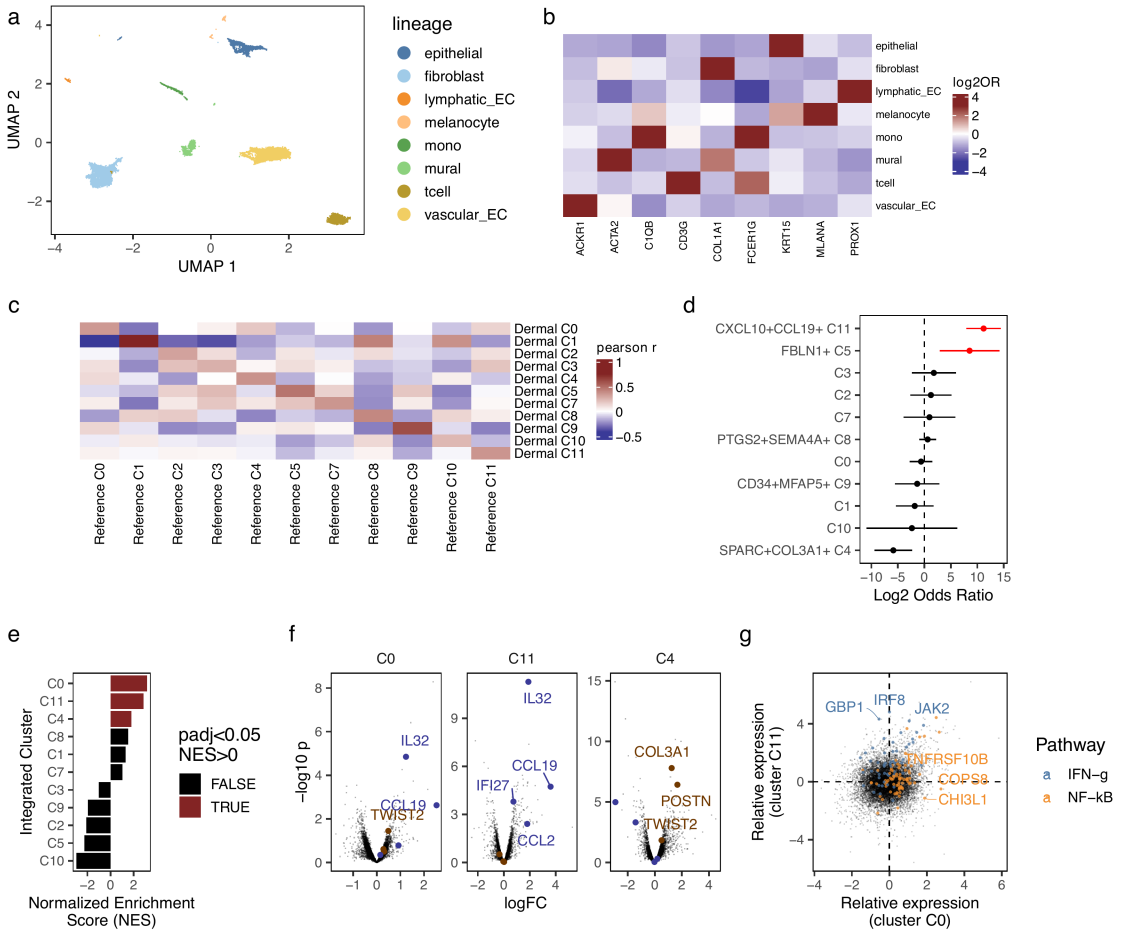

Figure S10. Dermal fibroblast scRNAseq profiles mapped to cross-tissue fibroblast atlas; related to Figure 6. (a) UMAP embedding of scRNAseq profiles of skin biopsies, colored by major cell types, using (b) canonical markers: KRT15+ epithelial cells, COL1A1+ fibroblasts, PROX1+ lymphatic endothelial cells, MLANA+ melanocytes, C1QB+ myeloid cells, ACTA2+ mural cells, CD3G+ T cells, and ACKR1+ vascular endothelial cells. (c) Correlation of gene expression profiles of dermal fibroblast clusters (y-axis) against reference clusters in multi-tissue atlas (x-axis). Color denotes Pearson's correlation coefficient. (d) Differential abundance of mapped dermal fibroblast clusters with inflammation score, with 95% confidence intervals. Red denotes FDR<5%. (e) Enrichment of published COL6A5+COL18A1+ dermal fibroblast signature in our cross-tissue reference clusters. Color denotes one-tailed statistical significance of enrichment. (f) Volcano plots of selected leading edge genes responsible for enrichment in cross-tissue clusters C0, C11, and C4. Log fold change (x-axis) and p value (y-axis) is for cluster marker association in skin fibroblasts. (g) Comparison of cluster markers for C0 and C11 clusters in skin fibroblasts. Highlighted genes show that C0 is enriched for NF-kB pathway genes (orange), while C11 is more enriched for IFN-g pathway genes (blue).

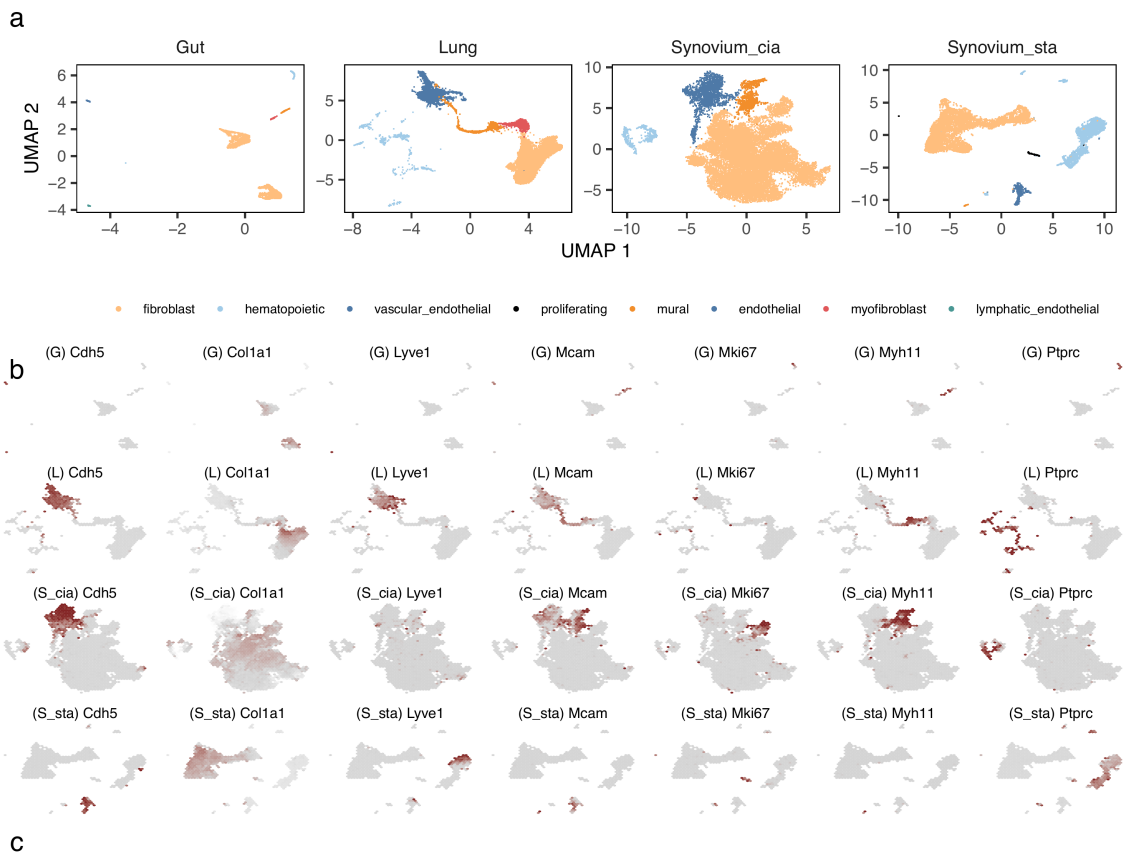

Figure S11. Replication in disease models; related to Figure 7. (a) UMAP embedding of mouse scRNAseq libraries from CD45- sorted colon samples, unsorted synovial samples from serum transfer arthritis (Synovium_sta) and collagen induced arthritis (Synovium_cia), and Col1a1+ sorted lung samples, colored by major cell types, identified with (b) canonical markers: Cdh5+ vascular endothelial cells, Col1a1+ fibroblasts, Lyve1+ lymphatic endothelial cells, Mcam+ mural cells, Myh11+ myofibroblasts, Ki67 proliferating cells, and Ptprc+ immune cells. (c) Relative abundance of inferred fibroblast clusters in each mouse dataset. (d) Comparison of mouse cluster gene expression profiles (y-axis) to human reference cluster profiles (x-axis). Heatmap color denotes Pearson's correlation coefficient. Columns and rows are colored first by cluster identity and then by tissue. (e) Differential abundance (inflamed vs non-inflamed) of cross-tissue fibroblast clusters in human data (x-axis) versus mapped fibroblast clusters in mouse samples (y-axis), with 95% confidence intervals. Red denotes FDR<20% and logFC>0.5.

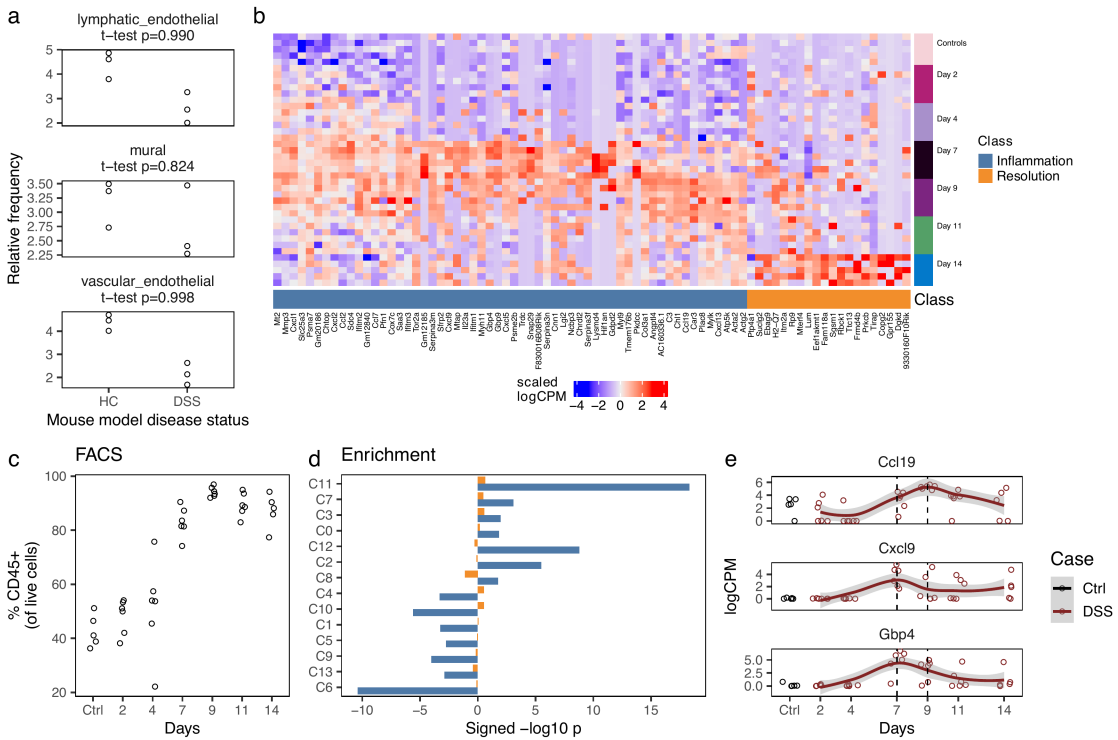

Figure S12. RNAseq DSS time course on sorted fibroblasts; related to Figure 7. (a) In the single-cell DSS dataset, DSS mice profiled at day 7 show a lack of evidence for vascular expansion. P-values derived from one-tailed t-test. (b) We found 52 genes whose expression was associated with the time course and higher than healthy controls at least one time point. Each row denotes on RNAseq profile, each column a gene that peaked either during acute inflammation (blue) or during resolution (orange), each the heatmap color denotes mean centered and scaled logCPM expression. (c) For the same samples profiled with RNAseq, proportion of CD45+ cells, as measured with flow sorting. (d) Gene set enrichment analysis results comparing single-cell cluster markers to the two gene sets defined in (b). (e) Dynamic gene expression plots for control mice and DSS treated mice for Ccl19, Cxcl9, and Gbp4, all genes upregulated in the C11 immuno-fibroblast markers and overexpressed in fibroblasts during the acute phase of inflammation in the DSS model.