Supplemental Material

Dissecting the polygenic basis of atherosclerosis using disease associated cell state signatures

Örd *et al.*

Supplementary Figures

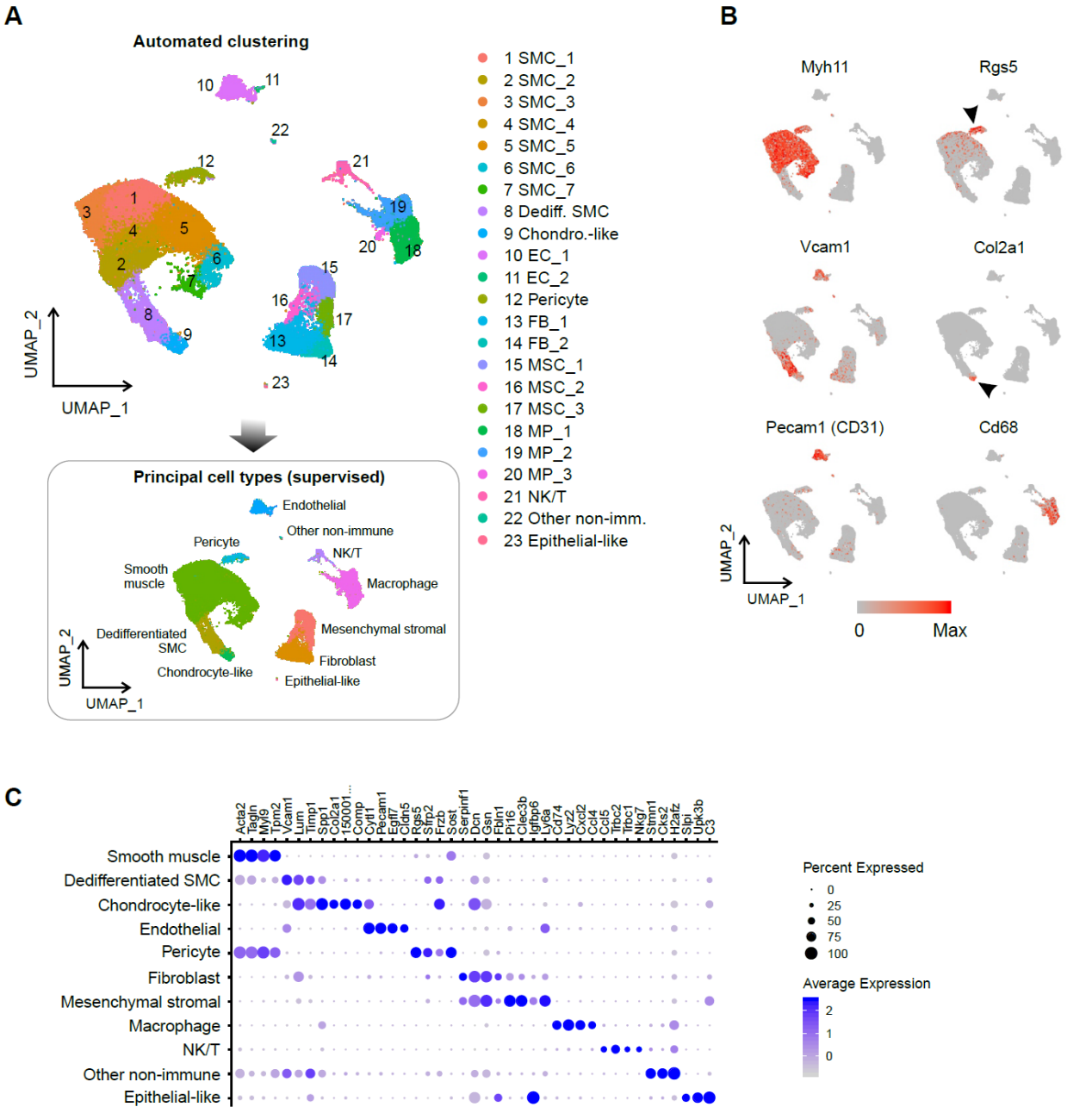
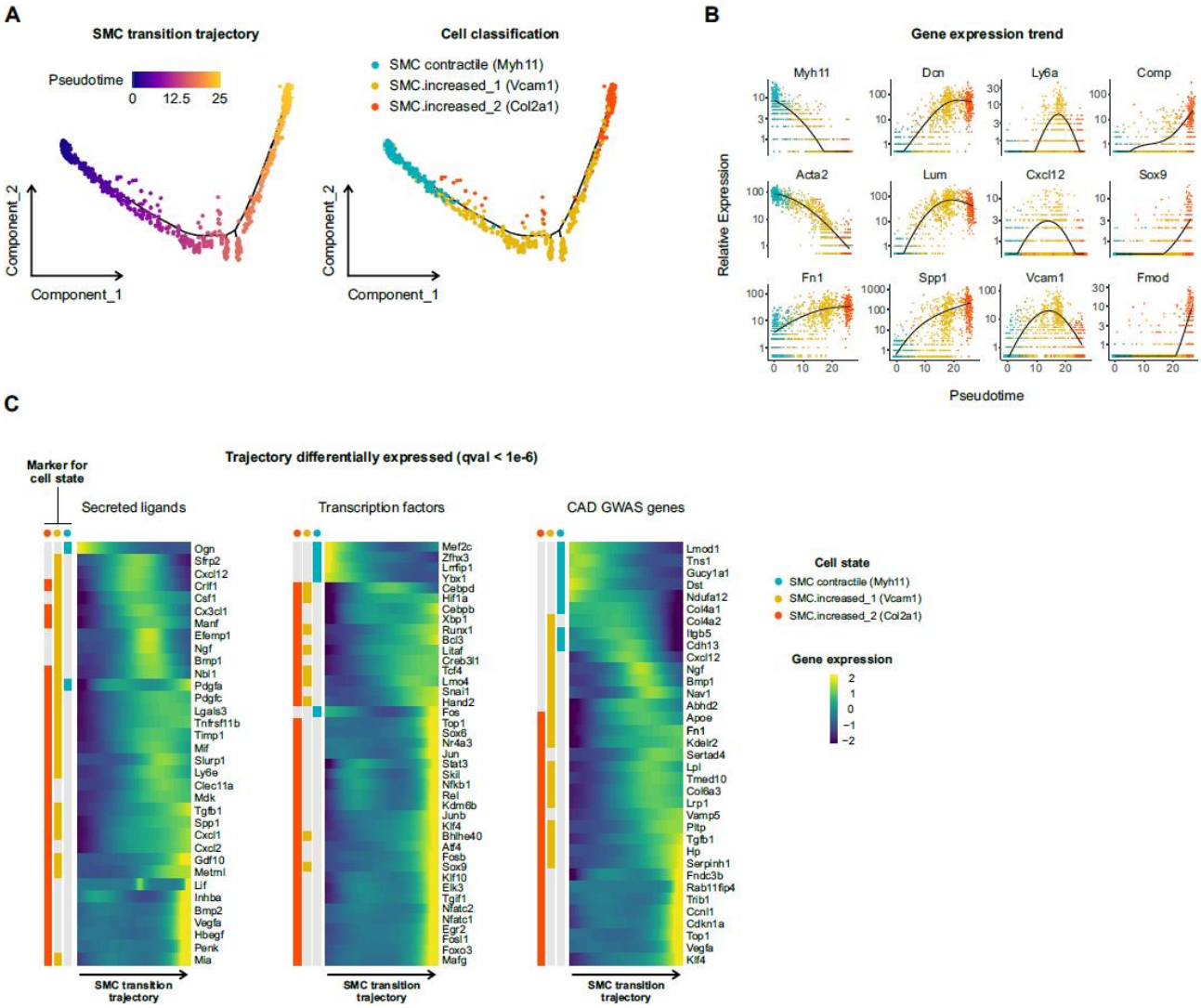
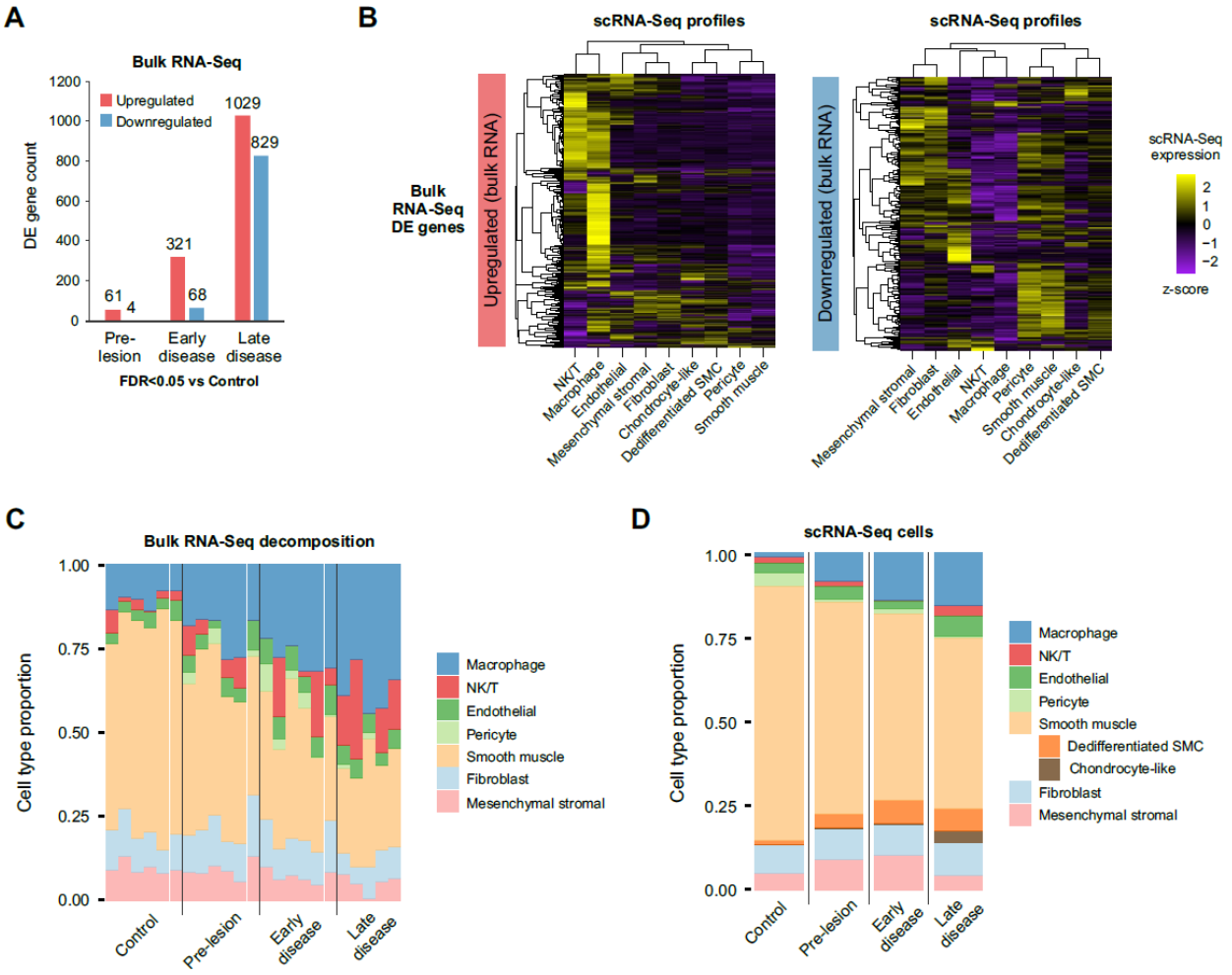


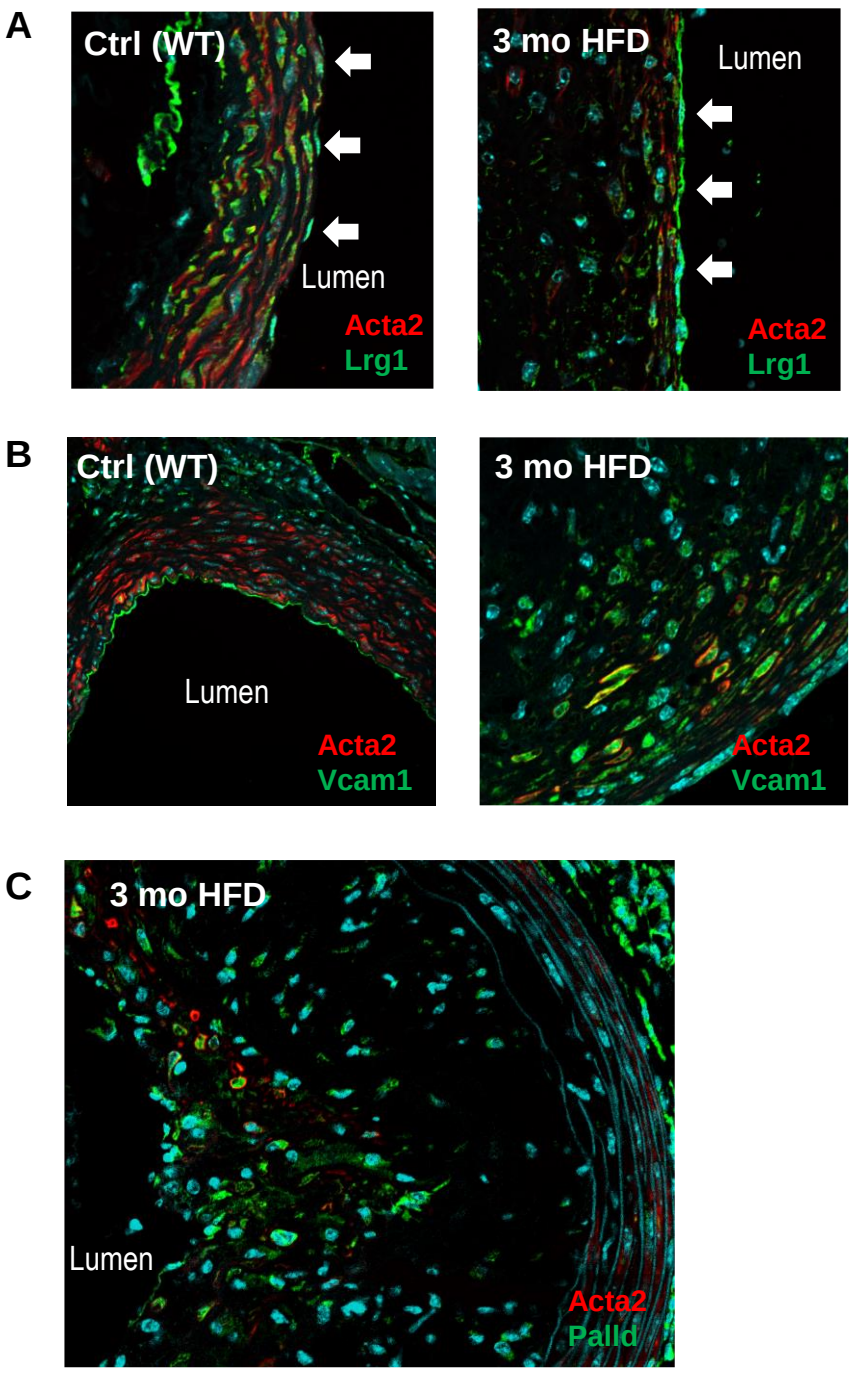
Figure 1. Clustering and identification of cell types in scRNA-Seq of mouse atherosclerotic lesion. (A) UMAP projection of the scRNA-Seq profiles represented as 23 clusters identified using automated clustering and the eleven manually annotated populations. (**B**) UMAP plots showing the expression of selected markers used to annotate the cell types. (**C**) Dot plot demonstrating the top four marker genes for each lesional cell type. Dot size corresponds to the proportion of cells within the cluster that expressed the gene, and dot color intensity corresponds to the average expression level.



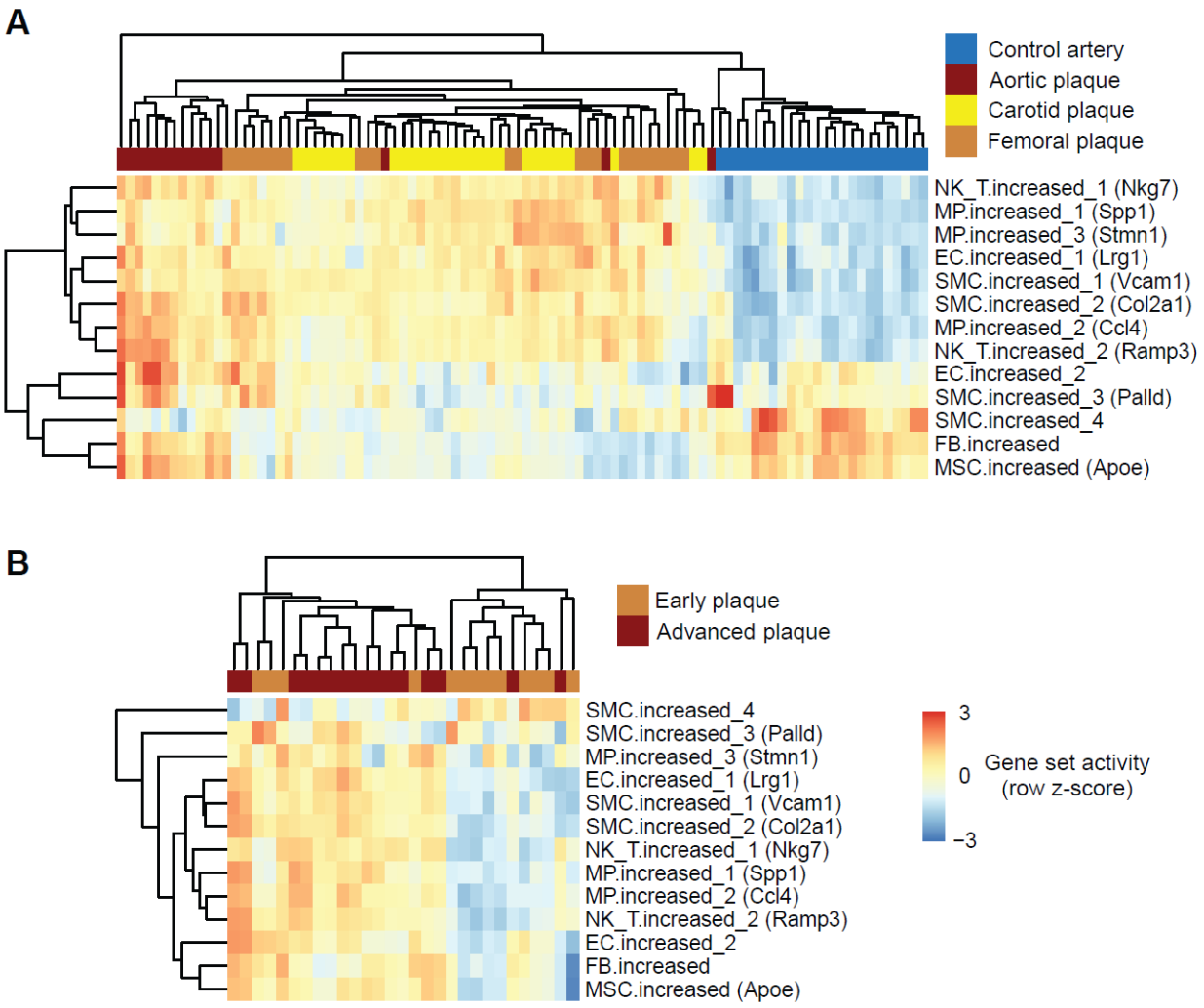
Supplementary Figure 2. Pseudotime analysis of SMC cell states from atherosclerotic aorta. (**A**) scRNA-Seq trajectory analysis of SMC states colored by pseudotime and cell state assignment. (**B**) Gene expression changes of selected marker genes along the pseudotime trajectory. Cells are colored by cell state as in panel A. (**C**) Expression changes of secreted ligands, transcription factors and CAD GWAS genes that display differential expression along the pseudotime trajectory.



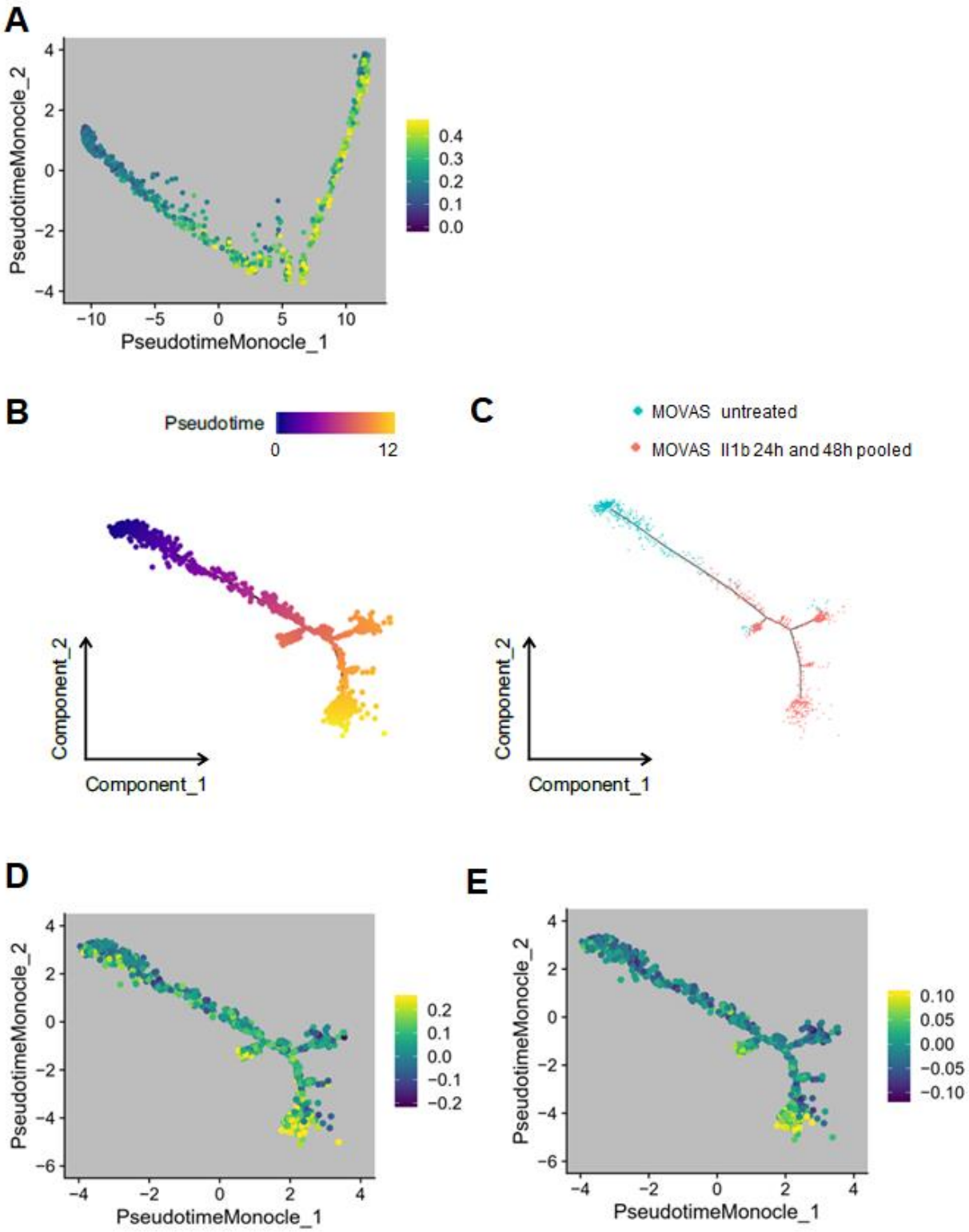
Supplementary Figure 3. Analysis of cell type proportions from bulk transcriptomics data. (**A**) Number of differentially expressed genes identified from bulk RNA-Seq analysis during different stages of disease progression (see Figure 1A). (**B**) Cell types expressing the genes detected as differentially expressed in bulk RNA-Seq comparison of disease stages. Expression levels of differentially expressed genes from bulk RNA-seq analysis (panel A) are plotted from the cell types identified in scRNA-Seq. (**C**) Computational prediction of cell type composition from the bulk RNA-Seq using the scRNA-Seq data as reference. (**D**) Cell type proportions detected in the scRNA-seq data.



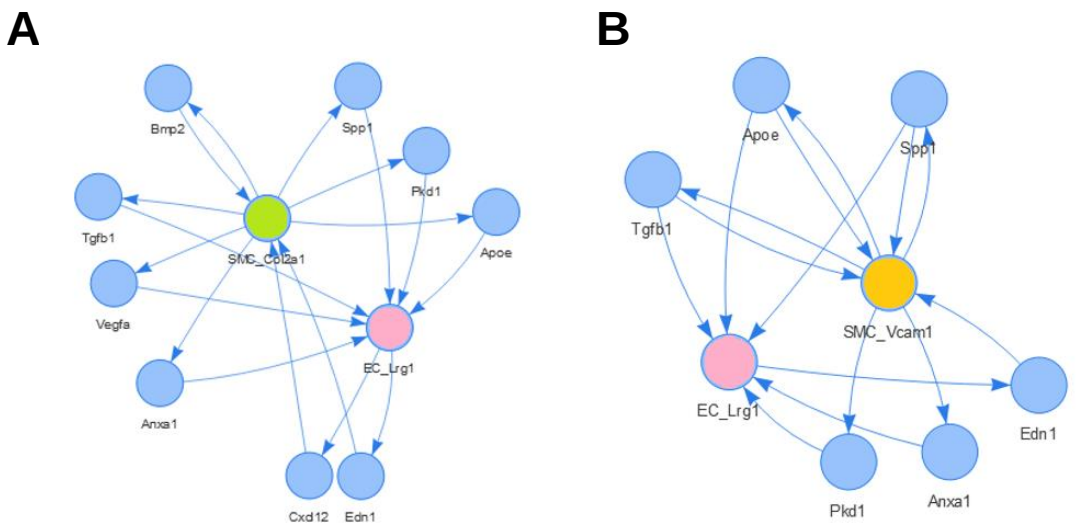
Supplementary Figure 4. Immunohistological validation of identified cell state marker genes. Representative images of (**A**) Lrg1, (**B**) Vcam1 and **(C**) Palld (all in green) staining in wild type chow diet-fed and LDLR$^{-/-}$/ApoB$^{100/100}$ 3-month high fat diet mice. The smooth muscle cells are stained with Acta2 antibody (red) and DAPI staining is shown in cyan.



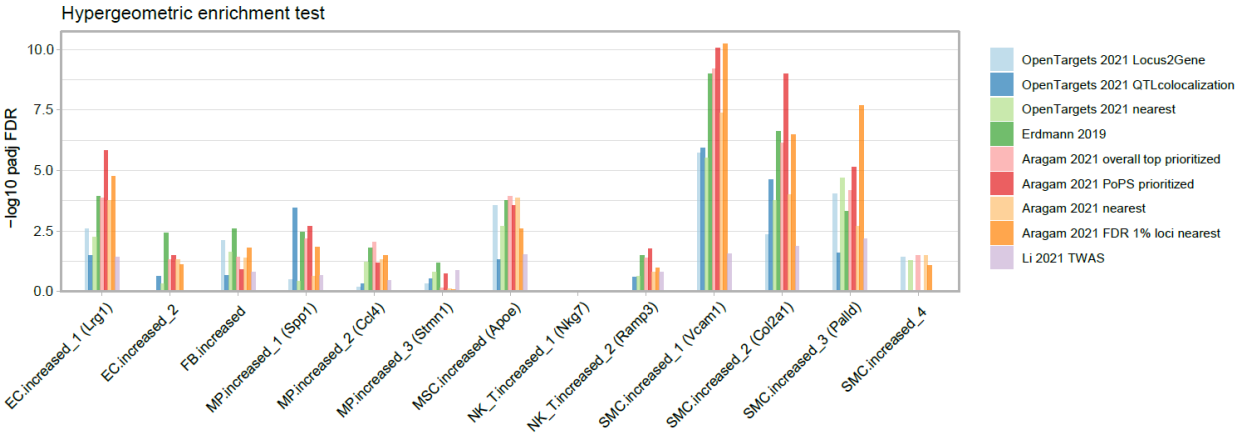
Supplementary Figure 5. Atherosclerosis associated cell state signatures are able to distinguish human atherosclerotic lesions in bulk RNA data. The gene set activity scores for each cell state investigated in the (**A**) Tampere Vascular Study[15] representing 68 advanced atherosclerotic plaques (15 aortic, 29 carotid and 24 femoral plaques) and 28 controls (left internal thoracic artery and (**B**) the Maastricht Pathology Tissue Collection[16] representing atherosclerotic carotid artery segments from 13 early intimal thickening/xanthoma lesions and from 16 advanced fibrous cap atheroma lesions.



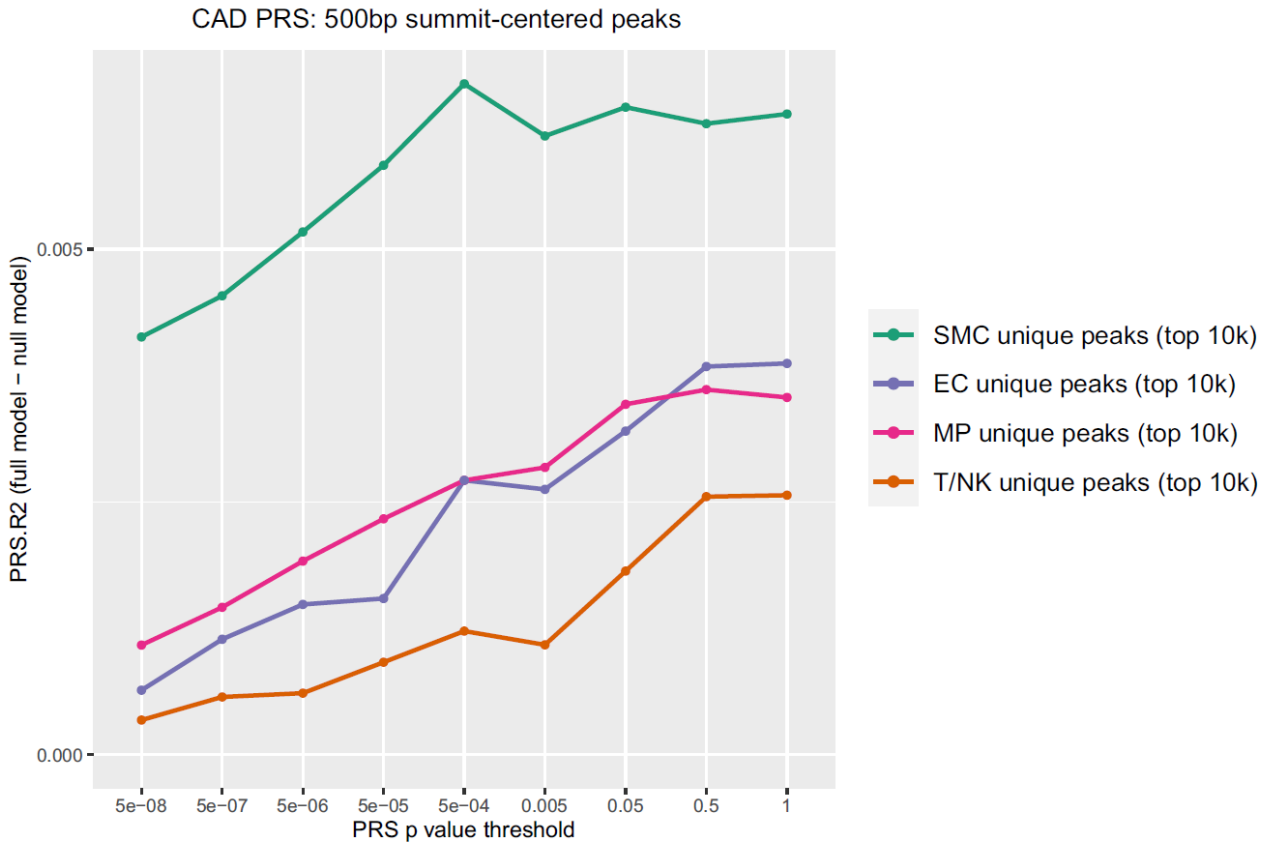
Supplementary Figure 6. Projection of IL-1β responsive genes on *in vivo* SMC disease response pseudotime trajectory. (A) Differentially expressed genes identified using scRNA-seq of *in vitro* SMCs (MOVAS cells) upon 24 and 48 h IL-1β treatment plotted on the *in vivo* SMC trajectory (see Supplementary Figure 2A) as a gene set activity score. scRNA-Seq trajectory analysis of *in vitro* SMCs under basal and IL-1β treatment conditions colored by (**B**) pseudotime and (**C**) treatment. (**D**) Vcam1+ SMC and (**E**) Col2a1+ SMC marker gene sets plotted as activity scores on the *in vitro* SMC IL-1β response trajectory (described in panels B-C).



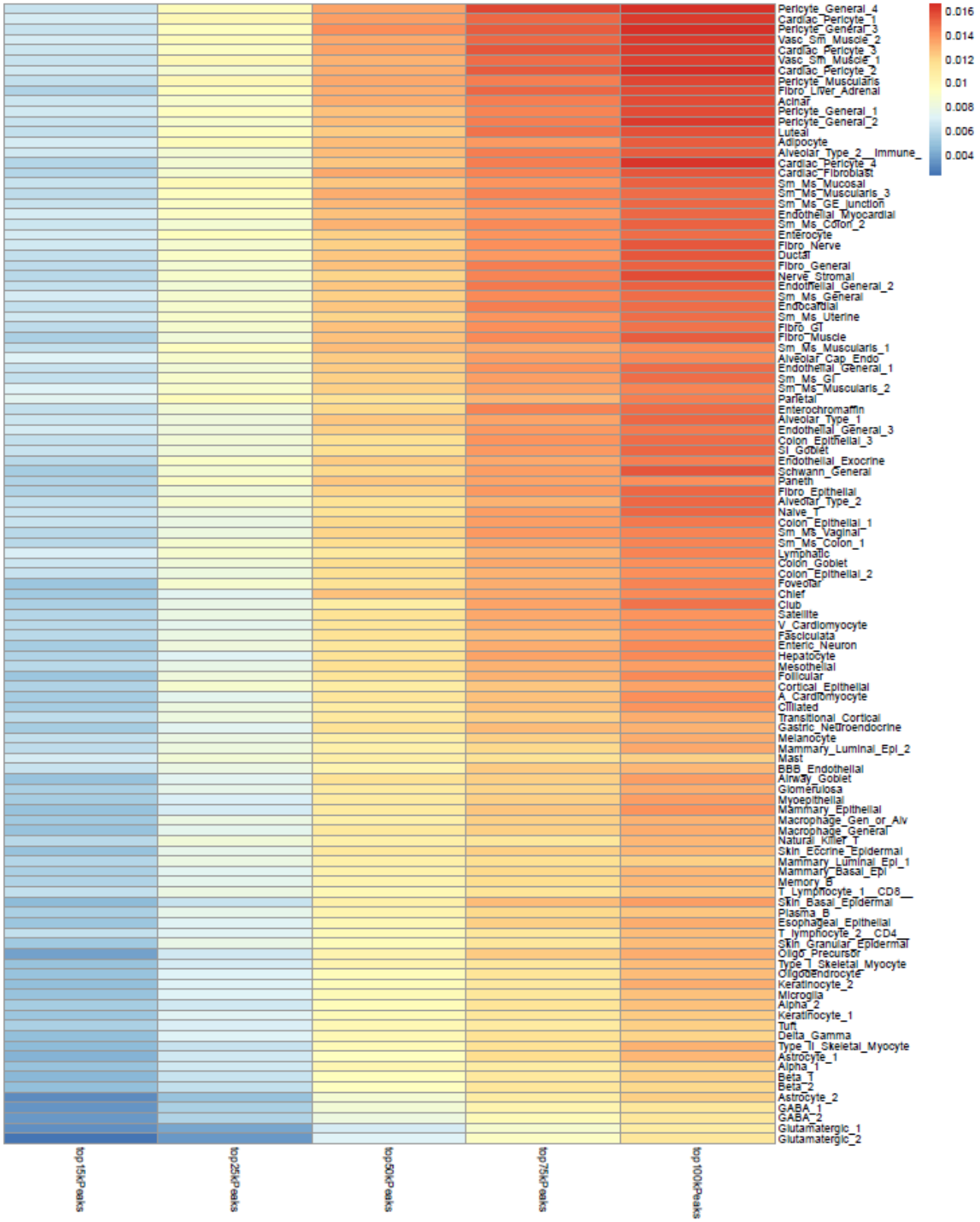
Supplementary Figure 7. Prediction of cell-cell signaling networks. Ligands predicted to mediate paracrine and autocrine signaling between Lrg1+ ECs and either (**A**) Col2a1+ SMCs or (**B**) Vcam1+ SMCs.



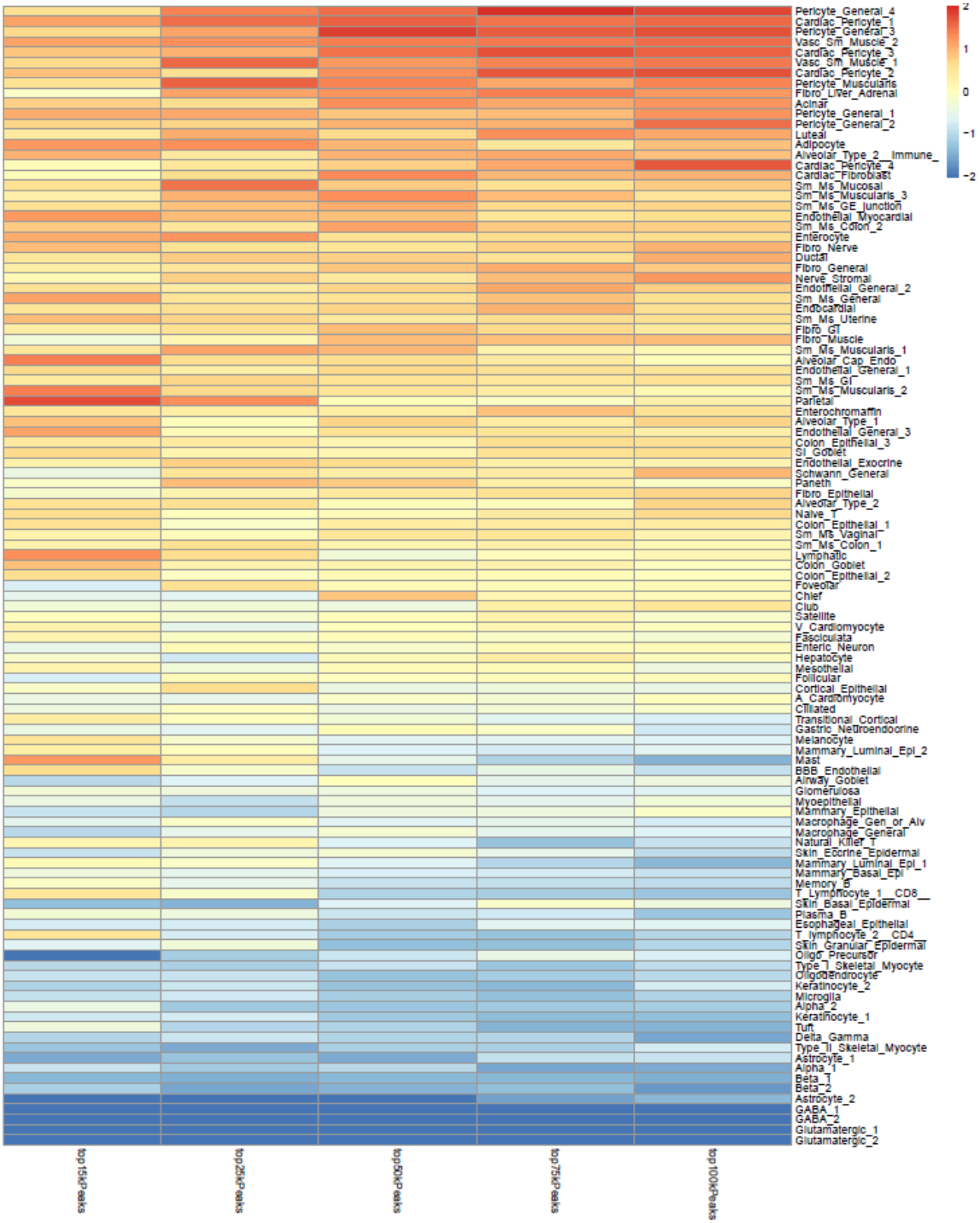
Supplementary Figure 8. Enrichment of CAD GWAS genes within the disease associated cell states gene signatures. Hypergeometric enrichment test results are shown for CAD GWAS candidate causal gene lists from 9 different sources (different colors).



Supplementary Figure 9. Proportion of variance of CAD explained by polygenic risk score (PRS) calculated using to 10,000 strongest cell type-specific peaks of each cell type. Peaks were ranked by ATAC signal score as calculated by the peak caller (MACS2). Peaks were each 500 bp in length. Thus, the fraction of genome included in the search space based for each cell type is equal-sized and non-overlapping between the cell types.



Supplementary Figure 10. Evaluation of CAD PRS strength among the 111 human cell types of the adult single-cell atlas of chromatin accessibility. PRS R2 (PRS strength) values are shown. For each cell type, the indicated number (either 100,000; 75,000; 50,000; 25,000 or 15,000) of strongest peaks were selected for PRS calculation. Peaks were equal-width.



Supplementary Figure 11. Evaluation of CAD PRS strength among the 111 human cell types of the adult single-cell atlas of chromatin accessibility. The results are as in Supplementary Figure 10, except the column z-score of the PRS R2 is shown.

Supplementary Table Legends

Table S1. Markers of the 13 atherosclerosis-associated cell states.

Table S2. Differentially expressed genes identified in bulk RNA-Seq experiment. Genes are sorted by effect direction followed by statistical significance.

Table S3. Complete gene ontology listing of the cell state marker genes used to generate Figure 2C.