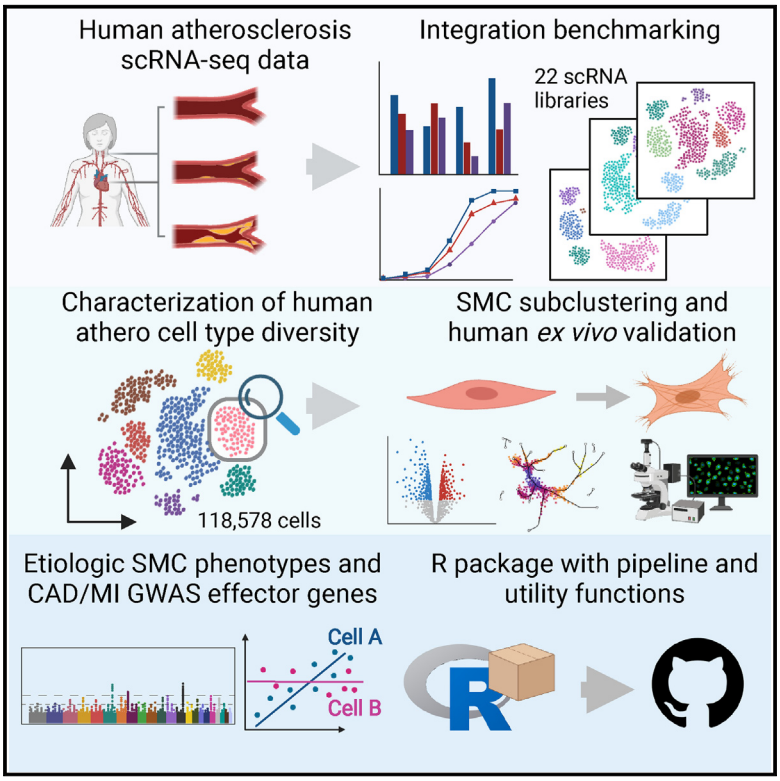
Cell Reports

Article

Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis

Graphical abstract



Highlights

- Integration benchmarking and generation of new scRNA atlas of human atherosclerosis
- GWAS data integration reveals etiologic SMC populations and effector genes for CAD
- *LTBP1* and *CRTAC1* as markers of modulated SMCs and atherosclerosis progression

Authors

Jose Verdezoto Mosquera, Gaëlle Auguste, Doris Wong, ..., Chongzhi Zang, Nathan C. Sheffield, Clint L. Miller

Correspondence

clintm@virginia.edu

In brief

The authors generate a comprehensive map of cell diversity in human atherosclerosis. This atlas provides insights into the heterogeneity of smooth muscle cells and their genetic contribution to atherosclerosis and reveals candidate markers of disease progression.

Mosquera et al., 2023, Cell Reports *42*, 113380
November 28, 2023
https://doi.org/10.1016/j.celrep.2023.113380

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis

Jose Verdezoto Mosquera,[1,2] Gaëlle Auguste,[2] Doris Wong,[1,2] Adam W. Turner,[2] Chani J. Hodonsky,[2] Astrid Catalina Alvarez-Yela,[3] Yipei Song,[2,4] Qi Cheng,[5] Christian L. Lino Cardenas,[6] Konstantinos Theofilatos,[7] Maxime Bos,[8] Maryam Kavousi,[8] Patricia A. Peyser,[9] Manuel Mayr,[7,10] Jason C. Kovacic,[11,12,13] Johan L.M. Björkegren,[14,15] Rajeev Malhotra,[6] P. Todd Stukenberg,[1] Aloke V. Finn,[5] Sander W. van der Laan,[16] Chongzhi Zang,[1,2,3,17] Nathan C. Sheffield,[1,2,3,17] and Clint L. Miller[1,2,3,17,18,*]

[1]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA
[2]Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA
[3]Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA
[4]Department of Computer Engineering, University of Virginia, Charlottesville, VA 22908, USA
[5]CVPath Institute, Gaithersburg, MD 20878, USA
[6]Cardiovascular Research Center, Cardiology Division, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129, USA
[7]King's British Heart Foundation Centre, King's College London, London WC2R 2LS, UK
[8]Department of Epidemiology, Erasmus University Medical Center, 3000 CA Rotterdam, the Netherlands
[9]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI 48019, USA
[10]National Heart and Lung Institute, Imperial College London, London SW3 6LY, UK
[11]Cardiovascular Research Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[12]Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010, Australia
[13]St. Vincent's Clinical School, University of New South Wales, Sydney, NSW 2052, Australia
[14]Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA
[15]Department of Medicine, Karolinska Institutet, 141 52 Huddinge, Sweden
[16]Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, the Netherlands
[17]Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA
[18]Lead contact
*Correspondence: clintm@virginia.edu
https://doi.org/10.1016/j.celrep.2023.113380

SUMMARY

Coronary artery disease (CAD) is characterized by atherosclerotic plaque formation in the arterial wall. CAD progression involves complex interactions and phenotypic plasticity among vascular and immune cell lineages. Single-cell RNA-seq (scRNA-seq) studies have highlighted lineage-specific transcriptomic signatures, but human cell phenotypes remain controversial. Here, we perform an integrated meta-analysis of 22 scRNA-seq libraries to generate a comprehensive map of human atherosclerosis with 118,578 cells. Besides characterizing granular cell-type diversity and communication, we leverage this atlas to provide insights into smooth muscle cell (SMC) modulation. We integrate genome-wide association study data and uncover a critical role for modulated SMC phenotypes in CAD, myocardial infarction, and coronary calcification. Finally, we identify fibromyocyte/fibrochondrogenic SMC markers (*LTBP1* and *CRTAC1*) as proxies of atherosclerosis progression and validate these through omics and spatial imaging analyses. Altogether, we create a unified atlas of human atherosclerosis informing cell state-specific mechanistic and translational studies of cardiovascular diseases.

INTRODUCTION

Cardiovascular diseases (CVDs), such as coronary artery disease (CAD), are the leading global causes of mortality and morbidity.[1] The pathological hallmark of CAD is atherosclerosis, a chronic build-up of plaque inside arterial walls, which can lead to thrombus formation and myocardial infarction (MI) or stroke.[2–5] This process involves a complex interplay of both immune and vascular cell types and cell state transitions along a continuum.[6,7] In response to injury of the inner vessel wall layer, contractile smooth muscle cells (SMCs) transition to a more proliferative and migratory state[8,9] and endothelial cells to a

mesenchymal state in early and advanced atherosclerosis.[10,11] Thus, a thorough assessment of cell heterogeneity and plasticity within the vessel wall is paramount to uncover new knowledge regarding atherosclerosis development and progression.

The advent of single-cell sequencing technologies has enabled study of gene expression and regulation in disease and development at the single-cell level. For instance, single-cell RNA sequencing (scRNA-seq) studies have resolved the cellular diversity and gene signatures in human and murine atherosclerotic lesions[12–16] as well as non-lesion arteries.[17] Lineage-tracing and scRNA-seq studies have shown that SMCs readily transform into a multipotent "pioneer" cell type in response to pro-atherogenic stimuli.[18–20] However, their post-transition fate remains controversial, with a few studies agreeing they can become fibroblast-like (fibromyocytes)[18] or osteogenic-like (fibrochondrocytes; FCs),[19] while others suggest they adopt pro-inflammatory or macrophage-like properties.[8,20] Biological interpretation of these individual studies could be potentially confounded by limited sample sizes, experimental design, or other technical factors. Thus, there remains a need for a consensus single-cell reference[21–23] that spans atherosclerotic disease stages in humans.

Here, we harmonize and meta-analyze four single-cell studies of human atherosclerosis using extensive benchmarking for optimal batch-correction method, encompassing both early and advanced lesion and non-lesion samples (Figure 1A; Table S1). We generate a streamlined analysis pipeline available as an R package (scRNAutils). The resulting atlas of 118,578 high-quality cells enables discovery of previously missed vascular and immune cell types and clarifies markers for known disease-relevant immune cells (e.g., foamy macrophages). We perform integrative downstream analyses and genome-wide association study (GWAS) trait enrichment to define CVD-relevant etiologic cell types with their corresponding effector genes. We further validate SMC phenotypes identified in lineage-tracing studies, reveal underrepresented SMC states from individual scRNA-seq studies, and highlight *CRTAC1* and *LTBP1* as candidate markers of synthetic/pro-calcifying SMCs and plaque stability in humans. This map of cell diversity in human atherosclerosis provides a critical step toward translating mechanistic knowledge and developing more targeted interventions.

RESULTS

Integration of lesion and non-lesion artery datasets

We sought to build a comprehensive single-cell reference to further investigate complex vascular processes such as SMC phenotypic modulation. To achieve proper representation of mural cells and immune types and to span the continuum of CAD risk, we used four human datasets including atherosclerotic lesions (Wirka et al.,[18] Pan et al.,[19] and Alsaigh et al.[16]) and from non-lesion coronary arteries[17] (Table S1). While carotid lesion samples[16,19] were obtained from carotid endarterectomies, lesion coronary samples[18] were dissected from explanted hearts of transplant recipients. Control coronary samples[17] were obtained from patients with end-stage heart failure but with no discernible atherosclerotic lesions.

We established a standardized pipeline for quality control (Figures 1 and S1A) and processing of the 22 raw sequencing libraries, and we observed improved separation and cohesion of cell clusters post filtering (Figure S1B). We benchmarked top-performing integration tools (Figure S1B; method details)[26,29,33–35] on a subset of the libraries[16,18,19] and found that reciprocal PCA (rPCA) and Harmony outperformed the other tools in terms of running time. Beyond confirming proper batch mixing through uniform manifold approximation and projection (UMAP) embedding visualization, we also evaluated the effectiveness of batch removal, biological variation conservation, and clustering purity using local inverse Simpson index (LISI) scores, k-nearest neighbors batch effect test (kBET), principal components regression,[30] and average width silhouette scores (cASW) (Figures S1B–S1G; method details).[28] We found that rPCA achieved the best balance in terms of running time, batch mixing, separation of unique cell types, and clustering purity across tested granularity parameters (Figure S1H). Integration of libraries with rPCA yielded a total of 118,578 high-quality cells and 41 Louvain clusters. We inspected reference UMAP embeddings and found mixing across the four studies for the major cell compartments confirming proper batch removal (Figure S1I).

We annotated our clusters by first labeling cells in UMAP space with broad cell lineage names (level 1) and resolved cell subtypes by combining manual and automated approaches (level 2). We defined level 1 annotations by reprocessing and transferring cell-type labels from the Tabula Sapiens (TS) vasculature single-cell atlas.[31] We found that labels were assigned with remarkably high confidence scores (Figure S2A). These annotations were supported by the expression of well-established marker genes in corresponding level 1 clusters (Figures 2A and 2B).

We observed a balanced number of cells labeled as macrophages and endothelial cells across studies. However, there were slightly more SMCs in Pan et al.[19] and T/natural killer (NK) cells in Alsaigh et al.[16] and slight biases from small clusters (e.g., plasma cells, B cells) between studies (Figures 2C and S2B). We observed overrepresented fibroblasts from coronary datasets (Wirka et al.[18] and Hu et al.[17]), as expected given the intact coronary vessel wall layers compared to carotid plaques. Across disease status, we observed a greater proportion of B cells, plasma cells, and plasmacytoid dendritic cells (pDCs) in lesion samples (Figures 2C and 2D). While cell-type variability across different studies or libraries of a single study could be biologically meaningful, it could also represent technical artifacts from library preparation or single-cell suspension procedures. For instance, foamy macrophages are highly sensitive to single-cell suspension procedures and could be underrepresented in some samples.[37]

Differentially expressed (DE) genes for each cluster (Table S2; method details) were used to run Gene Ontology (GO) enrichment analyses. We found enrichment for terms in their expected clusters, such as "muscle contraction" in SMCs and "endothelial development" in endothelial cells (ECs). We observed overrepresentation of terms such as "extracellular matrix organization" in SMCs, likely due to the presence of phenotypically modulated SMCs[9] (Figure S2C; Table S3). In contrast, myeloid and lymphoid clusters were enriched for immune-related terms such as "antigen processing and presentation" and "regulation

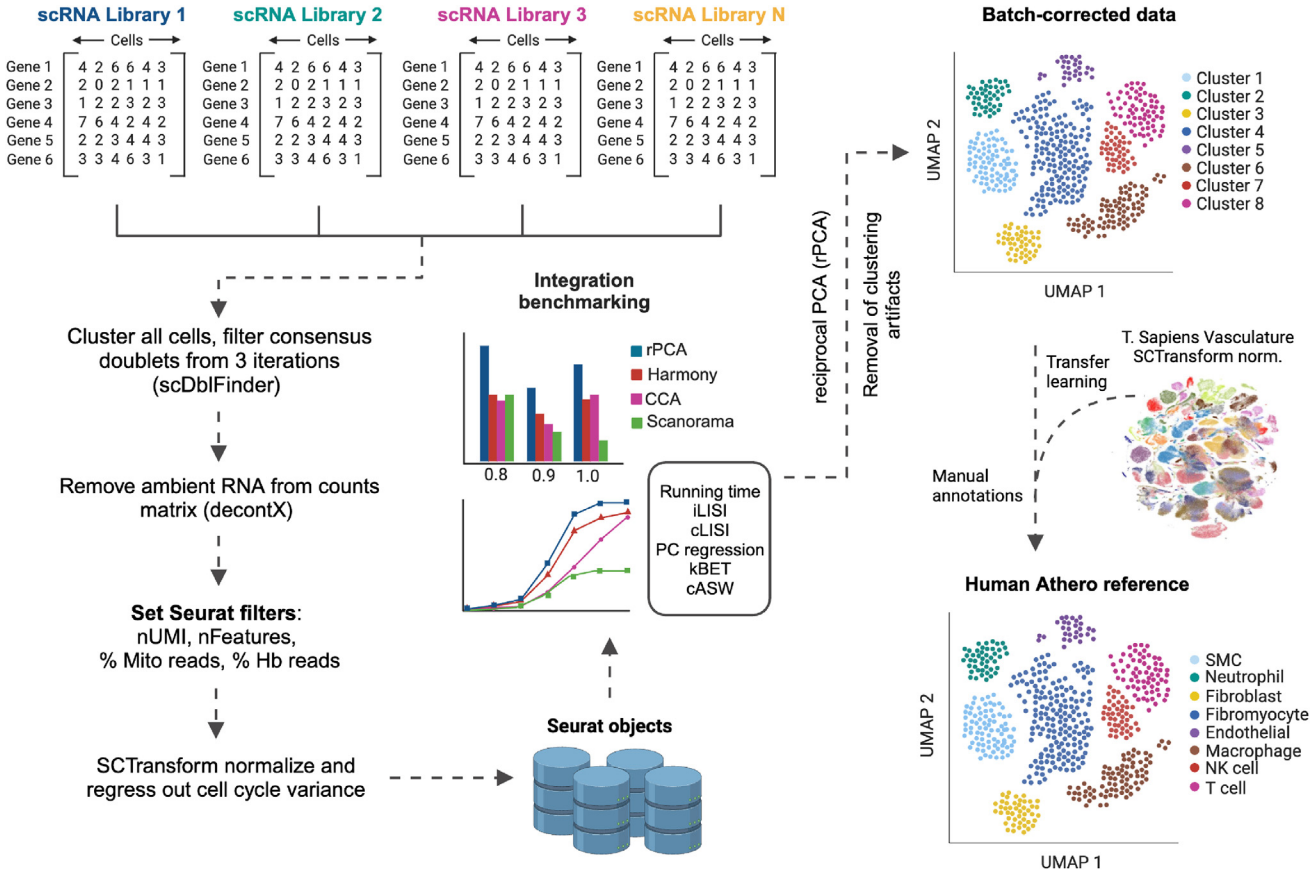


Figure 1. Pipeline devised to build the integrated scRNA atlas for human atherosclerosis

Doublets within individual scRNA-seq libraries were identified using scDblFinder,[24] while ambient RNA was removed using decontX.[25] Decontaminated matrices were used for downstream filtering of cells based on (1) detected gene number, (2) UMI number, (3) percent reads mapping to mitochondrial genome, and (4) percent reads mapping to hemoglobin genes using Seurat.[26] Libraries were normalized using SCTransform,[27] and a subset was used for benchmarking of four batch correction approaches: canonical correlation analysis + mutual nearest neighbors (CCA+MNN), reciprocal PCA (rPCA),[26] Harmony,[28] and Scanorama.[29] Besides running time, PCA embeddings were used for calculating the following metrics: integration and cell-type local inverse Simpson's Index (iLISI and cLISI) scores, k-nearest-neighbors batch effect test (kBET) rejection rates,[30] principal component (PC) regression coefficients,[30] and cluster average silhouette widths (cASW). rPCA was used to harmonize the 22 libraries, and level 1 annotations were added using transfer learning with the Tabula Sapiens vasculature dataset[31] as reference. Finally, automated[32] and manual curation enabled identification of granular cell subtypes (level 2 annotations). PCA, principal component analysis. See also method details, Figure S1, and Table S1.

of T cell activation." These results confirmed the accuracy of our level 1 annotations.

Defining candidate etiologic cell types for complex traits

As done previously,[38,39] we leveraged cell-type-specific gene signatures and GWAS summary statistics to define etiologic cell types for disease. The underlying assumption is that genetic association signals are enriched in genes for cell types closely related to the particular trait. We computed numeric estimates (ES_μ) of gene expression specificity[36] (method details) and show the robustness of this score by inspecting several canonical cell markers, e.g., *CD68* in macrophages and *PECAM1* in ECs (Figure 2E). We then applied two different genetic prioritization models: LD score regression (LDSC-SEG[40] and MAGMA[41]) to specifically expressed genes in level 1 annotated cell types using GWAS summary statistics for CVD and non-CVD traits[42–48] as input (method details; Table S4).

Both LDSC-SEG and MAGMA showed that SMC and pericyte gene signatures were significantly enriched (FDR < 0.05) for CV traits such as CAD,[42,48] coronary artery calcification (CAC),[49] and MI[43] (Figures 2F and S3A; Table S4), while EC signatures were enriched for carotid plaque associations (Figure 2F). Consistent with our previous work,[50] we observed level 1 macrophage annotations were the most highly enriched for Alzheimer's disease and white blood cell count (WBC) GWAS-tagged heritability (Figure 2E). While level 1 annotations from the lesion and non-lesion samples showed similar LDSC enrichments for most traits, we observed that SMCs from lesions were predominantly enriched for CAC (Figure S3B; Table S4). Lastly, by intersecting highly specific SMC genes (highest ES_μ values) with MAGMA-defined GWAS candidate genes, we identified both known and less-characterized SMC effector genes driving the CAD/MI enrichments[43] (Figure S3C; Table S4), thus providing further insights into the contribution of SMCs underlying CAD/MI.

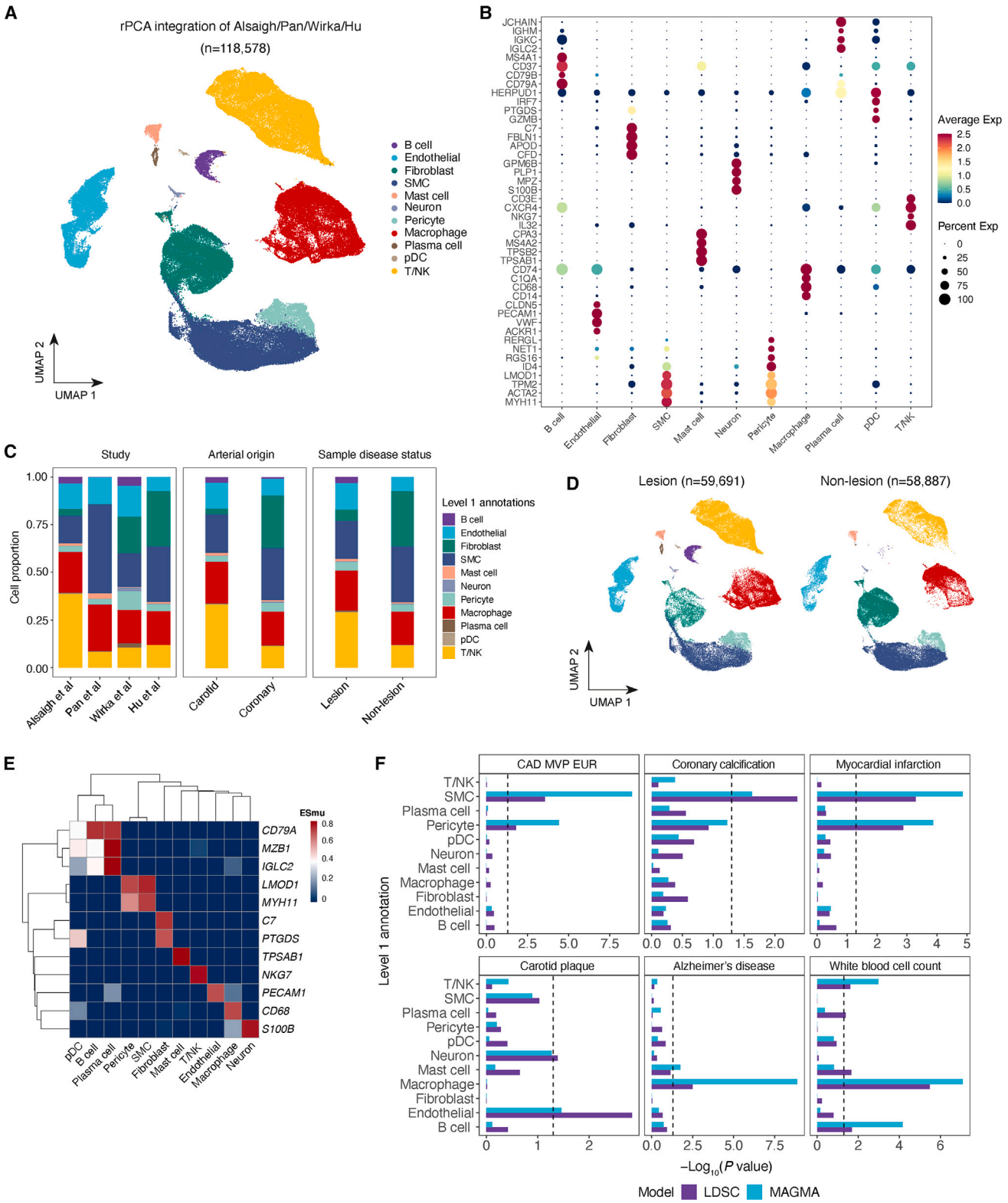


Figure 2. Integration of single-cell data identifies major cell compartments in human atherosclerosis

(A) UMAP embeddings of 118,578 cells based on reciprocal PCA (rPCA) integration of 22 sequencing libraries. Dot colors depict broad cell lineage annotations (level 1).

(legend continued on next page)

Defining cell subtype heterogeneity in human atherosclerosis

Next, we surveyed the 41 clusters using a combination of automated and manual annotation (method details). Manual annotations included markers of lymphoid, myeloid, and endothelial cell subtypes from the literature[21–23,51–56] that were verified using the CellTypist machine learning classifier[32] resulting in a more granular map of cell diversity in human atherosclerosis (Figure 3A). We summarize some of the most representative cell subpopulations and markers below.

Endothelial diversity

We identified cells highly expressing classical endothelial markers (*PECAM1*, *CLDN5*), intimal ECs defined by expression of homeostatic EC marker genes (e.g., *RAMP2*),[57] and pro-angiogenic ECs with upregulated *vasa vasorum* or angiogenic genes (e.g., *ACKR1,AQP1* and *FABP4*).[58–60] Interestingly, we identified a cluster likely reflecting a pro-inflammatory state with elevated expression of chemokine and adhesion molecules (*SELE*, *CCL2*) (Figures 3A and 3B).[61] Endothelial-to-mesenchymal transition (EndoMT) ECs[62] were defined by the expression of ECM genes (*COL1A2*, *FN1*) and contractile genes. Finally, we defined a small subcluster of lymphatic ECs based on expression of *LYVE1* and *CCL21*.[63]

Myeloid diversity

We distinguished inflammatory macrophages expressing known "M1-polarized" macrophage markers (e.g., *IL1B*, *TNF*) (Figures 3A and 3B) and foamy macrophages by higher expression of lipid genes (*APOE* and *FABP5*) and lower inflammatory genes.[15,21,23] We also identified resident macrophages (*LYVE1, FOLR2*), classical monocytes (*S100A8*, *S100A9*, *LYZ*), and conventional dendritic cells (*CD1C*, *CLEC10A*).[64–66] Importantly, we resolved critical smaller myeloid populations, including pDCs[67,68] and neutrophils (*NAMPT*, *S100A8*), overlooked by previous individual human scRNA-seq datasets. Consistently, we found more prevalent monocytes, foamy macrophages, and other myeloid populations in lesion-containing libraries (Figures 3C and 3D).

Lymphoid diversity

We identified NK and several subpopulations of T cells based on differential expression of *CD69*[69] and *CD8A/B*, among others (Figures 3A, 3B, and S2D). Specifically, we identified markers of early activated and memory/naive CD8 cytotoxic T cells (CTLs),[70,71] terminally differentiated CTLs, Th17 and Th2 helper cells, and regulatory T cells (Figure S2D). Finally, we defined clusters of B cells (*CD79A*, *CD79B*) and plasma cells (*IGLC2*, *IGHM*, *JCHAIN*). While all lymphoid populations showed larger frequencies in lesions, we found that B cells, plasma cells, and pDCs were highly depleted in non-lesion libraries (Figure S2E).

Fibroblast diversity

Defining fibroblast diversity in atherosclerosis is particularly challenging given the low specificity of known fibroblast markers.[55] Although most cells expressed traditional fibroblast ECM markers such as *LUM* and *DCN* (Table S2), we dissected a subset of fibroblasts that upregulated the contractile marker *ACTA2* (Figure 3B) and complement genes (*C3* and *C7*). These likely represent activated fibroblasts (myofibroblasts) known to adopt increased contractile, ECM-producing, and pro-inflammatory states in response to injury or atherosclerotic stimuli.[55,62] We also identified a group of cells strongly expressing *APOE* in addition to the chemokine ligands *CXCL12* and *CXCL14* and complement genes, which we termed APOE fibroblasts.

Characterization of SMC phenotypes in human atherosclerosis

To refine the role of human SMC phenotypes, we next compared our human scRNA reference to a recent scRNA meta-analysis of murine vascular SMCs.[22] We subsetted the full human atlas to include only SMCs, pericytes, and a subset of fibroblasts and assessed enrichment of lineage-traced murine SMC gene modules on a per-cell basis using the UCell R package[72] (method details). We showed a progressive loss of the murine SMC contractile signature within a portion of the human subset, coincident with a gain in the *Lgals3*$^{+}$ transitional gene signature (Figure 4A), supporting a transitional SMC signature in humans. Further, we detected an enriched signature of the murine calcification-promoting fibrochondrocytes distinct from non-SMC-derived fibroblasts. DE markers for each cluster and UCell module enrichment scores were used to guide SMC annotation as contractile, ECM-rich transitional SMCs, fibromyocytes, and FCs (Figures 4B, 4C, and S4B).

We observed similar proportions of contractile, transitional SMCs and fibromyocytes across arterial beds and lesion status, consistent with previous reports.[74] However, concordantly with their role in calcification, FCs predominated in lesions compared to non-lesion samples. The FC annotation was further supported by higher ES_{μ} values for SOX9 and RUNX2 (Figure S4C), known master regulators of SMC osteochondrogenic transitions.[75] Globally, SMCs, transitional SMCs, fibromyocytes, and FCs were enriched for relevant biological processes, thus validating our annotation approach (Table S3). We also identified a cluster enriched for the transitional SMC gene signature and expressing lipid metabolism genes (*APOE*, *APOC1*, *AGT*), which we termed "foam-like" SMCs (Figures 4B, 4C, and S4D and Table S2).

(B) Dot plot of top five marker genes by major cell lineage compartment. Dot size represents the portion of cells expressing the gene per level 1 compartment.
(C) Stacked bar plot showing the distribution of level 1-annotated cells across included studies, arterial beds (coronary/carotids), and lesion status (lesion, non-lesion).
(D) UMAP embeddings of level 1-annotated cells across lesion status.
(E) Heatmap showing ES_{μ} values for established markers of mural and immune level 1 cell annotations previously identified as differentially expressed (DE) using a Wilcoxon rank-sum test in (B).
(F) LD score regression applied to specific genes (LDSC-SEG) and MAGMA analyses prioritizing the contribution of level 1-annotated cell types to cardiovascular and non-cardiovascular GWAS traits. LDSC-SEG- and MAGMA-based regression analyses were carried out using an expression specificity matrix generated with CELLEX.[36] The black line depicts the FDR significance threshold (FDR < 0.05 at –log10(p) = 1.301). GWAS, genome-wide association study. See also Figures S2 and S3 and Tables S2 and S4.

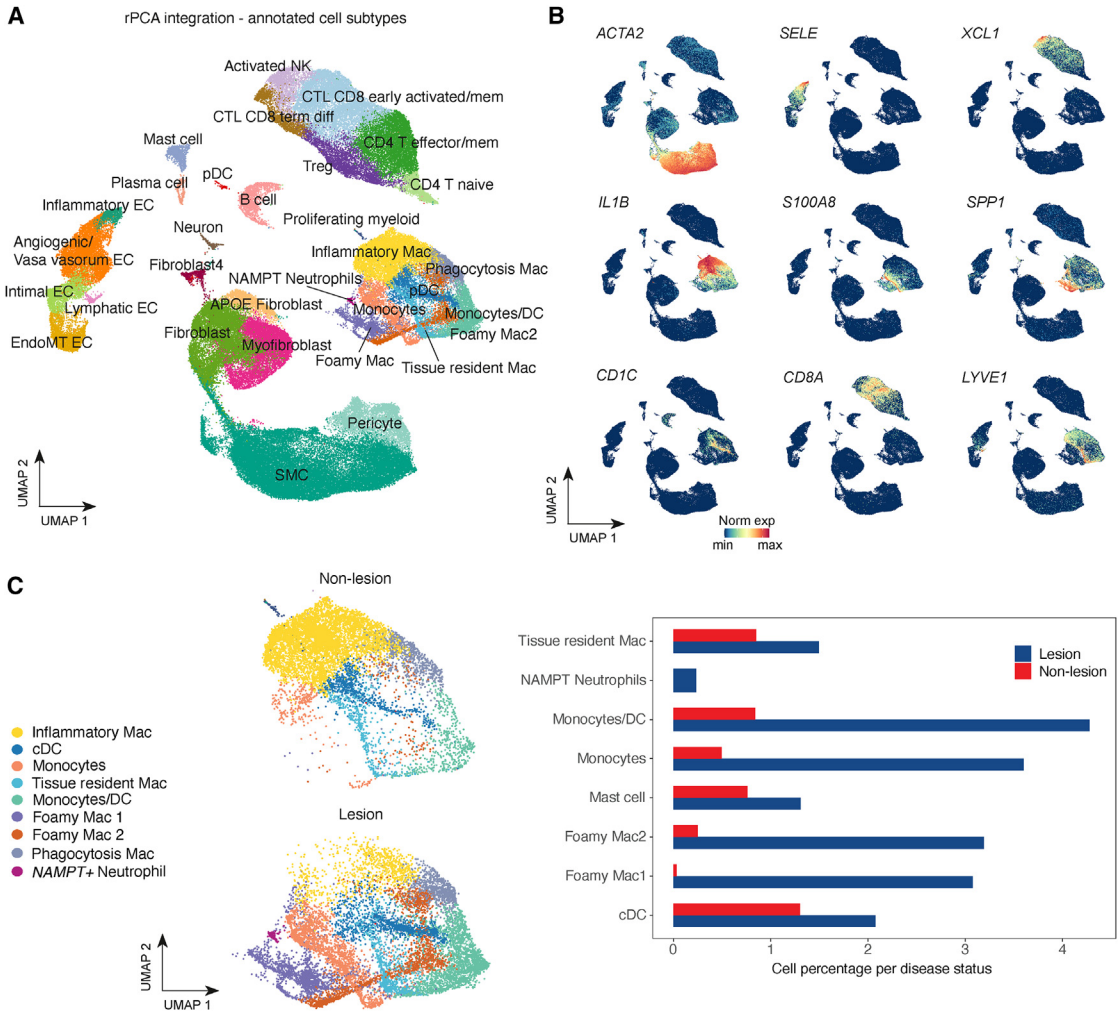


Figure 3. Atherosclerosis cell subpopulations (level 2) and distribution of myeloid subtypes across disease status

(A) UMAP embeddings of cell subtypes (level 2 annotations) within the largest level 1 cell compartments (T/NK, macrophage, endothelial, fibroblast).

(B) UMAP embeddings of canonical marker genes delineating immune and non-immune cell subtypes. SCTransform-normalized gene expression is indicated by color.

(C) UMAP embeddings and bar plot of level 2 myeloid subtypes according to lesion status. Frequencies for each subtype are normalized to the total number of cells in each condition (N lesion cells = 59,691; N non-lesion cells = 58,887) and shown as percentages. See also Figure S2.

These cells expressed ECM-remodeling genes such as *TIMP1* and pro-inflammatory genes *CCL19*, *CCL2*, and *IGFBP3*, consistent with a potential role in leukocyte recruitment.[76]

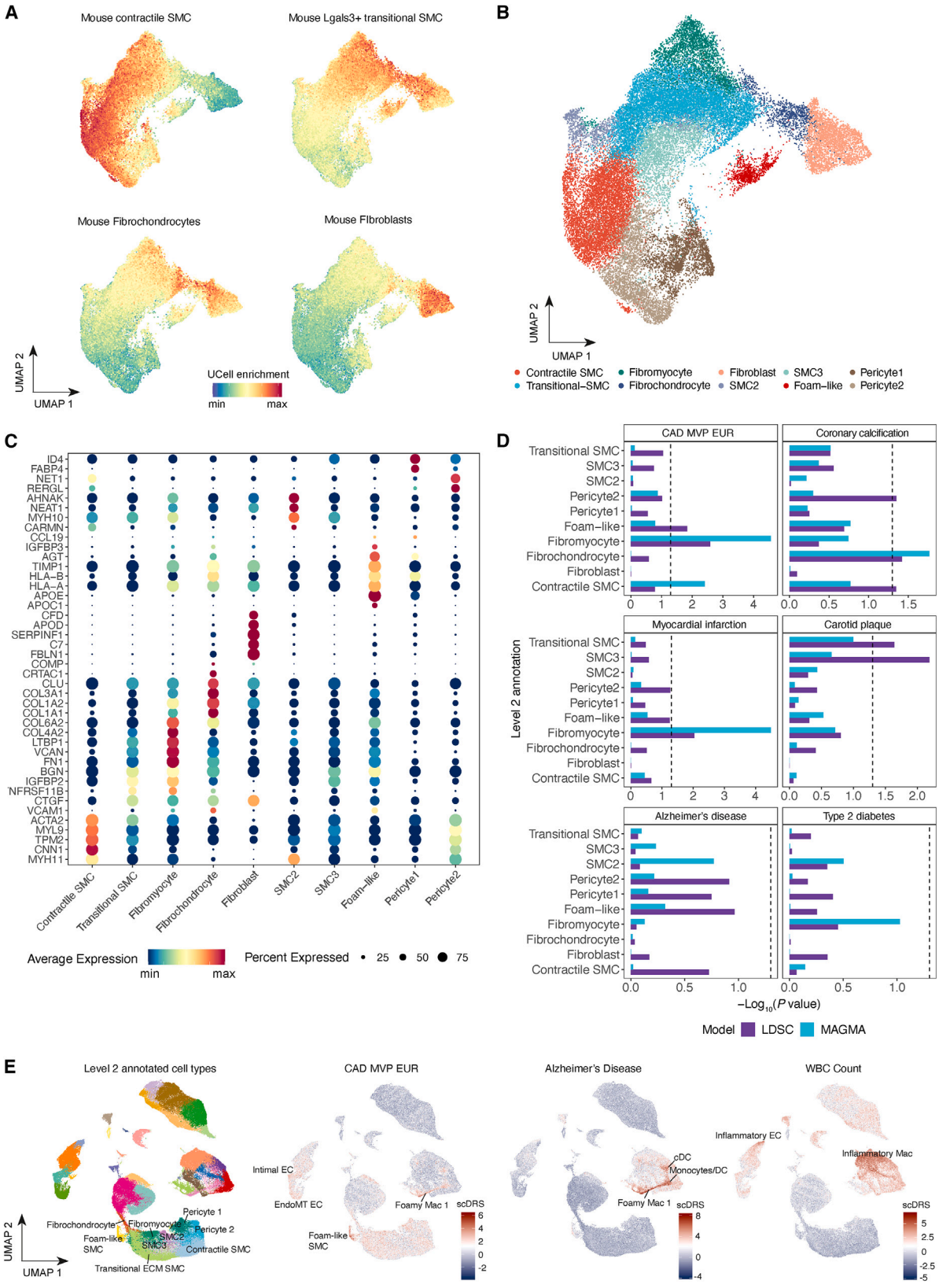
Definition of etiologic SMC and immune phenotypes

We leveraged the above-defined granular SMC labels to investigate the disease relevance of distinct modulated phenotypes. Using both LDSC-SEG and MAGMA, fibromyocytes were the most highly enriched SMC phenotype for CAD/MI traits (Figure 4D; Table S4). However, FCs were highly enriched for CAC,[49] further supporting the biology of CAC through this SMC phenotype. While not significant, we observed trends of CAD and CAC enrichment in foam-like SMC and contractile SMCs, respectively.

Given the limitations of relying solely on cluster-specific genes, we used an orthogonal approach to determine GWAS trait enrichment at single-cell resolution. Overall, estimation of single-cell disease relevance scores (scDRSs)[73] corroborated our findings using level 1 annotations but also prioritized pro-inflammatory macrophages and foamy macrophages in WBC counts and Alzheimer's disease (AD), respectively (Figure 4E). Beyond showing that fibromyocytes and foam-like SMCs had slightly larger weights compared to other SMC clusters, this analysis also revealed CAD enrichments in ECs and foamy macrophages, which could have been previously missed when surveying only highly specific genes. Among ECs, we found a gradient in CAD where intimal and EndoMT ECs were slightly more enriched than pro-inflammatory ECs.

Cell crosstalk in human atherosclerosis

To interrogate paracrine signaling in atherosclerosis, we dissected cellular crosstalk from our level 1 and 2 annotations



(legend on next page)

across lesion status using CellChat.[77] We observed strong interactions between SMCs and fibroblasts in non-lesion samples, while SMC and EC interactions with macrophages and T/NK were stronger in lesions (Figure 5A). Differential signaling pathway enrichment analysis between SMCs and macrophages revealed tumor necrosis factor alpha (TNFα) and platelet-derived growth factor signaling pathways enriched in non-lesion samples, whereas tumor necrosis factor-like weak inducer of apoptosis (TWEAK)- and osteopontin (SPP1)-mediated signaling pathways were enriched in lesion samples (Figure 5B; Table S5). Signaling involving SPP1 specifically targeted SMCs and was mostly driven by macrophage foam cell clusters (Figure 5C). This is consistent with previous studies implicating these signals in myeloid-SMC-mediated vascular inflammation and calcification.

Leveraging our SMC subtype annotations, we observed more TWEAK-mediated interactions between contractile/transitional SMCs and distinct myeloid subtypes in lesion samples (Figure S5A). Instead, SPP1 signals from foamy macrophage cells (foamy mac1) specifically targeted contractile and transitional SMCs (Figures 5D and S5B). Using ligand-receptor interaction analyses, we also found that myeloid-expressing genes encoding SPP1 ligand preferentially signal to fibromyocyte and foam-like SMCs via the heterodimeric ITGA8/ITGB1 receptor (Figures 5E and S5C; Table S5).

Modeling of SMC gene expression across pseudotime

Current evidence suggests that SMCs transition into fibromyocytes/FCs through an *Lgals3*$^+$ transitional state.[19,20] We modeled SMC de-differentiation via pseudotime analysis using Monocle 3,[78] in which we defined *MYH11*-expressing contractile SMCs as the starting point of phenotypic modulation, and we predicted transitional SMCs to adopt either a fibromyocyte or FC fate (Figure 6A). Moreover, we observed more FCs in lesion samples toward late pseudotime, consistent with calcification in advanced lesions (Figure S6A). We found similar densities of SMCs from carotid/coronary arteries across pseudotime, suggesting that SMCs from different vascular beds adopt a common transitional state before ECM-rich phenotypes (Figure S6A).

We next investigated changes in gene expression programs during SMC de-differentiation (method details) and found that canonical contractile markers, such as *ACTA2*, were active early in pseudotime. This coincided with expression of CAD GWAS effector genes expressed early (e.g., *TNS1* and *FHL5*), while effector genes such as *COL4A1/2*, *CDH13*, and *AGT* were expressed late in pseudotime (Figures 6B and S3C). Pseudotime DE genes were then grouped into modules specific to transitional SMC (modules 5 and 10), fibromyocytes (module 4), and FCs (module 9) to investigate genes relevant for each phenotypic transition (Figure S6B; method details). Transitional SMC modules harbored genes involved in early SMC investment in atherosclerotic lesions (e.g., *LGALS3*),[20] as well as cell division and proliferation (e.g., *TUBA1B* and *SIRT6*)[81] and ECM remodeling (e.g., *KRT8* and *SPARC*). As expected, fibromyocyte module 4 included known markers (e.g., *FN1*, *VCAN*, *COL4A1/2*, *PDGFRB*). In contrast, the FC module 9 harbored chondrocyte-related genes such as *BMP4,*[82] *WISP2*, and *SPRY1*[83] in addition to known ECM genes *LUM* and *DCN*.

We modeled the expression dynamics of pseudotime-variable fibromyocyte and FC genes along the SMC-transitional-fibromyocyte and SMC-transitional-FC axis (Figures 6B, 6C, and S6C). Some fibromyocyte markers (e.g., *FN1*, *AEBP1*, and *LTBP1*) and pioneer marker *LGALS3* showed a steady increase with adoption of the transitional state (Figure 6B), while genes such as *PDGFRB* were active later, suggesting a role in SMC transition toward fibromyocytes (Figure 6B). In parallel, evaluation of FC markers and genes from module 9 revealed a steady increase in *COL1A2* and *MMP2*, whereas *IBSP*, *CRTAC1*, and *COMP* were increased at later pseudotime stages, presumably as transitional SMCs adopt an FC fate (Figure 6C).

TF activity inference analysis

We next investigated the upstream transcription factors (TFs) driving cell-specific expression changes using TF activity inference with VIPER[79] and the DoRothEA collection of human regulons.[80] This analysis corroborated known regulators of fibromyocytes and FCs such as TCF21 and SOX9 (Figure 6D). We also detected changes in AP-1 (e.g., JUN), TEAD, ETV, and ETS factor activity in fibromyocytes vs. FCs. We observed increased regulon activity of the TGF-β signaling mediator SMAD3 in fibromyocytes and FCs compared to contractile and transitional SMCs, with FCs showing the highest activity. We then interrogated our coronary artery single-nucleus assay for transposase-accessible chromatin using sequencing (snATAC-seq) data[84] using ArchR[85] and confirmed increased accessibility of AP1 factors; notably, we found that accessible regions in the ECM-rich SMC cluster were specifically and highly enriched for SMAD3 motifs compared to contractile SMCs (Figures S6D

Figure 4. Characterization of etiologic SMC phenotypes for cardiovascular traits and diseases

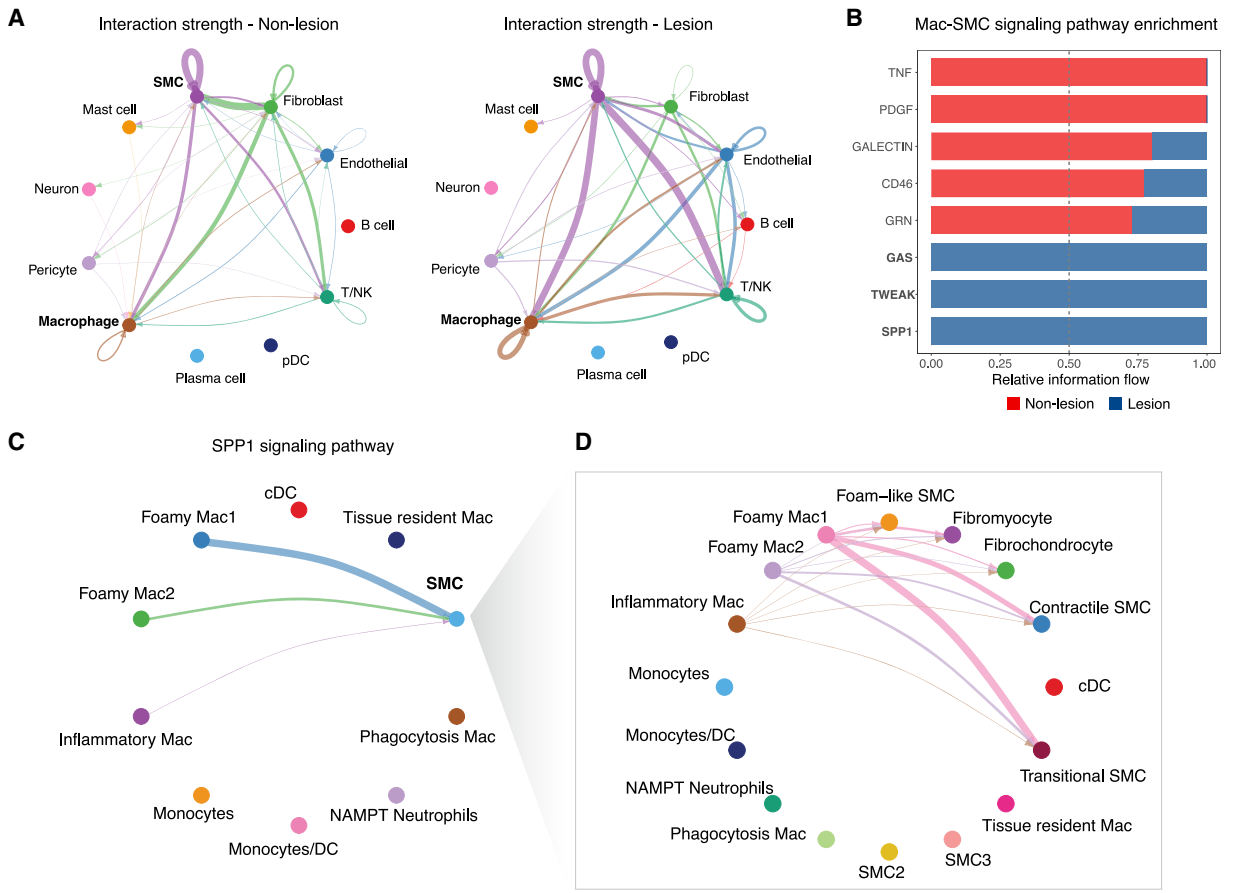
(A) UCell[72] enrichment of meta-analyzed SMC murine gene modules (contractile, *Lgals3*$^+$ transitional, fibrochondrocytes) and non-SMC-derived fibroblasts in the level 1 SMC compartment. A subset of human fibroblasts and pericytes were used as negative enrichment controls for murine SMC modules. UCell scores were calculated using the Mann-Whitney U statistic.

(B) UMAP embeddings of SMC level 2 annotations in addition to pericytes and subset of fibroblasts. Annotations were defined using UCell scores to guide SMC differentiation state in addition to differentially expressed markers from Louvain clusters.

(C) Dot plot showing expression of top marker genes after SCTransform normalization for SMC level 2 annotations. Dot size represents the portion of cells expressing the gene.

(D) LDSC-SEG and MAGMA analyses prioritizing the contribution of SMC phenotypes, pericytes, and fibroblasts to cardiovascular GWAS traits. Type 2 diabetes and AD were used as negative controls. The black line depicts the FDR significance threshold (FDR < 0.05 at –log10(p) = 1.301).

(E) Meta-analysis UMAP embeddings showing normalized scDRS[73] scores for CAD and immune traits (WBC count and AD) previously shown as highly enriched in level 1 myeloid annotations. Red indicates cells enriched for the above-mentioned traits, while non-relevant cells are denoted in dark blue. SMC, smooth muscle cell; GWAS, genome-wide association study; WBC, white blood cell; AD, Alzheimer's disease; scDRS, single-cell disease relevance score. See also Figure S4 and Tables S2 and S3.



(legend on next page)

and S6E). This suggests that SMAD3 activity is critical as SMCs transdifferentiate toward more synthetic phenotypes.

CRTAC1 and *LTBP1* as candidate markers of FC and fibromyocytes in human atherosclerosis

Fibromyocytes and FCs play major roles in atherosclerotic lesion stability.[18,20] However, there is limited characterization of distinct markers for these SMC subtypes in humans. In addition to validating known markers of FCs (*CYTL1*, *COMP, COL1A1/2*),[22,86] we found that cartilage acidic protein (*CRTAC1*) was expressed 3-fold higher relative to other SMC clusters (Table S2). *CRTAC1* is a specific marker of human chondrocytes during ossification,[82,87] and it is implicated in osteoarthritis.[88] However, it has not been identified in mouse atherosclerosis studies, potentially suggesting specificity for human SMCs. The top fibromyocyte markers (Figure 4C; Table S2) also revealed latent transforming growth factor binding protein 1 (*LTBP1*), a key regulator of TGF-beta secretion and activation,[89] which may also contribute to SMC modulation and plaque stability but remains underexplored in atherosclerosis.

Both *CRTAC1* and *LTBP1* are enriched in artery tissues compared to other tissues in GTEx and predominantly expressed in SMCs from level 1 annotations (Figure S7A). We also observed expression of *CRTAC1* in SMCs enriched for the murine FC gene signature (Figures 4A and S7B) along with the calcification marker, *IBSP*,[20,75,86,90] while *LTBP1* showed gradual upregulation from contractile SMCs to fibromyocytes, consistent with our previous pseudotime results (Figure 6B). Both *CRTAC1* and *LTBP1* were positively correlated with known osteochondrogenic or fibromyocyte markers, respectively (Figures 7B and 7C), but *CRTAC1* was more negatively correlated with canonical SMC markers. This suggests SMC expression of these genes is associated with a progressive loss of the contractile phenotype and gain of synthetic and pro-calcification gene programs.[91,92]

To validate *CRTAC1 and LTBP1* markers of human atherosclerosis, we queried in-house and public bulk transcriptomic datasets.[93] We found that *CRTAC1* expression was significantly upregulated in advanced coronary lesions compared to non-lesion control/subclinical samples, while *LTBP1* expression was downregulated in advanced lesions, similar to known fibromyocyte genes *TCF21* and *VCAN* (Figure 7D). Based on these findings, we also performed differential expression analysis in meta-analyzed SMCs according to lesion category (method details). Similar to our bulk coronary results, *LTBP1* was significantly downregulated in lesion SMCs, whereas *CRTAC1* was upregulated in SMCs from samples with evidence of atherosclerosis (Figure S7C). Furthermore, *CRTAC1* was upregulated in unstable versus stable carotid plaque regions in a bulk RNA-seq carotid dataset,[93] whereas *LTBP1* and other fibromyocyte genes showed the opposite trend (Figure S7D).

Next, we queried the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) gene regulatory networks across seven cardiometabolic tissues.[94] *CRTAC1* was identified as a significant key driver within its co-expression module (Table S6), which was highly associated with CAD genes, C-reactive protein, and LDL cholesterol (Figure 7E). The *CRTAC1* module was enriched for calcification-related GO terms, whereas the *LTBP1* module was enriched for cholesterol traits and ECM remodeling terms (Figure S7E; Table S6). Thus, our findings implicate *LTBP1* in earlier stages of SMC modulation as well as plaque stability concomitant with lesion progression. Our analyses also postulate *CRTAC1* as a candidate human-specific marker of osteogenic SMCs in advanced atherosclerotic lesions.

Ex vivo validation of *LTBP1* and *CRTAC1* in atherosclerotic coronary arteries

We validated the protein localization of these candidate markers by immunostaining in atherosclerotic human coronary arteries. Immunofluorescence results showed LTBP1 localized in medial and fibrous neointimal regions of early lesions (Figures 7F and S8A–S8C). Remarkably, we observed a more uniform neointimal expression in subclinical lesions compared to advanced lesions, which showed more heterogeneous expression along the vulnerable and lipid-rich shoulder regions from representative samples. We also observed LTBP1^{+} cells in the fibrous cap of advanced coronary lesions (Figure S8C). Immunohistochemistry and RNAscope *in situ* hybridization further confirmed *LTBP1* localization to medial and intimal regions of control/subclinical coronary segments, while *CRTAC1* was only detected in the lesion area of advanced atherosclerotic samples (Figures 7G, S9A, and S9B).

Finally, we validated the expression of *LTBP1* and *CRTAC1* in contractile and modulated SMCs, respectively, by reprocessing an external human carotid plaque scRNA-seq dataset[95] (Figure S9C), further confirming the relevance of these genes across atheroprone arterial beds. These data together support the potential role of *LTBP1* and *CRTAC1* as markers of plaque stability and/or lesion progression in fibromyocytes and FC, respectively, which could be further investigated in genetic models of atherosclerosis or plaque rupture.

Figure 5. Summary of cell crosstalk in human atherosclerosis

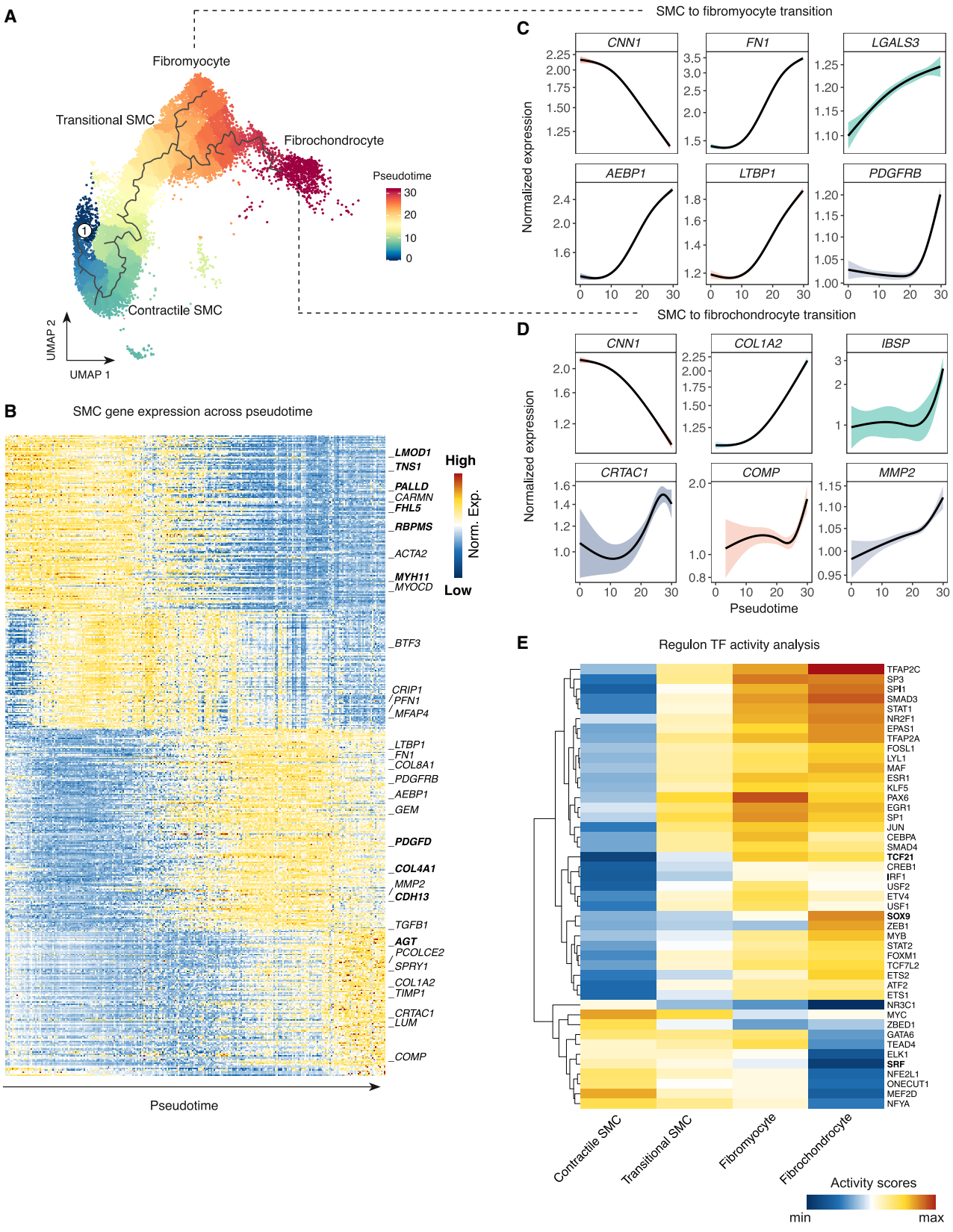
(A) Circle plots depicting aggregated cell-cell communication network for level 1 cell compartments leveraging the CellChat[77] human database. Interactions include secreted signaling and ECM-receptor and cell-cell contacts. Interactions were calculated separately across disease status (lesion vs. non-lesion). The top 30% of interactions are shown in the plot.

(B) Stacked bar plot showing conserved and disease status-specific signaling pathways. Signaling enrichment is based on pathway information flow changes (sum of communication probability among all pairs of cell groups in the inferred network). Pathways in bold denote those that showed statistically significant ($p < 0.05$) enrichments in each disease condition.

(C) Circle plot depicting sources and targets for SPP1 signaling using level 2 annotations for myeloid cells and level 1 SMC annotations.

(D) Circle plot depicting SPP1 signaling sources and targets for level 2 myeloid and SMC annotations.

(E) Summary dot plot of ligand-receptor interactions for level 2 myeloid and SMC annotations. Myeloid subtypes were defined as signaling sources, while SMCs were defined as signaling targets. Width of the edges in the circle plot depicts the weight/strength of the interactions in (A), (C), and (D). ECM, extracellular matrix. See also Figure S5 and Table S5.



(legend on next page)

DISCUSSION

This study generates a comprehensive single-cell transcriptomic atlas of human atherosclerosis including 118,578 high-quality cells from atherosclerotic coronary and carotid arteries. By performing systematic benchmarking of integration methods, we mitigated data overcorrection while separating major cell lineages (Figures S1C–S1H; method details). Notably, we define cell subtypes that have not been previously identified from individual human atherosclerosis scRNA-seq studies. We leverage this scRNA-seq reference to investigate SMC phenotypic modulation in humans and identify etiologic SMC subtypes for CVD GWAS traits, including CAD and CAC. We derive insights into SMC-myeloid cell crosstalk and gene regulatory networks in atherosclerotic lesions. Finally, we uncover two candidate SMC subtype-specific markers of plaque stability and calcification.

An inherent challenge in single-cell studies is cell cluster labeling, particularly when defining cell states/phenotypes. Our systematic labeling approach allowed corroboration of reported SMC murine phenotypes,[18–20] but it also unveiled rare SMC clusters including a "foam-like" state, supporting previous evidence of an SMC-derived foam-like phenotype upon exposure to lipoproteins[96,97] and in human lesions.[98] Tissue digestion and/or single-cell isolation procedures may explain the low abundance of foam-like SMCs in individual studies. Foam-like cells expressed lipid metabolism genes (e.g., *APOE*, *APOC1*) but not other traditional foamy macrophage markers, in line with previous findings.[98] Expression of ECM genes such as *TIMP1* suggests SMC-derived foam cells may acquire a unique gene signature, suggesting that modulated SMCs accumulate lipids in plaques without adopting monocyte-derived foam macrophage transcriptional signatures.[18]

Previous work from our group and others[50,95,99] established a substantial contribution of SMCs toward CAD risk. Our granular SMC annotations now refine etiologic SMC phenotypes for CVD traits. We further deconvolve the SMC signal to prioritize fibromyocytes, transitional SMCs, and foam-like SMCs underlying CAD. Using GWAS-based heritability, we demonstrate enrichment of a fibromyocyte-specific transcriptional signature for MIs, supporting their role in plaque stability, and the importance of targeting this SMC phenotype for future therapeutic interventions. We also link SMC-derived FCs to CAC, an established pathological hallmark of subclinical and advanced atherosclerotic lesions.[100] These findings highlight the power of integrating single-cell and human genetic data to discover atherosclerosis-relevant cell states/phenotypes such as fibromyocytes and FCs.

Though fibromyocytes and FCs originate from SMCs,[18–20,101] these two ECM-producing phenotypes play distinct roles in plaque stability[9] and may follow shared or unique fates. Using pseudotime analysis, we infer transitional SMCs could adopt either fibromyocyte or FC fates. This, however, does not preclude the possibility that fibromyocytes could be primed to become FCs as Cheng et al.[76] and our TF activity inference results suggest, but it implies an alternative modulation path. However, additional lineage-tracing experiments are needed to address this possibility. To refine the SMC regulatory landscape, we validated known vascular TFs such as TCF21 and SOX9 but also mitogenic TFs such as MYC in transitional SMCs and AP-1 and TEAD factors across fibromyocytes and FCs. We also observed higher motif accessibility of the CAD-associated gene *SMAD3*[99,102] in ECM-rich SMCs (fibromyocytes/FCs) compared to contractile SMCs in our snATAC data.[84]

The precise causal mechanism of *SMAD3* in SMC calcification and CAD risk remains inconclusive. It has been shown that *SMAD3* can stimulate chondrogenesis in mesenchymal stem cells by enhancing SOX9 transcriptional activity[103,104] and is upregulated in murine FCs along with other key bone factors such as BMP2 and SOX9.[86] Concordantly, our TF regulon and snATAC-seq analyses confirmed increased SMAD3 activity along the transitional SMC-fibromyocyte-FC continuum. We note that although these results point toward a potential association between increased *SMAD3* expression and CAD risk through an SMC osteochondrogenic transition, prior murine studies showed that an SMC-specific *Smad3* knockout increased plaque burden.[76] This discrepancy may be attributed to distinct roles across different states of the SMC modulation continuum and differences intrinsic to species and/or vascular beds, which further reinforces the importance of more in-depth human single-cell studies.

Our integrative analyses also revealed a potential role for *LTBP1* in SMCs. Early studies demonstrated a critical role for *LTBP1* in intimal thickening and promoting SMC migration in diabetic rat aortas.[105] Expression quantitative trait loci (eQTL) studies have also identified regulatory variants associated with *LTBP1* expression in cultured SMCs[106] and aortic tissue.[107] Here, we showed *LTBP1* expression in native contractile SMCs in the medial layer of coronary arteries and also in the fibrous neointima. Furthermore, *LTBP1* expression in the fibrous cap of coronary lesions in addition to upregulation in stable regions of carotid plaques[93] and the meta-analyzed fibromyocyte cluster suggests a protective role

Figure 6. Pseudotime and TF inference activity for ECM-rich SMC phenotypes

(A) UMAP embeddings showing the contractile-to-modulated SMC pseudotime trajectory calculated with Monocle3.[78] SMC phenotypes for this analysis included contractile, transitional SMCs, fibromyocytes, and fibrochondrocytes (FCs). The numbered circle depicts the trajectory root defined as the contractile SMC subset with highest *MYH11* expression.

(B) Heatmap showing the top 500 SMC differentially expressed (DE) genes across pseudotime. Bolded genes depict GWAS CAD SMC effector genes differentially expressed across pseudotime. The black arrow indicates a gradual increase in pseudotime toward the right.

(C and D) Cubic spline interpolation of SCTransform-normalized gene expression as a function of pseudotime. Genes plotted include hits from Monocle3 and Seurat DE tests (FDR < 0.05). Genes from SMCs to fibromyocytes trajectory: *FN1*, *LGALS3*, *AEBP1*, *LTBP1*, *PDGFRB*. From SMCs to FC trajectory: *COL1A2*, *IBSP*, *CRTAC1*, *COMP*, *MMP2*.

(E) TF activity prediction with VIPER[79] leveraging DoRothEA regulons[80] for SMC phenotypes. Only regulons with high confidence scores (A–C) were used for this analysis. Highly variable TFs were selected for plotting, and scale indicates relative predicted activity. SMC, smooth muscle cell; GWAS, genome-wide association study; CAD, coronary artery disease; TF, transcription factor. See also method details and Figures S3 and S6.

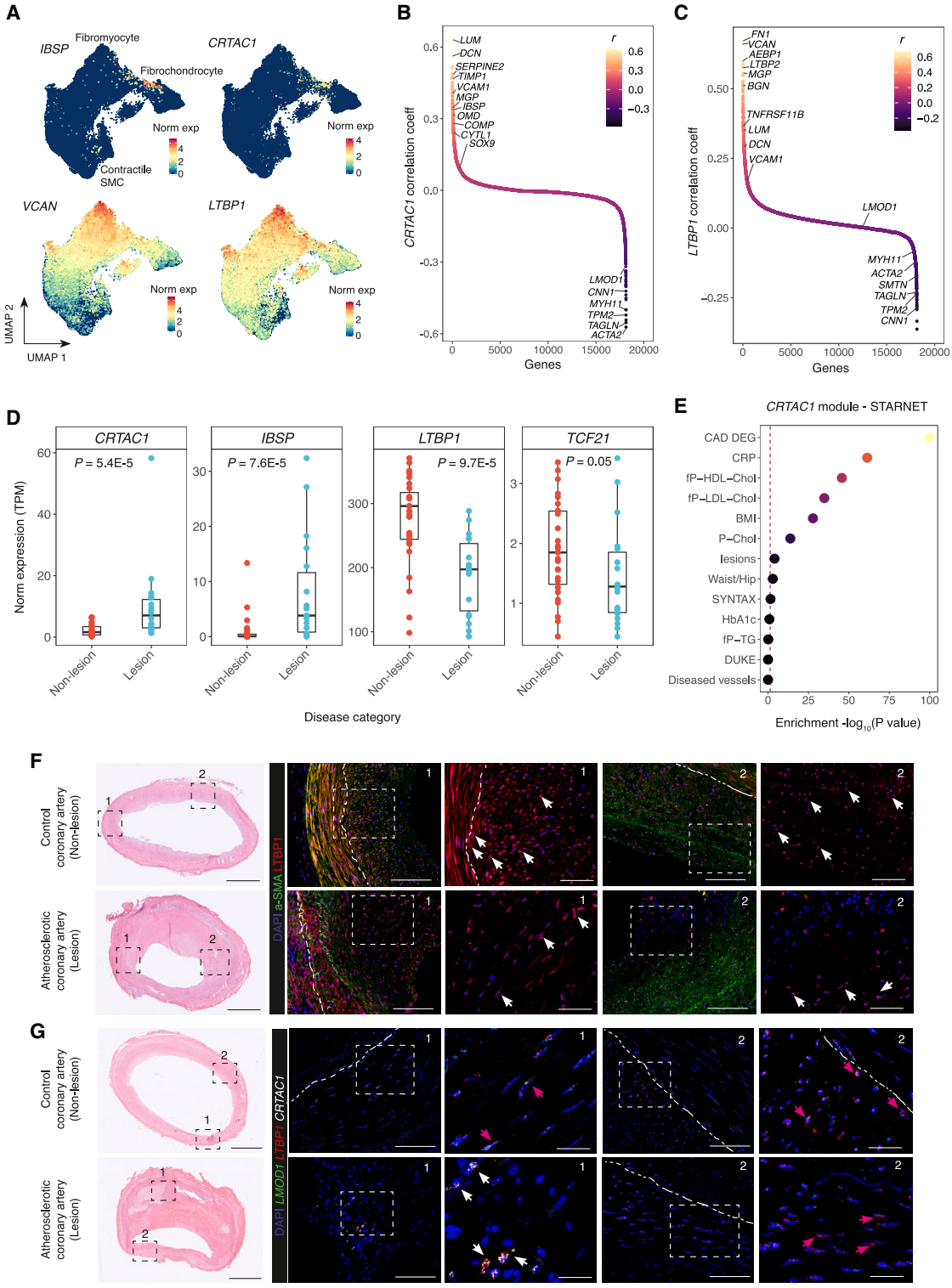


Figure 7. *CRTAC1* and *LTBP1* are candidate markers of atherosclerosis progression

(A) UMAP embeddings of *CRTAC1* and *LTBP1* expression within the subset defined in Figure 4A. *IBSP* and *VCAN* are provided as calcification and fibromyocyte markers, respectively.

(B) Pearson correlation plot of *CRTAC1* versus all genes across contractile SMC and FC clusters. Examples of osteochondrogenic and contractile genes are shown.

(legend continued on next page)

mediated by SMCs transitioning toward a more synthetic state. However, we note *LTBP1* is not an established CAD GWAS locus, and regulatory variants found in QTL studies may be associated with disease progression rather than initiation. We also identified *CRTAC1* as a previously uncharacterized FC marker with increased RNA and protein expression observed in advanced coronary lesions. Importantly, *CRTAC1* has been recently established as a novel CAD GWAS locus in participants of European ancestry,[48] and our results suggest a potential mechanism of action through SMC modulation into an osteochondrogenic phenotype.

Limitations of the study

We acknowledge some limitations in our study related to lesion status comparisons and the challenges of annotating cell subtypes using transcriptomics data. The limitation arises from the sourced datasets included in this meta-analysis (e.g., non-lesion samples came from patients with non-ischemic dilated cardiomyopathies,[17] and inflammatory cell populations could be consequences of myocardial inflammation or secondary subclinical diffuse intimal thickening), and while the majority of the cell types were balanced across samples, it is difficult to disentangle biologically meaningful processes or technical factors. The inherent sparsity of scRNA-seq data presents a challenge when annotating cell types with lower RNA, and future iterations of this reference could explore the effects of imputing meta-analyzed data and refining markers for less abundant cell types. Furthermore, while we performed careful level 2 annotations to identify *CRTAC1* and *LTBP1* as promising SMC biomarkers, we acknowledge the need for systematic protein-level and experimental validation to confirm their precise functions in atherosclerosis. Nonetheless, our atlas represents a valuable resource of the most relevant atherosclerotic arterial beds providing a comprehensive map of cell diversity in human lesions. With newly generated single-cell datasets, there will be a need to address the variability of reported phenotypes, and single-cell meta-analyses are a powerful tool to do so. Capture of both robust and subtle signals is expected to improve with future iterations. Ultimately, this will catalyze mechanistic and translational studies contributing toward innovative therapeutic strategies for CAD.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Quality control and normalization of scRNA-seq sequencing libraries
  - Integration benchmark of scRNA libraries
  - Metrics used for benchmarking
  - Cell type annotations
  - Genetic prioritization of etiologic cell types for GWAS traits
  - Cell communication analyses
  - Pseudotime analyses for SMCs
  - TF activity inference using DorothEA regulons
  - Human coronary artery tissue procurement
  - Coronary artery calcification GWAS meta-analysis data
  - Pearson correlation calculations and gene set enrichment analyses
  - Gene expression analysis in coronary artery datasets
  - STARNET regulatory networks and clinical trait enrichment analysis
  - Differential expression in SMCs
  - Image analysis of human coronary artery tissues
  - Human carotid plaque scRNA-seq data validation
  - Coronary artery snATAC-seq tissue processing and data analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
- ADDITIONAL RESOURCES

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113380.

(C) Pearson correlation plot of *LTBP1* versus all genes across contractile SMC and fibromyocyte clusters. Examples of contractile and ECM-related genes upregulated during SMC modulation[18] are shown.
(D) Bulk RNA-seq expression of *CRTAC1*, *IBSP*, *LTBP1*, and *TCF21* in human coronary arteries from lesion (n = 27) and non-lesion samples (n = 21). The y axis represents normalized expression counts (TPMs). p values were calculated using a non-parametric Wilcoxon rank-sum Test. Boxplots represent the median and the inter-quartile (IQR) range with upper (75%) and lower (25%) quartiles shown, and each dot represents a separate individual.
(E) Clinical trait enrichment for *CRTAC1*-containing module in subclinical mammary artery STARNET gene regulatory network. Pearson's correlation p values (gene level) were aggregated for each co-expression module using a two-sided Fisher's exact test. Case/control differential gene expression enrichment was estimated by a hypergeometric test.
(F) Representative H&E and LTBP1 immunofluorescence staining in human atherosclerotic coronary artery segments from normal (control/subclinical, n = 4) and stage IV–V lesions (n = 4) (Stary classification). Two regions of interest (ROIs) per sample stained with LTBP1 are shown on the right, respectively, at 10× and 20× magnification, with LTBP1 (red), alpha-smooth muscle actin (a-SMA; green), and DAPI-stained nuclei (blue). Arrows point to LTBP1$^+$ cells. Scale bars represent 0.1 mm (left) and 30 μm (right) in ROIs.
(G) Representative H&E and RNAscope staining of *LTBP1* and *CRTAC1* in human coronary artery segments from normal (control/subclinical, n = 2) and stage IV–V lesions (n = 4) (Stary classification). Two ROIs per condition are shown on the right at 40× magnification, with *LTBP1* (red), *LMOD1* (green), *CRTAC1* (white), and DAPI-stained nuclei (blue). Pink arrows point to *LTBP1*$^+$/*LMOD1*$^+$ cells, while white arrows point to *CRTAC1*$^+$ cells. Scale bars represent 0.1 mm (left) and 30 μm (right) in ROIs in RNAscope. Scale bars represent 1 mm for H&E images in (G) and (H). ECM, extracellular matrix; H&E, hematoxylin and eosin. See also Figures S4 and S7–S9 and Table S6.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (C.L.M.: R01HL148239 and R01HL164577; C.Z.: R35GM133712; R.M.: R01HL142809 and R01HL159514; C.L.L.C.: K01HL164687; J.L.M.B: R01HL125863; J.C.K.:R01HL130423, R01HL135093, and R01HL148167-01A1; A.V.F.: R01 HL141425), the American Heart Association (A.W.T.: 20POST35120545; J.V.M.: AHA909150; J.L.M.B.: A14SFRN20840000), the Swedish Research Council and Heart Lung Foundation (J.L.M.B.: 2018-02529 and 20170265), the Fondation Leducq "PlaqOmics" (C.L.M., S.W.v.d.L., J.L.M.B., M.M., and A.V.F.: 18CVD02), the European Union H2020 TO_AITION (S.W.v.d.L.: 848146), and the Single-Cell Data Insights award from the Chan Zuckerberg Initiative, LLC, and Silicon Valley Community Foundation (C.L.M., C.Z., and S.W.v.d.L.). We thank Dr. Timothy Bullock and Dr. Sanja Arandjelovic for their insights regarding myeloid and lymphoid subtype annotations.

AUTHOR CONTRIBUTIONS

C.L.M. supervised research primarily related to the study. M.M., J.C.K., J.L.M.B., R.M., N.C.S., C.Z., and S.W.v.d.L. jointly supervised research secondarily related to the study. J.V.M. and C.L.M. conceived and designed the experiments. J.V.M., G.A., A.W.T., K.T., C.L.L.C., and Q.C. performed the experiments. J.V.M. performed the statistical analyses. J.V.M., A.W.T., C.J.H., C.A.Y., and K.T. analyzed the data. M.B., M.K., P.P., M.M., J.C.K., J.L.M.B., S.W.v.d.L., A.V.F., C.A.Y., and P.T.S. contributed reagents/materials/analysis tools. J.V.M., D.W., G.A., A.W.T., C.J.H., N.C.S., and C.L.M. wrote the paper.

DECLARATION OF INTERESTS

J.L.M.B. is a shareholder in Clinical Gene Network AB and has a vested interest in STARNET. S.W.v.d.L. has received funding from Roche for unrelated work. C.L.M. has received funding from AstraZeneca for an unrelated project. R.M. has received funding from Angea Biotherapeutics and Amgen and serves as a consultant for Myokardia/BMS, Renovacor, Epizon Pharma, and Third Pole, all unrelated to the current project. J.C.K. is the recipient of an Agilent Thought Leader Award, which includes funding for research that is unrelated to the current project. A.V.F. has received institutional research support from 480 Biomedical; 4C Medical; 4Tech; Abbott; Accumedical; Amgen; Biosensors; Boston Scientific; Cardiac Implants; Celonova; Claret Medical; Concept Medical; Cook; CSI; DuNing, Inc; Edwards LifeSciences; Emboline; Endotronix; Envision Scientific; Lutonix/Bard; Gateway; Lifetech; Limflo; MedAlliance; Medtronic; Mercator; Merill; Microport Medical; Microvention; Mitraalign; Mitra assist; NAMSA; Nanova; Neovasc; NIPRO; Novogate; Occlutech; OrbusNeich Medical; Phenox; Profusa; Protembis; Qool; Recor; Senseonics, Shockwave; Sinomed; Spectranetics; Surmodics; Symic; Vesper; W.L. Gore; and Xeltis. A.V.F. has received honoraria from Abbott Vascular; Biosensors; Boston Scientific; Celonova; Cook Medical; CSI; Lutonix Bard; and Sinomed; Terumo Corporation and is a consultant to Amgen; Abbott Vascular; Boston Scientific; Celonova; Cook Medical; Lutonix Bard; and Sinomed.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: February 16, 2023
Revised: September 12, 2023
Accepted: October 20, 2023
Published: November 10, 2023

REFERENCES

1. Tsao, C.W., Aday, A.W., Almarzooq, Z.I., Alonso, A., Beaton, A.Z., Bittencourt, M.S., Boehme, A.K., Buxton, A.E., Carson, A.P., Commodore-Mensah, Y., et al. (2022). Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association. Circulation *145*, e153–e639. https://doi.org/10.1161/CIR.0000000000001052.
2. Cassar, A., Holmes, D.R., Rihal, C.S., and Gersh, B.J. (2009). Chronic Coronary Artery Disease: Diagnosis and Management. Mayo Clin. Proc. *84*, 1130–1146. https://doi.org/10.4065/mcp.2009.0391.
3. Khera, A.V., and Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. Nat. Rev. Genet. *18*, 331–344. https://doi.org/10.1038/nrg.2016.160.
4. Stary, H.C., Chandler, A.B., Dinsmore, R.E., Fuster, V., Glagov, S., Insull, W., Rosenfeld, M.E., Schwartz, C.J., Wagner, W.D., and Wissler, R.W. (1995). A Definition of Advanced Types of Atherosclerotic Lesions and a Histological Classification of Atherosclerosis. Circulation *92*, 1355–1374. https://doi.org/10.1161/01.cir.92.5.1355.
5. Virmani, R., Kolodgie, F.D., Burke, A.P., Farb, A., and Schwartz, S.M. (2000). Lessons from sudden coronary death: a comprehensive morphological classification scheme for atherosclerotic lesions. Arterioscler. Thromb. Vasc. Biol. *20*, 1262–1275. https://doi.org/10.1161/01.atv.20.5.1262.
6. Libby, P. (2002). Inflammation in atherosclerosis. Nature *420*, 868–874. https://doi.org/10.1038/nature01323.
7. Libby, P. (2021). The changing landscape of atherosclerosis. Nature *592*, 524–533. https://doi.org/10.1038/s41586-021-03392-8.
8. Shankman, L.S., Gomez, D., Cherepanova, O.A., Salmon, M., Alencar, G.F., Haskins, R.M., Swiatlowska, P., Newman, A.A.C., Greene, E.S., Straub, A.C., et al. (2015). KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. Nat. Med. *21*, 628–637. https://doi.org/10.1038/nm.3866.
9. Espinosa-Diez, C., Mandi, V., Du, M., Liu, M., and Gomez, D. (2021). Smooth muscle cells in atherosclerosis: Clones but not carbon copies. JVS. Vasc. Sci. *2*, 136–148. https://doi.org/10.1016/j.jvssci.2021.02.002.
10. Evrard, S.M., Lecce, L., Michelis, K.C., Nomura-Kitabayashi, A., Pandey, G., Purushothaman, K.-R., d'Escamard, V., Li, J.R., Hadri, L., Fujitani, K., et al. (2016). Endothelial to mesenchymal transition is common in atherosclerotic lesions and is associated with plaque instability. Nat. Commun. *7*, 11853. https://doi.org/10.1038/ncomms11853.
11. Andueza, A., Kumar, S., Kim, J., Kang, D.-W., Mumme, H.L., Perez, J.I., Villa-Roel, N., and Jo, H. (2020). Endothelial Reprogramming by Disturbed Flow Revealed by Single-Cell RNA and Chromatin Accessibility Study. Cell Rep. *33*, 108491. https://doi.org/10.1016/j.celrep.2020.108491.
12. Winkels, H., Ehinger, E., Vassallo, M., Buscher, K., Dinh, H.Q., Kobiyama, K., Hamers, A.A.J., Cochain, C., Vafadarnejad, E., Saliba, A.-E., et al. (2018). Atlas of the Immune Cell Repertoire in Mouse Atherosclerosis Defined by Single-Cell RNA-Sequencing and Mass Cytometry. Circ. Res. *122*, 1675–1688. https://doi.org/10.1161/CIRCRESAHA.117.312513.
13. Cochain, C., Vafadarnejad, E., Arampatzi, P., Pelisek, J., Winkels, H., Ley, K., Wolf, D., Saliba, A.-E., and Zernecke, A. (2018). Single-Cell RNA-Seq Reveals the Transcriptional Landscape and Heterogeneity of Aortic Macrophages in Murine Atherosclerosis. Circ. Res. *122*, 1661–1674. https://doi.org/10.1161/CIRCRESAHA.117.312509.
14. Fernandez, D.M., Rahman, A.H., Fernandez, N.F., Chudnovskiy, A., Amir, E.-A.D., Amadori, L., Khan, N.S., Wong, C.K., Shamailova, R., Hill, C.A., et al. (2019). Single-cell immune landscape of human atherosclerotic plaques. Nat. Med. *25*, 1576–1588. https://doi.org/10.1038/s41591-019-0590-4.
15. Depuydt, M.A.C., Prange, K.H.M., Slenders, L., Örd, T., Elbersen, D., Boltjes, A., de Jager, S.C.A., Asselbergs, F.W., de Borst, G.J., Aavik, E., et al. (2020). Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. Circ. Res. *127*, 1437–1455. https://doi.org/10.1161/CIRCRESAHA.120.316770.
16. Alsaigh, T., Evans, D., Frankel, D., and Torkamani, A. (2022). Decoding the transcriptome of calcified atherosclerotic plaque at single-cell

resolution. Commun. Biol. *5*, 1084. https://doi.org/10.1038/s42003-022-04056-7.

17. Hu, Z., Liu, W., Hua, X., Chen, X., Chang, Y., Hu, Y., Xu, Z., and Song, J. (2021). Single-Cell Transcriptomic Atlas of Different Human Cardiac Arteries Identifies Cell Types Associated With Vascular Physiology. Arterioscler. Thromb. Vasc. Biol. *41*, 1408–1427. https://doi.org/10.1161/ATVBAHA.120.315373.
18. Wirka, R.C., Wagh, D., Paik, D.T., Pjanic, M., Nguyen, T., Miller, C.L., Kundu, R., Nagao, M., Coller, J., Koyano, T.K., et al. (2019). Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. Nat. Med. *25*, 1280–1289. https://doi.org/10.1038/s41591-019-0512-5.
19. Pan, H., Xue, C., Auerbach, B.J., Fan, J., Bashore, A.C., Cui, J., Yang, D.Y., Trignano, S.B., Liu, W., Shi, J., et al. (2020). Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. Circulation *142*, 2060–2075. https://doi.org/10.1161/CIRCULATIONAHA.120.048378.
20. Alencar, G.F., Owsiany, K.M., Karnewar, S., Sukhavasi, K., Mocci, G., Nguyen, A.T., Williams, C.M., Shamsuzzaman, S., Mokry, M., Henderson, C.A., et al. (2020). Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. Circulation *142*, 2045–2059. https://doi.org/10.1161/CIRCULATIONAHA.120.046672.
21. Zernecke, A., Winkels, H., Cochain, C., Williams, J.W., Wolf, D., Soehnlein, O., Robbins, C.S., Monaco, C., Park, I., McNamara, C.A., et al. (2020). Meta-Analysis of Leukocyte Diversity in Atherosclerotic Mouse Aortas. Circ. Res. *127*, 402–426. https://doi.org/10.1161/CIRCRESAHA.120.316903.
22. Conklin, A.C., Nishi, H., Schlamp, F., Örd, T., Õunap, K., Kaikkonen, M.U., Fisher, E.A., and Romanoski, C.E. (2021). Meta-Analysis of Smooth Muscle Lineage Transcriptomes in Atherosclerosis and Their Relationships to In Vitro Models. Immunometabolism *3*, e210022. https://doi.org/10.20900/immunometab20210022.
23. Vallejo, J., Cochain, C., Zernecke, A., and Ley, K. (2021). Heterogeneity of immune cells in human atherosclerosis revealed by scRNA-Seq. Cardiovasc. Res. *117*, 2537–2543. https://doi.org/10.1093/cvr/cvab260.
24. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W., and Robinson, M.D. (2021). Doublet identification in single-cell sequencing data using scDblFinder. F1000Res. *10*, 979. https://doi.org/10.12688/f1000research.73600.1.
25. Yang, S., Corbett, S.E., Koga, Y., Wang, Z., Johnson, W.E., Yajima, M., and Campbell, J.D. (2020). Decontamination of ambient RNA in single-cell RNA-seq with DecontX. Genome Biol. *21*, 57. https://doi.org/10.1186/s13059-020-1950-6.
26. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.
27. Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296. https://doi.org/10.1186/s13059-019-1874-1.
28. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.-R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
29. Hie, B., Bryson, B., and Berger, B. (2019). Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. Nat. Biotechnol. *37*, 685–691. https://doi.org/10.1038/s41587-019-0113-3.
30. Büttner, M., Miao, Z., Wolf, F.A., Teichmann, S.A., and Theis, F.J. (2019). A test metric for assessing single-cell RNA-seq batch correction. Nat. Methods *16*, 43–49. https://doi.org/10.1038/s41592-018-0254-1.
31. Tabula Sapiens Consortium*; Jones, R.C., Karkanias, J., Krasnow, M.A., Pisco, A.O., Quake, S.R., Salzman, J., Yosef, N., Bulthaup, B., Brown, P., et al. (2022). The Tabula Sapiens: A multiple-organ, single-cell transcriptomic atlas of humans. Science *376*, eabl4896. https://doi.org/10.1126/science.abl4896.
32. Domínguez Conde, C., Xu, C., Jarvis, L.B., Rainbow, D.B., Wells, S.B., Gomes, T., Howlett, S.K., Suchanek, O., Polanski, K., King, H.W., et al. (2022). Cross-tissue immune cell analysis reveals tissue-specific features in humans. Science *376*, eabl5197. https://doi.org/10.1126/science.abl5197.
33. Tran, H.T.N., Ang, K.S., Chevrier, M., Zhang, X., Lee, N.Y.S., Goh, M., and Chen, J. (2020). A benchmark of batch-effect correction methods for single-cell RNA sequencing data. Genome Biol. *21*, 12. https://doi.org/10.1186/s13059-019-1850-9.
34. Luecken, M.D., Büttner, M., Chaichoompu, K., Danese, A., Interlandi, M., Mueller, M.F., Strobl, D.C., Zappia, L., Dugas, M., Colomé-Tatché, M., and Theis, F.J. (2022). Benchmarking atlas-level data integration in single-cell genomics. Nat. Methods *19*, 41–50. https://doi.org/10.1038/s41592-021-01336-8.
35. Meyer, K.B., Wilbrey-Clark, A., Nawijn, M., and Teichmann, S.A. (2021). The Human Lung Cell Atlas: a transformational resource for cells of the respiratory system. Lung Stem Cells Develop. Health Disease. *91*, 158–174. https://doi.org/10.1183/2312508x.10010920.
36. Timshel, P.N., Thompson, J.J., and Pers, T.H. (2020). Genetic mapping of etiologic brain cell types for obesity. Elife *9*, e55851. https://doi.org/10.7554/eLife.55851.
37. Winkels, H., Ehinger, E., Ghosheh, Y., Wolf, D., and Ley, K. (2018). Atherosclerosis in the single-cell era. Curr. Opin. Lipidol. *29*, 389–396. https://doi.org/10.1097/MOL.0000000000000537.
38. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.-R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235. https://doi.org/10.1038/ng.3404.
39. Skene, N.G., Bryois, J., Bakken, T.E., Breen, G., Crowley, J.J., Gaspar, H.A., Giusti-Rodriguez, P., Hodge, R.D., Miller, J.A., Muñoz-Manchado, A.B., et al. (2018). Genetic identification of brain cell types underlying schizophrenia. Nat. Genet. *50*, 825–833. https://doi.org/10.1038/s41588-018-0129-5.
40. Finucane, H.K., Reshef, Y.A., Anttila, V., Slowikowski, K., Gusev, A., Byrnes, A., Gazal, S., Loh, P.-R., Lareau, C., Shoresh, N., et al. (2018). Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat. Genet. *50*, 621–629. https://doi.org/10.1038/s41588-018-0081-4.
41. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.
42. van der Harst, P., and Verweij, N. (2018). Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease. Circ. Res. *122*, 433–443. https://doi.org/10.1161/CIRCRESAHA.117.312086.
43. Hartiala, J.A., Han, Y., Jia, Q., Hilser, J.R., Huang, P., Gukasyan, J., Schwartzman, W.S., Cai, Z., Biswas, S., Trégouët, D.A., et al. (2021). Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. Eur. Heart J. *42*, 919–933. https://doi.org/10.1093/eurheartj/ehaa1040.
44. Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M., et al. (2018). GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat. Commun. *9*, 5141. https://doi.org/10.1038/s41467-018-07340-5.
45. Giri, A., Hellwege, J.N., Keaton, J.M., Park, J., Qiu, C., Warren, H.R., Torstenson, E.S., Kovesdy, C.P., Sun, Y.V., Wilson, O.D., et al. (2019).

Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat. Genet. *51*, 51–62. https://doi.org/10.1038/s41588-018-0303-9.

46. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. *51*, 404–413. https://doi.org/10.1038/s41588-018-0311-9.

47. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med. *12*, e1001779. https://doi.org/10.1371/journal.pmed.1001779.

48. Tcheandjieu, C., Zhu, X., Hilliard, A.T., Clarke, S.L., Napolioni, V., Ma, S., Lee, K.M., Fang, H., Chen, F., Lu, Y., et al. (2022). Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. Nat. Med. *28*, 1679–1692. https://doi.org/10.1038/s41591-022-01891-3.

49. Kavousi, M., Bos, M.M., Barnes, H.J., Lino Cardenas, C.L., Wong, D., Lu, H., Hodonsky, C.J., Landsmeer, L.P.L., Turner, A.W., Kho, M., et al. (2023). Multi-ancestry genome-wide study identifies effector genes and druggable pathways for coronary artery calcification. Nat. Genet. *55*, 1651–1664. https://doi.org/10.1038/s41588-023-01518-4.

50. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E., et al. (2022). Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nat. Genet. *54*, 804–816. https://doi.org/10.1038/s41588-022-01069-0.

51. Dawson, A., Wang, Y., Li, Y., LeMaire, S.A., and Shen, Y.H. (2021). New Technologies With Increased Precision Improve Understanding of Endothelial Cell Heterogeneity in Cardiovascular Health and Disease. Front. Cell Dev. Biol. *9*, 679995. https://doi.org/10.3389/fcell.2021.679995.

52. Tombor, L.S., John, D., Glaser, S.F., Luxán, G., Forte, E., Furtado, M., Rosenthal, N., Baumgarten, N., Schulz, M.H., Wittig, J., et al. (2021). Single cell sequencing reveals endothelial plasticity with transient mesenchymal activation after myocardial infarction. Nat. Commun. *12*, 681. https://doi.org/10.1038/s41467-021-20905-1.

53. Kalucka, J., de Rooij, L.P.M.H., Goveia, J., Rohlenova, K., Dumas, S.J., Meta, E., Conchinha, N.V., Taverna, F., Teuwen, L.-A., Veys, K., et al. (2020). Single-Cell Transcriptome Atlas of Murine Endothelial Cells. Cell *180*, 764–779.e20. https://doi.org/10.1016/j.cell.2020.01.015.

54. Willemsen, L., and de Winther, M.P. (2020). Macrophage subsets in atherosclerosis as defined by single-cell technologies. J. Pathol. *250*, 705–714. https://doi.org/10.1002/path.5392.

55. Tillie, R.J.H.A., van Kuijk, K., and Sluimer, J.C. (2020). Fibroblasts in atherosclerosis: heterogeneous and plastic participants. Curr. Opin. Lipidol. *31*, 273–278. https://doi.org/10.1097/MOL.0000000000000700.

56. Chowdhury, R.R., D'Addabbo, J., Huang, X., Veizades, S., Sasagawa, K., Louis, D.M., Cheng, P., Sokol, J., Jensen, A., Tso, A., et al. (2022). Human Coronary Plaque T Cells Are Clonal and Cross-React to Virus and Self. Circ. Res. *130*, 1510–1530. https://doi.org/10.1161/CIRCRESAHA.121.320090.

57. Tanaka, M., Koyama, T., Sakurai, T., Kamiyoshi, A., Ichikawa-Shindo, Y., Kawate, H., Liu, T., Xian, X., Imai, A., Zhai, L., et al. (2016). The endothelial adrenomedullin-RAMP2 system regulates vascular integrity and suppresses tumour metastasis. Cardiovasc. Res. *111*, 398–409. https://doi.org/10.1093/cvr/cvw166.

58. Jiang, Y., Liu, H., Liu, W.-J., Tong, H.-B., Chen, C.-J., Lin, F.-G., Zhuo, Y.-H., Qian, X.-Z., Wang, Z.-B., Wang, Y., et al. (2016). Endothelial Aquaporin-1 (AQP1) Expression Is Regulated by Transcription Factor Mef2c. Mol. Cell. *39*, 292–298. https://doi.org/10.14348/molcells.2016.2223.

59. Rutkovskiy, A., Bliksøen, M., Hillestad, V., Amin, M., Czibik, G., Valen, G., Vaage, J., Amiry-Moghaddam, M., and Stensløkken, K.O. (2013). Aquaporin-1 in cardiac endothelial cells is downregulated in ischemia, hypoxia and cardioplegia. J. Mol. Cell. Cardiol. *56*, 22–33. https://doi.org/10.1016/j.yjmcc.2012.12.002.

60. Elmasri, H., Ghelfi, E., Yu, C.-W., Traphagen, S., Cernadas, M., Cao, H., Shi, G.-P., Plutzky, J., Sahin, M., Hotamisligil, G., and Cataltepe, S. (2012). Endothelial cell-fatty acid binding protein 4 promotes angiogenesis: role of stem cell factor/c-kit pathway. Angiogenesis *15*, 457–468. https://doi.org/10.1007/s10456-012-9274-0.

61. Xu, S., Ilyas, I., Little, P.J., Li, H., Kamato, D., Zheng, X., Luo, S., Li, Z., Liu, P., Han, J., et al. (2021). Endothelial Dysfunction in Atherosclerotic Cardiovascular Diseases and Beyond: From Mechanism to Pharmacotherapies. Pharmacol. Rev. *73*, 924–967. https://doi.org/10.1124/pharmrev.120.000096.

62. Newman, A.A.C., Serbulea, V., Baylis, R.A., Shankman, L.S., Bradley, X., Alencar, G.F., Owsiany, K., Deaton, R.A., Karnewar, S., Shamsuzzaman, S., et al. (2021). Multiple cell types contribute to the atherosclerotic lesion fibrous cap by PDGFRβ and bioenergetic mechanisms. Nat. Metab. *3*, 166–181. https://doi.org/10.1038/s42255-020-00338-8.

63. Johnson, L.A., and Jackson, D.G. (2010). Inflammation-induced secretion of CCL21 in lymphatic endothelium is a key regulator of integrin-mediated dendritic cell transmigration. Int. Immunol. *22*, 839–849. https://doi.org/10.1093/intimm/dxq435.

64. Lim, H.Y., Lim, S.Y., Tan, C.K., Thiam, C.H., Goh, C.C., Carbajo, D., Chew, S.H.S., See, P., Chakarov, S., Wang, X.N., et al. (2018). Hyaluronan Receptor LYVE-1-Expressing Macrophages Maintain Arterial Tone through Hyaluronan-Mediated Regulation of Smooth Muscle Cell Collagen. Immunity *49*, 1191. https://doi.org/10.1016/j.immuni.2018.12.009.

65. Boltjes, A., and van Wijk, F. (2014). Human dendritic cell functional specialization in steady-state and inflammation. Front. Immunol. *5*, 131. https://doi.org/10.3389/fimmu.2014.00131.

66. Villani, A.-C., Satija, R., Reynolds, G., Sarkizova, S., Shekhar, K., Fletcher, J., Griesbeck, M., Butler, A., Zheng, S., Lazo, S., et al. (2017). Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. Science *356*, eaah4573. https://doi.org/10.1126/science.aah4573.

67. Swiecki, M., and Colonna, M. (2015). The multifaceted biology of plasmacytoid dendritic cells. Nat. Rev. Immunol. *15*, 471–485. https://doi.org/10.1038/nri3865.

68. Jahrsdörfer, B., Vollmer, A., Blackwell, S.E., Maier, J., Sontheimer, K., Beyer, T., Mandel, B., Lunov, O., Tron, K., Nienhaus, G.U., et al. (2010). Granzyme B produced by human plasmacytoid dendritic cells suppresses T-cell expansion. Blood *115*, 1156–1165. https://doi.org/10.1182/blood-2009-07-235382.

69. Ziegler, S.F., Ramsdell, F., and Alderson, M.R. (1994). The activation antigen CD69. Stem Cell. *12*, 456–465. https://doi.org/10.1002/stem.5530120502.

70. Schluns, K.S., Kieper, W.C., Jameson, S.C., and Lefrançois, L. (2000). Interleukin-7 mediates the homeostasis of naïve and memory CD8 T cells in vivo. Nat. Immunol. *1*, 426–432. https://doi.org/10.1038/80868.

71. Li, J., Huston, G., and Swain, S.L. (2003). IL-7 promotes the transition of CD4 effectors to persistent memory cells. J. Exp. Med. *198*, 1807–1815. https://doi.org/10.1084/jem.20030725.

72. Andreatta, M., and Carmona, S.J. (2021). UCell: Robust and scalable single-cell gene signature scoring. Comput. Struct. Biotechnol. J. *19*, 3796–3798. https://doi.org/10.1016/j.csbj.2021.06.043.

73. Zhang, M.J., Hou, K., Dey, K.K., Sakaue, S., Jagadeesh, K.A., Weinand, K., Taychameekiatchai, A., Rao, P., Pisco, A.O., Zou, J., et al. (2022). Polygenic enrichment distinguishes disease associations of individual cells in single-cell RNA-seq data. Nat. Genet. *54*, 1572–1580. https://doi.org/10.1038/s41588-022-01167-z.

74. Dobnikar, L., Taylor, A.L., Chappell, J., Oldach, P., Harman, J.L., Oerton, E., Dzierzak, E., Bennett, M.R., Spivakov, M., and Jørgensen, H.F. (2018).

Disease-relevant transcriptional signatures identified in individual smooth muscle cells from healthy mouse vessels. Nat. Commun. *9*, 4567. https://doi.org/10.1038/s41467-018-06891-x.

75. Speer, M.Y., Yang, H.-Y., Brabb, T., Leaf, E., Look, A., Lin, W.-L., Frutkin, A., Dichek, D., and Giachelli, C.M. (2009). Smooth muscle cells give rise to osteochondrogenic precursors and chondrocytes in calcifying arteries. Circ. Res. *104*, 733–741. https://doi.org/10.1161/CIRCRESAHA.108.183053.
76. Cheng, P., Wirka, R.C., Kim, J.B., Kim, H.-J., Nguyen, T., Kundu, R., Zhao, Q., Sharma, D., Pedroza, A., Nagao, M., et al. (2022). Smad3 regulates smooth muscle cell fate and mediates adverse remodeling and calcification of the atherosclerotic plaque. Nat. Cardiovasc. Res. *1*, 322–333. https://doi.org/10.1038/s44161-022-00042-8.
77. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. *12*, 1088. https://doi.org/10.1038/s41467-021-21246-9.
78. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.
79. Alvarez, M.J., Shen, Y., Giorgi, F.M., Lachmann, A., Ding, B.B., Ye, B.H., and Califano, A. (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat. Genet. *48*, 838–847. https://doi.org/10.1038/ng.3593.
80. Garcia-Alonso, L., Holland, C.H., Ibrahim, M.M., Turei, D., and Saez-Rodriguez, J. (2019). Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. *29*, 1363–1375. https://doi.org/10.1101/gr.240663.118.
81. Grootaert, M.O.J., Finigan, A., Figg, N.L., Uryga, A.K., and Bennett, M.R. (2021). SIRT6 Protects Smooth Muscle Cells From Senescence and Reduces Atherosclerosis. Circ. Res. *128*, 474–491. https://doi.org/10.1161/CIRCRESAHA.120.318353.
82. Steck, E., Bräun, J., Pelttari, K., Kadel, S., Kalbacher, H., and Richter, W. (2007). Chondrocyte secreted CRTAC1: a glycosylated extracellular matrix molecule of human articular cartilage. Matrix Biol. *26*, 30–41. https://doi.org/10.1016/j.matbio.2006.09.006.
83. Yang, X., Harkins, L.K., Zubanova, O., Harrington, A., Kovalenko, D., Nadeau, R.J., Chen, P.-Y., Toher, J.L., Lindner, V., Liaw, L., and Friesel, R. (2008). Overexpression of Spry1 in chondrocytes causes attenuated FGFR ubiquitination and sustained ERK activation resulting in chondrodysplasia. Dev. Biol. *321*, 64–76. https://doi.org/10.1016/j.ydbio.2008.05.555.
84. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E., et al. (2022). Author Correction: Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nat. Genet. *54*, 1259. https://doi.org/10.1038/s41588-022-01142-8.
85. Granja, J.M., Corces, M.R., Pierce, S.E., Bagdatli, S.T., Choudhry, H., Chang, H.Y., and Greenleaf, W.J. (2021). ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. Nat. Genet. *53*, 403–411. https://doi.org/10.1038/s41588-021-00790-6.
86. Kim, J.B., Zhao, Q., Nguyen, T., Pjanic, M., Cheng, P., Wirka, R., Travisano, S., Nagao, M., Kundu, R., and Quertermous, T. (2020). Environment-Sensing Aryl Hydrocarbon Receptor Inhibits the Chondrogenic Fate of Modulated Smooth Muscle Cells in Atherosclerotic Lesions. Circulation *142*, 575–590. https://doi.org/10.1161/CIRCULATIONAHA.120.045981.
87. Sanchez, C., Bay-Jensen, A.-C., Pap, T., Dvir-Ginzberg, M., Quasnichka, H., Barrett-Jolley, R., Mobasheri, A., and Henrotin, Y. (2017). Chondrocyte secretome: a source of novel insights and exploratory biomarkers of osteoarthritis. Osteoarthritis Cartilage *25*, 1199–1209. https://doi.org/10.1016/j.joca.2017.02.797.
88. Szilagyi, I., Vallerga, C., Waarsing, J.H., Schiphof, D., Bierma-Zeinstra, S.M.A., and Van Meurs, J. (2021). OP0111 PLASMA PROTEOMICS IDENTIFIES CRTAC1 AS BIOMARKER FOR OSTEOARTHRITIS SEVERITY AND PROGRESSION. Ann. Rheum. Dis. *80*, 61.1–62. https://doi.org/10.1136/annrheumdis-2021-eular.1888.
89. Troilo, H., Steer, R., Collins, R.F., Kielty, C.M., and Baldock, C. (2016). Independent multimerization of Latent TGFβ Binding Protein-1 stabilized by cross-linking and enhanced by heparan sulfate. Sci. Rep. *6*, 34347. https://doi.org/10.1038/srep34347.
90. Woo, S.-H., Kyung, D., Lee, S.H., Park, K.S., Kim, M., Kim, K., Kwon, H.-J., Won, Y.-S., Choi, I., Park, Y.-J., et al. (2023). TXNIP Suppresses the Osteochondrogenic Switch of Vascular Smooth Muscle Cells in Atherosclerosis. Circ. Res. *132*, 52–71. https://doi.org/10.1161/CIRCRESAHA.122.321538.
91. Pustlauk, W., Westhoff, T.H., Claeys, L., Roch, T., Geißler, S., and Babel, N. (2020). Induced osteogenic differentiation of human smooth muscle cells as a model of vascular calcification. Sci. Rep. *10*, 5951. https://doi.org/10.1038/s41598-020-62568-w.
92. Kim, J.-S., Ryoo, Z.Y., and Chun, J.-S. (2007). Cytokine-like 1 (Cytl1) regulates the chondrogenesis of mesenchymal cells. J. Biol. Chem. *282*, 29359–29367. https://doi.org/10.1074/jbc.M700965200.
93. Mahmoud, A.D., Ballantyne, M.D., Miscianinov, V., Pinel, K., Hung, J., Scanlon, J.P., Iyinikkel, J., Kaczynski, J., Tavares, A.S., Bradshaw, A.C., et al. (2019). The Human-Specific and Smooth Muscle Cell-Enriched LncRNA SMILR Promotes Proliferation by Regulating Mitotic CENPF mRNA and Drives Cell-Cycle Progression Which Can Be Targeted to Limit Vascular Remodeling. Circ. Res. *125*, 535–551. https://doi.org/10.1161/CIRCRESAHA.119.314876.
94. Koplev, S., Seldin, M., Sukhavasi, K., Ermel, R., Pang, S., Zeng, L., Bankier, S., Di Narzo, A., Cheng, H., Meda, V., et al. (2022). A mechanistic framework for cardiometabolic and coronary artery diseases. Nat. Cardiovasc. Res. *1*, 85–100. https://doi.org/10.1038/s44161-021-00009-1.
95. Slenders, L., Landsmeer, L.P.L., Cui, K., Depuydt, M.A.C., Verwer, M., Mekke, J., Timmerman, N., van den Dungen, N.A.M., Kuiper, J., de Winther, M.P.J., et al. (2022). Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. Eur. Heart J. Open *2*, oeab043. https://doi.org/10.1093/ehjopen/oeab043.
96. Mietus-Snyder, M., Gowri, M.S., and Pitas, R.E. (2000). Class A Scavenger Receptor Up-regulation in Smooth Muscle Cells by Oxidized Low Density Lipoprotein. J. Biol. Chem. *275*, 17661–17670. https://doi.org/10.1074/jbc.275.23.17661.
97. Beyea, M.M., Reaume, S., Sawyez, C.G., Edwards, J.Y., O'Neil, C., Hegele, R.A., Pickering, J.G., and Huff, M.W. (2012). The oxysterol 24(s),25-epoxycholesterol attenuates human smooth muscle-derived foam cell formation via reduced low-density lipoprotein uptake and enhanced cholesterol efflux. J. Am. Heart Assoc. *1*, e000810. https://doi.org/10.1161/JAHA.112.000810.
98. Allahverdian, S., Chehroudi, A.C., McManus, B.M., Abraham, T., and Francis, G.A. (2014). Contribution of intimal smooth muscle cells to cholesterol accumulation and macrophage-like cells in human atherosclerosis. Circulation *129*, 1551–1559. https://doi.org/10.1161/CIRCULATIONAHA.113.005015.
99. Liu, B., Pjanic, M., Wang, T., Nguyen, T., Gloudemans, M., Rao, A., Castano, V.G., Nurnberg, S., Rader, D.J., Elwyn, S., et al. (2018). Genetic Regulatory Mechanisms of Smooth Muscle Cells Map to Coronary Artery Disease Risk Loci. Am. J. Hum. Genet. *103*, 377–388. https://doi.org/10.1016/j.ajhg.2018.08.001.
100. Jinnouchi, H., Sato, Y., Sakamoto, A., Cornelissen, A., Mori, M., Kawakami, R., Gadhoke, N.V., Kolodgie, F.D., Virmani, R., and Finn, A.V. (2020). Calcium deposition within coronary atherosclerotic lesion: Implications for plaque stability. Atherosclerosis *306*, 85–95. https://doi.org/10.1016/j.atherosclerosis.2020.05.017.
101. Naik, V., Leaf, E.M., Hu, J.H., Yang, H.-Y., Nguyen, N.B., Giachelli, C.M., and Speer, M.Y. (2012). Sources of cells that contribute to atherosclerotic intimal calcification: an in vivo genetic fate mapping study. Cardiovasc. Res. *94*, 545–554. https://doi.org/10.1093/cvr/cvs126.

102. Iyer, D., Zhao, Q., Wirka, R., Naravane, A., Nguyen, T., Liu, B., Nagao, M., Cheng, P., Miller, C.L., Kim, J.B., et al. (2018). Coronary artery disease genes SMAD3 and TCF21 promote opposing interactive genetic programs that regulate smooth muscle cell differentiation and disease risk. PLoS Genet. *14*, e1007681. https://doi.org/10.1371/journal.pgen.1007681.

103. Furumatsu, T., Tsuda, M., Taniguchi, N., Tajima, Y., and Asahara, H. (2005). Smad3 induces chondrogenesis through the activation of SOX9 via CREB-binding protein/p300 recruitment. J. Biol. Chem. *280*, 8343–8350. https://doi.org/10.1074/jbc.M413913200.

104. Furumatsu, T., Ozaki, T., and Asahara, H. (2009). Smad3 activates the Sox9-dependent transcription on chromatin. Int. J. Biochem. Cell Biol. *41*, 1198–1204. https://doi.org/10.1016/j.biocel.2008.10.032.

105. Kanzaki, T., and Otabe, M. (2003). Latent transforming growth factor-beta binding protein-1, a component of latent transforming growth factor-beta complex, accelerates the migration of aortic smooth muscle cells in diabetic rats through integrin-beta3. Diabetes *52*, 824–828. https://doi.org/10.2337/diabetes.52.3.824.

106. Aherrahrou, R., Guo, L., Nagraj, V.P., Aguhob, A., Hinkle, J., Chen, L., Yuhl Soh, J., Lue, D., Alencar, G.F., Boltjes, A., et al. (2020). Genetic Regulation of Atherosclerosis-Relevant Phenotypes in Human Vascular Smooth Muscle Cells. Circ. Res. *127*, 1552–1565. https://doi.org/10.1161/CIRCRESAHA.120.317415.

107. Consortium, G.; GTEx Consortium (2017). Genetic effects on gene expression across human tissues. Nature *550*, 204–213. https://doi.org/10.1038/nature24277.

108. Durinck, S., Spellman, P.T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat. Protoc. *4*, 1184–1191. https://doi.org/10.1038/nprot.2009.97.

109. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15. https://doi.org/10.1186/s13059-017-1382-0.

110. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J., and Peterson, H. (2020). gprofiler2 – an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. F1000Res. *9*, ELIXIR-709. https://doi.org/10.12688/f1000research.24956.2.

111. Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics *26*, 976–978. https://doi.org/10.1093/bioinformatics/btq064.

112. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

113. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat. Methods *12*, 1061–1063. https://doi.org/10.1038/nmeth.3582.

114. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. *12*, 323. https://doi.org/10.1186/1471-2105-12-323.

115. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559. https://doi.org/10.1186/1471-2105-9-559.

116. Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One *5*, e12776. https://doi.org/10.1371/journal.pone.0012776.

117. Van den Berge, K., Perraudeau, F., Soneson, C., Love, M.I., Risso, D., Vert, J.-P., Robinson, M.D., Dudoit, S., and Clement, L. (2018). Observation weights unlock bulk RNA-seq tools for zero inflation and single-cell applications. Genome Biol. *19*, 24. https://doi.org/10.1186/s13059-018-1406-4.

118. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

119. Schep, A.N., Wu, B., Buenrostro, J.D., and Greenleaf, W.J. (2017). chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. Nat. Methods *14*, 975–978. https://doi.org/10.1038/nmeth.4401.

120. Hodonsky, C.J., Turner, A.W., Khan, M.D., Barrientos, N.B., Methorst, R., Ma, L., Lopez, N.G., Mosquera, J.V., Auguste, G., Farber, E., et al. (2023). Integrative multi-ancestry genetic analysis of gene regulation in coronary arteries prioritizes disease risk loci. Preprint at medRxiv. https://doi.org/10.1101/2023.02.09.23285622.

121. Xi, N.M., and Li, J.J. (2021). Protocol for executing and benchmarking eight computational doublet-detection methods in single-cell RNA sequencing data analysis. STAR Protoc. *2*, 100699. https://doi.org/10.1016/j.xpro.2021.100699.

122. Sikkema, L., Strobl, D., Zappia, L., Madissoon, E., Markov, N.S., Zaragosi, L., Ansari, M., Arguel, M., Apperloo, L., Bécavin, C., et al. (2022). An integrated cell atlas of the human lung in health and disease. Preprint at bioRxiv, 2022–03. https://doi.org/10.1101/2022.03.10.483747.

123. Molecular Architecture of the Mouse Nervous System (2018). Cell *174*, 999–1014.e22. https://doi.org/10.1016/j.cell.2018.06.021.

124. Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. Nucleic Acids Res. *38*, 4218–4230. https://doi.org/10.1093/nar/gkq130.

125. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195. https://doi.org/10.1126/science.1222794.

126. 1000 Genomes Project Consortium; Abecasis, G.R., Auton, A., Brooks, L.D., DePristo, M.A., Durbin, R.M., Handsaker, R.E., Kang, H.M., Marth, G.T., and McVean, G.A. (2012). An integrated map of genetic variation from 1,092 human genomes. Nature *491*, 56–65. https://doi.org/10.1038/nature11632.

127. Shu, L., Zhao, Y., Kurt, Z., Byars, S.G., Tukiainen, T., Kettunen, J., Orozco, L.D., Pellegrini, M., Lusis, A.J., Ripatti, S., et al. (2016). Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. BMC Genom. *17*, 874. https://doi.org/10.1186/s12864-016-3198-9.

128. Jaakkola, M.K., Seyednasrollah, F., Mehmood, A., and Elo, L.L. (2017). Comparison of methods to detect differentially expressed genes between single-cell populations. Briefings Bioinf. *18*, 735–743. https://doi.org/10.1093/bib/bbw057.

129. McMurdie, P.J., and Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS One *8*, e61217. https://doi.org/10.1371/journal.pone.0061217.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit polyclonal anti-LTBP1 (Latent transforming growth factor β binding protein 1)	Proteintech	Cat# 26855-1-AP; RRID: AB_2880658
Mouse monoclonal anti-ACTA2-FITC (Actin, α-Smooth Muscle-FITC)	Sigma-Aldrich	Cat# F3777; RRID: AB_476977
Rabbit polyclonal anti-CRTAC1 (Cartilage acidic protein 1)	Proteintech	Cat# 13001-1-AP; RRID: AB_10638787
Donkey anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 555	Thermo Fisher Scientific	Cat# A-31572; RRID: AB_162543
Critical commercial assays		
RNAscope Multiplex Fluorescent Assay v2	Advanced Cell Diagnostics	Cat# 323270
LMOD1 RNAscope probe	Advanced Cell Diagnostics	Cat# 444141
LTBP1 RNAscope probe	Advanced Cell Diagnostics	Cat# 523281
CRTAC1 RNAscope probe	Advanced Cell Diagnostics	Cat# 1094551
HybEz oven unit	Advanced Cell Diagnostics	Cat# 321710
Vulcan Fast Red Chromogen Kit 2	Biocare Medical	Cat# FR805S
Peroxidase 1	Biocare Medical	Cat# PX968H
MACH4 MP AP-Polymer	Biocare Medical	Cat# MRAP536G
Deposited data		
scRNA-seq data for human coronary and carotid tissues	Wirka et al.[18]; Pan et al.[19]; Alsaigh et al.[16]; Hu et al.[17]; Slenders et al.[95]	GEO: GSE131778 (Wirka), GSE155512 (Pan), GSE159677 (Alsaigh), Zenodo: https://doi.org/10.5281/zenodo.6032099 (Hu) and Dataverse: https://dataverse.nl/dataverse/umculab (Slenders)
Bulk human coronary RNA sequencing (raw files and TPMs)	In-house	GEO: GSE225650
snATAC sequencing (raw and processed data)	In-house	GEO: accession IDs GSE175621 and GSE188422
Bulk human carotid RNA sequencing (FPKMs)	Mahmoud et al.[93]	GEO: accession ID GSE120521
GWAS summary statistics for coronary artery disease (CAD) - Multi-ancestry	Tcheandjieu et al.[48]	GWAS catalog: https://www.ebi.ac.uk/gwas/studies/GCST90132305
GWAS summary statistics for CAD - European ancestry	van der Harst et al.[42]	GWAS catalog: https://www.ebi.ac.uk/gwas/studies/GCST005194
GWAS summary statistics for myocardial infarction	Hartiala et al.[43]	GWAS catalog: https://www.ebi.ac.uk/gwas/studies/GCST011365
GWAS summary statistics for coronary artery calcification	Kavousi et al.[49]	GWAS catalog: https://www.ebi.ac.uk/gwas/studies/GCST90278456
GWAS summary statistics for carotid plaque	Franceschini et al.[44]	GWAS catalog: https://www.ebi.ac.uk/gwas/studies/GCST001231
GWAS summary statistics for pulse pressure	Giri et al.[45]	dbGaP: accession ID phs001672.v1.p1
GWAS summary statistics for Alzheimer's disease	Jansen et al.[46]	https://ctg.cncr.nl/software/summary_statistics
GWAS summary statistics for white blood cell count	Sudlow et al.[47]	http://www.nealelab.is/uk-biobank

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
GWAS summary statistics for type 2 diabetes	Sudlow et al.[47]	http://www.nealelab.is/uk-biobank
GWAS summary statistics for body mass index	Sudlow et al.[47]	http://www.nealelab.is/uk-biobank
Software and algorithms		
Custom analysis scripts	In-house	https://github.com/MillerLab-CPHG/Human_athero_scRNA_meta https://doi.org/10.5281/Zenodo.8418736
Custom scRNA processing pipeline	In-house	https://github.com/MillerLab-CPHG/scRNAutils https://doi.org/10.5281/Zenodo.8418743
R 4.0.3	R Core Team	https://www.r-project.org/
R package Seurat v4.1.0	Stuart et al.[26]	https://github.com/satijalab/seurat
R package sctransform v0.3.3	Hafemeister et al.[27]	https://github.com/satijalab/sctransform
R package scDblFinder v1.4.0	Germain et al.[24]	https://github.com/plger/scDblFinder
R package celda v1.6.1	Yang et al.[25]	https://github.com/campbio/celda
R package harmony v1.0.0	Korsunsky et al.[28]	https://github.com/immunogenomics/harmony
R package scanorama v1.7.1	Hie et al.[29]	https://github.com/brianhie/scanorama
R package reticulate v1.18	CRAN	https://rstudio.github.io/reticulate/
R package lisi v1.0.0	Korsunsky et al.[28]	https://github.com/immunogenomics/LISI
R package kBET v0.99.6	Buttner et al.[30]	https://github.com/theislab/kBET
R package DescTools v0.99.49	CRAN	https://andrisignorell.github.io/DescTools/
R package cluster v2.1.0	CRAN	https://cran.r-project.org/web/packages/cluster/index.html
R package SeuratDisk v0.0.0.9019	Stuart et al.[26]	https://github.com/mojaveazure/seurat-disk
Celltypist command line tool v1.0.0	Dominguez Conde et al.[32]	https://github.com/Teichlab/celltypist
R package biomaRt v2.46	Durinck et al.[108]	https://bioconductor.org/packages/release/bioc/html/biomaRt.html
R package UCell v1.3.1	Andreatta et al.[72]	https://github.com/carmonalab/UCell
Python package CELLECT v1.3.0	Timshel et al.[36]	https://github.com/perslab/CELLECT
S-LDSC v1.0.0	Finucane et al.[40]	https://github.com/bulik/ldsc
MAGMA v1.0.7	De Leeuw et al.[41]	https://ctg.cncr.nl/software/magma
scDRS command line tool v1.0.2	Zhang et al.[73]	https://github.com/martinjzhang/scDRS
Python package scanpy v1.9.3	Wolf et al.[109]	https://github.com/scverse/scanpy
R package CellChat v1.5.0	Jin et al.[77]	https://github.com/sqjin/CellChat
R package Monocle3 v1.0.0	Cao et al.[78]	https://cole-trapnell-lab.github.io/monocle3/
R package pheatmap v1.0.12	CRAN	https://github.com/raivokolde/pheatmap
R package DoRothEA v1.8.0	Garcia-Alonso et al.[80]	https://saezlab.github.io/dorothea/
R package VIPER v1.24.0	Alvarez et al.[79]	https://bioconductor.org/packages/release/bioc/html/viper.html
R package gProfiler2 v0.2.1	Kolberg et al.[110]	https://cran.r-project.org/web/packages/gprofiler2/index.html
R package gosemsim v2.16.1	Yu et al.[111]	https://bioconductor.org/packages/release/bioc/html/GOSemSim.html
STAR v2.7.3a	Dobin et al.[112]	https://github.com/alexdobin/STAR/releases
WASP v0.3.4	van de Geijn et al.[113]	https://github.com/bmvdgeijn/WASP
RSEM package	Li and Dewey[114]	https://deweylab.github.io/RSEM/README.html

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
WGCNA	Langfelder and Horvath[115]	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/
GENIE3	Huynh et al.[116]	https://github.com/vahuynh/GENIE3
R package zinbwave v1.12.0	Van den Berge et al.[117]	https://github.com/drisso/zinbwave
R package DESeq2 v1.30.1	Love et al.[118]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
CellRanger ATAC v1.2.0	10x Genomics	https://support.10xgenomics.com/single-cell-atac/software/pipelines/latest/installation
R package ArchR v1.0.2	Granja et al.[85]	https://www.archrproject.com/
R package ChromVAR v1.12.0	Schep et al.[119]	https://github.com/GreenleafLab/chromVAR

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Dr. Clint Miller (clintm@virginia.edu).

Materials availability

This study did not generate new unique reagents. Commercially available reagents are listed in the key resources table.

Data and code availability

- This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table. All raw and processed bulk RNA-sequencing data are available in the Gene Expression Omnibus (GEO) database and are publicly available as of the date of the publication under the GEO accession number GSE225650.
- All original code has been deposited on GitHub and is publicly available as of the date of publication. DOIs for the analysis and pipeline repositories (Zenodo: https://doi.org/10.5281/Zenodo.8418736 and https://doi.org/10.5281/Zenodo.8418743) are also listed in the key resources table.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Collection of human coronary artery samples for bulk RNA-seq data generation as well as histology and immunofluorescence/RNA-scope analyses described in this manuscript complies with ethical guidelines for human subjects research under approved Institutional Review Board (IRB) protocols at Stanford University (no. 4237 and no. 11925) and the University of Virginia (no. 20008), for the procurement and use of human tissues and information, respectively. The mean age for individuals in the paired bulk transcriptomic analyses was 51 ± 14 years. Furthermore, 27% of the included samples were female. Additional details about age, sex, and ancestry for each individual can be found in Hodonsky et al.[120] Additional information on coronary artery tissue collection can be found in method details. Metadata from human samples used in the scRNA-seq meta-analysis can be found in Table S1.

METHOD DETAILS

Quality control and normalization of scRNA-seq sequencing libraries

Raw count matrices from each library across the 4 studies were downloaded from GEO and Zenodo (Table S1). Processing for each of the 22 sequencing libraries was standardized in the following manner: Each library was loaded into the R programming environment (v4.0.3) using Seurat[26] (v4.1.0). For each library we did a first pass of clustering with SCTransform normalization[27] without removing low-quality cells.

In order to remove doublets, we referred to a recent benchmark of doublet-removal tools[121] and chose the scDblFinder R package[24] (v1.4.0) given its superior accuracy compared to other tools. Seurat objects for each library were converted to SingleCellExperiment objects and used as input to generate artificial doublets using the cluster-based modality of scDblFinder. Briefly, scDblFinder creates a k-nearest-neighbors graph using the union of real cells and artificial doublets and estimates the density of artificial doublets in the neighborhood of each cell. Since artificial-doublet generation approaches tend to display slight variance

across different runs, we only kept consensus doublets from 3 iterations of the above-described process. Cell-barcodes that were marked as doublets were then removed from each raw counts matrix.

Ambient RNA contamination is a key issue during 10x protocols and can negatively impact clustering and extraction of gene markers. To filter out reads from ambient RNA, we ran DecontX[25] within the celda R package (v1.6.1) in doublets-filtered raw counts matrices using default parameters. The decontaminated raw count matrices output by DecontX were then added into each Seurat object. We then set quality filters to keep cells that had 1) $\geq$ 200 and $\leq$ 4000 uniquely expressed genes 2) $\geq$ 200 and $\leq$ 20000 UMIs 3) $\leq$ 10% of reads mapped to the mitochondrial genome; cells with high percentages of reads mapped to mitochondrial genomes are considered to be low quality as this indicates cell membrane breaches and 4) $\leq$ 5% of reads mapped to hemoglobin genes since these cells likely depict contaminating erythrocytes as done in Alencar et al., 2021.

Raw count matrices were then normalized using SCTransform[27] with parameter (vst.flavor = v2), which accounts for sequencing depth variability across cells. This omits the need for heuristic steps such as log-transformation and it has been shown to improve variable gene selection, dimensionality reduction and differential expression.[27] To avoid clustering results confounded by cell cycle state, cell cycle variance was regressed out during SCTransform normalization. We then carried out dimensionality reduction of the normalized counts matrix using PCA. The first 30 principal components (PCs) were used as input for clustering in Seurat, which relies on a Shared-Nearest-Neighbors (SNN) and Louvain community detection approach. We then applied Uniform Manifold Approximation and Projection (UMAP) non-linear dimensionality reduction using the first 30 PCs. UMAP embeddings were used for visualization of Louvain clustering results. Processed matrices were then stored as seurat objects for batch-correction.

Integration benchmark of scRNA libraries

The choice of single-cell integration approach highly depends on the context of the individual datasets, the magnitude of batch effect, and cell number. In order to harmonize processed sequencing libraries, we selected the following methods recommended from three recent benchmarks[33,34,122] of single-cell transcriptomic data integration: Canonical Correlation Analysis + Mutual Nearest Neighbors (CCA+MNN), reciprocal PCA (rPCA)[26] (v4.1.0), Harmony[28] (v1.0) and Scanorama[29] (v1.7.1). For the benchmark, we used a subset of the data including 3 studies: Wirka et al., 2019,[18] Alsaigh et al., 2022[16] and Pan et al., 2020[19] Libraries from these studies were integrated as follows.

CCA+MNN

We created a list of selected Seurat objects and then selected 3000 most highly variable genes. Integration with those variable genes was done using the PrepSCTIntegration, FindIntegrationAnchors and IntegrateData functions. The batch-corrected expression matrix was then used for PCA dimensionality reduction, creation of the shared-nearest-neighbors (SNN) graph using 30 PCs and visualization with UMAP embeddings.

Harmony

Libraries were first stored into a list and highly variable genes extracted using the function SelectIntegrationFeatures. Libraries were merged into a single Seurat object and the list of highly variable genes was used for PCA dimensionality reduction. We used the first 30 PCs as input for RunHarmony from the harmony package (v1.0.0), setting sequencing libraries (sample column in metadata) as the variables to correct for batch effects. Harmony embeddings were used for subsequent generation of the SNN graph, Louvain clustering and visualization with UMAP by setting reduction = ”Harmony” within the FindNeighbors and RunUMAP Seurat functions and using the first 30 PCs.

Reciprocal PCA (rPCA)

we created a list of processed Seurat objects and extracted the 3000 most highly variable genes using SelectIntegrationFeatures. We then ran PCA across each library using the 3000 variable genes, identified integration anchors using FindIntegrationAnchors setting reduction = ”rpca” and harmonized datasets using IntegrateData. As done for CCA, the batch-corrected expression matrix was then used for PCA dimensionality reduction, creation of the shared-nearest-neighbors (SNN) graph using 30 PCs, and Louvain clustering followed by visualization with UMAP embeddings.

Scanorama

We used the reticulate R package (v1.18) to import the Scanorama python module (v1.7.1) into the R environment. We created a list with Seurat objects containing the datasets to be integrated and stored normalized SCTransform-normalized counts and gene names for each dataset into a new list. We then batch-corrected the data using the function using the correct function from the Scanorama package setting the following parameters (return_dimred = TRUE and return_dense = TRUE). The batch-corrected expression matrix output by correct() was used to create a new Seurat object and Scanorama-produced dimensionality-reduced embeddings were inserted into the Seurat object using the CreateDimReducObject function. Scanorama embeddings were subsequently used to create a shared-nearest-neighbors (SNN) graph for Louvain clustering and for visualization with UMAP using the first 30 PCs.

Metrics used for benchmarking

After integrating libraries as described above, we then evaluated the efficiency of batch removal and conservation of biological signal for each method using the following metrics.

Running time measurements

Running times for each integration task were then measured using base R Sys.time functions. Sys.time was defined at the beginning and the end of each integration task and then the time difference was calculated as end_time - start_time. From this benchmark, we found that Harmony substantially outperformed the other 3 algorithms followed by rPCA (Figure S1C).

Local Inverse Simpson's Index (LISI) scores

Briefly, integration LISI (iLISI) scores are a measure of the batch diversity within each cell neighborhood on a k-nearest-neighbor (KNN) graph. Higher iLISI scores depict increased mixing of batches within a cell neighborhood and therefore suggest improved removal of batch effects. For each of the integration methods above mentioned, we extracted PCA embeddings (30 first PCs) from the corresponding integrated Seurat object. We then created a data frame with each row corresponding to one cell and columns depicting batch variables (''Study''). We then computed iLISI scores for each cell using the compute_lisi function from the lisi R package[28] (v1.0.0). Mean iLISI values were plotted and compared across different integration methods. iLISI scores showed CCA as the top performer followed by rPCA (Figure S1D).

We also took into account conservation of biological signals since batch removal can sometimes lead to overcorrection of the data (e.g., SMCs and Fibroblasts clustering together). We assumed that Louvain clusters should represent different cell types or subtypes, and thus calculated cell type LISI (cLISI) scores to measure conservation of biological variation. With the assumption that each cluster should generally harbor cells from the same broad lineage or subtype, we created a dataframe with each row corresponding to each cell and a column depicting Louvain cluster identities. cLISI scores for each cell were calculated and plotted as described above. Computation of cLISI scores revealed rPCA as the method with the highest conservation of biological signal (Figure S1E).

Principal component regression

Principal component analysis (PCA) is a widely used dimensionality reduction technique in single-cell analyses. PCA operates through an orthogonal transformation of the counts matrix into a set of linearly correlated variables. PCs correspond to the eigenvectors of the covariance matrix and are sorted according to the amount of explained variance of the data. The assumption of this approach is that if a strong batch effect is present in the data, it contributes to a substantial amount of variance. and the corresponding batch covariate should significantly correlate with some of the PCs. Therefore, we should obtain a strong effect size when regressing the batch covariate B and the ith PC. This effect size, or regression coefficient is an approximation of the variance explained by B in each PC (PC regression).

To perform PC regression, we extracted the PCA embeddings from each integrated object (n = 30 PCs) and created a batch vector B with batch labels for each cell. We then fit a linear model of PCs explained variance and a batch variable B using the pcRegression function from the kBET R package[30] (v0.99.6) setting n_top = 30. We report the total contribution of the batch effect to the variance in the data, which can be approximated by

$$Var\ (C|B) = \sum_{i=1}^{G} Var(C|PC_i) \cdot R^2(PC_i|B)$$

Where $Var(C|PC_i)$ is the variance of C explained by the ith PC and R^2 the effect size of regressing each principal component PCi on the batch variable B. Our PC regression analysis showed Harmony as the method where batch effects explained less of the variance in the input PC embeddings (highest $-Log_{10}(R^2 Var)$). Harmony was followed by CCA and rPCA. Out of the 4 methods, Scanorama displayed the largest amount of variance explained by batch effects (worst performer) (Figure S1F).

k-nearest-neighbors batch effect test (kBET)

The k-nearest-neighbors batch effect test (kBET) determines whether the label composition of the k-nearest-neighborhood of a cell is similar to the expected (global) label composition. This test is repeated for a random subset of cells and results are summarized as a rejection rate over all tested neighborhoods where lower rejection rates represent an improvement in the removal of batch effects. For each integration method, we extracted PCA embeddings (30 first PCs) from the corresponding Seurat object. We also created a matrix containing batch labels for each cell. We calculated observed rejection rates by using PCA embeddings and the batch label matrix as input using a custom wrapper for the kBET function from the kBET package[30] (v0.99.6). It is known that observed rejection rates output by kBET are highly sensitive to the choice of the cell's neighborhood size (k). Therefore we controlled for this variable by calculating rejection rates across a wide range of k values from a minimum (k = 10) to a maximum k 0 defined by:

$$k0 = \frac{1}{4}(Avg(Batch\ size))$$

This range was used as input for the k0 parameter of the kBET function. Since we provided our own k range and used PCA embeddings as input we set the parameters heuristics = FALSE and do.pca = FALSE. We found that for all the integration methods tested, as k approached $k0$, rejection rates approached 1, an expected behavior according to the kBET authors. To get a more accurate performance estimate across this k range, we further calculated the area under the curve (AUC) for each method separately. AUC was determined through a natural cubic spline interpolation of the observed rejection rate across the k range and subsequent numerical integration using the AUC function from the DescTools R package (v0.99.49) setting method = ''spline''. As lower rejection rates imply well-mixed batches, we reasoned that a lower AUC would in turn represent a similar conclusion. We found that CCA showed the smallest AUC (5.02). While Harmony and rPCA showed fairly comparable AUC values (5.18 and 5.22 respectively),

Scanorama yielded the largest AUC (5.39), depicting a reduced performance in comparison to the other 3 methods (Figure S1G). This result is consistent with our iLISI and PC regression results.

Cluster average silhouette widths (cASW)

Silhouette coefficients are widely used to measure the quality or ''goodness'' of clusters. Again, under the assumption that each cluster should represent a cell type or subtype, we used this metric as another way to measure biological conservation. To compute cluster average silhouette width (cASW) values, we extracted PCA embeddings from Seurat objects with CCA+MNN, rPCA, Harmony, and Scanorama integration outputs keeping the first 30 PCs. We then used these embeddings to compute an Euclidean distance matrix using the stats R package. Cluster IDs for each cell were obtained using the FindNeighbors and FindClusters functions from Seurat across a range of clustering resolutions (0.8–1.8). We then used the Euclidean distance matrix and cluster IDs as input to the silhouette function from the cluster R package (v2.1.0). Computing cASW values across a wide range of clustering resolutions allowed us to 1) control for the granularity parameter and 2) find an optimal resolution range to avoid under/over clustering of the data. We reasoned that a resolution of 0.8 was an appropriate starting point given the recommendation of the Seurat authors of using a resolution of 0.2–0.4 when dealing with very small numbers of cells (2000–4000). Strikingly, rPCA attained the highest cASW values across all of the resolutions tested (Figure S1H). On the other hand, Harmony obtained the lowest cASW scores across all clustering resolutions, suggesting that this approach might be overcorrecting the data included in this study.

Aggregated benchmark results

While we previously highlighted Harmony and CCA+MNN as the top performers in terms of batch effects removal, their lower cLISI and cASW scores suggest these methods might be overcorrecting the data to a higher degree. rPCA, on the other hand, yielded both the highest cLISI score and cASW across all the resolutions tested, suggesting improved conservation of biological signal. Therefore, we chose rPCA as it achieved the best balance between running time, removal of batch effects, and conservation of biological signal.

Cell type annotations

To annotate cell types in the integrated reference, we used a systematic approach to define broad labels (level 1) as well as more granular cell subtype labels (level 2).

Level 1 annotations

To define broad cell type partitions, we accessed public data from the Tabula Sapiens (TS) consortium (https://tabula-sapiens-portal.ds.czbiohub.org/organs). To improve the specificity of annotations, we downloaded the vasculature subset of this transcriptomic atlas. Upon downloading the TS vasculature h5ad file, this dataset was converted into a Seurat-compatible format using the SeuratDisk R package (v0.0.0.9019). To match the normalization workflow described in the scRNA sequencing library processing section, we extracted the TS vasculature raw counts matrix and normalized gene expression data using SCTransform. We then applied Seurat's reference-based transfer learning using the FindIntegrationAnchors and TransferData functions to annotate cells in our meta-analyzed reference. In this case, the TS vasculature Seurat object with author-provided cell type annotations was defined as reference for label transfer. Confidence scores of predicted labels ranging from 0 to 1 (where 1 indicates that labels were annotated in a fully unambiguous manner) were extracted from the output of TransferData and are shown in the Data Supplement. Gene markers for level 1 annotations were obtained using the PrepSCTMarkers and FindAllMarkers functions from Seurat (v4.1.0) setting the following thresholds: logFC = 0.25 and min.pct = 0.25.

Level 2 annotations for endothelial, fibroblasts and immune cells

To define more granular cell subtypes for the meta-analyzed data, we used a combination of automated and manual annotations. We first annotated cell subtypes for endothelial, myeloid and lymphoid lineages using markers from atherosclerosis murine scRNA meta-analyses of SMCs and immune cells as well as relevant human atherosclerosis scRNA studies.[21–23,51–56] Annotations using curated markers from the literature were corroborated with the assistance of experts at UVA. To further confirm and inspect immune cell subtype annotations in our reference, we used a logistic-regression with stochastic gradient-descent framework implemented by the command-line tool CellTypist.[32] CellTypist leverages a database of 20 different tissues and 19 reference datasets with a focus on myeloid and lymphoid cells. Specifically, we applied CellTypist low-hierarchy classifiers (using the Immune_All_Low.pkl and Immune_All_AddPIP.pkl models which harbor 90 and 101 cell types, respectively) to our SCT-normalized reference counts matrix using both default settings as well as the majority voting classifier. Gene markers for level 2 annotations were obtained using the PrepSCTMarkers and FindAllMarkers functions from Seurat (v4.1.0) setting the following thresholds: logFC = 0.25 and min.pct = 0.25.

Level 2 annotations for SMCs

To explore SMC diversity in human atherosclerosis, we subset the main meta-analyzed reference to include only the pericyte-SMC-fibroblast level1 partitions. This subset was then reclustered using Seurat (v.4.1.0) with a resolution of 0.9 based on an additional silhouette width benchmark. Next, gene modules (encompassing top markers from differential expression analyses) specific to contractile (n = 50), *Lgals3*$^{+}$ pioneer (n = 50), and fibrochondrocyte (n = 50) SMC phenotypes were extracted from a recent SMC lineage-traced murine scRNA meta-analysis. We also extracted a non-SMC-derived fibroblast module (n = 50) as a negative enrichment control. Genes in each module were ranked by Log2FC and then converted to human homologs nomenclature and filtered to keep those with a one-to-one orthology relationship using custom wrapper functions with the biomaRt R package[108] (v2.46). We then calculated the enrichment of murine gene modules on individual cells within the pericyte-SMC-Fibroblast human subset using the UCell R package[72] (v1.3.1). In addition to the enrichment of murine gene modules, we also obtained gene markers for each of the 17 SNN-derived clusters using the PrepSCTMarkers and FindAllMarkers functions from Seurat (v4.1.0) setting the following thresholds: logFC = 0.25

and min.pct = 0.1. Final annotations for SMC subtypes were derived based on the UCell enrichment scores along UMAP coordinates and cluster markers.

Genetic prioritization of etiologic cell types for GWAS traits

Generation of gene expression specificity matrices

Integration of scRNA and GWAS summary statistics was performed using the CELLECT snakemake workflow.[36] GWAS enrichment per cell type was carried out using gene expression specificity values for each cell type as done by other studies.[38,39] However, we derived expression specificity estimates for each gene using the approach proposed by Timshel et al. within the CELLEX snakemake workflow.[36] While previous approaches for genetic prioritization of cell types have used binary or discrete representations of cell type expression,[38,39] the CELLEX workflow uses a continuous representation of cell type expression. CELLEX combines 4 different Expression Specificity (ES) metrics into a single specificity estimate. Briefly, gene expression specificity weights (ES_W) were calculated separately for each of the following ES metrics: Gene Enrichment Score (GES),[123] Expression Proportion (EP),[39] Normalized Specificity Index (NSI)[124] and Differential Expression T-statistic (DET). ES_W values were then averaged into a single ES estimate (ES_μ), assuming equal weights for each metric. An ES_μ value represents the score that a gene is specifically expressed in a given cell type. This combined metric has been shown to be more robust than single-expression specificity measures. Accordingly, we extracted SCTransform-normalized counts for the counts matrix of interest and converted into sparse format to load into Python 3.8. Along with each matrix, we extracted a metadata matrix containing a label for each cell (level 1 or 2 annotations). We then used the counts and metadata matrices to create a CELLEX ESObject and computed ES_μ values using the eso.compute function with default settings.

Collection and pre-processing of GWAS summary statistics

We downloaded GWAS summary statistics for: CAD from Million Veterans Program[48] (N sample size = 773268; N significant loci = 194) and van der Harst et al.[42] (N sample size = 296525; N significant loci = 161); Myocardial infarction[43] (N sample size = 639000; N significant loci = 80); Coronary Artery Calcification (CAC) multi-ancestry GWAS meta-analysis[49] (N sample size = 35776; N significant loci = 11), carotid artery plaques[44] (N sample size = 48434; N significant loci = 5); pulse pressure from the Million Veterans Program[45] (N sample size = 318890; N significant loci = 331); Alzheimer disease[46] (N sample size = 455258; N significant loci = 29) type 2 diabetes (UK Biobank)[47] (N sample size = 455607; N significant loci = 92); body mass index (UK Biobank)[47] (N sample size = 457824; N significant loci = 52); White blood cell count (UK Biobank)[47] (N sample size = 444502; N significant loci = 46). UK Biobank summary statistics were downloaded from https://alkesgroup.broadinstitute.org/UKBB/.

We used custom R scripts (https://github.com/MillerLab-CPHG/Human_athero_scRNA_meta) as well as the provided mtag_munge.py python script (https://github.com/pascaltimshel/ldsc/tree/d869cfd1e9fe1abc03b65c00b8a672bd530d0617) to convert GWAS summary statistics to a format compatible with that of the CELLECT S-LDSC and MAGMA wrappers.

LD score regression applied to specifically expressed genes (LDSC-SEG) prioritization of cell types

We then performed LDSC-SEG with the gene expression specificity matrix for level 1 annotations across the above-described GWAS studies using the established CELLECT snakemake workflow as shown in https://github.com/perslab/CELLECT/wiki/CELLECT-LDSC-Tutorial. Briefly, we used stratified S-LDSC[40] (v.1.0.0) to prioritize cell types after transforming cell type ES_μ vectors into S-LDSC annotations, which broadly follows three steps: generation of annotation files, computation of annotation LD scores and fitting of annotation model coefficients. We created annotations for each cell type by assigning genes' ES_μ values to genetic variants utilizing a 100 kilobase (kb) window of the genes' transcribed regions. When a variant overlapped with multiple genes within the 100 kb window, the maximum ES_μ value was assigned. Given that the majority of trait-associated variants have been shown to be located in non-coding regions,[125] we chose a relatively large window size to capture the effects of nearby regulatory variants.

In line with recommendations from Finucane et al., we used an "all genes" annotation for each dataset tested, by assigning the value 1 to variants within 100 kb windows of all genes. We used hg19 (Ensembl v.91) reference genome coordinates to map SNPs to genes. For constructing annotations, we used 1000 Genomes Project SNPs[126] as in the default S-LDSC baseline model. We computed LD scores for cell-type annotations by using a 1-cM window (–ld-wind-cm 1). As recommended by S-LDSC authors, we restricted computation of LD scores to Hapmap3 SNPs and excluded the major histocompatibility complex (MHC) region due to its high gene density and unusual LD architecture.

For primary cell type prioritization, we jointly fit the following annotations: a) each cell type annotation b) all genes annotations c) the S-LDSC baseline model. We then ran S-LDSC with default settings and reported p values for the one-tailed test of positive association between trai heritability and cell type annotation ES_μ. We note that highly significant p values occur due to correlated cell types with true signal, whereas those negatively correlated with true signal have p values near 1. For all results, we used Bonferroni correction within a trait and dataset to control for multiple testing. The regression effect size estimate for each cell type is reported, which can be interpreted as the change in per-SNP heritability given the cell type annotation, beyond what can be explained by the "all genes" and baseline model.

LDSC-SEG across disease categories

To determine whether heritability enrichment was driven by lesion status, we split our entire meta-analyzed count matrix according to lesion category (lesion and non-lesion samples) along with the corresponding metadata. We then generated separate expression specificity (level 1 annotations) matrices for each condition (N lesion cells = 59691; N non-lesion cells = 58887) and carried out S-LDSC as detailed above.

LDSC-SEG for SMC level 2 cell type annotations

We subset the whole meta-analyzed reference Seurat object to include only cells along the pericyte-SMC-Fibroblast partitions. Metadata of this subset was used to generate the gene expression specificity matrix for level 2 annotations. Munging of GWAS summary statistics and subsequent S-LDSC analyses were performed as described above.

MAGMA-based prioritization of cell types

We used Multi-marker Analysis of GenoMic Annotation MAGMA (v1.0.7),[41] a leading software for gene set analysis to compute gene-level association p values from GWAS summary statistics. We first mapped SNPs to genes using a 10-kb window size around the gene body. We tested different window sizes (5, 10 and 100-kb) using data from Wirka et al. 2019 and found results to be consistent. Gene-level association statistics were then computed using the default test of mean SNP association (model: *snp-wise=mean*). This model allows one to combine SNP association p values in the specified window surrounding the gene into gene-level p values and Z-statistics, while accounting for LD (computed using the 1000 Genomes Project phase 3 European panel). Gene-level Z-statistics were corrected for the default MAGMA covariates: gene size, gene density, and inverse mean minor allele count as well as the log value of these variables. Next, we fit a linear regression model using the CELLECT snakemake pipeline (https://github.com/perslab/CELLECT/blob/master/cellect-magma.snakefile). This linear regression model used MAGMA gene-level Z-statistics as the dependent variable and cell type ES_{μ} scores as the independent variable. Cell prioritization p values (from the linear regression model) can be interpreted as the positive contribution of a cell type-specific expression program to a GWAS trait/disease gene-level Z-statistics.

scDRS enrichment of GWAS-tagged heritability

We used the single-cell disease relevance score (scDRS) framework[73] to identify cell subtypes driving the GWAS enrichments previously shown in level 1 annotations. scDRS integrates gene expression profiles from scRNA-seq with polygenic disease information from GWASs to associate individual cells to disease by assessing the excess expression of GWAS putative disease genes in a given cell relative to other genes with similar expression across all cells.

For these analyses, we selected CAD as a representative cardiovascular trait and immune traits showing enrichment of level 1 Macrophage annotations (WBC count and Alzheimer's disease) to better guide the specificity of CAD enrichments. To run scDRS, we selected the top 1,000 MAGMA genes weighted by their Z-scores as putative disease gene sets. Munging of MAGMA gene sets for scDRS was done through a combination of custom (https://github.com/MillerLab-CPHG/Human_athero_scRNA_meta/blob/main/scDRS/prep_MAGMA_gs_for_munging.R) and scripts provided in the scDRS command line tool (v1.0.2). We ran the enrichments in the entire meta-analyzed reference as it is recommended to use data harboring a diverse set of cells with varying relevance to the GWAS traits of interest.

To prepare the count matrix, we extracted raw counts and metadata from the meta-analyzed Seurat object and created a.h5ad object. Preprocessing of the data (scaling and log normalization) was done using Scanpy (v1.9.3).[109] Prior to the computation of normalized disease scores, we generated 250 control gene sets for each putative disease gene set (matching the mean and variance of the disease gene set). Finally, we calculated normalized disease scores using the CLI with "scdrs compute-score" adding a covariate matrix including *sex* and *disease status* variables. Normalized disease scores for each cell in the reference were then added back into the meta-analyzed Seurat object and plotted with Seurat's FeaturePlot function.

MAGMA-based gene effector analysis

To identify effector genes driving the heritability enrichment by cell type, we utilized the MAGMA gene sets for CAD and MI generated as described above and ranked genes according to their p values. We selected the top 1000 genes according to that ranking. Next, we ranked genes in prioritized cell types according to their ES_{μ} values and retained the most highly specific genes (>90% percentile). We then termed genes that were found in these two sets as "effector genes" driving the enrichment in a prioritized cell type.

Cell communication analyses

Cell communication analyses were carried out using the Cellchat R package[77] (v1.5.0). We selected the CellChat human database (Interactions considered include secreted signaling, ECM-receptor, and cell-cell contacts). First, we extracted SCTransform-normalized counts from the integrated Seurat object. For the first round of analyses, we separated cells from each disease status (lesion and non-lesion) and grouped them according to level 1 labels. We created a Cellchat object for matrices from each disease status using the createCellChat() function. We subsequently identified overexpressed genes in each condition using the identifyOverExpressedInteractions. Communication probabilities were estimated with computeCommunProb and aggregated cell communication networks calculated with the aggregateNet function. We then merged lesion and non-lesion cellchat objects using the mergeCellChat function. In order to identify pathways between Myeloid cells and SMCs that were enriched in each condition compared to the other, we input the merged Cellchat object to the function rankNet with parameters (mode = "comparison, sources.use = "Macrophage", targets.use = "SMC"). Significantly enriched pathways were denoted as those with $p < 0.05$. To further explore differentially enriched pathways with increased granularity, we created a new CellChat object using normalized counts from Macrophages and SMCs from lesions and grouped them using their respective level 2 annotations. We computed communication probabilities and aggregated cell communication networks as described above. Circle plots for specific signaling pathways were generated with the netvisualAggregate function. The top 30% of interactions (based on interaction weights/strength from computed communication probability) were used for plotting interactions between level 1-annotated cell types. Given that we had a larger number of cell types when deriving networks with level 2 labels, we chose to plot the top 15% of interactions.

Pseudotime analyses for SMCs

Cells within the pericyte-SMC-fibroblast axis were subset to contain only contractile SMCs, transitional-ECM-SMCs, fibromyocytes and fibrochondrocytes. Single-cell transcriptomic pseudotime analyses were performed using monocle3[78] (v1.0.0). Given that gene expression within this subset was normalized, the SCTransform-normalized expression matrix and corresponding metadata were extracted from the corresponding Seurat object. Metadata and SCT counts were used to create a cell_data_set object. In order to preserve clustering structure from previous analyses, we also extracted PCA/UMAP embeddings, cluster IDs, and cell type annotations from the processed Seurat object and inserted those into the corresponding slots of the cell_data_set object. A trajectory was then inferred using the learn_graph and order_cells functions setting contractile SMCs with the highest expression of *MYH11* as the root of the trajectory. DEGs across the trajectory were calculated with grapth_test and grouped into modules using the find_gene_modules function. To model gene expression dynamics across pseudotime, we extracted pseudotime assignment values for each cell in the trajectory as well as SCTransform-normalized expression values and cell type annotations from the cell_data_set object. We then wrote a custom script to plot gene expression changes as a function of pseudotime where we applied cubic spline interpolation to expression values using the geom_smooth function with parameters (method = "lm", formula = $y \sim$ splinesns(x, 3)).

To better resolve changes in gene expression as SMCs undergo phenotypic modulation, we identified genes that were differentially expressed across pseudotime using the graph_test function from monocle3 which leverages a statistic from spatial autocorrelation analysis called Moran's I. We kept only genes passing a significance threshold (FDR < 0.05). Genes were then ranked according to their significance and we selected the top 500 genes. Pseudotime values were added back into the subclustered SMC Seurat object. We then used the AverageExp function from Seurat to calculate the mean expression of the top 500 DE genes across different points in pseudotime. Finally, we performed hierarchical clustering of the pseudotime-variable genes using Euclidean distances and plotted their expression across pseudotime using the R package pheatmap (v1.0.12).

TF activity inference using DorothEA regulons

For inference of TF activity, we also used a subset of the main reference only including SMCs, transitional SMCs, fibromyocytes, and FCs. We downloaded a collection of curated TF regulons from the DoRothEA R package[80] (v1.8.0). We accessed human regulons using the dorothea_hs function and only kept those with A, B and C confident scores for a more accurate prediction of regulon activity on each cell. Confidence scores had been previously defined based on the number of supporting evidence for each regulon.[80] TF activities for each cell were then estimated with the R package VIPER (v1.24.0)[79] providing the list of filtered regulons and the processed Seurat object as input. Mean TF activities were then calculated across the SMC annotations of interest and the most variable TFs were selected for plotting.

Human coronary artery tissue procurement

Freshly explanted hearts from orthopedic heart transplant recipients were obtained at Stanford University under approved Institutional Review Board (IRB) protocols with the respective informed consents. Hearts were arrested in cardioplegic solution and rapidly transported from the operating room to an adjacent laboratory on ice. The proximal 5–6 cm of three major coronary arteries (LAD, LCX, RCA) were dissected from the epicardium, trimmed of surrounding adipose, rinsed in cold PBS, and snap-frozen in liquid nitrogen. Human coronary artery tissue specimens were also obtained at Stanford University from non-diseased donor hearts rejected for orthotopic heart transplantation and processed following the same protocol as hearts for transplant. Reasons for rejected hearts included size incompatibility, risk for cardiotoxicity or comorbidities. Tissues were de-identified and clinical and histopathology information was used to classify ischemic, non-ischemic hearts and lesion- and non-lesion-containing arteries. All normal arteries originated from hearts with left ventricular ejection fraction (LVEF) greater than 50%. Frozen tissues were transferred to the University of Virginia through a material transfer agreement and IRB-approved protocols.

Coronary artery calcification GWAS meta-analysis data

The GWAS meta-analysis for coronary artery calcification (CAC) was conducted on 16 cohorts including 26,909 participants of European ancestry and 8,867 participants of African ancestry. CAC scores were calculated from computed tomography imaging at baseline or first examination as described.[49] Genotyping quality control, imputation (1000 Genomes Phase 3), and variant filtering was performed as described. A joint meta-analysis of all available CAC GWAS was performed using a fixed-effects meta-analysis in METAL, using sample size weighted SNP *P* values. The summary statistics from each study were combined using an inverse variance-weighted meta-analysis.

Pearson correlation calculations and gene set enrichment analyses

Normalized counts for cell types of interest were extracted from the corresponding Seurat object. Matrices were transposed to define genes as variables and then we calculated pairwise Pearson correlations for a gene of interest (e.g., *CRTAC1*) with all of the other genes across the cell types of interest using apply and cor.test functions with parameters (method = "pearson") from the stats R package (v4.0.3).

For gene set enrichment analyses, we calculated DE genes as described in the above section. We ranked genes by log2 fold change values (log2FC) and extracted the top 100 hits per cell annotation. We then use the gost function within the R gProfiler2

package[110] (v.0.2.1) with parameters (order = TRUE) to weight genes according to their log2FC values. We then selected significant GO:BP ontology terms (FDR < 0.05) and ranked them according to their adjusted p values for plotting using custom functions from our scRNA_processing_utils.R script (https://github.com/MillerLab-CPHG/Human_athero_scRNA_meta). We found that the top GO:BP terms for fibrochondrocytes were highly redundant. Therefore, we used the gosemsim package[111] (v2.16.1) and a custom script adapted from (https://github.com/YuLab-SMU/clusterProfiler/blob/master/R/simplify.R) in order to calculate semantic similarity between GO:BP terms. We removed highly redundant terms accordingly.

Gene expression analysis in coronary artery datasets

RNA extraction and library construction

Total RNA was extracted from frozen coronary artery segments using the Qiagen miRNeasy Mini RNA Extraction kit (catalog #217004). Approximately 50 mg of frozen tissue was pulverized using a mortar and pestle under liquid nitrogen. Tissue powder was then further homogenized in Qiazol lysis buffer using stainless steel beads in a Bullet Blender (Next Advance) homogenizer, followed by column-based purification. RNA concentration was determined using Qubit 3.0 and RNA quality was determined using Agilent 4200 TapeStation. Samples with RNA Integrity Number (RIN) greater than 5.5 and Illumina DV_{200} values greater than 75 were included for library construction. Total RNA libraries were constructed using the Illumina TruSeq Stranded Total RNA Gold kit (catalog #20020599) and barcoded using Illumina TruSeq RNA unique dual indexes (catalog # 20022371). After re-evaluating library quality using TapeStation, individually barcoded libraries were sent to Novogene for next-generation sequencing. After passing additional QC, libraries were multiplexed and subjected to paired-end 150 bp read sequencing on an Illumina NovaSeq S4 Flowcell to a median depth of 100 million total reads (>30 G) per library.

RNA-seq processing and analysis

The raw passed filter sequencing reads obtained from Novogene were demultiplexed using the bcl2fastq script. The quality of the reads was assessed using FASTQC and the adapter sequences were trimmed using trimgalore. Trimmed reads were aligned to the hg38 human reference genome using STAR[112] (v2.7.3a) according to the GATK Best Practices for RNA-seq. To increase mapping efficiency and sensitivity, novel splice junctions discovered in a first alignment pass with high stringency, were used as annotation in a second pass to permit lower stringency alignment and therefore increase sensitivity. PCR duplicates were marked using Picard and WASP[113] (v0.3.4) was used to filter reads prone to mapping bias. Total read counts and Transcripts per million normalization (TPM) for both genes and isoforms was calculated from individual bam files using the RSEM (https://deweylab.github.io/RSEM/README.html) rsem-calculate-expression command with the paired-end option and gencode version 32 as a reference.[114]

Definition of disease categories for transcriptomic comparisons

Disease status of coronary artery segments was determined based on our previous publication.[84] For the present study, samples that are lesion-free (no evidence of atherosclerosis) or harbored subclinical/early lesions were included within the ''non-lesion group''. Samples with evidence of advanced/complex atherosclerotic lesions were included in the ''lesion'' group for comparative transcriptomic analyses.

STARNET regulatory networks and clinical trait enrichment analysis

Based on STARNET[94] multi-tissue bulk RNA-seq data, tissue-specific and cross-tissue co-expression modules were inferred using WGCNA.[115] Enrichment for clinical traits was computed by aggregating Pearson's correlation p values by co-expression module using Fisher's method. Enrichment for DE genes was calculated with the hypergeometric test using DESeq2 called genes (30% change, FDR < 0.01) adjusting for age and sex covariates. The gene regulatory network for *CRTAC1* and *LTBP1* co-expressed genes was inferred using GENIE3.[116] Weighted key driver analysis was then applied to identify hub or highly influential genes in the regulatory network using the Mergeomics R package.[127]

Differential expression in SMCs

Differential expression analyses in single-cell data are often challenging due to the substantial sparsity of the data. It has been shown that dedicated DE tools for scRNA-seq data do not provide additional benefits compared to traditional bulk RNA-seq tools.[128] Thus, we used the Zero-Inflated-based Negative Binomial Wanted Variation Extraction (ZINB-WaVE) strategy proposed by Van der Berge et al., 2018[117]. This approach identifies excess zero counts and generates gene- and cell-specific observation weights to unlock bulk RNA-seq DE pipelines such as DESeq2[118] for zero-inflated data.[117]

To model zero inflation in our SMC scRNA-seq counts matrix, we generated observation-level weights using the R zinbwave package (v1.12.0). Given that popular bulk RNA-seq methods use generalized linear models (GLMs), which readily accommodate inference based on observation-level weights, we use ZINB-WaVE weights as input for DESeq2 (v1.30.1). This approach allows for the DESeq2 inference to apply to the part of the data which is not due to zero inflation.

Prior to calculating ZINB-WaVE weights, we identified the 3000 most highly variable genes in the SMC subclustered data. We then converted the SMC Seurat object into a SingleCellExperiment object and retained only variable features to tailor the analysis to SMCs and speed up weights computation. Subsequently, we performed an additional round of filtering to remove genes with very low counts. After this filtering step, we calculated ZINB-WaVE weights using the zinbwave function with the following parameters: K = 0 and observationalWeights = TRUE. Weights were computed using a design matrix with *sex*, *arterial origin* and *lesion category* as covariates. We then used the previously computed ZINB-WaVE weights to create an object of class *DESeqDataSet*. As with the

generation of observation weights, we used a design formula where we added *sex* and *arterial origin* as covariates to avoid confounding by these factors. Given that DESeq2′ normalization procedure is based on the geometric means of counts, which are zero for genes with at least a zero count, this constrains the number of genes that can be used as input for normalization in scRNA-seq data.[117] Therefore we used the PHYLOSEQ normalization procedure[129] by setting argument type = "poscounts" within the DESeq2 function estimateSizeFactors.Taking into account recommendations for DE analysis for single-cell data, for which expected counts might be very low, we used the likelihood ratio test as implemented in the nbinomLRT function setting the minimum expected count parameter minmu = 1e-6 to identify DE genes (N lesion SMCs = 8725; N non-lesion SMCs = 12423). *LTBP1* and *CRTAC1* expression across lesion categories were then plotted using custom scripts.

Image analysis of human coronary artery tissues

Histological staining

Human coronary artery tissues were obtained as described in the above section.

Briefly, coronary artery segments were isolated from control/subclinical and advanced atherosclerotic left and right coronary artery branches. Tissues were embedded in OCT blocks, snap-frozen in liquid nitrogen, and stored at −80°C. Segments for histological staining were chosen based on sample availability and bulk RNA-seq information. OCT-embedded human coronary artery segments were cryosectioned at −20°C and 8-μm thickness. A minimum of two sections per sample were placed on each slide and then blindly stained with: Oil Red O (ORO), Picro-Sirius red (PSR) and Hematoxylin and Eosin (H&E) at the UVA Research Histology Core. Briefly, for ORO staining, frozen sections were fixed in 10% Neutral Buffered Formalin solution, washed and stained in Oil Red O solution (Polysciences #s2120) for 5 min. After washing, slides were stained in Hematoxylin solution (Richard Allen #7221) for 1 min before rinsing and mounting with aqueous mounting medium. For H&E, slides were stained using Hematoxylin 360 reagents manufactured by Leica in an automated Gemini Stainer. For PSR staining, slides were placed in Picro-Sirius red solution (Direct Red 80 in saturated aqueous solution of picric acid) for 1 h, rinsed in deionized water and washed twice in acidified water. Slides were then dehydrated in ethanol, cleared in xylene and mounted. Whole slide images were then captured at approximately 100,000 x 30,000 pixel resolution using a Hamamatsu NanoZoomer S360 Digital Slide Scanner C13220 at the Biorepository and Tissue Research Facility at UVA.

Immunofluorescence staining

We performed immunofluorescence staining in tissue sections adjacent to those used for histology. Sections were processed for immunostaining as follows: Sections were retrieved from the −80°C freezer and allowed to briefly come to room temperature for 1 min. Sections were then rehydrated in 1X PBS and then fixed in 4% formaldehyde for 10 min at room temperature (RT). This was followed by three PBS washes. Sections were then permeabilized using 0.1% Triton X-100 for 10 min at room temperature. Upon permeabilization, sections were washed with PBS and then protein blocking was done with 10% normal donkey serum for 1 h at RT. After blocking, slides were incubated with an anti-LTBP1 rabbit polyclonal antibody (Proteintech, 26855-1-AP; 1:250 in antibody dilution solution (1% BSA)) overnight at 4°C or no antibody negative control (antibody dilution solution). Optimal dilution concentration was determined with previous titrations with control tissues. Each slide had at least two sections stained with primary antibody and one section used for the negative control. Sections were washed with PBS and incubated with donkey anti-rabbit Alexa Fluor 555 conjugated secondary antibody (Thermo Fisher, A31572; 1:1000) and aSMA conjugated to FITC (Sigma, F3777; clone 1A4; 1:500) for 1 h at room temperature. Slides were then washed with PBS and nuclear counter-staining was performed with DAPI (0.1 μg/ml) (MiltenyiBiotec, 130-111-570). Slides were subsequently coverslipped with aqueous mounting media. Stained sections were visualized using an Olympus BX41 microscope under x10 and ×20 objective magnifications. Images were obtained using a Nikon color camera DS-Fi3 at an NIS Elements imaging software (v110.04.3707.E9).

Chromogen-based immunohistochemistry

Immunohistochemistry was performed on 7 um thick-frozen coronary artery sections using the Vulcan red Chromogen kit 2 (FR805S). Briefly, sections were fixed using 4% formaldehyde for 10 min, then peroxidase block was performed used peroxidase 1 (Biocare PX968H) for 5 min before a 15 min permeabilization step using 0.1% Triton X-100 in TBS. Sections were then blocked using a solution of 1% donkey serum in TBS for 1 h and then Background punisher (Biocare BP97AH) for 7 min. Sections were then incubated with primary antibodies overnight at 4°C (1:500 CRTAC1, ProteinTech 13001-1-AP; 1:800 LTBP1 26855-1-AP). The day after, the sections were incubated with the MACH4 MP AP-Polymer (Biocare MRAP536G) for 30 min before incubation with the Vulcan Fast Red solution (Biocare FR805S) for 5 to 10 min. Sections were then counterstained using Hematoxylin and mounted before microscopic evaluation. Whole slide images were captured at approximately 100,000 x 30,000 pixel resolution using a Hamamatsu NanoZoomer S360 Digital Slide Scanner C13220 and visualized and annotated within the PathcoreFlow workspace.

RNAscope in situ *hybridization*

Reagents for the RNAscope Multiplex Fluorescent Assay v2 were obtained from ACD Bio (Cat. No. 323270). Tissue section slides were fixed with 4% Paraformaldehyde for 1 h at 4°C. Slides were then washed twice with 1X PBS for 5 min (each wash). Slides were sequentially incubated in 50%, 70%, and 100% EtOH (5 min per incubation). Slides were then incubated in fresh 100% EtOH for 5 additional minutes. Slides were air-dried for 5 min before drawing a hydrophobic barrier around each tissue section. Next, hydrogen peroxide was added into each slide per manufacturer instructions and incubated for 10 min at RT. Slides were then washed twice with Milli-Q water. Sections were incubated with protease III reagent for 15 min at RT and then washed twice with Milli-Q water (2 min per wash). Probes for human *LMOD1* (Cat. No. 444141), *LTBP1* (Cat. No. 523281), and *CRTAC1*

(Cat. No. 1094551) as well as negative control probes (*dapB;* Cat. No. 321831) were acquired from ACD Bio and diluted per manufacturer instructions (1:50). Probes were then incubated for 2 h at 40°C within the proprietary HybEz oven unit from ACD Bio (Cat. No. 321710). Slides were then washed twice with 1X Wash buffer and then stored slides in 5X Sodium Saline Citrate (SSC) overnight at RT. Signal was amplified per manufacturer instructions by performing 3 subsequent incubations of Amplification (Amp1, Amp2, and Amp3) reagents. Signal was then developed by incubating sections with horseradish peroxidase reagent (HRP) specific for each channel for 15 min at 40°C, followed by incubation of the respective TSA vivid dye (TSA Vivid 650, TSA Vivid 570 and TSA Vivid 520) for 30 min at 40°C and then HRP-blocker to stop each reaction for 15 min at 40°C. Prior to the development of the signal, fluorophores were diluted as follows: TSA Vivid 650 (1:1000), TSA Vivid 570 (1:1500), and TSA Vivid 520 (1:1500) in Multiplex TSA buffer. Sections were then incubated with DAPI for 30 s at RT to counterstain nuclei and mounted using ProLong Gold Antifade Mountant (Invitrogen, P36930). Image acquisition was carried out with a Nikon ECLIPSE Ti2 with a Plan Fluor 40x/1.30 oil-immersion objective and four diode lasers centered at 405, 488, 561, and 640 nm. DIC H N2 (0.86 μm/px @ 512x512). Images were processed using the NIS-Element and ImageJ software.

Human carotid plaque scRNA-seq data validation

Raw human carotid plaque scRNA-seq data was accessed through https://dataverse.nl/dataverse/umculab. Quality Control (QC), normalization and dimensionality reduction of the matrix was done as described above using the pipeline within our scRNAutils R package. Clusters were then labeled according to annotations provided in the original publication.[95]

Coronary artery snATAC-seq tissue processing and data analysis

Coronary artery samples processing and nuclei isolation for snATAC-seq

snATAC-seq data for human coronary arteries was generated as described in our previous publication.[84] Briefly, isolated nuclei were processed using the 10X Genomics Chromium Single Cell ATAC and fastq files were preprocessed using the 10X Genomics Cellranger pipeline (CellRanger ATAC v1.2.0) using the hg38 reference genome and default parameters.

Dimensionality reduction, clustering of snATAC-seq data and generation of gene activity scores

Outputs from Cellranger were analyzed with the ArchR pipeline[85] (v1.0.2). Fragment files for each of the 41 patients were used to generate ArchR arrow files. We filtered low-quality cells with the following parameters: TSS enrichment >7, unique number of fragments >10000 and a doublet ratio <1.5. The genome was then divided into 500bp windows and then fragments within each window were used to generate a tile matrix (28316 cells x ~ 6 million tiles). Iterative latent semantic index (LSI) was then used to reduce dimensionality of the tile matrix. We checked for batch effects using Harmony (v1.0) and did not observe major differences in the data clustering structure (clusters driven by individual samples). We then used the first 30 components output by LSI for running non-linear dimensionality reduction (UMAP). Subsequent cell clustering was performed using the SNN modularity optimization-based algorithm as implemented in Seurat (v4.1.0). Chromatin accessibility (defined as the number of fragments within each tile) within gene bodies as well as proximally/distally from the TSS was used to infer gene expression by means of a gene activity score model. In this model, the number of fragments inside tiles of gene bodies are considered as well as surrounding tiles. To account for the activity of putative distal regulatory elements, an exponential weighting function is applied where tiles that reside further from genes TSS are assigned lower weights. Additionally, this model imposes gene boundaries to minimize the contribution of unrelated regulatory elements to a specific gene score.

TF motif enrichments

Enriched TF motifs for each cell type were predicted using the addMotifAnnotations() function in ArchR. Z deviation scores for each TF were then estimated with the chromVAR R package[119] (v1.12.0).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistics for bulk transcriptomics data were performed using the stats R package. The mean age for individuals in bulk transcriptomics analyses was 51 years with a standard deviation of 14 years. Furthermore, 27% of the included samples were female. To compare levels of normalized gene expression in human coronary arteries across disease conditions (non-lesion vs. lesion), we used a non-parametric Wilcoxon Rank-Sum Test. Significance tests for comparing stable and unstable carotid plaque regions from the same donor were carried out using paired Student's *t*-tests. The number of samples used in each analysis is indicated within each figure legend. For each statistical test, normality of the data was checked by generating quantile-quantile plots and inspection of sample variance. Variance of the data was inspected by plotting the distribution of gene expression values across samples using the geom_density function from the ggplot R package. Wilcoxon Rank-Sum Tests and Student *t*-tests were performed with a significance threshold of $p < 0.05$.

Differentially expressed markers for each cell cluster in scRNA-seq analyses were obtained using a Wilcoxon Rank-Sum Test as implemented in Seurat (v4.1.0). Differential expression analyses for SMCs across lesion categories were performed by generating cell and gene-specific weights to account for zero inflation and then using the DESeq2 likelihood ratio test (LRT) with a significance threshold of FDR < 0.05. The Pearson's product moment correlation coefficients for comparing normalized gene expression were calculated using the cor.test function within the stats R package. For calculations of Pearson correlations across the entire transcriptome in the scRNA-seq data, p values from were adjusted for multiple testing using the Benjamini Hochberg correction as

implemented in the stats package with the p.adjust function and parameters (method = ”fdr”). Additional sample size quantification and details of each statistical analysis with their respective parameters are described in method details and Figure legends.

ADDITIONAL RESOURCES

The scRNA analysis pipeline and other utility functions used in this manuscript have been wrapped as a public R package available at: https://github.com/MillerLab-CPHG/scRNAutils.