SUPPLEMENTAL MATERIAL

Supplemental Methods

Single cell preparation of mouse arterial specimens

Mice were sacrificed by CO_2 asphyxiation and then perfused intravenously with 10 mL PBS. Arterial tissues, including ascending aorta, brachiocephalic artery (BCA) and thoracic aorta, were isolated, placed into enzyme cocktail (4 U/mL Liberase TM (Sigma-Aldrich, 5401127001), 60 U/mL hyaluronidase (Sigma-Aldrich, H3506), and 120 U/mL DNase I (Worthington Biochemical Corporation, LS006333) in RPMI-1640)[8], and minced to ~1 mm pieces. After incubation at 37°C for 45 min, cell suspensions were filtered through 70-μm strainers and centrifuged at 500x g for 5 min. The cell pellets were washed once with FACS buffer (2% FBS, 5 mM EDTA, 20 mM HEPES and 1 mM Sodium Pyruvate in PBS) for further analysis.

Fluorescence-activated cell sorting and flow cytometry analysis

Mouse aortic single cells for scRNA-seq were prepared from three SMC-lineage tracing mice at the same timepoint of Western diet (WD) feeding (0, 8, 16, 26 weeks). After washing with FACS buffer, cell pellets were resuspended in FACS buffer with DAPI (0.5 μg/mL). Cells were gated on forward/side scatter parameters to exclude small debris and on forward scatter height versus forward scatter area to exclude doublets. Dead cells ($DAPI^+$) were also excluded. Live $ZsGreen1^+$ cells (SMC-lineage cells) and $ZsGreen1^-$ cells were sorted separately via BD Influx instrument into two collection tubes and were immediately subjected to scRNA-seq using Chromium Single Cell Gene Expression system (10x Genomics) in parallel.

Mice bone marrow (BM) cells and blood cells were isolated from SMC-lineage tracing mice fed tamoxifen diet for indicated timepoints. BM cells were flushed with DMEM basal medium by

27G needle connected syringe and through 40-μm nylon cell strainers. BM cells were achieved by centrifuging at 1,400 rpm for 5 mins at room temperature (RT) and resuspended with BMDM medium (DMEM + 20% L-cell supernatant). The BM cells were subsequently subjected to flow cytometry analysis. Blood cells were isolated from whole blood collected into Blood Collection Tubes with Sodium Heparin (BD 368037) via retro-orbital sinus from mice. Blood was then incubated with red blood cell lysis buffer (BD Pharm Lyse 555899) following the manufacturer's instruction. For flow cytometry analysis of BM cells and blood cells, small debris, doublets, and dead cells were excluded as described above. ZsGreen1^{+} and ZsGreen1^{-} cells were quantified on an BD Influx cytometer and analyzed via FCS Express software.

For SMC-derived macrophages analysis and sorting, aortic single cells were prepared as described above. After washing with FACS buffer, cell pellets were resuspended in FACS buffer and incubated with purified rat anti-mouse CD16/CD32 (eBioscience, 16-0161, 1: 200) at 4°C for 20 min to block unspecific binding of antibodies to Fc receptors. Then, resuspended cells were stained with rat anti-mouse CD11b-PE-Cy7 (eBioscience, 25-0112, 1:150) and DAPI (0.5 μg/mL) for 20 min at 4°C. Cells were then washed three times with FACS buffer before sorting. Small debris, doublets and dead cells were excluded as described above. SMC-derived macrophages (ZsGreen1^{+}CD11b^{+}) from three mice at each timepoint of WD feeding (0, 16, 20, 26 weeks) were analyzed on BD Influx instrument. SMC-derived macrophages from atherosclerotic aortas of three mice fed WD for 16 weeks were sorted and cultured in DMEM+10% FBS for 3 days *in vitro*.

For SEM cells sorting, after washing with FACS buffer, single cell pellets from atherosclerotic aortas of mice fed WD for 26 weeks were resuspended in FACS buffer with blocking antibody rat anti-mouse CD16/CD32 (eBioscience, 16-0161, 1: 200) at 4°C for 20 min to block unspecific binding of antibodies to Fc receptors. Resuspended cells were incubated with DAPI (0.5 μg/mL), rat anti-mouse CD11b-PE-Cy7 (eBioscience, 25-0112, 1:150), rat anti-mouse LY6A-APC (eBioscience, 17-5981, 1:200) and rat anti-mouse LY6C-PE (Biolegend,128007,

1:100) for 20 min at 4°C. Cells were then washed three times with FACS buffer before sorting on BD Influx instrument. Small debris, doublets and dead cells were excluded as described above. SMC-derived SEM cells (ZsGreen1^{+}CD11b^{-}LY6A^{+}LY6C1^{+}) and non-SEM cells (ZsGreen1^{+}CD11b^{-}LY6A^{-}LY6C1^{-}) were separately sorted for further analysis.

Single cell preparation of human atherosclerotic carotid arteries

Human atherosclerotic carotid arteries were obtained from patients undergoing endarterectomy (Table I in the Data Supplement) as part of a clinical research protocol approved by the Institutional Review Board (IRB) of Columbia University Irving Medical Center. All patients were fully informed of the research and signed IRB-approved informed consent forms. Fresh human atherosclerotic specimens (1-2 cm) were placed in Medium 199 supplemented with 10% FBS. The plaque was cleared of blood by washing with PBS, minced into ~1 mm pieces, and digested with RPMI-1640 medium including 10 U/mL Liberase TM (Sigma-Aldrich, 5401127001), 8 U/mL Elastase (Sigma-Aldrich, E7885), and 120 U/mL DNase I (Worthington Biochemical Corporation, LS006333) at 37°C for 1 hour with agitation by ThermoMixer (Eppendorf) for preparation of single cells. The cell suspensions were filtered through 70-μm strainers (Corning, 431751). Dead cells were removed by staining cells with Hoechst 33342, CellTrace calcein green AM (Invitrogen, C34852) and Propidium Iodide (PI, BD, 556463). Cells were then sorted on BD Influx instrument, gated to removed debris and exclude doublets, and Hoechst 33342^{+}PI^{-}clacein green AM^{+} live cells were sorted for scRNA-seq using Chromium Single Cell Gene Expression system (10x Genomics).

Single-cell RNA sequencing

All scRNA-seq experiments were performed at the Single Cell Core Facility of JP Sulzberger Columbia Genome Center. Samples were prepared using the 10x Genomics Chromium Single Cell 3' Reagent Kits v2 (mouse samples) or v3 (human samples) according to manufacturer's

instructions. 5000 cells and 200M reads per sample were targeted. 12 cycles for cDNA amplification and 12 cycles for sample index PCR were used. Sequencing was performed on a NovaSeq 6000 (R1 = 26 or 28, index = 8, R2 = 91).

Analysis of mouse scRNA-seq data

FASTQ files from each sample were processed by Cell Ranger v3.0.2 where reads were aligned to mouse reference genome mm10 and transcriptome based on Ensembl 93 annotation. A unique molecular identifier (UMI) count matrix was obtained from each sample. The datasets were analyzed using Seurat v3.1.1 in R[48]. Genes that were expressed in less than 10 cells were excluded from downstream analysis. The following filters were applied to cells in each dataset: number of genes, number of UMIs, and the percentage of reads that mapped to mitochondrial genes. Due to differences in sequencing depth among samples, we applied slightly different filtering parameters on each dataset, summarized in Table II in the Data Supplement. Raw UMI counts from each cell were normalized by the total counts in that cell, multiplied by 10,000, and then natural-log transformed by adding a pseudo-count of 1. Highly variable genes were identified in each dataset based on mean expression and dispersion. Genes that have mean expression between 0.05 and 10 and dispersion between 1.5 and 20 were selected as highly variable genes. To integrate datasets across timepoints and SMC lineages while minimizing batch effect, we utilized the integration pipeline in Seurat built on canonical correlation analysis (CCA). To remove batch effect across samples, we performed integration analysis implemented in Seurat v3.1.1. Integration anchors were identified based on the union of highly variable genes across datasets (maximum of 1,500 genes used in total) and the first 20 dimensions from the canonical correlation analysis (CCA). Integration was then performed using the anchors identified. Two-dimensional UMAP was used for visualization using the first 20 principal components (PC) built on the integration assay. Graph-based clustering was performed on the integrated dataset. In *Ldlr*$^{-/-}$ mouse scRNA-seq, a Shared Nearest Neighbor (SNN) graph was constructed with 50 nearest

neighbors and 20 dimensions of PCs as input. Clusters were identified using the above graph with a resolution parameter of 0.4. In *ApoE*$^{-/-}$ mouse scRNA-seq, SNN graph was constructed with 50 nearest neighbors and 10 dimensions of PCs as input. Clusters were identified with a resolution parameter of 0.5. Differential gene expression between clusters were performed using MAST test implemented in Seurat with each gene required to be expressed in at least 25% of cells in either of the two groups. Genes with fold change ≥1.5 and Bonferroni corrected *P*-value <0.05 were considered differentially expressed. In *Ldlr*$^{-/-}$ mouse scRNA-seq, 102 cells (0.37%) of cells were excluded from visualization as they were SMC/SMC-derived cells appearing in macrophage clusters or macrophages appearing in SMC/SMC-derived clusters, due to apparent UMAP visualization artifact. 16 cells (0.055%) labeled as macrophages from 0 week of HFD, ZsGreen1^{+} dataset was also excluded from visualization as they were considered SMCs based on expression profile. FASTQ files and expression matrices from mouse scRNA-seq data are available from the NCBI Gene Expression Omnibus (GEO) database under the accession number GSE155513.

GO enrichment analysis

Differentially expressed genes (DEGs) in SEM cells compared to SMC were identified in the ZsGreen1^{+} scRNA-seq dataset of *Ldlr*$^{-/-}$ mice fed 16-week WD (n=149 upregulated and 159 downregulated genes in SEM cells versus SMC, fold change ≥ 1.5, Bonferroni corrected *P*-value < 0.05). Background set was defined as genes expressed in at least 10 cells in either SEM or SMC. Enriched GO Biological Pathways were identified for up-regulated genes and down-regulated genes using clusterProfiler package in R. Redundant GO terms were removed using *simplify* function with similarity cutoff of 0.7.

IPA analysis

DEGs profile above was employed for Ingenuity Pathway Analysis (IPA). The list of 308 DEGs was uploaded into Qiagen's IPA system for core analysis and then overlaid with the global

molecular network in the Ingenuity pathway knowledge base. Canonical pathways, diseases and functions, and gene networks were opted to explore upstream regulators and downstream effects for mechanistic analysis and to categorize DEGs in specific diseases and functions.

Analysis of human atherosclerotic carotid artery scRNA-seq data

Human carotid artery scRNA-seq data from three subjects were processed following the same steps as in the mouse data. Reads were aligned to human reference genome GRCh38 and transcriptome based on Ensembl 93 annotation. Genes expressed in less than 10 cells were excluded from downstream analysis. Filtering on number of genes, number of UMIs, and the percentage of reads mapped to mitochondrial genes were summarized in Table III in the Data Supplement. Genes that have mean expression between 0.1 and 10 and dispersion between 1.25 and 20 were selected as highly variable genes. SNN graph was constructed with 30 nearest neighbors and 10 dimensions of PCs as input. Clustering was performed with a resolution parameter of 0.5. FASTQ files and expression matrices from human atherosclerotic carotid artery scRNA-seq data are available from the NCBI GEO database under the accession number GSE155512.

Integration analysis of human and mouse scRNA-seq data

Reference-based integration analysis was performed on the ZsGreen1^{+} scRNA-seq data of *Ldlr*$^{-/-}$ mouse fed 16-week WD and human atherosclerotic carotid artery scRNA-seq data, as well as a public human coronary artery scRNA-seq dataset[12] using Seurat v3.1.1. Human gene symbols were first mapped to mouse gene symbols using biomaRt package in R and Ensembl 93 annotation. Integration anchors were identified using the same approach as described above, except that the mouse dataset was specified as the reference. Cluster labels from mouse scRNA-seq joint analysis were used to identify human cells that clustered with mouse cells.

To investigate the counterpart of "fibromyocyte" population identified by Wirka et al.[12] in our mouse data, we first did clustering analysis of the human coronary scRNA-seq to identify fibromyocyte population, then annotated each cell from the human coronary scRNA-seq dataset using SingleR package in R with our scRNA-seq data of *Ldlr*$^{-/-}$ mice fed WD for 16 weeks as reference.

Analysis of human atherosclerotic carotid artery RNA-seq data

FASTQ files from human stable (macroscopically normal adjacent areas, n=4) and unstable (visible zone of plaque rupture, n=4, paired) atherosclerotic carotid artery plaques were obtained from Gene Expression Omnibus (accession number GSE120521)[30]. Quantification was performed by Salmon v0.11.2 using GENCODE 28 annotation. Differential gene expression was analyzed by DESeq2 v1.18.1. Genes with absolute fold change ≥ 1.2 between two conditions (unstable versus stable) and FDR adjusted *P*-value < 0.05 were considered differentially expressed genes.

Master regulators analysis

To infer master regulators (MRs) involved in cell state transition during atherosclerosis development, we applied metaVIPER[28, 29] on the ZsGreen1^{+} scRNA-seq data of *Ldlr*$^{-/-}$ mice fed 16-week WD, following the pipeline designed for MRs inference in scRNA-seq (https://github.com/califano-lab/single-cell-pipeline). Briefly, we first built a regulatory network using ARACNe on all single cells. ARACNe was run separately on each set of regulators: 1458 transcription factors, 821 transcriptional cofactors, and 3193 signaling proteins in mouse. Gene expression signature was generated on each cell using the median gene expression as internal reference. We then inferred protein activity based on the combined regulatory network and gene expression signature. Partitioning around medoids (PAM) clustering was applied to distance matrix built on protein activity with the number of clusters ranging from 2 to 10. The optimal

number of clusters was determined by silhouette score. In this study, we have identified 3 clusters in mouse data based on protein activity. Next, we generated cluster-specific ARACNe networks using meta-cell profiles built from integrating gene expression profiles of cells with similar protein activity profiles within each cluster. Counts from 5 nearest cells were integrated to generate each meta-cell. Protein activity was inferred again using cluster-specific networks as input to VIPER. T-test was conducted on transcription factor and cofactor activities between SMC and SEM populations with 100 bootstraps. Top activated and repressed MRs and their target genes are visualized using Cytoscape v3.3.0.

Protein activity in human unstable (visible zone of plaque rupture, n=4) and stable (macroscopically normal adjacent areas, n=4, paired) atherosclerotic carotid artery plaques[30] were estimated using metaVIPER. Prebuilt ARACNe networks from 45 tissues in GTEx (https://github.com/califano-lab/GTEx-Networks) were used to infer protein activity in each sample. metaVIPER is designed to accurately assess protein activities in tissues that do not have a tissue-matched ARACNe network. Thus, networks built from hundreds of samples per tissue in GTEx data are expected to provide accurate estimation of protein activities in atherosclerotic plaque samples.

Gene expression from early (intimal thickening and xanthoma, n=13) and advanced (fibrous cap atheroma, n=16) human atherosclerotic lesions of carotid arteries was downloaded from Gene Expression Omnibus (accession number GSE28829)[31]. Protein activities in each sample were estimated using metaVIPER with the same ARACNe networks from 45 tissues in GTEx as described above.

Human GWAS and eQTLs analyses

RA signaling target genes that harbored genome-wide association studies (GWAS) signals were identified in public summary data of CARDIoGRAMplusC4D (Coronary ARtery DIsease Genome wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D)

Genetics) (http://www.cardiogramplusc4d.org/)[38], a GWAS meta-analysis of 185,000 CAD cases and controls for 9.4 million SNPs imputed based on the 1000 Genomes Project. For incremental sets of increasingly stringently-defined RA signaling target genes[34-37], we examined their overlap with GWAS association results. A gene was declared to overlap with GWAS association if at least one SNP located within 100-500 kb upstream/downstream from the transcript start/end of the gene reached significance level of 0.001. GWAS SNPs that overlapped with the RA signaling target genes were visualized using LocusZoom. Sets of increasing stringent RA genes were chosen based on (i) RA target gene set[34-37], retinoic acid metabolism related genes[49] plus RA signaling transducers[50]; (ii) RA target gene set[34-37] plus RA signaling transducers[50] and (iii) RA target gene subset[35] plus RA signaling transducers[50].

RA signaling target genes were further examined for enrichment of GWAS association. We hypothesized that RA genes are more likely to show association with CAD/MI than a randomly selected gene set of equal size. To test this hypothesis, we randomly sampled an equal number of genes as the RA genes from the rest of the genome and calculated the percentage of genes in this control gene set overlapping with GWAS association. We repeated this resampling procedure 10,000 times and recorded the GWAS overlapping percentages. The enrichment p-value was calculated as the proportion of times that the overlapping percentage for the RA genes is bigger than those obtained from resampled control gene sets.

To assess the overlap of GWAS SNPs and eQTLs, we queried all SNPs in these loci with GWAS *P*-value <0.05 in aorta and tibial artery eQTL results from GTEx dataset (GTEx V8). Violin plots were generated for each eQTL SNP-gene pair on GTEx portal (https://www.gtexportal.org/home/), showing the distribution of normalized expression value across genotypes.

Preparation of frozen sections

Mouse spleen, BCA and aortic root along with the base of the heart were fixed in Histochoice Tissue Fixative for 3 hours at RT. Tissues were washed with PBS three times, embedded in Tissue Frozen Medium (General Data Healthcare, TFM-C) and stored at -80°C for sectioning. The blocks were sectioned to 10-μm thickness in the Histology Service of the Molecular Pathology Shared Resource at the Herbert Irving Comprehensive Cancer Center (HICCC), Columbia University Irving Medical Center.

Immunohistochemistry staining

Frozen sections were fixed in Histochoice Tissue Fixative for 45 min, washed three times with PBS, blocked with 10% Goat Serum at RT for 1 hour, and incubated with primary antibodies, including mouse anti-ACTA2 (1:200, Abcam, ab7817) and rabbit anti-MYH11 (1:200, Abcam, ab53219), at 4°C, overnight. Sections were washed three times with PBS, and then incubated with Hoechst 33342 (Invitrogen, H3570) and secondary antibodies, including Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594 (Invitrogen, A-11032) and Goat anti-Rabbit IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor 594 (Invitrogen, A-11037), for 1 hour at RT. After washing with PBS three times, sections were mounted with ProLong Diamond Antifade Mountant (Invitrogen, P36970).

RNAscope assay

RNAscope Multiplex Fluorescent Reagent Kit v2 (ACD Bio., 323100) was used for RNAscope staining. The assay was performed following the manufacturer's instruction with several modifications. To preserve the original fluorescence of ZsGreen1, slides were removed from -80°C and fixed with Histochoice Tissue Fixative for 1 hour at RT, which can leave nucleic acids in their native state. After washing with PBS three times, sections were dehydrated using 50%, 70% and 100% ethanol, and pretreated with RNAscope Hydrogen Peroxide for 10 min at RT. Then, the protease digestion step was omitted, and sections were incubated with probe set

targeting mouse *Vcam1* (ACD Bio., 438641) or a negative control probe set for 2 hours at 40°C. The slides were then mounted with ProLong Diamond Antifade Mountant (Invitrogen, P36970).

Differentiation of SEM cells to macrophage-like cells

FACS sorted SEM cells and non-SEM cells were induced with macrophage differentiation medium (DMEM+10% FBS+100 ng/mL M-CSF (GoldBio, 1320-09-10)) for 4 days without disturbance. Later, equal volume of fresh macrophage differentiation medium was added, and the cells were cultured for another 2 days, with medium replacement every other day. After 10 days of induction, the cells were subjected to immunofluorescence (IF) staining of the macrophage marker, CD68. After triple washes with PBS, cells were fixed with Histochoice Tissue Fixative for 20 min at RT, washed in PBS, permeabilized with 0.1% TritonX-100/PBS for 10 min, and blocked with 10% Goat serum for 1 hour at RT before incubating with rat anti-mouse CD68 antibody (1:100, Bio-Rad, MCA1957) at 4°C, overnight. The cells were subsequently washed three times with PBS and incubated with Hoechst 33342 (Invitrogen, H3570) and secondary antibody Goat anti-Rat IgG (H+L) Cross-Adsorbed Goat anti-Rat, Alexa Fluor 594 (Invitrogen, A-11007) for 1 hour at RT. After washing with PBS, cells were visualized under Nikon Eclipse Ti-S microscope.

Differentiation of SEM cells to fibroblast-like cells

SEM cells to fibroblast-like cells induction was performed according to a published method[24] with some modifications. FACS sorted SEM cells were resuspended in regular culture medium (DMEM+10% FBS). After growth to 80%-90% confluence, the cells were induced with fibroblast differentiation medium (DMEM+10% FBS+100 ng/mL CTGF (BioVendor, RD272589025)) for 4 weeks. RT-qPCR was performed to detect expression of fibroblast markers. SEM cells cultured in regular medium served as controls.

Differentiation of SEM cells to SMC phenotype

FACS sorted SEM cells and non-SEM cells were cultured in DMEM+10% FBS for 4 days. Cells were then treated with/without 10 ng/mL TGFβ1 (Biolegend, 763102) for 3 days. Subsequently, IF staining of SMC marker, ACTA2, using mouse anti-ACTA2 antibody (1:200, Abcam, ab7817) was performed as described above. Cells were visualized by Nikon Eclipse Ti-S microscope.

Macrophage efferocytosis

BM-derived macrophages harvested after being cultured in BMDM medium for 12 days. PKH26-labeled apoptotic cells (ACs) were added to macrophages for 60 min, followed by vigorous washing three times with PBS to remove unbound ACs. ACs were Jurkat cells induced by 0.5 μg/mL Staurosporine.

Tissue and cell imaging

IHC and RNAscope stained sections were visualized using Nikon Eclipse Ti-E inverted microscope at 10x and 20x objective magnifications. The confocal pinhole was set at 1 Airy unit, to produce an optical section of approximately 0.5 μm. DAPI/Hoechst 33342 (Nuclei), ZsGreen1, Vcam1/LY6A/ACTA2/MYH11 were imaged sequentially with excitation at 405, 488 and 561, respectively, and emission was collected with standard emission filters. Atherosclerotic lesions images were analyzed with Fiji/ImageJ software (NIH).

IF stained cells and ZsGreen1^{+} cells in efferocytosis assay were visualized using Nikon Ti-S microscope. The images were also analyzed with Fiji/ImageJ software (NIH).

Quantitative RT-PCR

RNA was extracted using Quick-RNA Miniprep kit (Zymo Research, R1054) and cDNA was synthetized by reverse transcription with High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, 4368813). Quantitative PCR (qPCR) was performed with specific gene primers and

2x PowerUp SYBR Green Master Mix (Applied Biosystems, A25777) using QuantStudio 7 Flex Real-Time PCR System (Applied Biosystems). All qPCR primers used in the study can be found in Table IV in the Data Supplement.

In vitro studies of ATRA effects on mouse SMC

Mouse SMC (ATCC, CRL-2797) cultured in DMEM+10% FBS medium were treated with 25 ng/mL recombinant TNFα (PeproTech, 31501A5UG), 40 μg/mL cholesterol (Sigma-Aldrich, C4951) and 10 μM ATRA (Sigma-Aldrich, R2625) for 72 hours, as indicated in Figure 6A. After that, RT-qPCR was performed to detect the expression of *Ly6a* and *Vcam1*.

In vivo studies of ATRA effects on atherosclerosis, SMC transition and vascular injury

Atherosclerosis and SMC transition studies. After induction with tamoxifen and feeding WD for 4 weeks, *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11-CreER*^{T2} mice were administrated with vehicle (corn oil (Sigma-Aldrich, C8267), control) or ATRA (2.5 mg/kg mice) 3 times per week via gavage. The mice were continued on WD for another 12 weeks. After total of 16-week WD feeding, the mice were sacrificed for further analysis. For atherosclerosis analysis, apical parts of the hearts were collected from the mice, fixed in Histochoice Tissue Fixative for 3 hours and embedded in Tissue Frozen Medium. H&E staining was performed using aortic sinus sections, and areas of atherosclerotic lesions were measured using Fiji/ImageJ software. For analysis of the proportion of SEM cells in atherosclerotic arteries, mouse aortic single cells were prepared as described above and the proportion of ZsGreen1^{+}LY6A^{+}LY6C1^{+} SEM cells and ZsGreen1^{+}LY6A^{-}LY6C1^{-} non-SEM cells were analyzed by flow cytometry. For RNAscope analysis of ZsGreen1^{+}*Vcam1*$^{+}$ SEM cells in BCA sections, three sections of each mouse 300 μm, 450 μm and 600 μm from the start of the BCA branch were analyzed. RNAscope analysis using probe set against *Vcam1* mRNA was performed as described above.

Vascular injury. To detect the direct effects of ATRA on SMC, carotid artery ligation model was performed in *ROSA26*$^{ZsGreen1/+}$; *Myh11-CreER*^{T2} mice. One day prior to carotid injury surgery, mice were randomly divided into control group or ATRA-treated group and fed corn oil or ATRA (20 mg/kg), respectively. Mice were anaesthetized with ketamine/xylazine. The left common carotid artery (LCCA) was ligated with 6-0 silk suture (Covidien, s1172) proximal to the carotid bifurcation, and the unmanipulated right carotid artery served as a control. Post-surgery, mice were fed chow diet and administered corn oil or ATRA (20 mg/kg) 3 times per week for 4 weeks[46]. Mice were sacrificed by CO_2 asphyxiation and then perfused intravenously with PBS. Carotid arteries were then collected from the mice, fixed with Histochoice Tissue Fixative for 3 hours and embedded in Tissue Frozen Medium for sectioning (10-μm thickness) and H&E staining. Morphometric analysis of the carotid arteries was performed using Fiji/ImageJ software. The area of the lumen, the area inside the internal elastic lamina, and the area inside the external elastic lamina were measured in pixels. The areas of the intima and media, as well as the ratio of intima area:media area, were calculated.

Hematoxylin and eosin (H&E) staining

H&E staining of aortic sinus sections and ligated LCCA sections was performed in the Histology Service of the Molecular Pathology Shared Resource at the HICCC, Columbia University Irving Medical Center.

Serum cholesterol measurement

Whole blood was collected into Serum Separation Tubes (BD 368013) via retro-orbital sinus plexus from mice. Following 2 hours at RT, serum was obtained by centrifuging for 15 min at 1,000x g and stored in aliquots at -80°C. Serum total cholesterol was measured using Infinity Cholesterol Reagent (ThermoFisher Scientific, TR13421) according to the manufacturer's instruction.

Supplemental Tables

Data Supplement Table I. Patients' characteristics of human atherosclerotic carotid artery specimens for scRNA-seq

Patient	10x Genomics ID	Gender	Age at surgery (years old)	Race	Symptomatic/Asymptomatic
Patient 1	RPE004	Male	83	White/European	Symptomatic
Patient 2	RPE005	Male	67	White/European	Asymptomatic
Patient 3	RPE006	Female	76	White/European	Asymptomatic

Data Supplement Table II. Filtering criteria for *Ldlr*$^{-/-}$ and *ApoE*$^{-/-}$ mouse scRNA-seq dataset. The maximum number of UMIs, minimum number of genes, maximum number of genes, and maximum percentage of mitochondrial reads allowed per cell are listed for each dataset, followed by the number of genes and the number of cells after filtering.

Genotype	Weeks of WD	ZsGreen1 status	Maximum # UMIs	Minimum # genes	Maximum # genes	Maximum % mitochondrial reads	# genes	# cells
Ldlr$^{-/-}$	0	+	15000	300	3500	10	13011	3094
	0	-	10000	300	3000	10	13614	3513
	8	+	15000	300	3500	10	12833	2498
	8	-	10000	300	3000	10	12625	1714
	16	+	20000	300	4000	10	14232	4287
	16	-	15000	300	3500	10	13760	2742
	26	+	15000	300	3500	10	14408	6041
	26	-	10000	300	3000	10	13609	3403
ApoE$^{-/-}$	8	+	20000	200	4000	5	13817	2932
	8	-	15000	200	3500	5	13627	1935
	16	+	20000	200	4000	5	14086	3150
	16	-	15000	200	3500	5	13381	1575
	22	+	20000	200	4000	5	14106	3426
	22	-	20000	200	4000	5	13511	2062

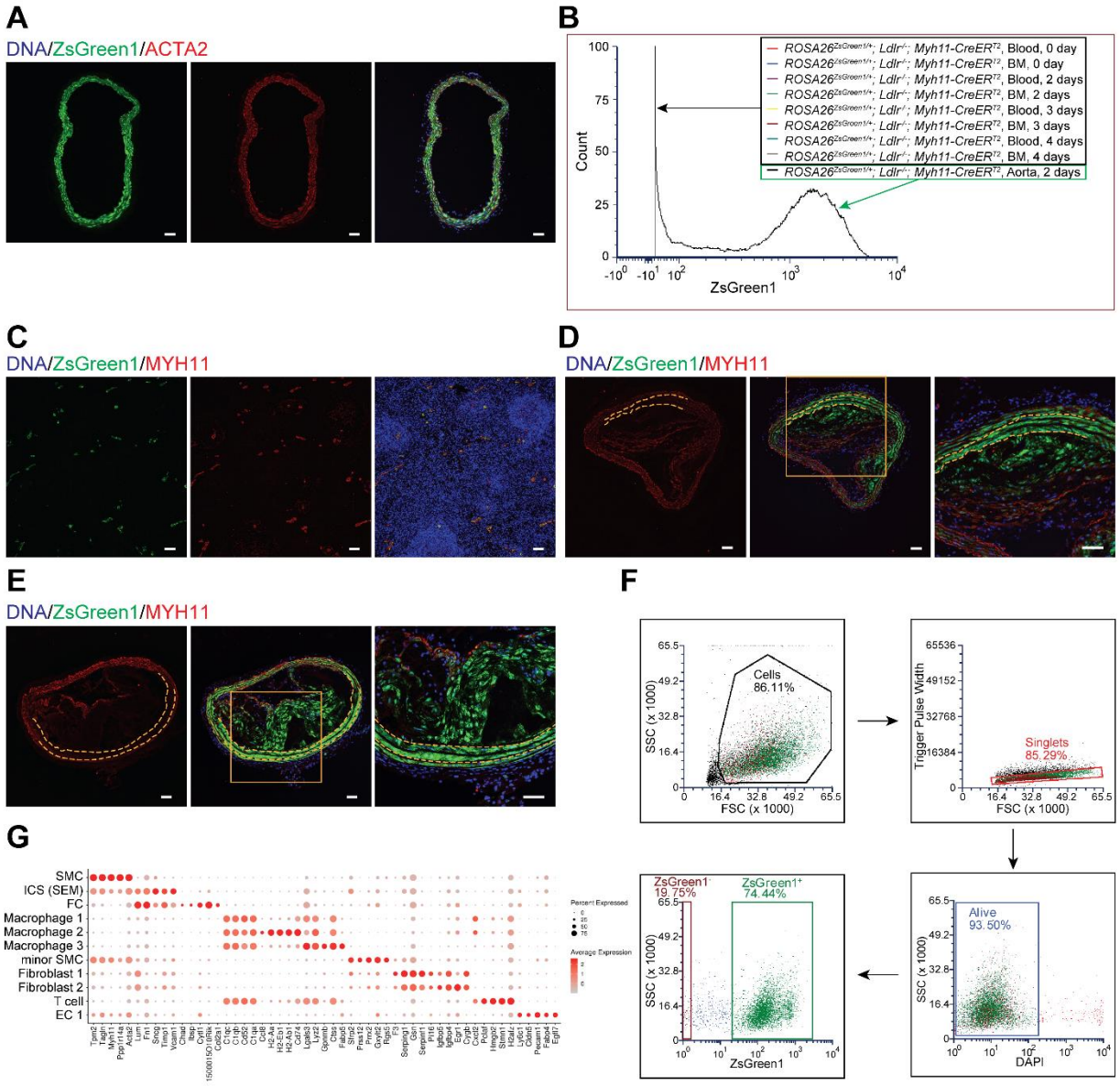
Data Supplement Table III. Filtering criteria for human carotid artery scRNA-seq dataset. The maximum number of UMIs, minimum number of genes, maximum number of genes, and maximum percentage of mitochondrial reads allowed per cell are listed for each dataset, followed by the number of genes and the number of cells after filtering.

10x Genomics ID	Maximum # UMIs	Minimum # genes	Maximum # genes	Maximum % mitochondrial reads	# genes	# cells
RPE004	20000	200	4000	10	15796	2614
RPE005	20000	200	4000	10	17397	3486
RPE006	20000	200	4000	10	15687	2767

Data Supplement Table IV. qPCR primer sequences used in the study

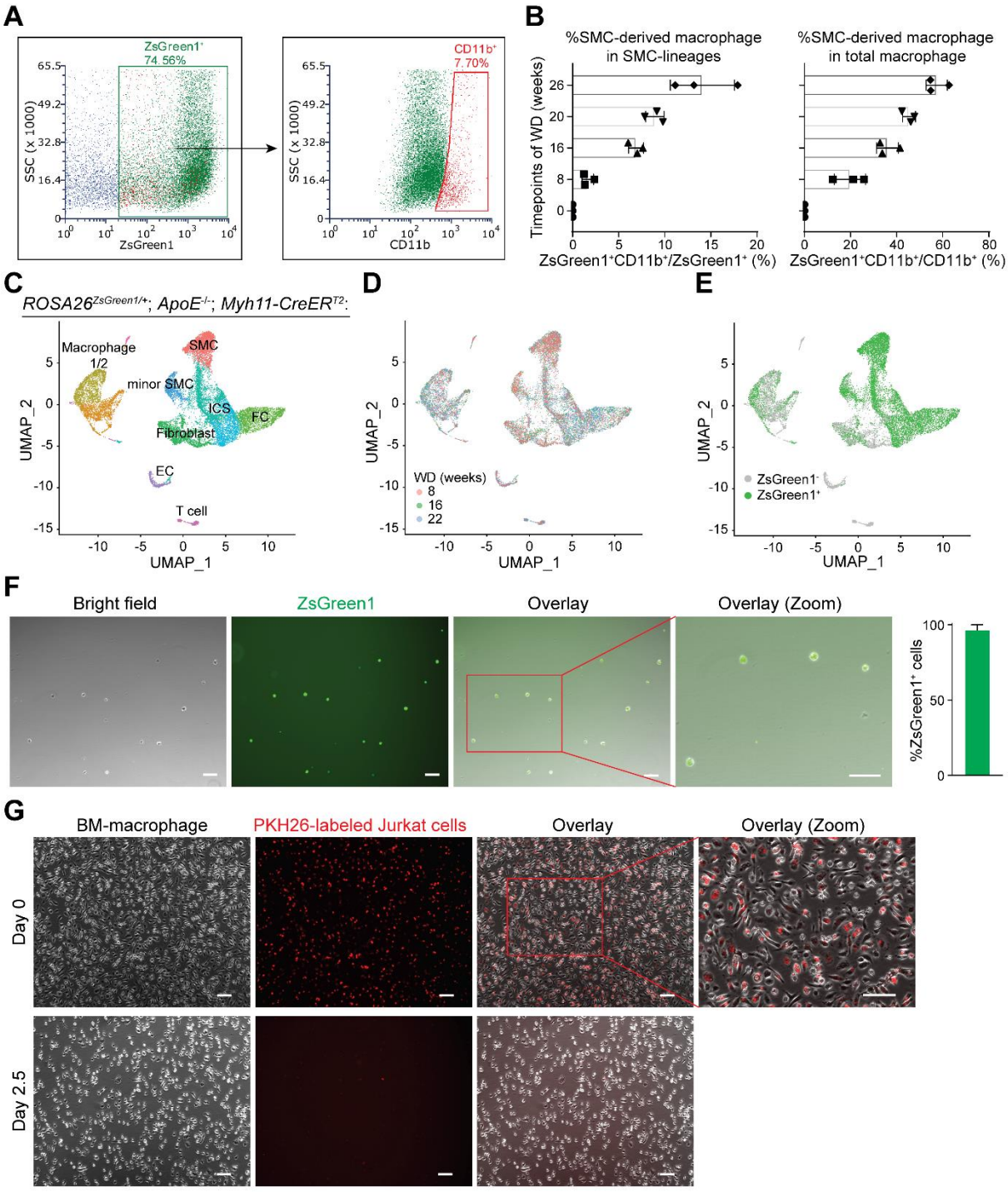
Primer	Sequence
qPCR mActb F	GGCTGTATTCCCCTCCATCG
qPCR mActb R	CCAGTTGGTAACAATGCCATGT
qPCR mCol1a1 F	CTGGCGGTTCAGGTCCAAT
qPCR mCol1a1 R	TTCCAGGCAATCCACGAGC
qPCR mCol3a1 F	CCTGGCTCAAATGGCTCAC
qPCR mCol3a1 R	CAGGACTGCCGTTATTCCCG
qPCR mFn1 F	ATGTGGACCCCTCCTGATAGT
qPCR mFn1 R	GCCCAGTGATTTCAGCAAAGG
qPCR mLy6a F	AGGAGGCAGCAGTTATTGTGG
qPCR mLy6a R	CGTTGACCTTAGTACCCAGGA
qPCR mS100a4/Fsp1 F	TCCACAAATACTCAGGCAAAGAG
qPCR mS100a4/Fsp1 R	GCAGCTCCCTGGTCAGTAG
qPCR mTnc F	ACGGCTACCACAGAAGCTG
qPCR mTnc R	ATGGCTGTTGTTGCTATGGCA
qPCR mVcam1 F	AGTTGGGGATTCGGTTGTTCT
qPCR mVcam1 R	CCCCTCATTCCTTACCACCC
qPCR mVim F	CGTCCACACGCACCTACAG
qPCR mVim R	GGGGGATGAGGAATAGAGGCT

Supplemental Figures and Figure Legends

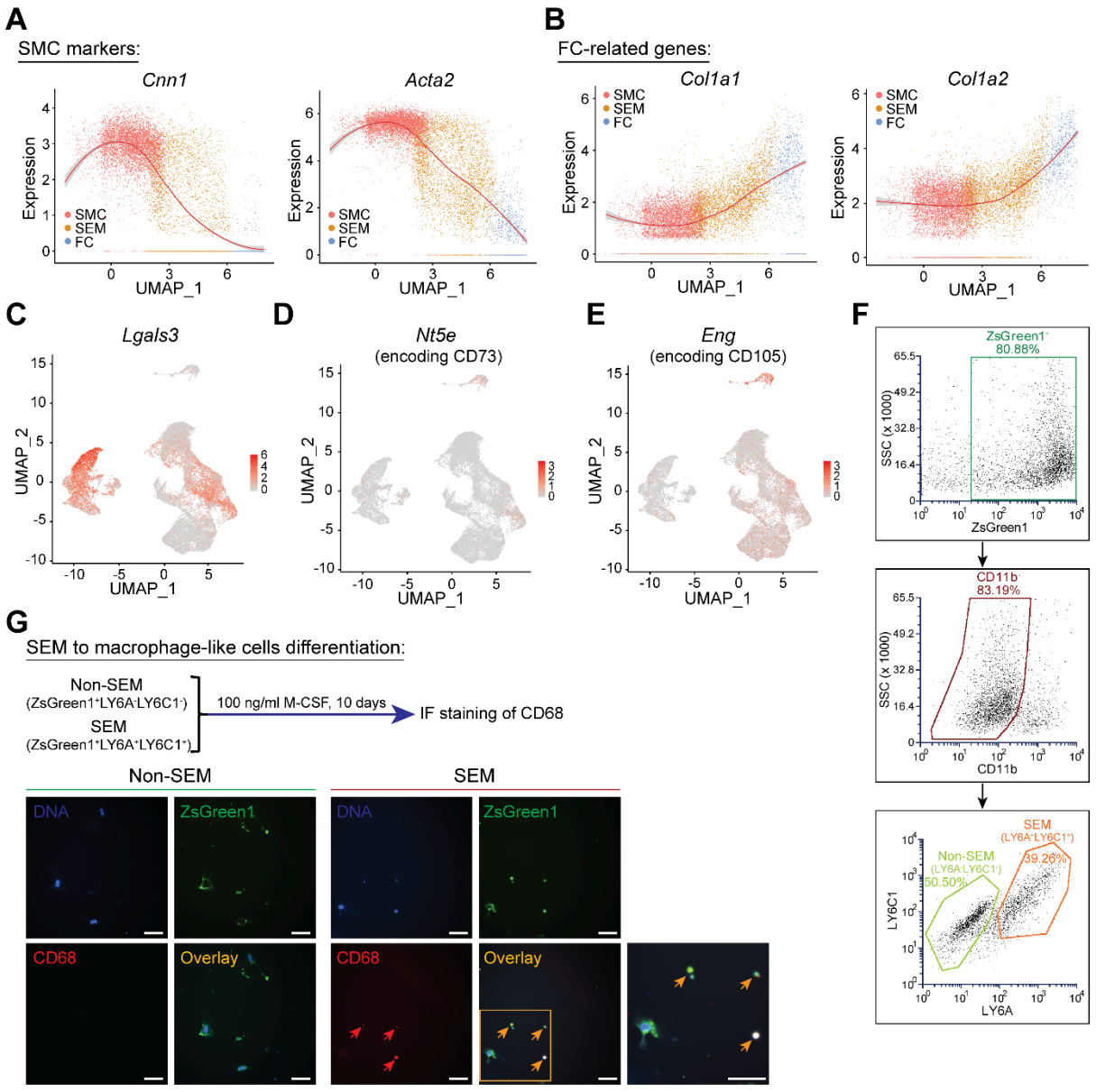


Data Supplement Figure I. Transdifferentiation of SMC into other cell types/states is found in the SMC-lineage tracing murine model. A, IHC staining of SMC marker, ACTA2, in BCA sections from tamoxifen (TAM)-induced 8-week *ROSA26*$^{ZsGreen1/+}$; *Myh11-CreER*^{T2} mice. Scale bars, 50 μm. **B**, Flow cytometry analysis of blood and bone marrow (BM) cells isolated from *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11-CreER*^{T2} mice induced with TAM for various timepoints (0, 2, 3, 4 days) as indicated, followed by 2-day chow diet. Single cells from the aorta of a mouse induced with TAM for 2 days are used as positive control. **C**, IHC staining of MYH11 in spleen sections from *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11-CreER*^{T2} mice induced with TAM for 2 days. Scale bars, 50 μm. **D** and **E**, IHC staining of SMC marker, MYH11, in BCA sections from TAM-induced *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11-CreER*^{T2} mice fed WD for 15 weeks (**D**) and 16 weeks (**E**). Yellow broken lines highlight ZsGreen1^{+} cells lacking or with reduced MYH11 expression in regions of

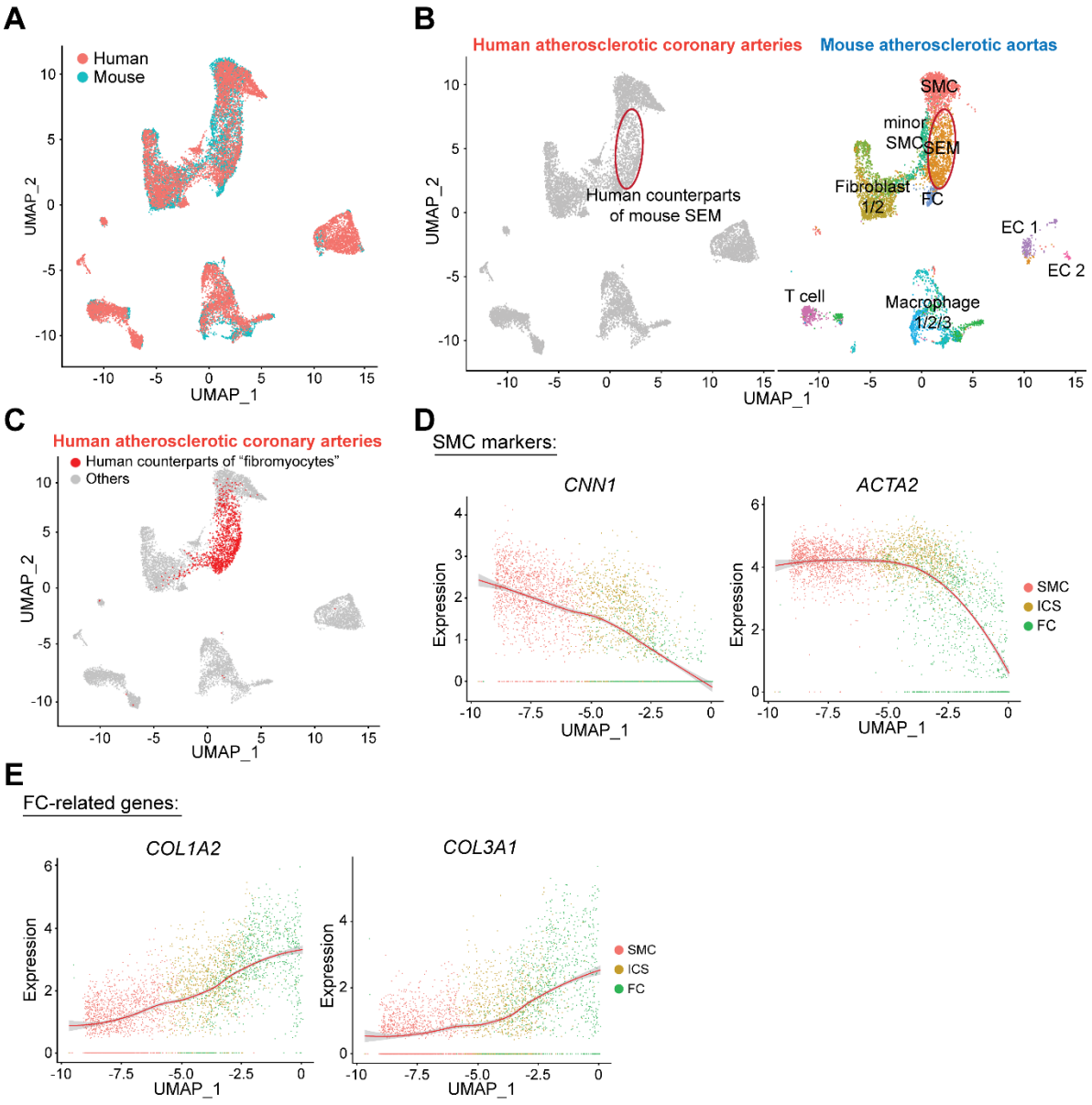
media. Scale bars, 50 μm. **F**, FACS gating strategy of ZsGreen1^{+} and ZsGreen1^{-} cells sorting for scRNA-seq. **G**, Dotplot showing top five differentially expressed genes (DEGs) for each ZsGreen1^{+} cell type/state in atherosclerotic lesions identified via scRNA-seq.



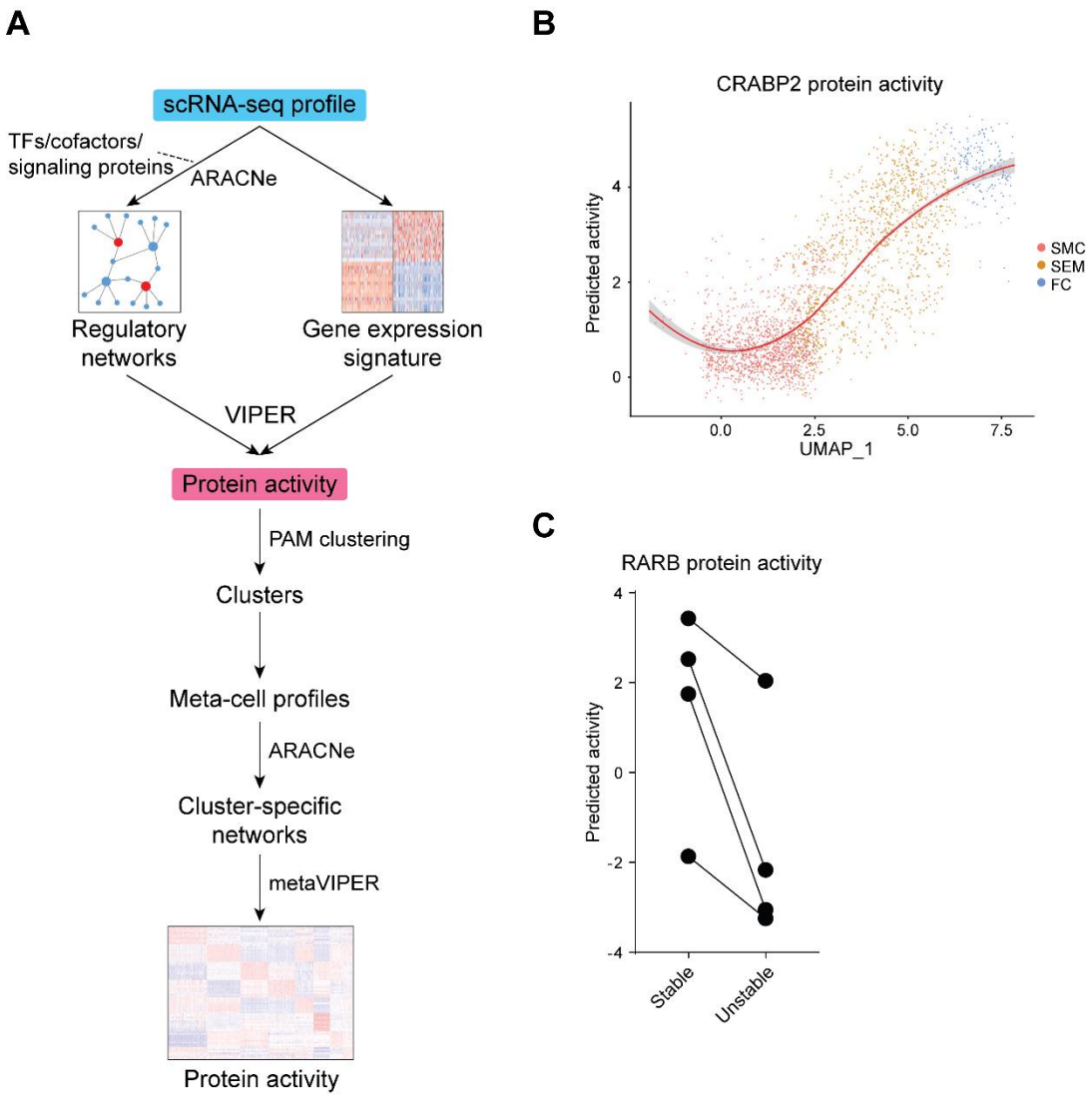
Data Supplement Figure II. SMC-derived macrophage-like cells appear in atherosclerotic lesions of both *Ldlr*$^{-/-}$ and *ApoE*$^{-/-}$ mice. A, Gating strategy of ZsGreen1^{+}CD11b^{+} SMC-derived macrophage-like cells for flow cytometry analysis and sorting. The cells were from mice fed WD for 16 weeks. **B**, Change in flow cytometry-estimated proportion of ZsGreen1^{+}CD11b^{+} SMC-derived macrophage-like cells among all ZsGreen1^{+} SMC-derived cells (left) and ZsGreen1^{+}CD11b^{+} SMC-derived macrophage-like cells among all CD11b^{+} macrophages (right) during WD feeding (0, 8, 16, 20, 26 weeks). Values are shown as mean ± s.d. **C**-**E**, UMAP visualization of all scRNA-seq data from atherosclerotic aortas of *ROSA26*$^{ZsGreen1/+}$; *ApoE*$^{-/-}$; *Myh11-CreER*^{T2} mice fed WD for various timepoints (8, 16, 22 weeks). Representative cell type/state of each cluster shows cell composition of atherosclerotic lesions from *ApoE*$^{-/-}$ mice (**C**), which is similar to that in *Ldlr*$^{-/-}$ mice (Figure 1B). Cells of each timepoint (**D**) and ZsGreen1 status (ZsGreen1^{+} and ZsGreen1^{-}) in all scRNA-seq data (**E**) are visualized by UMAP. **F**, Morphology of FACS sorted ZsGreen1^{+}CD11b^{+} SMC-derived macrophage-like cells cultured *ex vivo* for 3 days. The proportion of cells remaining ZsGreen1 expression after 3-day culture is calculated. The cells are isolated from aortas of *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11*-CreER^{T2} mice fed WD for 16 weeks (n=3 mice), Scale bars, 100 μm. **G**, Efferocytosis assay is performed using wild-type mice BM-derived macrophages incubated with PKH26-labled Jurkat cells for 1 hour. Following incubation, cells are washed with PBS three times and engulfed PKH26^{+} Jurkat cells are shown (Day 0). After 2.5-day culture, almost all PKH26^{+} Jurkat cells are degraded (Day 2.5). Scale bars, 100 μm.



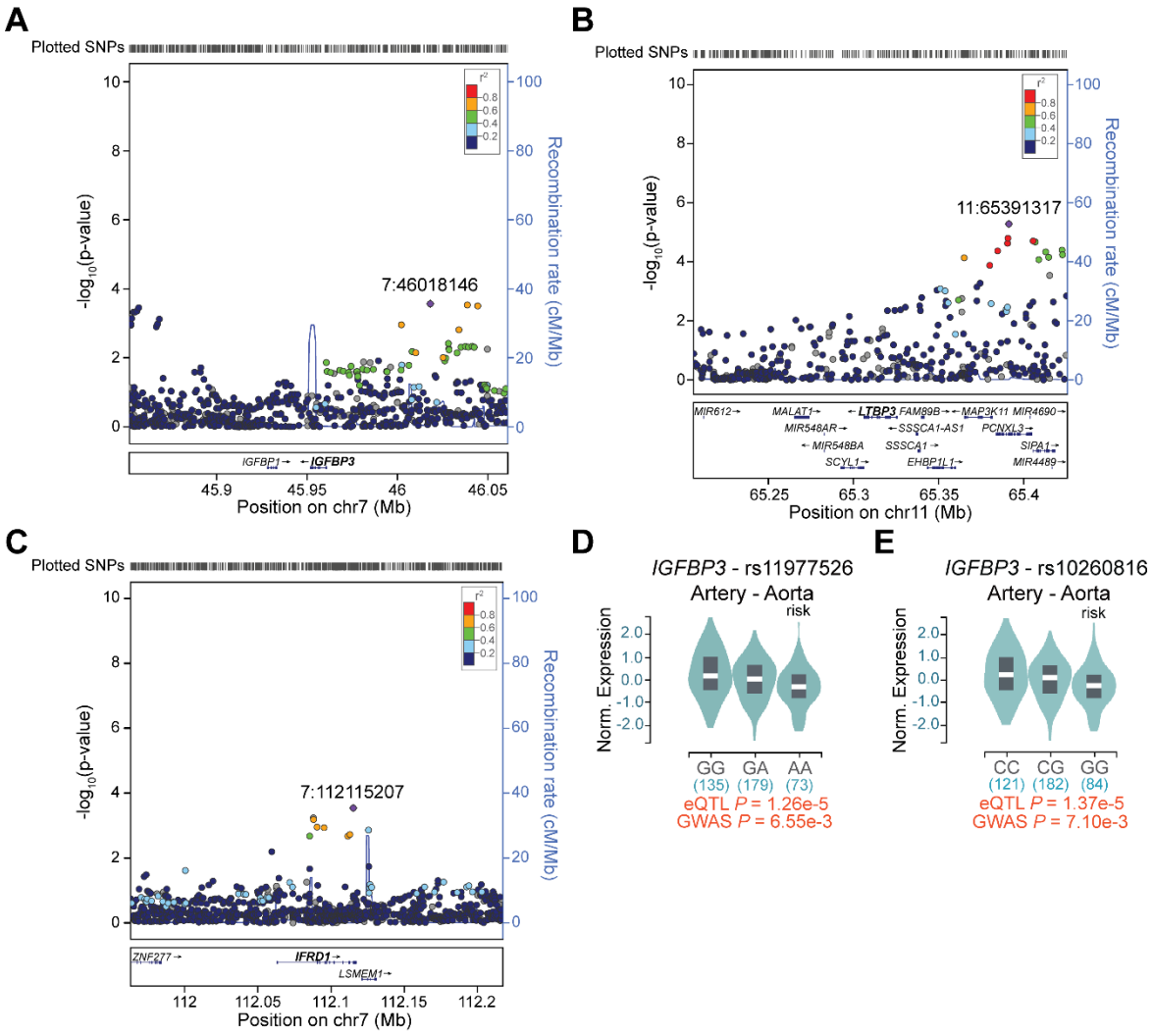
Data Supplement Figure III. Interrogation of molecular and cellular features reveals a unique SMC-derived intermediate SEM cell state formed during atherosclerosis. A and **B**, Gene expression tendencies of SMC marker genes (*Cnn1*, *Acta2*) (**A**) and FC-related genes (*Col1a1, Col1a2*) (**B**) through SMC-SEM-FC axis are shown as normalized expression versus UMAP_1. **C**-**E**, The expression levels of previously known macrophage marker, *Lgals3* (**C**), and mesenchymal stem cell markers, *Nt5e* (encoding CD73) (**D**) and *Eng* (encoding CD105) (**E**), in all cell types/states are indicated by color scales. **F**, Flow cytometry gating strategy for ZsGreen1^{+}LY6A^{+}LY6C1^{+} SEM cells and ZsGreen1^{+}LY6A^{-}LY6C1^{-} non-SEM cells sorting from aortas of *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$; *Myh11-CreER*^{T2} mice fed WD for 26 weeks. **G**, Following 10-day treatment with 100 ng/mL macrophage colony-stimulating factor (M-CSF), isolated ZsGreen1^{+}LY6A^{+}LY6C1^{+} SEM cells, but not ZsGreen1^{+}LY6A^{-}LY6C1^{-} non-SEM cells, differentiate into macrophage-like cells *ex vivo,* assessed by IF staining of CD68, a macrophage cell marker. The cells were from mice fed WD for 26 weeks. Scale bars, 100 μm.



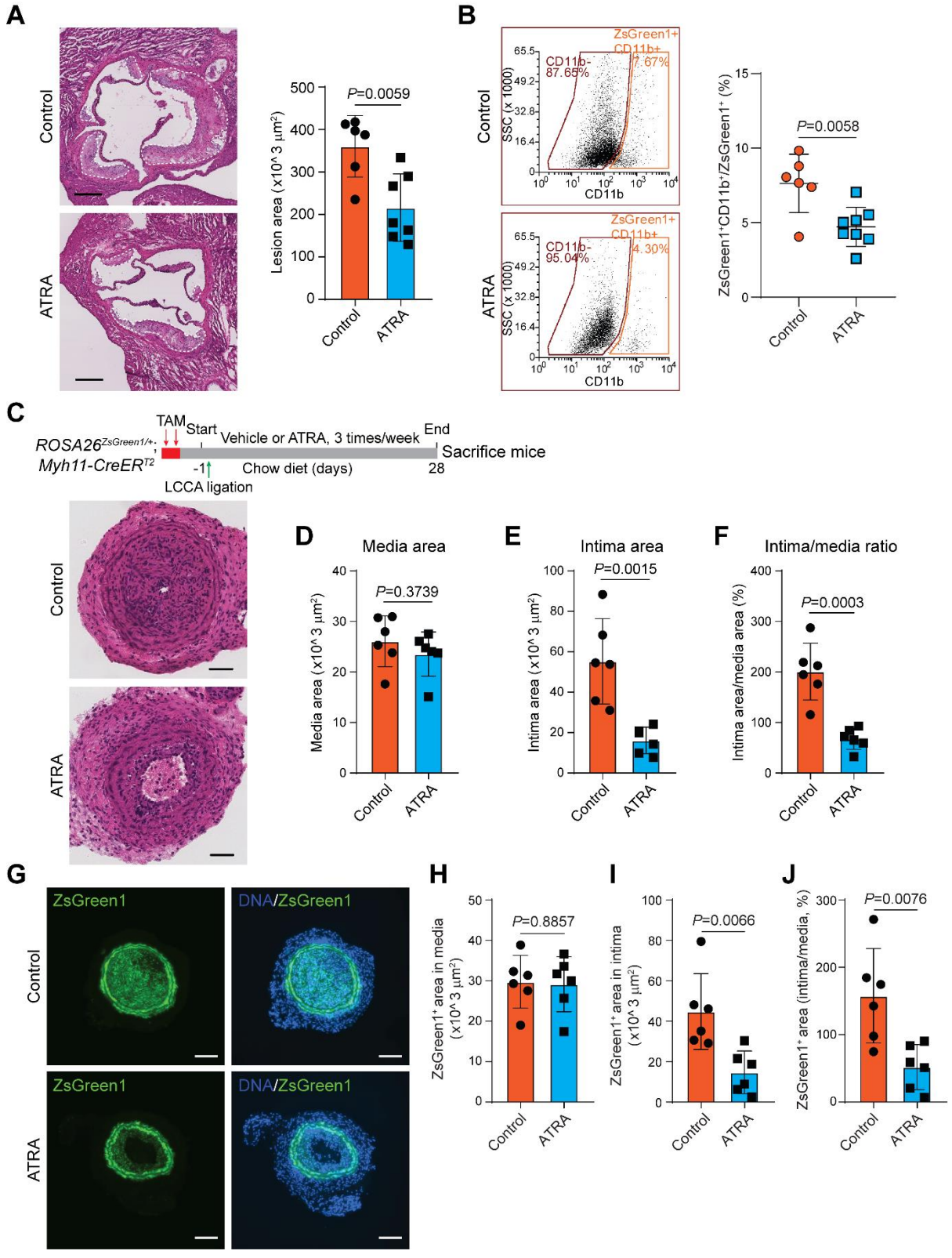
Data Supplement Figure IV. Counterparts of mouse SEM cells are found in human coronary artery atherosclerotic plaques. A and **B**, Reference-based integration analysis of combined scRNA-seq data of *Ldlr*$^{-/-}$ mice fed 16-week WD (n=3 mice; 7029 cells) and a public scRNA-seq data[12] of human atherosclerotic coronary arteries (n=4 patients; 11585 cells). Mouse cell clusters are used as reference, human scRNA-seq data are projected to mouse data. Representative mouse cell types/states are indicated. Combination of mouse and human scRNA-seq data is visualized by UMAP (**A**). Mouse and human clusters are shown separately and a cell population in human data suggests the counterparts of mouse SEM cells (red circled) in human plaques (**B**). **C**, Reference-based integration analysis of human coronary artery scRNA-seq data shows that human counterparts of "fibromyocytes" identified by Wirka et al.[12] contain both SEM and fibrochondrocytes identified in our study. Human counterparts of "fibromyocytes" are indicated with red dots. **D** and **E**, Gene expression tendencies of SMC markers (*CNN1*, *ACTA2*) (**D**) and FC-related genes (*COL1A2*, *COL3A1*) (**E**) through SMC-ICS-FC axis are shown as normalized expression versus UMAP_1.



Data Supplement Figure V. metaVIPER analysis predicts protein activities of RA signaling transducers, CRABP2 and RARB, in SMC-SEM cell transition and human atherosclerotic plaques, respectively. A, Workflow of metaVIPER analysis. **B**, Predicted protein activity of CRABP2 (cellular retinoic acid binding protein 2) through SMC-SEM-FC axis, shown as protein activity versus UMAP_1, suggests increased CRABP2 protein activity during SMC transition to intermediate SEM cell state in atherogenesis. **C**, Predicted protein activity of RA signaling transducer, RARB (retinoic acid receptor beta), estimated via metaVIPER in 4 pairs of stable and unstable plaque specimens[30]. Dots and lines represent paired plaque specimens from individual patients.



Data Supplement Figure VI. Moderate GWAS signals for CAD are found in multiple RA gene loci and CAD risk alleles are correlated with reduced expression of RA gene in human CAD-relevant tissue. A-C, Regional association plots of 1000-Genomes based GWAS of CAD/MI in several RA signaling target gene loci, *IGFBP3* (**A**), *LTBP3* (**B**) and *IFRD1* (**C**). $-\log_{10}$(p-value) of SNP association for CAD/MI is shown for each plot. Linkage disequilibrium with the top SNP from each region is color coded by r^2. **D** and **E**, Representative violin plots showing normalized expression of *IGFBP3* by genotype in CAD-relevant tissue, aorta. Numbers below the genotypes indicate sample size. Risk genotypes and eQTL and GWAS *P*-values are indicated. The association between reduced *IGFBP3* expression and risk alleles is observed in aorta.



Data Supplement Figure VII. ATRA treatment reduces atherosclerosis in *Ldlr*$^{-/-}$ mice and attenuates vascular injury in carotid artery ligation mouse model. A, Representative H&E stained aortic sinus sections from control (n=6) and ATRA-treated (n=7) *ROSA26*$^{ZsGreen1/+}$; *Ldlr*$^{-/-}$;

Myh11-CreER^{T2} mice (left), and quantification of aortic sinus lesion area (right). Scale bars, 250 μm. Values are shown as mean ± s.d. *P*-value is indicated. **B**, Flow cytometry analysis of the proportion of ZsGreen1$^+$CD11b$^+$ SMC-derived macrophage-like cells among total ZsGreen1$^+$ cells (control, n=6 mice; ATRA-treated, n=8 mice). Values are shown as mean ± s.d. *P*-value is indicated. **C**-**F**, *ROSA26*$^{ZsGreen1/+}$; *Myh11-CreER*^{T2} mice are induced by 2-day TAM diet, followed with chow diet. Carotid artery ligation is performed on the left common carotid arteries of control (n=6) and ATRA-treated (n=6) mice. Vehicle (corn oil, control) or ATRA (20 mg/kg) is administrated one day before surgery and continued 3 times/week to 28 days post-surgery. H&E stained ligated carotid artery sections (**C**), and quantification of media area (**D**), intima area (**E**), and ratio of intima area/media area (**F**) are shown. Scale bars, 50 μm. Values are shown as mean ± s.d. *P*-values are indicated. **G**, Representative images of ligated carotid artery sections from control (n=6) and ATRA-treated (n=6) mice show ZsGreen1$^+$ areas in both media and intima regions. Scale bars, 100 μm. **H**-**J**, Quantification of ZsGreen1$^+$ areas in media (**H**), intima area (**I**), and ratio of ZsGreen1$^+$ areas in intima/media (**J**) are shown. Values are shown as mean ± s.d. *P*-values are indicated.