Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk

Adam W. Turner1,19, Shengen Shawn Hu1,19, Jose Verdezoto Mosquera2,12, Wei Feng Ma1,3,4, Chani J. Hodonsky1,15, Doris Wong1,2,5, Gaëlle Auguste1, Yipei Song1, Katia Sol-Church4,6, Emily Farber1,7, Soumya Kundu8, Anshul Kundaje8,9, Nicolas G. Lopez10, Lijiang Ma11, Saikat Kumar B. Ghosh12, Suna Onengut-Gumuscu1,7,13, Euan A. Ashley1,2,13, Aloke V. Finn12, Nicholas J. Leeper1,2,17,18, Euan A. Ashley1,2,13,19, and Clint L. Miller1,2,5,13

Coronary artery disease (CAD) is a complex inflammatory disease involving genetic influences across cell types. Genome-wide association studies have identified over 200 loci associated with CAD, where the majority of risk variants reside in noncoding DNA sequences impacting cis-regulatory elements. Here, we applied single-nucleus assay for transposase-accessible chromatin with sequencing to profile 28,316 nuclei across coronary artery segments from 41 patients with varying stages of CAD, which revealed 14 distinct cellular clusters. We mapped ~320,000 accessible sites across all cells, identified cell-type-specific elements and transcription factors, and prioritized functional CAD risk variants. We identified elements in smooth muscle cell transition states (for example, fibromyocytes) and functional variants predicted to alter smooth muscle cell- and macrophage-specific regulation of MRAS (3q22) and LIPA (10q23), respectively. We further nominated key driver transcription factors such as PRDM16 and TBX2. Together, this single-nucleus atlas provides a critical step towards interpreting regulatory mechanisms across the continuum of CAD risk.

CAD is the leading cause of death globally and results from injury to the vessel wall and atherosclerotic plaque buildup. Atherosclerotic coronary arteries are complex due to the propensity of multiple cell types to undergo cell phenotypic switching, including endothelial cells, smooth muscle cells (SMCs), fibroblasts and various immune cells1–3. This has hindered efforts to combat the disease process itself, as currently approved therapies only treat the traditional risk factors, such as elevated blood pressure or cholesterol levels. Recent single-cell RNA-sequencing (scRNA-seq) analyses have yielded numerous cellular insights into atherosclerosis4–11. In particular, lineage-traced scRNA-seq approaches have shown that SMCs transdifferentiate to several distinct phenotypes during the disease process itself, as currently approved therapies only treat the traditional risk factors, such as elevated blood pressure or cholesterol levels. Recent single-cell RNA-sequencing (scRNA-seq) analyses have yielded numerous cellular insights into atherosclerosis4–11. In particular, lineage-traced scRNA-seq approaches have shown that SMCs transdifferentiate to several distinct phenotypes during atherosclerosis: (1) ‘fibromyocytes’ with fibroblast-like signatures7; (2) an intermediate cell state that can become fibrochondrocyte- or macrophage-like; or (3) a transitional state giving rise to multiple plaque cell types10. Together, these studies demonstrate that SMC-derived cells can elicit beneficial or detrimental effects depending on the stage of CAD and/or plaque environment. Despite these advances, the underlying cell-specific regulatory mechanisms remain elusive.

As a complex disease, CAD involves an interplay of environmental and genetic factors over the life course. Genome-wide association studies (GWASs) have now identified over 200 independent CAD loci12–16. Many of these are predicted to function in vessel wall processes such as regulation of vascular remodeling, vaso-motor tone and inflammation17. The majority of CAD-associated SNPs reside in noncoding regions and are enriched in cis-regulatory elements (CREs)18, pointing towards regulatory functions19. Since CREs are commonly cell-type-specific20,21, understanding CAD regulatory mechanisms at the cellular level is required to fully interpret the functional impact of risk variants. The assay for transposase-accessible chromatin with sequencing (ATAC-seq) is a widely adopted approach to systematically detect CREs22 and has been conducted in CAD-relevant cultured human coronary
artery SMCs and aortic endothelial cells. However, cultured cell models often do not fully recapitulate the complex cellular and regulatory landscape in vivo. Thus, single-nucleus ATAC-seq (snATAC-seq) of primary human coronary artery samples has the potential to provide a more complete regulatory map to unravel disease mechanisms in vivo.

Single-nucleus epigenomic profiling has recently been applied across various human tissues, including carotid arteries, however, to date there is still no large reference dataset spanning CAD progression in coronary arteries. In this study, we performed snATAC-seq-based chromatin profiling to uncover ~320,000 candidate CREs in human coronary arteries from 41 patients with varying clinical presentations of CAD. In generating this cell-specific chromatin atlas of the human coronary artery, we identified candidate CREs and transcription factors (TFs) for the major cell types or transition states in the coronary artery. We then applied these profiles to associate CAD risk variants with specific cell types by linking CREs to target gene promoters. Finally, we employed both allele-specific mapping and sequence-based predictive modeling to resolve genetic regulatory mechanisms that could help inform preclinical studies of CAD targets in the vessel wall.

Results

snATAC-seq profiling in human coronary artery. We performed snATAC-seq on coronary arteries (left anterior descending artery, left circumflex artery or right coronary artery) from 41 patients with various presentations of atherosclerosis using a droplet-based protocol (Fig. 1a, Supplementary Tables 1–4 and Extended Data Fig. 1). We isolated nuclei from a total of 44 frozen coronary segments using a protocol optimized for frozen tissues. After sequencing, we performed stringent quality control to retain highly informative nuclei (Supplementary Fig. 1). The libraries showed the expected insert size distributions and enrichment of reads at transcription start sites (TSSs) (Supplementary Fig. 1). Aggregating reads from all nuclei approximated bulk coronary ATAC-seq profiles derived from the same patient, further illustrating the quality of the single-nucleus dataset (Supplementary Fig. 2).

After filtering, we obtained a total of 28,316 high-quality nuclei and identified 14 clusters using iterative latent semantic indexing in ArchR for dimensionality reduction (Fig. 1b). Importantly, the identified clusters distinguished biological cell types rather than individual donor or other covariates (for example, age, sex) (Supplementary Figs. 1 and 3). We assigned each cluster to a coronary artery cell type using gene activity scores, which infer gene expression based on chromatin accessibility at established marker genes. Accessibility at SMC marker genes MYOC, MYH11, CNN1, TAGLN and ACTA2 defined four distinct clusters of SMCs, the most abundant cell type in our dataset (57.8 ± 17.6% of cells, Fig. 1b). We further identified clusters of endothelial cells (CLDN5), fibroblasts (TCF21, LUM), macrophages (CSF1R) and T cells/natural killer cells (TBX21) (Fig. 1c). Additional cluster annotations included pericytes, plasma (B) cells, mast cells and ‘unknown’ immune cells (resembling macrophages/mast cells). Data integration (Methods) showed our gene activity score-based annotations were in high agreement with recently reported scRNA-seq annotations from human coronary artery (Fig. 1d and Extended Data Figs. 2 and 3). In general, we observed higher immune cell proportions (cells in clusters 8–14) in atheroma and fibrocalcific coronary artery samples (44.1 ± 18.8%) relative to nonlesion or healthy controls (17.7 ± 8.3%), which is consistent with the cellular etiology of atherosclerosis progression (Fig. 1e and Supplementary Fig. 3). In samples devoid of adventitia (n = 11), we observed nearly absent fibroblasts as well as depletion of endothelial cells and pericytes (Fig. 1e). This is consistent with the expected cell composition of the outer adventitial layer and vasa vasaorum and supports the specificity of our cell-type annotations.

Characterization of cell-type-specific regulatory profiles. We next applied this coronary artery snATAC-seq dataset to characterize cell-type-specific cis-regulatory profiles. Using snATAC-seq gene scores we identified 5,121 marker genes across all cell types, which revealed both cell identity and/or disease response genes (Fig. 2a and Supplementary Data 1). By aggregating reads from all nuclei, we generated a master set of 323,727 peaks (Supplementary Data 2), which mostly map to intronic and intergenic sequences as expected (Extended Data Fig. 4 provides peak annotations). Notably, 54% were uniquely accessible in only one or limited cell types (Fig. 2b), emphasizing the benefits of single-nucleus profiling to define context-specific regulatory profiles that could be missed in bulk studies.

We next compared these marker genes and peaks with cultured human coronary artery SMC super enhancers detected by previous H3K27ac chromatin immunoprecipitation followed by sequencing (ChIP-seq) data. SMC ATAC-seq peak clusters (ATAC-seq peaks longer than 10 kb, Methods) showed the highest association with SMC super enhancers (Supplementary Fig. 4). These SMC super enhancers showed substantially higher regulatory potential for the identified SMC marker genes compared with the marker genes from all other cell types.

To investigate the TFs potentially driving the regulatory profiles/programs in each coronary artery cell type, we performed HOMER motif enrichment analysis for these marker peaks (Fig. 2c and Supplementary Data 3). Top enriched motifs in SMCs (MEF2 family, TEAD family, CARG box binding myocardin-serum response factor (Fig. 2d)) strongly agree with established SMC TFs in the literature. Similarly, we observed enrichment of ETS and SOX family motifs in endothelial cells, PU.1/SIPB and IRF motifs in macrophages, CEBP and AP-1 family members in fibroblasts, RUNX family motifs in T cells and GATA family motifs in mast cells (Fig. 2c). Besides defining established cell-type-specific TFs, we also discovered a number of lesser-known coronary TF motifs (for example, SMC21 in SMCs and fibroblasts) (Fig. 2e). Importantly, the TCF21 motif was also highly enriched in fibrocyocytes (Fig. 3c).

---

Fig. 1 | snATAC-seq profiling of 28,316 nuclei from human coronary arteries reveals cell-type chromatin accessibility patterns across 41 individuals. a, snATAC-seq was performed on nuclei isolated from frozen human coronary artery samples taken from explanted hearts from 41 unique patients. Samples came from segments of the left anterior descending coronary artery (LAD), left circumflex artery (LCX) or right coronary artery (RCA). After isolation using density gradient centrifugation, nuclei were transposed in bulk and mixed with barcoded gel beads and partitioning oil to generate gel beads in emulsions (GEMs). b, UMAP and clustering based on single-nucleus chromatin accessibility identifies 14 distinct coronary artery clusters. Each dot represents an individual cell colored by cluster assignment. c, UMAP plot of b colored by gene score for coronary artery cell-type marker genes, including myocardin (MYOC, SMCs), TCF21 (SMCs and fibroblasts), LUM (fibroblasts), CLDN5 (endothelial cells), CSF1R (macrophages) and TBX21 (T cells). d, Heatmap representing the contingency table highlighting correspondence between snATAC-seq and scRNA-seq cell-type assignments. e, Distribution of cell types across all of the snATAC-seq samples, divided by whether or not the corresponding sample had an adventitial layer. The schematic in a was created using BioRender. NK, natural killer cells.
Human coronary artery samples (n = 41)

Left circumflex artery (LCX)
Left anterior descending artery (LAD)
Right coronary artery (RCA)

Coronary artery disease progression

a

n = 28,316 nuclei

Endothelial (1)
Macrophage (9–10)
Smooth muscle (4–7)
Mast cell (12)
Pericyte (3)
Fibroblast (2)
Plasma/stem cell (11)
T/NK cell (13–14)

b

Candidate cell type-specific gene scores

MYOCD (0.1–1.0)
TCF21 (0.1–0.8)
LUM (0.1–0.6)
CLDN5 (0.1–1.0)
CSF1R (0.1–1.3)
TBX21 (0.2–1.1)

C

scRNA-seq

Row normalized co-occurrence

-0.5 0 3

scATAC-seq

Endothelial
Fibroblast
Macrophage
Mast
Pericyte
Plasma
SMC
T/NK
Unknown

Adventitia

No adventitia

Category 1
Category 2
Category 3

Category 1: Sample has no evidence of atherosclerosis
Category 2: Individual has evidence of atherosclerosis but sample is lesion free
Category 3: Sample has evidence of atherosclerosis/presence of lesion

d

e

Isolate and transpose nuclei
Generate snATAC libraries

Link peaks to target genes

Prioritize candidate functional non-coding GWAS (e.g. CAD) variants

Identify key coronary artery cell type
Specific transcription factors

Identify cell type specific chromatin accessibility QTLs

Candidate cell type-specific gene scores

Cell proportion

Cell type

Sample
providing epigenomic-based support for this TF previously shown to drive SMC modulation.\textsuperscript{7,62}

To determine whether cell-type-specific accessible regions were enriched for GWAS variants for CAD and other vessel wall phenotypes, we performed cell-type linkage disequilibrium (LD) score regression (LDSC).\textsuperscript{63} CAD and blood pressure GWAS variants were highly enriched in SMC, endothelial cell and macrophage peaks (Fig. 2f), while variants for pulse pressure and carotid...
Articles

NATURE GENETICS

Characterization of gene regulatory programs in SMCs. Given the extensive phenotypic plasticity of SMCs, we next investigated differences in the cis-regulatory profiles between contractile and modulated SMCs. While the studies from Alencar et al.9 and Pan et al.10 both provide compelling evidence of modulated SMC populations, we expanded upon the Wirka et al. study. This study included human coronary arteries and identified a modulated ‘fibromyocyte population’ (markers TNFRSF11B and FN1). SMCs in our dataset partition into four subclusters (Fig. 1b), referred to as C4–C7 (Fig. 3a). Clusters C5 and C6 have greater accessibility in differentiated SMC genes (MYH11 and CNN1), whereas clusters C4 and C7 have greater accessibility in phenotypically modulated SMC marker genes (TNFRSF11B and FN1) (Fig. 3a,b). To address potential noise due to sparsity of snATAC-seq-based gene scores, we also derived integrated RNA scores using mutual nearest neighbor integration and label transfer between identified anchors in the snATAC-seq data and Wirka et al. scRNA-seq data (Methods and Extended Data Figs. 2 and 3). This approach provided higher resolution to clearly delineate the SMC-derived fibromyocyte population based on TNFRSF11B, FN1 and other modulation markers (Fig. 3a and Extended Data Fig. 2). Consistently, chromVAR-based TF enrichment revealed highly enriched motifs for AP-1 family members (for example, ATF3) and TCF21 in the fibromyocyte cluster, which was depleted of differentiated SMC CArG box motif (Fig. 3c and Extended Data Fig. 2). Other TFs such as TEAD4 were enriched in all SMC clusters (Fig. 3c). Together, these results suggest that we can leverage single-nucleus accessibility profiles to understand regulatory drivers of SMC phenotypic modulation.

In a similar approach, we leveraged chromVAR motif deviations to perform trajectory analyses in SMC clusters. By assigning a path of accessibility from differentiated SMCs towards fibromyocytes (Fig. 3d), we identified enriched MEF2 and CArG motifs at the start of the trajectory, followed by enrichment of ETS and NFY motifs, then AP-1 and RUNX motifs in fibromyocytes (Fig. 3d). Using the scRNA-seq integrated data, we further identified 7,802 differentially accessible peaks (5,681 upregulated and 2,121 downregulated) between cells annotated as fibromyocytes versus traditional/differentiated SMCs (Supplementary Data 4). In particular, we identified 170 significantly upregulated and 108 downregulated promoter peaks in fibromyocytes (Fig. 3e and Supplementary Data 4). Promoters with higher accessibility in fibromyocytes include several extracellular matrix genes (for example, VCAN, COL4A3/4 and TNFAIP6), previously identified using scRNA-seq. However, we also revealed a number of candidate fibromyocyte markers such as the Rho GTPase effector gene, CDC42EP5, linked to actin-mediated migration/proliferation. Using HOMER de novo motif enrichment, we again observed AP-1 (FRA1), RUNX, TEAD and TCF21 motifs in upregulated fibromyocyte peaks, but also motifs for inflammatory response factors STAT3 and ARID5A (ref. 16) (Fig. 3f and Supplementary Data 4). The RUNX motif includes RUNX2, suggesting this fibromyocyte population may include osteochondrogenic cells17,18. Conversely, MEF2A and CArG box motifs were the top enriched motifs in the downregulated peaks. Genomic region enrichment analysis identified extracellular matrix organization and cell migration processes in upregulated fibromyocyte peaks, compared with enrichment for SMC contraction and related processes in downregulated peaks (Fig. 3g). Together, these snATAC-based results confirm the role of TCF21 and also identify candidate TFs underlying SMC phenotypic modulation during CAD.

Annotation of target cell types and genes at CAD GWAS loci. Noncoding GWAS variants are enriched in CREs and often operate in a cell-type-specific manner19–21. We thus prioritized candidate functional CAD GWAS variants using a multi-tiered approach. To first identify variants in CREs, we overlapped CAD lead variants (and variants in high LD (r² > 0.8; EUR)) from two recent CAD GWAS meta-analyses22,23 with snATAC peaks (Fig. 4a). This resolved a subset of variants (±50 bp) overlapping both shared and marker peaks (Fig. 4b and Supplementary Data 5), with the majority of CAD SNPs residing within SMC peaks, followed by macrophage, fibroblast and endothelial cell peaks. Based on these overlaps, we then highlighted target cell types for CAD regulatory variants (Fig. 4c). Several top candidate CAD variants map to cell-type-specific peaks, including rs3918226 at the NOS3 (endothelial nitric oxide synthase) locus and rs9337951 at the KIAA1462/JCAD (junctural cadherin 5-associated) locus24, both within endothelial peaks, and rs7500448 at the CDH13 (T-cadherin) locus within an SMC peak (Fig. 4c). Other CAD variants map to peaks shared across SMC, fibroblast and pericyte cell types such as rs1537373 at the 9p21 locus (CDKN2B-AS1/ANRIL),

Fig. 3 | Subcluster analysis of SMC accessible chromatin identifies fibromyocyte regulatory programs. a, snATAC-seq UMAP for the four SMC clusters (C4–C7). The UMAP was colored by snATAC-seq cluster (top) and by cell-type labels assigned by scRNA-seq label transfer (bottom). Dashed lines demarcate boundary of cells with increased SMC marker gene scores (clusters C5 and C6), or decreased SMC marker gene scores (C4 and C7). Integrated snATAC/scRNA UMAP highlights both fibromyocyte SMCs and traditional SMCs (demarcated by dashed lines and colors) within clusters C4–C7. ‘Pericyte 1’ and ‘Pericyte 2’ labels from scRNA-seq were also mixed in clusters C4 and C5. b, Quantification of imputed snATAC-seq gene scores highlights higher chromatin accessibility at differentiated SMC marker genes MYH11 and CNN1 in clusters C5 and C6, and higher accessibility at modulated SMC/fibromyocyte marker genes TNFRSF11B and FN1 in clusters C4 and C7. P values were calculated using a one-sided Wilcoxon test. The exact P values are as follows: MYH11: P = 0; CNN1: P = 0; TNFRSF11B: P = 0; FN1: P = 1.7 x 10⁻⁶. The N values for nuclei in each cluster are as follows: C4: 6,275; C7: 1,971; C5: 6,134; C6: 1,988. c, ChromVAR TF motif enrichment for differentiated SMC CArG box in traditional SMC and enrichment for ATF3 and TCF21 motifs in modulated SMC/fibromyocyte and fibroblast clusters. The TEAD4 motif is enriched in both contractile and modulated SMCs. d, Left, scatter plot overlay of SMC UMAP depicts the trajectory path from differentiated SMC to modulated SMC/fibromyocyte subclusters (left). Motif enrichment heatmap shows the top enriched motifs across the trajectory pseudotime (right). Values represent accessibility gene Z-scores. e, Volcano plot of differential peak analysis (subset to promoter peaks) comparing fibromyocytes and traditional SMCs. Fibromyocyte and SMC annotated cells were defined based on RNA label transferring (Methods) and significant peaks determined by a Wilcoxon test as implemented in ArchR. Peaks with significant differences at FDR < 0.05 and log2 fold change > 1 were colored light red (fibromyocyte upregulated) and blue (fibromyocyte downregulated). f, Top enriched motifs within the total upregulated fibromyocyte peaks (5,681) detected using HOMER de novo enrichment analysis with the hypergeometric distribution test. P values shown are unadjusted for multiple comparisons. g, Functional annotation of fibromyocyte upregulated (light red) and downregulated (blue) peaks conducted using GREAT with the binomial distribution test. Top enriched biological processes functional terms are listed. P values shown are unadjusted for multiple comparisons.
rs2327429 (upstream of TCF21) and rs9515203 at COL4A2 (Fig. 4c). The lead variant rs9349379 (PHACTR1/EDN1), disrupting an MEF2 binding site71, is located within SMC and macrophage peaks (Fig. 4c). We identified other CAD variants in macrophage peaks (for example, rs7296737 at SCARB1 and rs12246441 at TSPAN14) (Extended Data Fig. 7). We also prioritized a number of CAD loci within CREs acting through more than one cell type and confirmed SMC-specificity for previously validated loci such as LMOD1 (ref. 72) (Fig. 4d).

Since noncoding variants do not always regulate the nearest gene(s), we also linked candidate variants to target promoters through co-accessibility and scRNA-seq integration (Methods and Fig. 4a). For instance, the SMC peak-containing variant, rs7500448, shows high co-accessibility with the CDH13 promoter (Extended Data Fig. 5), which is also an artery-specific expression quantitative trait locus (eQTL) for CDH13 in the Genotype-Tissue Expression project (GTEx). Another relevant example is rs998584, located within a strong fibroblast peak 3′ of VEGFA, which is highly
Fig. 4 | Single-nucleus chromatin accessibility further resolves mechanisms for functional CAD GWAS loci. a. To prioritize candidate CAD-associated GWAS variants we used a multi-tiered strategy, first by taking variants in moderate to high LD with the reported lead variants ($r^2 \geq 0.8$). We next prioritized variants overlapping snATAC-seq peaks and narrowed down the cell type(s) whereby these variants are potentially functioning. Finally, we determined whether candidate variants are within TF motifs and linked to target genes through co-accessibility and links to gene expression through scRNA-seq integration (Peak2Gene). b. Overlap of LD-expanded ($r^2 \geq 0.8$; EUR) CAD GWAS variants (±50 bp) with coronary artery cell-type peaks (both from the total peak set and marker peaks). LD-expanded SNPs were obtained from two recent CAD GWAS studies (van der Harst et al.\(^1\)) and Koyama et al.\(^2\)) that performed trans-ancestry meta-analysis. c. Examples of the benefits of snATAC-seq for pinpointing cell types whereby candidate CAD regulatory variants are acting. Highlighted are candidate functional variants at the 9p21 (CDKN2B-AS1/ANRIL), TARID-TCF21, NOS3, KIAA1462/JCAD, CDH13, COL4A2 and PHACTR1 loci. d. Heatmap showing number of peaks per cell type overlapping CAD GWAS variants for 100 of the CAD loci (van der Harst et al.\(^3\)). Full overlaps of CAD GWAS variants with snATAC-seq peaks are provided in Supplementary Data 5. The schematic in a was created using BioRender.
co-accessible with the VEGFA promoter (Extended Data Fig. 5). In an orthogonal approach, we combined co-accessibility and RNA integration to identify peaks where accessibility correlates with target gene expression (referred to as Peak2Gene links). We identified a total of 148,617 Peak2Gene links when aggregating all cell types (Extended Data Fig. 5), including for many CAD risk variants (Supplementary Data 5). Together, these single-nucleus chromatin annotations refine candidate regulatory mechanisms at CAD GWAS loci, for future functional validation in the appropriate cell types.

Prioritizing cell-type-specific CAD functional variants. Chromatin accessibility quantitative trait locus (caQTL) mapping is a powerful association analysis to resolve candidate GWAS regulatory mechanisms\(^\text{19-28}\). We thus calculated caQTLs in our dataset in four major coronary cell types (SMCs, macrophages, fibroblasts and endothelial cells) (Supplementary Data 6). Given our modest sample size \((n = \text{41})\), we used RASQUAL\(^\text{17}\) for caQTL mapping to capture both population and allele-specific effects (Methods), as done previously for cultured coronary artery SMCs\(^\text{80}\). As expected, the number of quantitative trait loci (QTLs) discovered per coronary cell type was proportional to the respective number of annotated nuclei (Fig. 5a), with the most belonging to SMCs \((1,984 \text{ at } 5\% \text{ false discovery rate (FDR)})\). Further, 26\% of these single-cell caQTLs were also observed in coronary artery bulk ATAC-seq libraries (Supplementary Data 7) from the same patients \((n = \text{35})\, \text{with } 86\% \text{ consistent effect size directions (Extended Data Fig. 6). To determine whether these caQTLs regulate gene expression, we queried these variants for eQTL signals in GTEx artery tissues (coronary, aorta and tibial). Out of the 1,984 unique SMC caQTLs \((5\% \text{ FDR})\), 47\% were significant eQTLs (GTEx 5\% FDR) in at least one GTEx arterial tissue. Most of the coronary SMC caQTLs that are GTEx eQTLs are shared across all artery types (Fig. 5b). We also identified 71\% concordant coronary artery caQTL and eQTL effect sizes (Fig. 5c), which is consistent with reported findings in human T cells\(^\text{8}\).

We next applied these cell-type caQTLs to further dissect CAD-regulated mechanisms in the vessel wall. One example is rs4450010 at the MEF2D migraine/cardiovascular-associated gene within an SMC peak. The rs4450010-T allele creates a TEF1 (TEAD) binding site and (TEAD) correlates with both increased peak accessibility (Fig. 5d) and increased MEF2D RNA expression in GTEx arterial tissues (Supplementary Fig. 5). Several CAD GWAS variants were significant caQTLs in SMCs or macrophages. For example, rs13324341 within intron 1 of MRAS (muscle RAS oncogene homolog), also in a DNase site\(^\text{41}\), is both an SMC caQTL and a strong eQTL in GTEx arterial tissues (Supplementary Fig. 5). Other top CAD GWAS-overlapping caQTLs include, among others, rs73551705 (BMP1) and rs17293632 (SMAD3) in SMCs and rs72844419 (GGCX) and rs10418535 (FCHO1) in macrophages (Extended Data Fig. 6 and Fig. 5e). At the MRAS locus, the rs13324341 minor allele T (increased CAD risk) creates an MEF2 binding site (Fig. 5f) and correlates with both increased accessibility and increased MRAS messenger RNA levels (Fig. 5g). These MEF2D, MRAS, BMP1, SMAD3 and FCHO1 SNPs (or highly linked SNPs, \(\text{r}^2 > 0.8\)) are all significant caQTLs \((5\% \text{ FDR})\) in bulk coronary artery ATAC-seq data (Supplementary Data 7).

To complement our QTL-based approach, we employed a machine learning-based strategy to assign sequence importance scores to CAD variants \((10,117 \text{ tested})\) with effects on chromatin accessibility\(^\text{82}\). Across three similar approaches (GkmExplain\(^\text{83}\), gkmpredict, deltaSVM\(^\text{84}\)), we identified 127 high- or moderate-confidence CAD variants with predicted functional effects on chromatin accessibility (Supplementary Data 8). Of these, 102 \((80\%)\) had functional probability scores \(> 0.6\) in RegulomeDb 2.0 and were annotated by enhancer, promoter and TF ChiP-seq enrichment as well as motif disruption (Supplementary Data 8). About half of these variants were predicted to be functional in a single cell type. One representative CAD variant, rs1320496 (LIPA), resides in a strong macrophage-specific peak, with the T allele \((\text{increased CAD risk})\) creating putative binding sites for SPIB, TBX21 and IRF4/8 (Fig. 5h). Another intergenic SNP, rs10418535-C allele \((\text{increased CAD risk})\) disrupts a PU.1/IRF motif and is predicted to attenuate chromatin accessibility (Extended Data Fig. 7). rs10418535 is also a macrophage caQTL with a positive effect for the T allele, consistent with the deltaSVM prediction (Fig. 5e and Extended Data Fig. 7). Together, we demonstrate cell-type caQTL mapping and machine learning are complementary approaches to pinpoint candidate functional disease risk variants at high resolution.

PRDM16 and TBX2 are top candidate CAD TFs. Epigenomic profiles in disease-relevant tissues have been shown to resolve the correct target gene(s) at GWAS risk loci, which are often incorrectly annotated to the nearest gene\(^\text{45}\). Importantly, our data nominate many targets of CAD risk variants in the vessel wall, including two previously unannotated TFs as candidate causal genes at their respective loci \((\text{Supplementary Data 5})\). The first locus on chr17 harbors dozens of tightly linked CAD variants within peaks in the BCAS3 gene (Extended Data Fig. 8). However, Peak2Gene analysis demonstrates stronger links between these peaks and TBX2 expression in coronary arteries. At the second locus on chr1, several linked CAD variants are located within SMC peaks \(5\% \text{ of the } \text{ACTRT2} \text{ gene and are highly correlated with PRDM16 and LINCO09982 (PRDM16 divergent transcript) expression, but not with other genes at the locus (Fig. 6a)\). This locus also harbors an independent missense CAD-associated SNP \((\text{rs2493292\; p.Pro634Leu})\) in exon 9 of PRDM16, suggesting both noncoding and coding effects on PRDM16 expression (Fig. 6a). These results are consistent

---

Fig. 5 | Identification of genetic variants that regulate chromatin accessibility within coronary artery cell types. a. The number of caQTLs identified using RASQUAL \((10\% \text{, } 5\% \text{ and } 1\% \text{ FDR cutoffs})\) within a cell type is proportional to the number of annotated cells. The color represents the cell type and shape represents the FDR cutoff. b. Upset plot for SMG caQTLs that are eQTLs in GTEx arterial tissues. Bars represent the intersection size for overlap of eQTLs between coronary artery, aorta and tibial artery. c. Comparison of RASQUAL effect sizes with GTEx effect sizes (beta). d. e. Boxplots highlighting SMC \((n = 40)\)-normalized accessibility for MEF2D and MRAS caQTL variants \((d)\) and macropage \((n = 39)\)-normalized accessibility for FCHO1 and MARCO caQTL variants \((e)\). Vst, variance stabilizing transformation. The Q values represent the lead caQTL SNP Q values \((\text{Benjamini-Hochberg})\) generated from the likelihood-ratio test for the respective peak in RASQUAL. Boxplots \((d, e)\) represent the median and interquartile range (IQR), with upper \((75\%)\) and lower \((25\%)\) quartiles shown and each dot representing a separate individual. f. Example genome browser tracks showing CAD-associated caQTL at the MRAS locus in an SMC-specific peak. The T allele for rs13324341 creates an MEF2 putative binding site. g. In GTEx artery \((\text{aorta shown here, } n = 387 \text{ unique individuals})\) the T allele for rs13324341 is highly associated with increased MRAS mRNA levels. The cis-eQTL \(P\) value is shown from the GTEx pipeline that performs linear regression between genotype and normalized gene expression levels. The boxplot \((\text{black})\) within the violin plot includes median \((\text{white line})\) and IQR from 25\% to 75\%. h. Example of prioritization of functional CAD variants using lsgkm machine learning-based prediction. The rs1320496 variant at the LIPA locus \((\text{chromosome } 10)\) resides in a strong macrophage peak. The T allele is predicted to create a putative SPIB binding site and increased chromatin accessibility. Feature importance score tracks for effect and noneffect alleles are visualized by gkmExplain \((\text{Methods})\). Alt, T; Ref, C.
with activity-by-contact enhancer-gene mapping of CAD SNPs in coronary artery in ENCODE (Supplementary Table 6), differential expression (Supplementary Table 7) and cis-eQTLs (TBX2) in artery tissues (Supplementary Data 10), supporting PRDM16 and TBX2 as target CAD genes.

Both PRDM16 and TBX2 are snATAC SMC marker genes and, remarkably, PRDM16 is one of the top SMC marker genes along with known SMC gene *LMOD1* (ref. 72) (Supplementary Data 1). Given the similar gene score enrichment of PRDM16 and *LMOD1* in SMC, we ranked PRDM16 by correlating all SMC gene scores and integrated RNA scores with *LMOD1* (Fig. 6b). Interestingly, PRDM16 was modestly correlated with traditional SMC markers, and negatively correlated with fibromyocyte marker genes. This may implicate PRDM16 as an SMC injury-response gene as opposed to an SMC identity...
**Fig. 6** | **PRDM16 is a CAD-associated key driver transcriptional regulator in SMCs.**

**a.** Genome browser track highlighting the association between CAD-associated SNPs and SMC marker genes through co-accessibility (peak2gene) detected by snATAC-seq data (Methods). The red loops represent the association between PRDM16 promoter and CAD-associated SNPs. **b.** Correlation coefficients of snATAC-seq/scRNA-seq integration scores gene expression levels between LMO1 and genome-wide coding genes in SMCs. Genes were ranked by Pearson's correlation coefficient with LMO1. Representative positive- and negative-correlated SMC gene names are labeled. **c.** Clinical trait enrichment for PRDM16-containing module in subclinical mammary artery in STARNET gene regulatory network datasets. 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 (DeG) enrichment was estimated by a hypergeometric test. **d.** Movat pentachrome staining and PRDM16 (red) and α-SMA (green) immunofluorescence staining of atherosclerotic human coronary artery segments—LAD from normal–Stage I, Stage III–IV and Stage V–VI lesions based on Stary classification stages. Whole-slide images captured from ×20 confocal microscopy stitched tiles. PRDM16/α-SMA co-staining (see arrows) depicted in yellow from merged images. DAPI (blue) marks nuclei. n = 4 per group. Scale bar, 1 mm; except for region of interest (ROI): scale bar, 100 µm.

In addition to scRNA-seq datasets, PRDM16 and TBX2 are enriched in mural cells (SMCs and pericytes) in both human coronary artery and mouse aorta (Supplementary Fig. 6). To gain further insight into these two TF genes, we queried the Stockholm–Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) gene regulatory networks across seven cardiometabolic tissues (n = 600), which revealed both PRDM16 and TBX2 as significant key driver genes in artery tissues (Supplementary Table 8 and Extended Data Fig. 8). In subclinical artery, the PRDM16-regulated module was highly enriched for the presence of atherosclerotic lesions and CAD severity, as well as metabolic clinical traits (Fig. 6c and Extended Data Fig. 8). Finally, we confirmed PRDM16 protein expression via immunofluorescence of normal, subclinical and advanced atherosclerotic coronary artery segments, with alpha-smooth muscle actin (α-SMA) and LMOD1 as positive controls (Fig. 6d and Extended Data Fig. 9). Similar to a-SMA, PRDM16 localized to the SMCs in the medial layer and small vessels in the vasa vasorum in healthy arteries; however, expression was more restricted to the vasa vasorum and endothelium in diseased arteries (Fig. 6d and Extended Data Fig. 9).

While we highlight these two examples, particularly PRDM16, this coronary dataset can be similarly utilized to prioritize mechanisms at many other CAD loci.

**Discussion**

In this study we have generated a single-nucleus atlas of human coronary artery chromatin accessibility for over 40 patients encompassing healthy and atherosclerotic samples, which captures gene regulation in vivo. Over half of the 323,767 identified CREs (54%) are unique to a specific cell type or a limited number of cell types, underscoring the power of single-nucleus epigenomics for resolving unique cell-type regulatory processes. Our snATAC-seq results also provide direct insights into SMC phenotypic modulation. More specifically, we discovered accessible regions, genes and putative TF motifs that may drive the transition of native SMCs towards modulated SMCs (for example, fibromyocytes). Finally, using an integrative statistical genetics and machine learning approach we prioritized cell-specific candidate regulatory variants and mechanisms underlying CAD loci.

There are now over 200 genetic loci associated with CAD risk, primarily located within noncoding genomic regions.

---

**Image Descriptions**

- **Fig. 6a:** Genome browser track highlighting association between CAD-associated SNPs and SMC marker genes through co-accessibility (peak2gene) detected by snATAC-seq data. The red loops represent the association between PRDM16 promoter and CAD-associated SNPs. Genes were ranked by Pearson's correlation coefficient with LMO1.
- **Fig. 6b:** Correlation coefficients of snATAC-seq/scRNA-seq integration scores gene expression levels between LMO1 and genome-wide coding genes in SMCs. Representative positive- and negative-correlated SMC gene names are labeled.
- **Fig. 6c:** Clinical trait enrichment for PRDM16-containing module in subclinical mammary artery in STARNET gene regulatory network datasets. Case/control differential gene expression (DeG) enrichment was estimated by a hypergeometric test.
- **Fig. 6d:** Movat pentachrome staining and PRDM16 (red) and α-SMA (green) immunofluorescence staining of atherosclerotic human coronary artery segments—LAD from normal–Stage I, Stage III–IV and Stage V–VI lesions based on Stary classification stages. Whole-slide images captured from ×20 confocal microscopy stitched tiles. PRDM16/α-SMA co-staining depicted in yellow from merged images. DAPI (blue) marks nuclei. n = 4 per group. Scale bar, 1 mm; except for region of interest (ROI): scale bar, 100 µm.
This single-nucleus coronary artery epigenomic landscape provides a valuable resource to disentangle the target cell type(s), candidate causal genes and variants at the expanding number of CAD risk loci in diverse populations. For example, we highlight a top eQTL, rs13324341, at the MRAS locus which alters an MEF2 binding site in SMCs. Given the role of MRAS in Noonan syndrome-associated cardiomyocyte hypertrophy\(^8\),\(^9\), these results may provide clues into SMC growth responses during CAD. We also highlight top predicted CAD regulatory variants acting in one or more cell types (for example, rs1320496 at LIPA in macrophages). This dataset can also be leveraged to interrogate GWAS loci for related common vascular diseases (for example, hypertension or coronary artery calcification).

By taking into account co-accessibility and scRNA-seq integration, we systematically link CAD risk variants to target gene promoters. This is critical given that GWAS variants are estimated to only target the nearest gene ~50% of the time. For example, using this approach we nominate two TF genes, PRDM16 and TBX2, as top candidate genes at their respective loci. PRDM16 is a top SMC marker gene from snATAC-seq gene scores; however, it may not be limited to marking SMC identity. PRDM16 (MEL1) is a TF known for roles in metabolism and controlling brown-fat-to-skeletal muscle switches\(^10\)–\(^12\). However, PRDM16 is enriched in GTEx arterial tissues and was identified as a key driver gene in STARNET artery tissue, consistent with our snATAC data. PRDM16 regulates TGFB\(^-\) signaling\(^13\) through direct interactions with Smad\(^-\) and SKI\(^-\) proteins, both of which are associated with CAD\(^14\). PRDM16 may play key roles in endothelial cells in arterial flow recovery\(^15\). Similarly, TBX2 is enriched in GTEx arterial tissues and SMC clusters in our dataset, consistent with earlier studies showing that Tbx2 activates SRF\(^16\). TBX2 also links to relevant CAD pathways such as BMP, TGFB and FGF signaling\(^17\). Functional follow-up studies to investigate target binding sites and affected SMC processes for these TFs may reveal additional mechanistic drivers of disease risk.

While this study provides high-resolution insights into coronary artery gene regulatory signatures using primary human tissue samples, there are some known limitations. Given the lack of available lineage-tracing snATAC-seq datasets, we cannot fully annotate intermediate cell types or precisely resolve their origins and fates during atherosclerosis\(^18\)–\(^20\). For example, some SMC-derived cells may be incorrectly annotated in T cell clusters, consistent with Alencar et al.\(^21\) and Hansson et al.\(^22\). Also, it is worth noting that we captured more nuclei from subclinical lesions compared with advanced atherosclerotic lesions, which potentially reflects higher difficulty in nuclei extraction for diseased samples. Finally, given our modest sample size for QTL-based studies, we were underpowered to discover a large number of eQTLs for less abundant cell types (for example, endothelial and T cells) or less frequent transition states. Future studies that can capture more nuclei per individual, especially in diseased coronary samples, will facilitate identification of additional context-specific regulatory mechanisms. Decreasing costs and adoption of single-nucleus and spatial sequencing technologies may further improve discovery of regulatory variants and mechanisms through multi-modal and integrative approaches\(^23\)–\(^25\).

In summary, we provide an atlas of chromatin accessibility in both healthy and atherosclerotic human coronary arteries. These cell-type-specific epigenomic profiles characterize cis-regulatory programs at base-pair resolution to further our understanding of cell plasticity and heritable disease risk in the coronary vessel wall. We anticipate this will provide a valuable resource for the field, and act as a key next step toward functionally interrogating causal disease processes and informing preclinical studies to treat atherosclerosis.

**Online content**

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-022-01069-0.

Received: 4 June 2021; Accepted: 31 March 2022; Published online: 19 May 2022

**References**

1. Libby, P. Inflammation in atherosclerosis. *Nature* **420**, 688–694 (2002).
2. Soulhol, C., Harmsen, M. C., Evans, P. C. & Krenning, G. Endothelial-mesenchymal transition in atherosclerosis. *Cardiovasc. Res.** **114**, 565–577 (2018).
3. Stary, H. C. et al. A definition of advanced types of atherosclerotic lesions and a histological classification of atherosclerosis. A report from the Committee on Vascular Lesions of the Council on Arteriosclerosis, American Heart Association. *Circulation* **92**, 1355–1374 (1995).
4. Winkels, H. et al. Atlas of the immune cell repertoire in mouse atherosclerosis defined by single-cell RNA-sequencing and mass cytometry. *Circ. Res.** **122**, 1675–1688 (2018).
5. Cochain, C. et al. Single-cell RNA-seq reveals the transcriptional landscape and heterogeneity of aortic macrophages in murine atherosclerosis. *Circ. Res.** **122**, 1661–1674 (2018).
6. Fernandez, D. M. et al. Single-cell immune landscape of human atherosclerotic plaques. *Nat. Med.* **25**, 1576–1588 (2019).
7. Wiria, R. C. et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019).
8. Depuydt, M. A. C. et al. Microanatomy of the human atherosclerotic plaque by single-cell transcriptomics. *Circ. Res.** **127**, 1437–1453 (2020).
9. Pan, H. et al. 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 (2020).
10. Alencar, G. E. et al. Stem cell pluripotency genes Klf4 and Oct4 regulate complex SMC phenotypic changes critical in late-stage atherosclerotic lesion pathogenesis. *Circulation* **142**, 2045–2059 (2020).
11. Wang, Y. et al. Clonally expanding smooth muscle cells promote atherosclerosis by escaping efferocytosis and activating the complement cascade. *Proc. Natl Acad. Sci. USA** **117**, 15818–15826 (2020).
12. Nikpay, M. et al. A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
13. van der Harst, P. & Verweij, N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.** **122**, 433–443 (2018).
14. Nelson, C. P. & al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391 (2017).
15. Koyama, S. et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177 (2020).
16. Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. Preprint at medRxiv https://doi.org/10.1101/2021.05.24.21257377 (2021).
17. Erdmann, J., Kessler, T., Munoz Venegas, L. & Schunkert, H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc. Res.* **114**, 1241–1257 (2016).
18. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
19. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
20. Heinz, S., Romanov, C. E., Benner, C. & Glass, C. K. The selection and function of cell type-specific enhancers. *Nat. Rev. Mol. Cell Biol.* **16**, 144–154 (2015).
21. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
22. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods** **10**, 1213–1218 (2013).
23. Miller, C. L. et al. Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **7**, 12092 (2016).
24. Liu, R. E. & Montag, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression commlicate GWAS follow-up. *Nat. Genet.* **51**, 768–769 (2019).
83. Shrikumar, A., Prakash, E. & Kundaje, A. GkmExplain: fast and accurate interpretation of nonlinear gapped k-mer SVMs. Bioinformatics 35, i173–i182 (2019).
84. Lee, D. et al. A method to predict the impact of regulatory variants from DNA sequence. Nat. Genet. 47, 955–961 (2015).
85. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. Nature 593, 238–243 (2021).
86. Higgins, E. M. et al. Elucidation of MRAS-mediated Noonan syndrome with cardiac hypertrophy. JCI Insight 2, e91225 (2017).
87. Seale, P. et al. PRDM16 controls a brown fat/skeletal muscle switch. Nature 454, 961–967 (2008).
88. Seale, P. et al. Transcriptional control of brown fat determination by PRDM16. Cell Metab. 6, 38–54 (2007).
89. Kajimura, S. et al. Initiation of myoblast to brown fat switch by a PRDM16-C/EBP-β transcriptional complex. Nature 460, 1154–1158 (2009).
90. Liu, D. et al. PRDM16 upregulation induced by microRNA-448 inhibition alleviates atherosclerosis via the TGF-β signaling pathway inactivation. Front. Physiol. 11, 846 (2020).
91. Warner, D. R. et al. PRDM16/MEL1: a novel Smad binding protein expressed in murine embryonic orofacial tissue. Biochim. Biophys. Acta 1773, 814–820 (2007).
92. Takahata, M. et al. SKI and MEL1 cooperate to inhibit transforming growth factor-β signal in gastric cancer cells. J. Biol. Chem. 284, 3334–3344 (2009).
93. Craps, S. et al. Prdm16 supports arterial flow recovery by maintaining endothelial function. Circ. Res. 129, 63–77 (2021).
94. Barron, M. R. et al. Serum response factor, an enriched cardiac mesoderm obligatory factor, is a downstream gene target for Tbx genes. J. Biol. Chem. 280, 11816–11828 (2005).
95. Shirai, M., Imanaka-Yoshida, K., Schneider, M. D., Schwartz, R. J. & Morisaki, T. T-box 2, a mediator of Bmp-Smad signaling, induced hyaluronan synthase 2 and Tgfβ2 expression and endocardial cushion formation. Proc. Natl Acad. Sci. USA 106, 18604–18609 (2009).
96. Hansson, G. K., Jonasson, L., Holm, J. & Claesson-Welsh, L. Class II MHC antigen expression in the atherosclerotic plaque: smooth muscle cells express HLA-DR, HLA-DQ and the invariant gamma chain. Clin. Exp. Immunol. 64, 261–268 (1986).
97. Cao, J. et al. Joint profiling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380–1385 (2018).
98. Ma, S. et al. Chromatin potential identified by shared single-cell profiling of RNA and chromatin. Cell 183, 1103–1116.e20 (2020).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022, corrected publication 2022
Methods

Ethics statement. All research described herein complies with ethical guidelines for human subjects research under approved Institutional Review Board (IRB) protocols at Stanford University under approved IRB protocols and with written, informed consent. Participants were not compensated for this study. Hearts were harvested in cardioplegic solution and rapidly transported from the operating room to the adjacent laboratory on ice. The proximal 5–6 cm of three major coronary vessels (left anterior descending, left circumflex and right coronary artery) were dissected from the epicardium on ice, trimmed of surrounding adipose (and in some samples the adventitia), rinsed at cold PBS and snap-frozen in liquid nitrogen. Coronary artery samples were also obtained at Stanford University (from Donor Network West and California Transplant Donor Network) from nondiseased donor hearts rejected by surgeons for heart transplantation and procured for research studies. All hearts were procured after written, informed consent from legal next-of-kin or authorized parties for the donors. Reasons for rejected hearts include size incompatibility, comorbidities or risks for cardiotoxicity. Hearts were arrested in cardioplegic solution and transported on ice following the same protocol as hearts used for transplant. Explanted hearts were generally classified as ischemic or nonischemic cardiomyopathy and previous ischemic events and evidence of atherosclerosis was obtained per the Stanford Heart Transplant Protocol and reviewed with the Department of Hospital Medicine and Clinical Hospital and Clinics. The disease status of coronary segments from donor and explanted hearts was also evaluated by gross inspection at the time of harvest (for presence of lesions), as well as histological analysis of adjacent frozen tissues embedded in OCT blocks. Frozen tissues were transferred to the University of Virginia through a material transfer agreement and IRB-approved protocols. All samples were then stored at −80°C until the day of processing.

Coronary artery sample processing and nuclei isolation. We performed snATAC-seq on four coronary artery samples per day. For nuclear isolation we used a similar protocol to Omni-ATAC that was optimized for frozen tissues and reported lower mitochondrial reads. We used approximately 50 mg of tissue per sample and the full nuclear isolation protocol is provided in the Supplementary Methods. After the iodixanol gradient step we then carefully took the band containing the nuclei (setting the pipette volume to 100 µl) and added the nuclei to 1.3 ml of cold Nuclei Wash Buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 1% BSA, 0.1% Tween 20) in a 1.5 ml Lo-Bind microcentrifuge tube. The microcentrifuge tube was inverted gently five times, nuclei gently mixed by pipetting (setting the pipette volume to 1 ml) and contents passed through a 40-µm Falcon cell strainer (Corning) into a new 1.5 ml Lo-Bind microcentrifuge tube (Eppendorf). Nuclei were pelleted by centrifugation for 5 min at 500 g at 4°C and supernatant carefully removed. Finally, this nuclei pellet was gently resuspended in 100 µl of Nuclei Buffer provided with the kit (diluted from 20X Stock to 1X working concentration with nuclease-free water) by gently pipetting up and down. Samples and nuclei were kept on ice for all steps of the nuclear isolation. For each sample we measured the nuclei concentration by taking the mean of two separate counts using Trypan blue (Thermo Fisher) and the Countess II instrument (Thermo Fisher). Post cell lysis, we generally observed less than 5% live cells when counts using Trypan blue (Thermo Fisher) and the Countess II instrument (described below). After the iodixanol gradient step we then carefully took the band containing the nuclei (setting the pipette volume to 100 µl) and added the nuclei to 1.3 ml of cold Nuclei Wash Buffer (10 mM Tris- HCl (pH 7.4), 10 mM NaCl, 3 mM MgCl₂, 1% BSA, 0.1% Tween 20) in a 1.5 ml Lo-Bind microcentrifuge tube. The microcentrifuge tube was inverted gently five times, nuclei gently mixed by pipetting (setting the pipette volume to 1 ml) and contents passed through a 40-µm Falcon cell strainer (Corning) into a new 1.5 ml Lo-Bind microcentrifuge tube (Eppendorf). Nuclei were pelleted by centrifugation for 5 min at 500 g at 4°C and supernatant carefully removed. Finally, this nuclei pellet was gently resuspended in 100 µl of Nuclei Buffer provided with the kit (diluted from 20X Stock to 1X working concentration with nuclease-free water) by gently pipetting up and down. Samples and nuclei were kept on ice for all steps of the nuclear isolation. For each sample we measured the nuclei concentration by taking the mean of two separate counts using Trypan blue (Thermo Fisher) and the Countess II instrument (Thermo Fisher). Post cell lysis, we generally observed less than 5% live cells when visualizing with the Countess, consistent with proper lysis.

snATAC-seq library preparation. We used the 10x Genomics Chromium Single Cell ATAC Kit for all snATAC-seq experiments (additional details are provided in the Supplementary Methods). The full protocols for the snATAC-seq data generation are available at the following link: https://support.10xgenomics.com/

snATAC-seq library sequencing. snATAC-seq libraries were shipped on dry ice to the Genome Core Facility at the Icahn School of Medicine at Mount Sinai (New York, NY, USA) for sequencing on an Illumina NovaSeq 6000. There, 40 libraries were sequenced using a NovaSeq S Prime flow cell (100 cycles, 2x 50bp) and 4 libraries were sequenced using a NovaSeq S Prime flow cell (100 cycles, 2x 50bp).

Raw snATAC-seq data processing and quality control. In-house snATAC-seq data were preprocessed using the 10x Genomics pipeline (Cell Ranger ATAC v1.2.0 (ref. 1)) using the hg38 genome and default parameters. Samples from different patients were processed separately. Individual cells with mitochondrial reads ≥ 5% were kept for downstream analysis (TSS enrichment ≥ 7, unique barcode number ≥ 10,000 and doublet ratio < 1.5). Quality control measurements and filtering were conducted using Cell Ranger (v1.0.1)(1).

Clustering of coronary artery snATAC-seq data. snATAC-seq reads from different individuals were combined at the single-nucleus level and then mapped to each 500-bp bin across the hg38 reference genome for dimensionality reduction and clustering. The dimensionality reduction was conducted using a latent semantic indexing algorithm and the 25,000 bins with the highest signal variance across individual cells were selected as input. The top 30 dimensions were selected for cell clustering. Cell clusters were identified by a shared nearest neighbor modularity optimization-based clustering algorithm from the Seurat (v4.0.0)(99) package. Batch effect removal was conducted using the Harmony (v1.0.1)(100) package. We did not observe any improvement after batch correction using Harmony(100).

snATAC-seq gene scores and cell-type-specific genes. The chromatin accessibility within a gene body was measured as the enrichment of reads in the TSS above background level. The TSS was used to infer gene expression via computation of a ‘Gene Score’ using the default method in ArchR (v1.0.1). The gene score profiles for all cells were subsequently used to generate a gene score matrix. The gene score matrix was also integrated with scRNA-seq expression data (described below). Finally, a cell-type annotation for each cluster was assigned using gene scores for cell-type marker genes and later validated or further refined through scRNA-seq label transfer (Fig. 1b–d).

Cell-type-specific marker genes in our snATAC-seq data (genes with significantly higher chromatin accessibility in a cluster than in other clusters) were identified using Wilcoxon rank-sum test and the genes with (Benjamini–Hochberg) adjusted P ≤ 0.01 and fold change ≥ 2 were selected. The z-normalized gene scores for the cell-type-specific genes were plotted as a heatmap (Fig. 2a).

scRNA-seq processing and snATAC-seq integrative analysis. We integrated the coronary snATAC-seq dataset (28,316 nuclei) with previously published human coronary artery scRNA-seq data from Wirka et al.(101). The preprocessed scRNA-seq data were downloaded from the Gene Expression Omnibus (accession code GSE131780) and processed using Seurat(102) as described in the study. Genes expressed in less than five cells were filtered out. Cells with ≤500 or ≥3,500 genes were also trimmed from the dataset as they likely represent defective cells or doublet/multiplet events. Moreover, cells containing ≥7.5% of reads mapping to the mitochondrial genome were discarded as low-quality/dying cells often exhibit high levels of mitochondrial contamination. Upon discarding poor-quality cells, 11,756 high-quality cells and 19,965 genes remained for further analysis. Read counts were normalized using Seurat's global-scaling method that normalizes gene expression measurements for each cell by the total expression, multiplies them by a 10,000-scaling factor and log-transforms them. Upon finding the 2,000 most variable genes in the data, dimensionality reduction was performed using principal component analysis. The top ten principal components were further used for uniform manifold approximation and projection (UMAP) visualization and cell clustering (using a shared nearest neighbor modularity optimization-based algorithm in the Seurat package). The cluster-specific genes (marker genes) for each cluster were identified with the Seurat default method. The cell types of clusters were assigned according to the comparison between the cluster-specific genes and the cell-type-specific gene lists provided in the previous study (Supplementary Table 6).

The cell-type annotated scRNA-seq expression matrix was then integrated with the snATAC-seq gene score matrix (described in the above section) using the ‘addGenes’ function from ArchR. This step identifies corresponding cells across datasets or ‘anchors’ using Seurat’s mutual nearest neighbors algorithm. To scale this step across thousands of cells, the total number of cells was divided into smaller groups and alignments were performed in parallel. Cell-type labels within the Seurat scRNA-seq object metadata were transferred to the corresponding mutual nearest neighbors in the snATAC-seq data along with their gene expression signatures. The output of the integration step resulted in snATAC-seq cells having both a chromatin accessibility and a gene expression profile. After integration, snATAC-seq cells were re-annotated in UMAP space using the scRNA-seq-transferred labels and these defined groups were used for downstream analyses as an alternative annotation in addition to the marker gene-based annotation. The scRNA-seq-transferred labels were also used in the fibrocyte versus SMC differential analysis (Fig. 3h–i).

Cell-type-specific peak and TF motif enrichment. Genome-wide chromatin accessible regions for each ‘pseudo bulk’ sample (reads from the same cluster were combined as a new sample) were detected using the ‘addReproduciblePeaks’ function in ArchR (with parameters cutOffSize = 100, cutOffFDR = 0.01, extendSummits = 200). Thereafter, 323,767 chromatin accessible regions (peaks) were detected. The cell-type-specific peaks (marker peaks) for each cluster/cell type were identified using a similar strategy as identification of cell-type-specific genes (with parameters FDR ≤ 0.01 and Log2FC ≥ 1). This resulted in a total of 173,357 cell-type-specific peaks for different cell types. The enriched motifs for each cell type were predicted using the ‘addMotifAnnotations’ function in the ArchR package based on the HOMER(103) motif database (v4.11.0). The chromatin accessibility variability and deviation of TFs were estimated by chromVAR (1.12.0)(104) with genome-wide motif sites provided as potential binding sites.

Peak pathway annotation. To perform functional annotation of cell-type marker peaks (Extended Data Fig. 4), we used GREAT(5) (v. 4.0.4) with default parameters. The top five functional annotation terms (from the Gene Ontology Biological Processes database) for each cell type were displayed as a dot plot. The colors and
sizes of the dots represent $-\log(\text{FDR})$ (from the hypergeometric gene-based test) and the percentage of associated genes, respectively.

**Trajectory analysis.** The trajectory analysis was performed using the "addTrajectory" function in ArchR and specifying the cluster order (cluster 6 – cluster 5 – cluster 7, Fig. 3f). We further visualized trajectory-dependent changes (of summarized) ATAC-seq motif signals using the 'plotTrajectoryHeatmap' function in ArchR (Fig. 3g).

**Peak to gene linkage and co-accessibility analysis.** We leveraged the integrated scRNA-seq and snATAC-seq data to explore correlations between co-accessible regions and gene expression. These candidate gene regulatory interactions were predicted using the 'getPeak2GeneLinks' function with default parameters in ArchR. The peakgene index were collected using the Sashimi package40 with red color highlighting SNP-associated loci and gray color for other loci (Fig. 6a,b). For the loops around VEGFA and CDH13 promoter (Extended Data Fig. 5), the loops were predicted using the ArchR 'addCoAccessibility' function with the additional parameter 'maxDist = 160'.

**Differential analysis between SMCs and fibromyocytes.** Details for differential analyses between SMCs and fibroblasts are provided in the Supplementary methods.

**LDSC.** We used the LDSC package (https://github.com/bulik/ldsc) to perform LDSC using our snATAC-seq peaks46. We first downloaded GWAS summary statistics for: CAD13; carotid intima–media thickness103; carotid artery plaques103; coronary cell types: SMCs, endothelial cells, fibroblasts and macrophages. We used the LDSC package (https://github.com/bulik/ldsc) to perform LDSC. For each coronary artery cell type, we lifted over bed file peak coordinates from hg38 to hg19. We then used these hg19 bed files to make annotation files for each cell type. We performed LDSC according to the cell-type-specific analysis tutorial (https://github.com/bulik/ldsc/wiki/Cell-type-specific-analyses).

**Calculation of caQTLs.** For each snATAC-seq peak we tested association for all variants within a ±10-kb window. We ran RASQUAL using the -f flag to output only the top associated SNP per peak. For all RASQUAL runs we adjusted for age, sex and the first three principal components of ancestry in the covariate file (–x flag). To obtain a null distribution of Q values we performed five separate permutation runs for each cell type using the --random-permutation flag to break the relationship between genotype and peak accessibility.

To adjust for multiple testing, we performed two FDR corrections. First, for each peak we obtained a Q value corresponding to the SNP-level FDR (Benjamini–Hochberg method) for that peak. Next, the permutation test in RASQUAL adjusts for genome-wide multiple testing. For each peak we averaged the Q values across the five RASQUAL permutation runs. This produced two vectors: one with real RASQUAL Q values and one from the permuted Q values for each peak. By comparing the real and permuted vectors of Q values we were then able to calculate the Q value corresponding to 10% FDR, 5% FDR or 1% FDR. For plotting RASQUAL caQTL results as boxplots, we took raw count files for each cell type, adjusted for library size and performed variance stabilization transformation in DESeq2 (ref. 11).

**Overlap of caQTLs with GTEx eQTLs.** We used the QTLr R package52 to query the significant SMC caQTL rsIDs for eQTL signals in GTEx v.8. We only retained GTEx eQTL signals at 5% FDR and subsequently filtered for relevant arterial tissues (coronary artery, aorta, tubial artery).

Publicly available gene expression data. Gene expression levels, eQTL data and eQTL boxplots were obtained from the GTEx v.8 portal website (https://www.gtexportal.org/home/). Differential gene expression data from the publicly available GEO in cardiovascular-relevant systems were obtained via the HeartBioPortal (https://heartbioportal.com/).

**Functional variant sequence-based predictive modeling.** We first downloaded CAD GWAS summary statistics from the study by van der Harst et al.13 and retained variants passing the genome-wide threshold (P < 5 × 10−8). This resulted in 10,117 variants that were tested. The variant scoring analysis (Fig. 5b) was conducted using the lsgkm package (https://github.com/kundajelab/lsgkm)114 and the GkmExplain package (https://github.com/kundajelab/gkmexplain)14. We used cell type (for example, SMC), the reads from all the individual cells assigned to the cell type were first collected as a pseudo bulk sample. The pseudo bulk snATAC-seq peaks were detected with MACS2 (ref. 11) (paired-end mode, with additional parameter –q 0.01). In the model-building step, peaks were split for cross-validation. For each fold, the top 60,000 peaks with highest --Q (Q value) were selected as the training set. The 5,000 peaks from the previous training set were used as a positive set, while sequences from a 1,000-kb region outside of peaks with matching GC content were used as a negative set. The importance score of all the positions around the target SNP (up to ±100 bp) were plotted as sequence logo (Fig. 5b).

**STARNET gene regulatory network analysis.** Based on STARNET multi-tissue gene expression data (bulk RNA-seq data), tissue-specific and cross-tissue co-expression modules were inferred using WGCNA42 as previously described42. Enrichment for clinical trait associations was computed by first extracted cell-type assigned reads from our ArchR analysis in bam format for each snATAC-seq library. For each individual cell type, we excluded individuals with less than 20 cells from caQTL analysis. We ended up with SMC bam files for 40 patients, endothelial cell bam files for 37 patients, fibroblast bam files for 26 patients (due to some samples lacking avanticia) and macrophage bam files for 39 patients. To obtain region sets we took the peak set across all cell types and converted these peaks from bed to saf format. We used these peak coordinates in saf format and cell-type bam files as input for featureCounts11 (v.1.6.4) with the -p flag for paired-end mode. This subsequently generated raw count matrices for SMC, endothelial, fibroblast and macrophage cells. For each cell type we only retained peaks with an average of 5 read counts across individuals.

We used RASQUAL (v.1.0.0) to calculate caQTLs, which leverages differences in read counts between individuals as well as allele differences within an individual at heterozygous sites. To simplify preparation of RASQUAL input files we used rasqalTools (https://github.com/kaularossou/rasqal/tree/master/rasqalTools) to prepare compatible snATAC-seq read count, metadata and sample-specific offset files. To calculate sample offsets, we adjusted for library size as well as the GC content of each peak.

For each individual we obtained genotypes from the low-pass whole-genome sequencing that was performed by Gencove. We used VCFtools112 to filter for variants with at least 5% minor allele frequency and select patients with corresponding snATAC-seq libraries. We used RASQUAL to create allele-specific vcf files (createASVCEsh) for each cell type, which contain genotype information plus counts for reference and alternative alleles. We bypassed the qFilterBam part of the RASQUAL script to incorporate incompatible data formats generated with our snATAC-seq bam files. However, our bam files extracted from each cell type were previously filtered using ArchR and contain high-quality cells and reads.

**Functional variant sequence-based predictive modeling.** We first downloaded CAD GWAS summary statistics from the study by van der Harst et al.13 and retained variants passing the genome-wide threshold (P < 5 × 10−8). This resulted in 10,117 variants that were tested. The variant scoring analysis (Fig. 5b) was conducted using the lsgkm package (https://github.com/kundajelab/lsgkm)114 and the GkmExplain package (https://github.com/kundajelab/gkmexplain)14. We used cell type (for example, SMC), the reads from all the individual cells assigned to the cell type were first collected as a pseudo bulk sample. The pseudo bulk snATAC-seq peaks were detected with MACS2 (ref. 11) (paired-end mode, with additional parameter –q 0.01). In the model-building step, peaks were split for cross-validation. For each fold, the top 60,000 peaks with highest --Q (Q value) were selected as the training set. The 5,000 peaks from the previous training set were used as a positive set, while sequences from a 1,000-kb region outside of peaks with matching GC content were used as a negative set. The importance score of all the positions around the target SNP (up to ±100 bp) were plotted as sequence logo (Fig. 5b).

**STARNET gene regulatory network analysis.** Based on STARNET multi-tissue gene expression data (bulk RNA-seq data), tissue-specific and cross-tissue co-expression modules were inferred using WGCNA42 as previously described42. Enrichment for clinical trait associations was computed by...
aggregating Pearson’s correlation $R$ values by co-expression module using Fisher’s method. Enrichment for differentially expressed genes was calculated using the hypergeometric test, with differentially expressed genes called by DESeq2 ($\leq 0.01$ change, FDR $\leq 0.01$) used for enrichment. Treatment for age and sex. The gene regulatory network was inferred among PRDM16 and TBX2 co-expressed genes using GENIE3 (ref. 10) with potential regulators restricted to eQTL genes or known TFs. To identify hub genes in the network, weighted key driver analysis was carried out using the Mergenomics R package 112.

Immunofluorescence of human coronary artery tissues. Human coronary artery tissues were obtained as described above. Briefly, coronary artery segments were isolated from healthy and subclinical atherosclerotic left main and right coronary artery branches. Tissues were embedded in OCT blocks, snap-frozen in liquid nitrogen and stored at $-80^\circ$C. Tissue blocks were cryosectioned at $-20^\circ$C and 6-μm thickness and processed for immunostaining. Sections were rehydrated in PBS at room temperature and fixed in 10% neutral buffered formaldehyde for 10 min at room temperature, followed by PBS washes, protein blocking in casein buffer for 1 h at room temperature and incubation overnight at $4^\circ$C with anti-LMOD1 rabbit polyclonal antibody (Proteintech, 15117-1-AP; 1:100), a-TRA-1-60 mouse monoclonal antibody (clones D1, D50, M08S1; 1:100) or anti-PRDM16 rabbit polyclonal antibody (Abcam, ab106410; 1:2,000) or no antibody negative control (PBS), with optional dilutions determined by titrations with control tissues. Sections were washed in PBS and incubated with donkey anti-rabbit Alexa Fluor 555 conjugated secondary antibody (Thermo Fisher, A31572) or donkey anti-mouse Alexa 488 conjugated secondary antibody (Thermo Fisher, A21202; 1:50) for 30 min at room temperature, washed in PBS and stained with DAPI (1:500), and coveredslipped using aqueous mounting media. Whole-slide images were captured at $\times 25$ magnification using a Zeiss LSM 880 Indimo, Axiolimmaer confocal microscope with a Plan-Apochromat $\times 2.5$ objective in Line Sequential unidirectional mode. Signals corresponding to the DAPI (channel 1) and the protein of interest (channel 2) were obtained using lasers (relative wavelengths of excitation of 405 and 561 nm); photomultiplier tubes and filters were used to collect the fluorescence emitted, respectively, at 410–480 nm and 561–597 nm. Images in both channels were merged with the Zeiss ZEN 3.3 Lite software (v.3.3.89). Brightness, gamma and contrast were uniformly adjusted. Corresponding regions of interest of the sections immunostained with both antibodies were numerically magnified. Whole-slide images were reconstructed from tiles acquired in brightfield using a high-resolution HV-F203SCi. Hitachi camera mounted on an Axiolimina microscope using a Plan-Apochromat $\times 100$/0.3 objective.

For histology analysis, adjacent sections were stained with hematoxylin and eosin and Movat pentachrome as previously described 113. Images were captured using a Zeiss 183 Axiolimina Z1 at $\times 20$ magnification. The resulting czi files were visualized for staining using Zeiss ZEN 3.3 Lite software (v.3.3.89).

Histological analysis and quantitation of atherosclerosis. Please refer to the Supplementary methods.

Sample size. No sample size calculations were performed a priori. Sample size ($n = 41$) was determined based on the availability of tissue materials. However, we also confirmed with a power analysis calculator that this sample size has 95% power to detect low-frequency cell types (5–10%) based on the average number of cells captured per sample. Additional descriptions of post hoc power calculations are provided in Supplementary Fig. 7 and the Supplementary methods.

Genome annotations and browser tracks. All sequence alignments and genome annotations and browser tracks are provided in Supplementary Fig. 7 and the Supplementary methods.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
All raw and processed single-nucleus chromatin accessibility sequencing datasets are made available on the Gene Expression Omnibus (GEO) database (accession codes GSE175621 and GSE188422). The processed and analyzed snATAC-seq data will also be made available on the PlaqView single-cell data portal (https://www.plaqview.com). All coQTL data are available in the Supplementary Data. Low-pass whole-genome sequencing-based genotyping data are available on dbGaP (accession code phs002855.v1.p1). The human coronary artery scRNA-seq dataset we used in this study from Wikra et al. is available through GEO (accession code GSE131778). The mouse atherosclerosis scRNA-seq dataset from Pan et al. is available through GEO (accession code GSE155513). The reprocessed and analyzed human and mouse datasets are also available on PlaqView. Gene expression levels, expression quantitative trait locus (eQTL) data and eQTL boxplots were obtained from the Genotype-Tissue Expression (GTEx) v8 portal website (https://www.gtexportal.org), GEO and HeartBioPortal (www.heartbioportal.com). Gene regulatory network analysis data from the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) are available at https://starnet.mssm.edu.

Code availability
Our results make use of published software tools with detailed parameters included in the Methods. All custom scripts used to generate these results are available on GitHub (https://github.com/MillerLab-CPHG/Coronary_snATAC) and https://github.com/MillerLab-CPHG/Coronary_histology).

References
99. Stuart, T. et al. Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21 (2019).
100. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019).
101. McLean, C. Y. et al. GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. 28, 495–501 (2010).
102. Phanstiel, D. H., Boyle, A. P., Arya, C. L. & Snyder, M. P. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. Bioinformatics 30, 2808–2810 (2014).
103. Franceschini, N. et al. GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. Nat. Commun. 9, 5141 (2018).
104. Giri, A. et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat. Genet. 51, 51–62 (2019).
105. Jansen, I. E. et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat. Genet. 51, 404–413 (2019).
106. Sudlow, C. et al. 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 (2015).
107. Nep, S. et al. BDOPS: high-performance genomic feature operations. Bioinformatics 28, 1919–1920 (2012).
108. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
109. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am. J. Hum. Genet. 81, 1084–1097 (2007).
110. Browning, B. L., Zhou, Y. & Browning, S. R. A one-penny imputed genome from next-generation reference panels. Am. J. Hum. Genet. 103, 338–348 (2018).
111. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930 (2014).
112. Danecek, P. et al. The variant call format and VCFtools. Bioinformatics 27, 2156–2158 (2011).
113. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550 (2014).
114. Lee, D. LS-GKM: a new gkm-SVM for large-scale datasets. Bioinformatics 32, 2196–2198 (2016).
115. Zhang, Y. et al. Model-based Analysis of ChiP-Seq (MACS). Genome Biol. 9, R137 (2008).
116. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 9, 559 (2008).
117. Talukdar, H. A. et al. Cross-tissue regulatory gene networks in coronary artery disease. Cell Syst. 2, 196–208 (2016).
118. Huynh-Thu, V. A., Irurham, A., Wehenkel, L. & Geurts, P. Inferring regulatory networks from expression data using tree-based methods. PLoS ONE 5, e12776 (2010).
119. Shu, L. et al. Mergeomics: multidimensional data integration to identify pathogenic perturbations to biological systems. BMC Genom. 17, 874 (2016).
120. Otsuka, E. et al. Natural progression of atherosclerosis from pathologic intimal thickening to late fibroatheroma in human coronary arteries: a pathology study. Atherosclerosis 241, 772–782 (2015).

Acknowledgements
This work was supported by grants from: the National Institutes of Health (grant nos. R01HL141425 and R01HL125912 to L.M.B.; grant no. R35GM133712 to C.Z.; grant no. R01HL141425 to A.V.F; grant no. R01HL125863 to J.L.M.B.; grant nos. R01HL130423, R01HL135939 and R01HL148671-01A1 to J.C.K.; grant nos. R35HL144475, R01HL135224, R01HL134817 and R01HL139478 to T.Q.; and grant no. R01HL13370-01 N.I., the American Heart Association (grant no. 20POST35120545 to A.W.T.; grant no. A14SFRN20840000 to J.L.M.B.; grant no. 19EIA34770065 to N.J.L.), the Swedish
Research Council and Heart Lung Foundation (grant nos. 2018-02529 and 20170265 to J.L.M.B.) and the Fondation Leducq (grant no. ‘PlaqOmics’ 18CVD02 to N.J.L., J.L.M.B., A.V.F. and C.L.M.). We thank P. Chu, P. Chang and M. Wong at Stanford University for surgical assistance and research donor heart procurement. We thank T. Koyano at Stanford University for assistance in extracting clinical information. We thank all of the transplant recipients and heart donors, family members, study coordinators and the transplant tissue procurement team at Stanford. We thank B. Liu and N. Kumasaka for helpful discussions on QTL scripts. Finally, we thank P. Pramoonjago and S. VanHoose at the University of Virginia for histological support and all staff of the core facilities for tissue processing, library construction and sequencing assistance.

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

Competing interests
J.L.M.B. is a shareholder in Clinical Gene Network AB who have a vested interest in STARNET. A.V.F. at CVPath also acknowledges receiving financial support from the following entities: 4C Medical, 4Tech, Abbott Vascular, Ablative Solutions, Absorption Systems, Advanced NanoTherapies, Aerwave Medical, Alivas, Amgen, Asahi Medical, Aurios Medical, AvanteC Vascular, BD, Biosensors, Biotronik, Biotyx Medical, Bolt Medical, Boston Scientific, Canon, Cardiac Implants, Cardiawave, CardioMech, Cardionomic, Celonova, Cerus, EndoVascular, Chansu Vascular Technologies, Children’s National, Concept Medical, Cook Medical, Cooper Health, Cormaxe, CRL, Crevalve, CSI, Dexcom, Edwards Lifesciences, Elucida Bioimaging, eLum Technologies, Emboline, Endotronics, Envision, Filterex, Imperative Care, Innovarole, Innovative, Cardiovascular Solutions, Intact Vascular, Interface Biologics, Intershunt Technologies, Invatin, Lahav, LimFlow, L&J Bio, Lutonix, Lyra Therapeutics, Mayo Clinic, Maywell, MDS, MedAlliance, Medanex, Medtronic, Mercator, Microporot, Microvention, Neovasc, Nephronyx, Nova Vascular, Nyra Medical, Occultech, Olympus, Ohio Health, OrbusNeich, Ossoio, Phenox, Pfi-Cardia, Polares Medical, Polysvascular, Profusa, ProKidney, LLC, Proteombs, Pulse Biosciences, Qcel Therapeutics, Recombinetics, Recor Medical, Regencor, Renata Medical, Restore Medical, Ripple Therapeutics, Rush University, Sanofi, Shockwave, SMT, SoundPipe, Spartan Micro, Spectrawave, Surmodics, Terumo Corporation, The Jacobs Institute, Transmural Systems, Transverse Medical, TruLeaf, UCSF, UPMC, Vascudyne, Vesper, Yestech Medical, Whiteswell, WL Gore, Xeltis. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The other authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41588-022-01069-0.
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-022-01069-0.
Correspondence and requests for materials should be addressed to Chongzhi Zang or Clint L. Miller.
Peer review information Nature Genetics thanks Vivek Swarup and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.
Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Histological characterization of human coronary artery sections. (a) Representative histology staining of adjacent frozen human coronary artery sections at different disease categories used for snATAC-seq profiling. Category 1 reflects normal to Stary atherosclerosis stage I/II lesions with adaptive intimal thickening and early lipid (Oil Red O (ORO)) and collagen (Sirius Red) accumulation in the subintimal layer. Category 2 reflects Stary stage III/IV early/intermediate atheroma lesions with increased lipid and collagen accumulation and proliferation (Hematoxylin & Eosin (H&E)). Category 3 reflects Stary stage V/VI advanced fibroatheroma or complex lesions with more severe lipid and collagen deposition as well as lipid core and thin media layer. (b) Whole slide quantitative results of ORO area (mm2) normalized to overall tissue area and (c) Sirius Red based quantitation of intima-media thickness (IMT) with maximum intima and average media width captured from >6 automatically defined measurements (Methods). (a-c) Similar results were observed from n = 3, n = 5, and n = 10 independent donor samples per lesion stage, respectively. ANOVA p-values shown for comparisons across lesion stages. Boxplots represent the median and interquartile range (IQR). Scale bar = 1 mm.
Extended Data Fig. 2 | See next page for caption.
Extended Data Fig. 2 | Coronary artery cell type marker genes from snATAC-seq gene scores. (a) Representative UMAP plots of snATAC-seq imputed gene activity scores and integrated RNA scores for SMC and fibromyocyte marker genes. (b) UMAP plots of imputed gene scores for additional cell type marker genes and CAD GWAS genes. (c) Top candidate genes at CAD GWAS loci with cell type enriched chromatin accessibility. Negative Log10 FDR enrichment values shown for CAD GWAS marker genes.
Extended Data Fig. 3 | Integration of human coronary artery snATAC-seq data with human coronary artery scRNA-seq (from Wirka et al.). (a) UMAP showing projection of scRNA-seq cluster labels onto cells in the snATAC-seq dataset. Colors represent the assigned cellular identities from scRNA-seq label transfer. Detailed parameters of the snATAC-seq/scRNA-seq integration are provided in the Methods section. (b) Heatmap of marker gene scores after ArchR scRNA-seq/snATAC-seq integration highlights 4,649 marker features. (c) Correlation of cell type specific scRNA-seq and snATAC-seq promoter accessibility (pseudo bulk reads from ATAC signal centered on TSS (+/− 3 kb) for each gene). Log2 transformed data is represented as scatter plots and Pearson correlation coefficients are shown for each cell type. White lines represent missing gene counts from scRNA-seq dataset, which is most apparent in the low abundant Mast cells.
Extended Data Fig. 4 | Coronary artery snATAC-seq peak cell type and functional annotation. (a) Pie chart showing genomic annotations of the consensus set of coronary peaks across all cell types (n = 323,767). Peaks were annotated using the ChIPseeker R/Bioconductor package (Yu et al. Bioinformatics 2015). (b) Pie chart of cell type annotation for peaks in the consensus peak set (n = 323,767) according to ArchR (Granja et al.41). Peaks were annotated with a cell type according to the group from which each peak originated according to ArchR’s iterative overlap procedure. (c) Functional enrichment analysis of cell type marker peaks using GReAT.
Extended Data Fig. 5 | snATAC-seq co-accessibility and integration with scRNA-seq link putative regulatory elements to target promoters. (a) Genome browser tracks highlighting CAD-associated SNPs located within peaks linked to the VEGFA promoter peak through co-accessibility. (b) Genome browser tracks highlighting the intronic CAD SNP rs7500448 located in a smooth muscle cell peak in the CDH13 gene linked to the CDH13 promoter peak through co-accessibility. (c) Heatmap summary of ArchR Peak2Gene links ($n = 148,617$) at 10 kb resolution where chromatin accessibility is highly correlated with target gene expression. Shown on the left are Z-scores for snATAC-seq peak accessibility and on the right are Z-scores for RNA expression.
Extended Data Fig. 6 | Additional CAD-associated variants that are coronary artery chromatin accessibility QTLs (caQTLs). (a-b) Smooth muscle cell caQTL boxplots for variants at the BMP1 (rs73551705) and SMAD3 (rs17293632) CAD loci (n = 40 unique individuals). (c) Macrophage caQTL boxplot for the rs72844419 variant at the GGcox CAD locus (n = 39). Chromatin accessibility reads were normalized using variance stabilizing transformation (vst) in DEseq2. Boxplots represent the median and interquartile range (IQR), while the whisker represent up to 1.5 X IQR. (d-e) Comparison of effect size directions between smooth muscle cell caQTLs (5% FDR) and bulk coronary artery caQTLs (5% FDR), as visualized in scatter plot (d) and donut plot (e). For this analysis, 503 caQTL peaks are shared between both datasets (peaks with a corresponding significant caQTL variant). The rsID reported in the SMC caQTL results (n = 40 individuals) was compared with the rsID reported in the bulk caQTL results (n = 35 individuals). Two variants were considered to be in linkage disequilibrium (LD) if the r² value between them was between 0.2 and 1 (in EUR population). If variants had an r² value < 0.2 (in EUR population), the variants were considered to be in low LD (blue). For the caQTL effect size direction, we considered the RASQUAL Pi statistic. The RASQUAL Pi statistic can range from 0–1, where Pi < 0.5 reflects lower peak accessibility for the alternative allele and Pi > 0.5 reflects higher accessibility for the alternative allele. The effect sizes for linked variants go in the same direction (green) if the Pi values in SMCs and bulk coronary artery are both < 0.5 or both > 0.5. Linear regression line and Pearson correlation coefficient shown in (d).
Extended Data Fig. 7 | Examples of candidate CAD functional variants within macrophage accessible chromatin. (a) CAD GWAS locus MAP1S/FCHO1 on chromosome 19 depicting multiple genome-wide significant variants (above dashed line). Highlighted variant rs10418535 is located within a macrophage/immune cell ATAC peak as shown in the genome browser tracks. gkm-SVM importance scores show the predicted effects of the T allele to form a functional binding site, while the C allele (non-effect) is predicted to disrupt TF binding. (b) Genome browser view showing 95% credible CAD SNPs (blue), highlighting rs7295737 located within a strong macrophage marker peak in the first intron of SCARB1 on chr12. (c) Genome browser view highlighting top credible CAD SNP rs17680741 residing in macrophage marker peak in the second intron of TSPAN14 on chr10.
Extended Data Fig. 8 | Co-accessibility and gene regulatory analyses prioritize transcriptional regulators TBX2 and PRDM16. (a) Genome browser track highlighting the association between CAD associated SNPs and SMC marker genes through co-accessibility (peak2gene) detected by snATAC-seq data (Methods). The red loops represent the association between TBX2 promoter and CAD associated SNPs. (b) Network visualization of TBX2 key driver target genes in STARNeT atherosclerotic aortic root (AOR) tissue. (c) Clinical trait enrichment for PRDM16 module 28 in STARNeT liver tissue. (d) Network visualization of PRDM16 key driver target genes in STARNeT mammary artery (MAM) and liver tissues.
Extended Data Fig. 9 | Immunostaining of PRDM16 protein in coronary atherosclerosis sections. (a) Representative negative control (no primary antibody) immunofluorescence (IF) staining in human coronary artery - left anterior descending (LAD). Positive staining of rabbit anti-PRDM16 in vessels in control kidney tissues. Similar results were observed from n = 4 independent donor samples per tissue. Scale bar = 100 um. (b) Representative IF staining of PRDM16 and LMOD1 in atherosclerotic human coronary artery (LAD) segments from normal-Stage II, Stage III-IV, and Stage V-VI lesions based on Stary classification stages. Red = PRDM16 or LMOD1, Green = alpha smooth muscle actin (α-SMA) and blue = DAPI (nuclei). Scale bar = 1mm (whole slide) or 100 um (highlighted regions of interest). (c) Representative hematoxylin & eosin (H&E) and MOVAT histology staining of distinct human coronary artery segments with similar lesion stages as (b). Scale bar = 1mm. (b-c) Similar results were observed from n = 4 (Normal-stage II), n = 6 (Stage III-IV), and n = 6 (Stage V-VI) independent donor samples per group.
Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- **n/a** Confirmed
- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
  - *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  - *Give P values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

*Our web collection on statistics for biologists contains articles on many of the points above.*

Software and code

Policy information about [availability of computer code](#)

**Data collection**

We used the Qtizer [v1.0.0] R package to collect cis-eQTL data from various eQTL databases to validate our caQTLs.

**Data analysis**

The names and versions of software packages used for data analysis are provided below as well as in the Methods section. More complete documentation and custom scripts used are available on GitHub [https://github.com/MillerLab-CPHG/Coronary_snATAC and https://github.com/MillerLab-CPHG/coronary_histology].

R packages:

- ArchR [v1.0.1]
- Seurat [v4.0.0]
- Harmony [v1.0.0]
- chromVAR [v1.12.0]
- DESeq2 [v1.26.1]
- rasqulTools [v1.0.0]
- ChipSeekeR [v1.26.0]
- GenomicRanges [v1.42.0]
- ensemblib [v2.14.0]
- EnsDb.Hsapiens.v86 [v2.99.0]
- TxDB.Hsapiens_UCSC_hg38.knownGene [v3.10.0]
- LDlinkR [v1.1.2]
- GENIE3 [v1.16.0]
- Mergeomics [v1.22.0]

Command line packages:
**Data**

Policy information about availability of data. All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All snATAC and bulk ATAC-seq data generated in this work are available online through the Gene Expression Omnibus (GEO) (accessions: GSE175621 and GSE188422). The processed and analyzed snATAC datasets are also available on PlaqView single-cell data portal [https://www.plaqview.com]. Chromatin accessibility QTL results are available in the Supplementary Data files. Low-pass whole genome sequencing based genotyping data are available on dbGaP (accession phs002855.v1.p1).

The human coronary artery scRNA-seq dataset we used in this study from Wirka et al. (Nature Medicine 2019) is available through GEO (accession: GSE131778). The mouse atherosclerosis scRNA-seq dataset we queried from Pan et al. (Circulation 2020) is available through GEO (accession: GSE155513). The reprocessed and analyzed human and mouse datasets are also available on PlaqView (https://www.plaqview.com).

Gene expression levels, expression quantitative loci (eQTL) data, and eQTL boxplots were obtained from the Genotype-Tissue Expression (GTEx) v8 portal website [https://www.gtexportal.org], GEO and HeartBioPortal (www.heartbiportal.com).

Gene regulatory network analysis data from the Stockholm-Tartu Atherosclerosis Reverse Network Engineering Task (STARNET) are available at http://starnet.mssm.edu.

**Field-specific reporting**

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑️ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-list.pdf

**Life sciences study design**

All studies must disclose on these points even when the disclosure is negative.

- **Sample size**: No sample size calculations were performed a priori. Sample size (n=41) was determined based on the availability of tissue materials. However, we also confirmed with a power analysis calculator that this sample size has 95% power to detect low frequency cell types (5-10%) based on the average number of cells captured per sample. We also performed post-hoc power analyses which determined that the number of nuclei obtained per cell-type was sufficient to detect differential accessibility between major cell types. Finally, assuming a standard linear model, this sample size has 90% power to detect eQTLs at ~5% minor allele frequency.

- **Data exclusions**: We used pre-established quality control thresholds to exclude low quality cells from the analysis. We excluded cells with fewer than 10,000
Data exclusions

Unique fragments or transcription start site (TSS) enrichment scores below 7 as these metrics have been shown to provide more biologically informative results that are not driven by technical artifacts. For cell-type caQTL analysis we excluded an individual if they had less than 20 cells belonging to that specific cell-type.

Replication

We created pseudo bulk accessible chromatin profiles from snATAC-seq data and correlated these with bulk chromatin profiles for the same patient sample. While there are no other external coronary artery single-nucleus ATAC-seq datasets, we integrated our dataset with a published coronary artery scRNA-seq dataset to validate the cell type clusters and genome-wide snATAC based gene scores. We replicated our caQTLs using coronary artery eQTLs from GTex, STARnet, and other relevant cells/tissues. Finally we replicated our SVM based functional variant predictions using functional probability scores (based on DeepSEA) as implemented in RegulomeDb 2.0. Overall our attempts at replication were successful.

Randomization

Samples from healthy and diseased coronary artery segments or hearts were randomly processed for single-cell library preparation and data analysis. Relevant patient and library covariates were visualized in UMAP embeddings and also adjusted for in the caQTL analysis.

Blinding

Prior to snATAC-seq experiments, coronary artery samples were randomly allocated based on tissue availability, encompassing various degrees of atherosclerosis. All samples were processed individually for nuclear isolation and snATAC-seq. Relevant patient and library covariates were visualized in UMAP embeddings and also adjusted for in the caQTL analysis.

Reporting for specific materials, systems, and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

| Materials & experimental systems | Methods |
|---------------------------------|---------|
| n/a Involved in the study       | n/a Involved in the study |
| ☑ Antibodies                   | ☑ ChIP-seq |
| ☑ Eukaryotic cell lines        | ☑ Flow cytometry |
| ☑ Palaeontology and archaeology| ☑ MRI-based neuroimaging |
| ☑ Animals and other organisms  |         |
| ☑ Human research participants  |         |
| ☑ Clinical data                |         |
| ☑ Dual use research of concern |         |

Antibodies

Primary antibodies used in the immunofluorescence analysis for PRDM16: 1) anti-PRDM16 rabbit polyclonal antibody (Abcam, catalog # ab106410, Lot # GR3314435-2), 2) anti-LMO1 rabbit polyclonal antibody (Proteintech, catalog # 15117-1-AP, Lot # 00006190), 3) anti-alpha-SMA mouse monoclonal antibody (clone 1A2; Agilent/Dako, catalog # M0851, Lot # 20031819). Secondary antibodies used include donkey anti-rabbit Alexa Fluor 555 conjugated secondary antibody (Thermo Fisher, catalog # A31577, Lot # 2286312) and donkey anti-mouse Alexa 488 conjugated secondary antibody (Thermo Fisher, catalog # A21202, Lot # 2147618).

Validation

Each of these antibodies were pre-validated by the manufacturer and previous labs to recognize native proteins in human tissues for immunofluorescence (IF) or immunohistochemistry (IHC) applications and we provide relevant citations. For LMO1, the manufacturer (Proteintech) identified positive IF result in human A375 cells and positive IHC result in human colon tissue. The recommended dilution for IF is 1:10-1:100. For PRDM16, the manufacturer (Abcam) identified positive tested immunocytochemistry (ICC) result in human adipocytes and human brain cells and positive IHC result in human brain tissue. The recommended final concentration for ICC is 20μg/ml and for IHC is 2 μg/ml. For anti-SMA, the manufacturer (Agilent/Dako) demonstrate application for both frozen and formalin fixed immunohistochemistry and provide positive staining results for various tissues including human colon, liver, and appendix.

We also performed our own detailed validation and antibody titration experiments for immunofluorescence using positive and negative control human kidney and brain tissues and compared these with no primary control and the same secondary concentrations to ensure specificity. These data are included in Extended Data Figure 5 and described in the methods.

Human research participants

Policy information about studies involving human research participants

Population characteristics

Coronary arteries were obtained from both heart transplant recipients and from organ donor hearts (rejected for transplantation). Individuals ranged in age from 21 to 71 (mean age 54 +/- 11 SD), with 56% representing males. This study population includes several ancestries, with European ancestry being the most common (54% of individuals).

Recruitment

The recruitment site (Stanford University Hospital and Clinics) for the original tissue collection obtained written informed consent under the individual IRB-approved protocols at the time of enrollment. There was no targeted number for the recruitment since all subjects undergoing orthotopic heart transplantation (or organ donors) were eligible to participate. Participants were not compensated for this study. Samples analyzed at the University of Virginia were randomly selected for inclusion in the enclosed study. Given the low nuclei yield from more advanced disease samples, there is potential self
selection bias towards the earlier disease samples. Thus our results may not fully capture profiles for cell types present only in advanced plaques, as noted in the manuscript.

Ethics oversight

All research described herein complies with ethical guidelines for human subjects research under approved Institutional Review Board (IRB) protocols at Stanford University (#4237 and #11925) and the University of Virginia (#20008), for the procurement and use of human tissues and information, respectively.

Note that full information on the approval of the study protocol must also be provided in the manuscript.