Article

Single-cell variant-to-enhancer-to-gene map for coronary artery disease

https://doi.org/10.1038/s41586-026-11275-z

Received: 13 November 2024

Accepted: 7 October 2026

Published online: 25 November 2026

Open access

Check for updates

Junedh M. Amrute[1,2,3,42]✉, Paul C. Lee[1,42], Ittai Eres[2,42], Tracy Yamawaki[2,42], Chang Jie Mick Lee[4,5], Maharshi Chakraborty[6], Andrea Bredemeyer[1], Maya U. Sheth[7,8,9], Quanyi Zhao[10], Trieu Nguyen[10], Eric Yuan[2], Hannah C. Plunkett Paletta[1], Yizhou Wang[1], Rijan Gurung[4,5], Chukwuemeka Anene-Nzelu[11,12], Wen-Ling Lin[2,13], Wei-Lin Qiu[14,15], Xue Liang[16], Yuntian Zhang[17], Cassia Williams-Rogers[1], Joohee Oh[1], Jixin Cui[2], Roland S. Wu[2], Ruth Gómez Graciani[6], Wei Qiang Seow[4,18,19], Soumya Kundu[7], Daniel Y. Li[10], Markus Ramste[10], Daniel R. Lu[2], Anthony Tan[7,8], Chul Joo Kang[1], Ryan E. Wagoner[1], Arturo Alisio[1], Sikander Hayat[20,21,22], Daniel Kreisel[23,24], Kory J. Lavine[1,23,25], Christoph Kuppe[26], Paul Cheng[10,27], Ira M. Hall[28,29], Rajat M. Gupta[30,31], Yi-Hsiang Hsu[2,29,32], Saptarsi M. Haldar[2], Rafael Kramann[20,26,33], Richard Sallari[6], Patrick Sulem[2,34], Howard Y. Chang[2], Clint L. Miller[35], Boxiang Liu[16,17,36,37,38], Scott Martin[2], Simon Jackson[2], Robin Andersson[14,15], Jesse M. Engreitz[7,8,15,27,39], Roger S-Y Foo[4,5], Chi-Ming Li[2], Brandon Ason[2], Thomas Quertermous[7,10,27] & Nathan O. Stitziel[1,40,41]✉

Although coronary artery disease (CAD) genome-wide association studies (GWAS) have identified hundreds of loci, most variants lie in non-coding regions, obscuring the causal gene programmes and cell types driving disease[1,2]. Here we performed single-nucleus multi-omic profiling of gene expression and chromatin accessibility in human coronary arteries from 44 people to construct a genome-wide single-cell variant-to-enhancer-to-gene map. We identified 11,269 chromatin accessibility quantitative trait loci (caQTLs) across principal vascular cell types, including CAD-associated caQTLs in smooth muscle cells (SMCs). Using dynamic single-cell caQTL modelling[3], we show that variant effects on chromatin vary across SMC states. For example, rs658956-mediated suppression of the *FGGY-DT* enhancer through disruption of an AP-1 motif is amplified as SMCs transition towards a fibromyocyte fate. We next performed genome-wide high-throughput chromosome conformation capture (Hi-C) in 15 human coronary arteries, showing that rs4887091 in the *ADAMTS7* CAD locus has allele-dependent effects on CCCTC-binding factor occupancy and loop strength. Integrating Hi-C, Hi-C with chromatin immunoprecipitation and in vitro clustered regularly interspaced short palindromic repeats interference (CRISPRi) nominated *AMOTL2* as a distal SMC-specific CAD gene whose loss promotes proliferation and suppresses contractile identity. Finally, arrayed Perturb–seq of GWAS-nominated enhancers converged on extracellular matrix remodelling as a core SMC disease programme. Together, these findings provide a framework for mapping non-coding GWAS variants to cell types, chromatin states and gene programmes in human disease.

Common diseases are driven by a multitude of factors and remain difficult to treat given the lack of a unified disease mechanism[4]. Coronary artery disease (CAD) is among the most prevalent common chronic diseases, and numerous genome-wide association studies (GWAS) have now identified hundreds of variants linked to CAD risk. Although these results may potentially pave the way for new therapeutic target discovery[1,2,5,6], much of the translation from GWAS to therapeutic targets for CAD has been restricted to lipid-lowering drugs[7,8]. Large-scale clinical studies have demonstrated a persistent residual risk of CAD independent of lipid levels[9,10], but we lack drugs that directly target disease pathways in the vessel wall due to an incomplete understanding of causal disease mechanisms. Leveraging human genetics to uncover new lipid-independent biological mechanisms and drug targets in CAD represents a tremendous unmet need and would be a principal milestone in translating population-scale GWAS findings to therapies for patients[11,12].

Although population-scale GWAS offer great opportunity for target discovery, most identified variants lie within non-coding regions of the genome, making it difficult to dissect the causal cell types and molecular pathways[13]. The advent of single-cell sequencing technologies now allows for the profiling of human tissues in health and disease at unprecedented granularity[14]. Previous studies have used

A list of affiliations appears at the end of the paper.

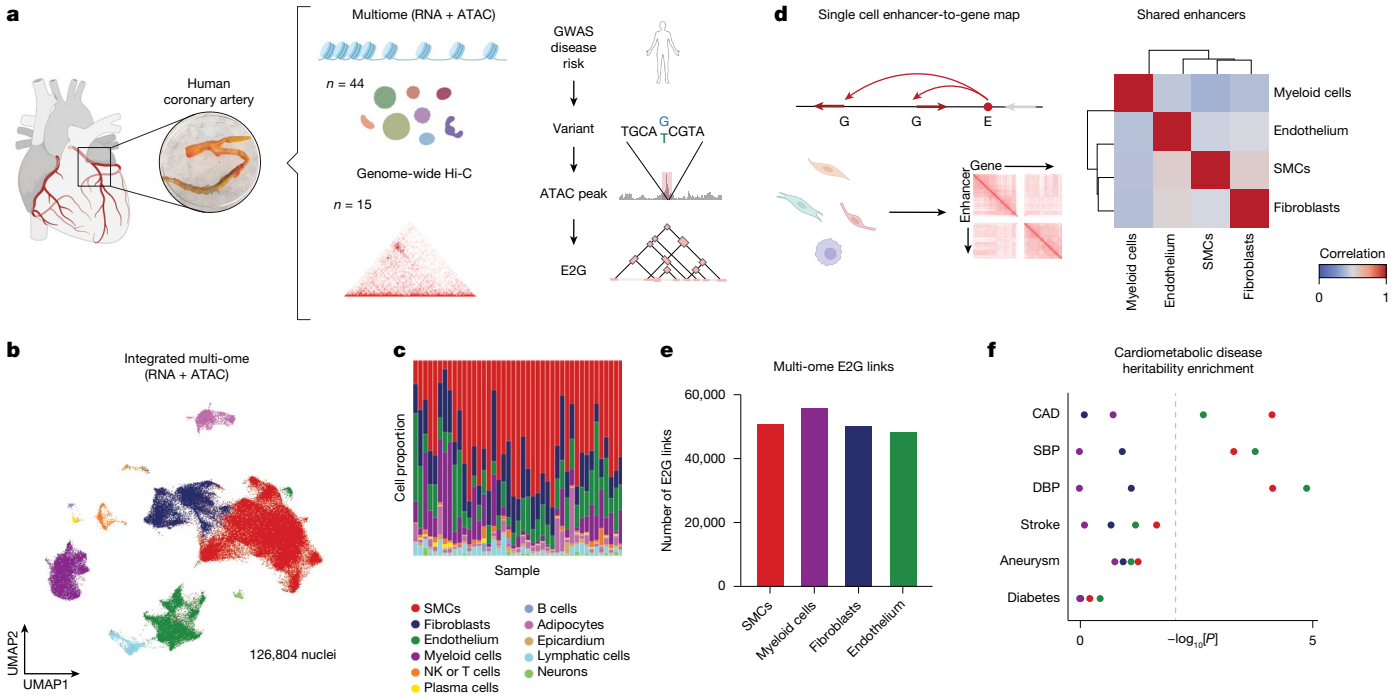


Fig. 1 | Multi-omic profiling of human coronary arteries. a, Study design. **b**, Uniform manifold approximation and projection (UMAP) embedding plot of integrated data clustered by RNA. **c**, Cell-type composition stack plot by sample. NK, natural killer. **d**, Number of shared enhancers heatmap (right) from scE2G map (left); defined as the number of enhancers where more than 50% of the peak width overlaps with other cell type-specific peaks. **e**, Number of scE2G links in the main cell types. **f**, Linkage disequilibrium score regression (LDSC) for cardiometabolic traits by cell type; dotted line, multiple-testing-adjusted *P* value threshold. *P* values were calculated using one-sided stratified LDSC, with Bonferroni correction for six conditions ($P < 8.33 \times 10^{-3}$). Illustrations in **a** and **d** created in BioRender: **a**, Lavine, K. https://biorender.com/ib7ak3x (2026); **d**, Lavine, K. https://biorender.com/cwi31gk (2026).

traditional single-cell RNA sequencing (scRNA-seq) for detailed characterization of cell states in human tissue that drive disease[15–18]; however, these studies do not profile the non-coding genome, making it challenging to link GWAS variants to observed molecular changes. More recently, single-nucleus assay for transposase-accessible chromatin (ATAC) with sequencing (snATAC–seq) in human coronary arteries was conducted to link CAD GWAS variants to regulatory elements[19]. Although this represented an important advance in the field, the lack of complementary transcriptomic profiling rendered it challenging to nominate causal genes in vascular cell types. Furthermore, isolated snATAC–seq precludes the discovery of distal gene regulation as a driver of disease[20], and there are currently no studies that profile the three-dimensional (3D) chromatin architecture in human coronary arteries to enable the investigation of these distal regulatory connections.

Herein we performed simultaneous multi-omic profiling of the transcriptome and epigenome in nuclei from 44 human coronary arteries to build a multi-omic enhancer–gene map of CAD. Collectively, these analyses identified CAD-associated regulatory variants and enhancer–gene links in smooth muscle cells (SMCs), motivating focused analysis of SMC state-dependent gene regulation. To expand insight into gene regulation across functionally heterogeneous SMC cell states, we used dynamic chromatin accessibility quantitative trait loci (caQTL) modelling to uncover cell-state-dependent pathogenicity of disease variants. To dissect the role of disease-associated variants in regulating larger scale gene networks, we performed genome-wide Hi-C in 15 human coronary arteries to link variants to distal genes and validated our findings using cell-specific high-throughput chromosome conformation capture (Hi-C) with chromatin immunoprecipitation (Hi-ChIP–seq) and in vitro CRISPRi. Finally, we performed Perturb–seq in human coronary artery SMCs (HCASMCs) to characterize functional mechanisms of repressing disease-associated enhancers and genes. Collectively, we provide a multi-omic map of human CAD by integrating transcriptomic, epigenetic and chromatin architecture to link disease-associated genetic variants with candidate disease-relevant cell types and target genes.

Multi-omic map of human CAD

We performed single-nucleus Multi-ome (RNA+ATAC) sequencing in 44 human coronary arteries from non-failing donors and patients with chronic heart failure isolated at the time of heart transplantation (Fig. 1a, Supplementary Fig. 1a and Supplementary Table 1). After quality control, we recovered 126,804 nuclei from 11 distinct cell types (Fig. 1b,c and Supplementary Fig. 1b–d). Using differential expression of canonical genes, we identified clusters corresponding to all main vascular cell types, including coronary artery SMCs (*MYH11*, *ACTA2*, *ITGA8*), endothelial cells (*VWF*, *LDB2*, *PECAM1*), fibroblasts (*LAMA2*, *DCN*, *FBLN1*) and myeloid cells (*F13A1*, *CD163*, *MRC1*) (Fig. 1b, Extended Data Fig. 1a–c and Supplementary Table 2). We used MACS2 to identify areas of open chromatin ('peaks') from the ATAC–seq data (Fig. 1e) and identified 143,743 cell-type-specific differentially accessible peaks (Extended Data Fig. 1d,e). Generally, cell-type clustering was concordant between the RNA and ATAC data. However, gene-expression data enabled the annotation of rarer cell types that were not captured by ATAC-based clustering alone (Supplementary Fig. 1e–g). Using differentially accessible peaks across cells, we also discovered cell-type-specific enrichment of transcriptional motifs such as SpiB and PU.1 in myeloid cells, MEF2B/C/D in SMCs and CCAAT/enhancer-binding protein in adipocytes (Supplementary Fig. 1h). We linked peaks to genes using a new supervised single-cell model called single-cell enhancer-to-gene (E2G) (scE2G)[21] (Fig. 1d, Extended Data Fig. 2a–g and Supplementary Tables 3–14; Methods). We found that the non-immune cells had the greatest number of shared enhancers relative to myeloid cells (Fig. 1d) and that all cell types had a similar number of E2G links (Fig. 1e and Supplementary Table 3).

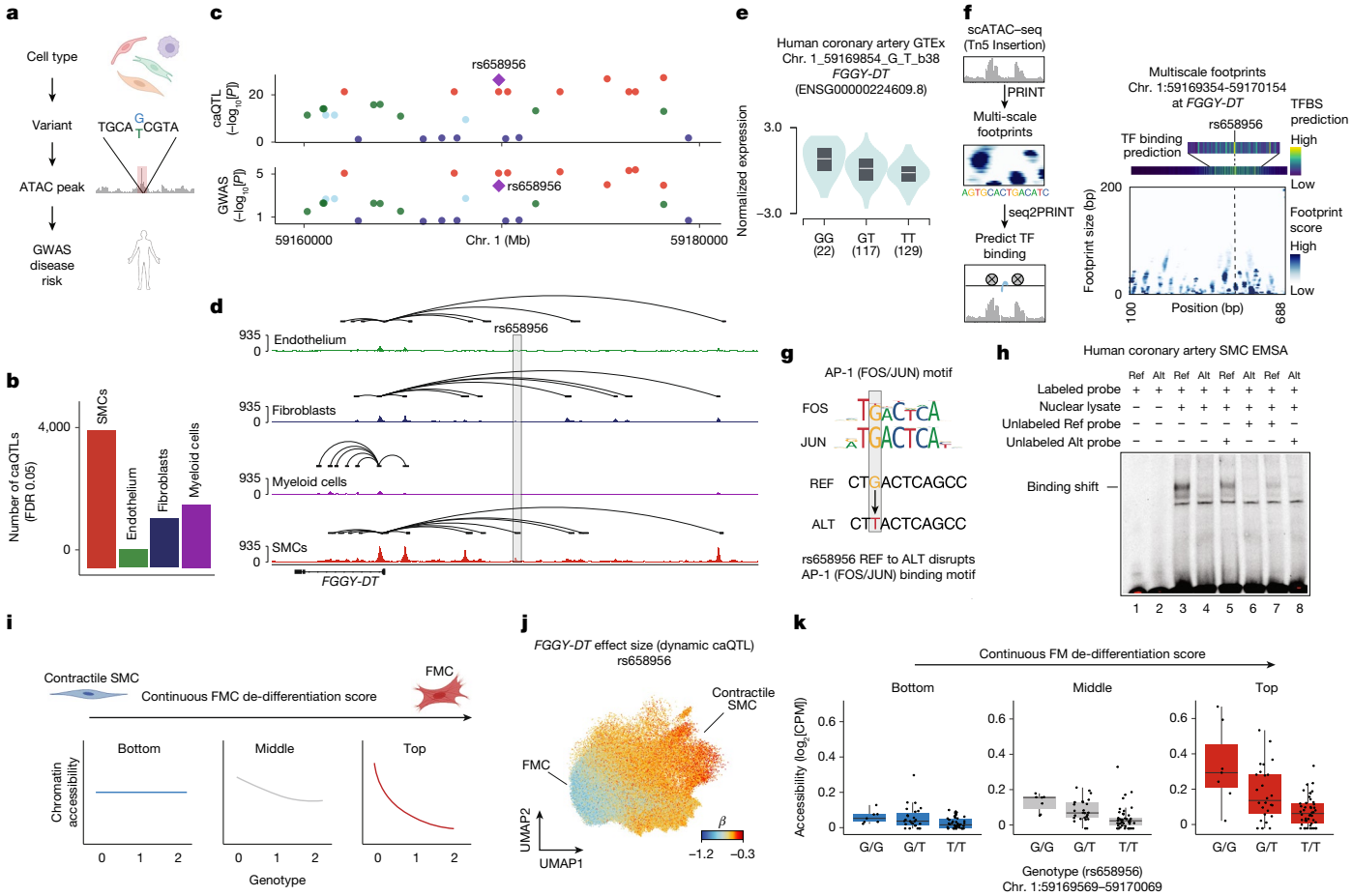


Fig. 2 | sc-caQTL discovery. a, Schematic of sc-caQTL discovery with disease enrichment. **b**, Number of sc-caQTLs by cell type at FDR 0.05. **c**, Regional Manhattan plot of CAD GWAS[1] and SMC caQTL results at the rs658956 locus. d, snATAC–seq tracks with E2G link centred at the rs658956 locus showing SMC-specific enrichment of the enhancer with an E2G link to *FGGY-DT*. **e**, Violin plot showing normalized expression by genotype for rs658956 and *FGGY-DT* in coronary arteries from GTEx ($P = 3.39 \times 10^{-12}$). The *P* value was calculated using a two-sided *t*-test of the GTEx eQTL linear regression coefficient against zero. **f**, PRINT and seq2PRINT transcription factor binding prediction pipeline (left). rs658956 overlaps region with high transcription factor binding probability from seq2PRINT model (top) and multiscale footprinting associates rs658956 with a transcription factor-specific footprint (bottom). **g**, Motif analysis finds rs658956 overlaps canonical AP-1 transcription factor (JUN, FOS) motifs. Ref, reference alelle; Alt, alternate allele. **h**, EMSA demonstrates a binding shift in the reference allele sequence (containing G) that was not present in the alternative allele sequence (containing T) when these allelic sequences were exposed to nuclear lysate from SMCs ($n = 3$ replicates, representative experiment shown). **i**, Dynamic caQTL modelling and discovery strategy (cell-state-dependent). **j**, rs658956 caQTL dynamic effect size. Net effect size depicted here represents the sum of β_g and $\beta_{FMC \times G}$ in our generalized linear mixed model across cells. **k**, rs658956 genotype association with normalized chromatin accessibility across FMC tertiles. Negative slope of association increases as FMC score increases; $n = 268$ (**e**) and $n = 79$ (**k**) biologically independent samples were used. Illustrations in **a** and **i** created in BioRender: **a**, Lavine, K. https://biorender.com/cwi31gk (2026); **i**, Lavine, K. https://biorender.com/m3r2gap (2026).

Notably, we found that cardiometabolic disease variants for CAD, systolic blood pressure (SBP) and diastolic blood pressure (DBP) were enriched in differentially accessible marker peaks specific to SMCs and endothelial cells (Fig. 1f). As a predominantly non-vascular trait, diabetes did not demonstrate enrichment of disease-associated variants (Fig. 1f).

Single-nuclei caQTL discovery

To boost our power for single-cell caQTL (sc-caQTL) discovery, we integrated our Multi-ome ATAC–seq data with available snATAC–seq data from human coronary arteries[19] to create a meta-map of 88 samples and 245,562 nuclei across four ancestry groups (Extended Data Fig. 3a–f). We then used RASQUAL to identify pseudobulk caQTLs in four principal vascular cell types (SMCs, endothelial cells, myeloid cells, fibroblasts) (Fig. 2a). At the 10% false discovery rate (FDR) threshold, we found 11,269 significant caQTL associations in SMCs, endothelial cells, myeloid cells and fibroblasts (Extended Data Fig. 3g–i, Supplementary Fig. 2a–d and Supplementary Tables 15–26). Power, sensitivity, subsampling and runtime analyses are shown in Supplementary Figs. 2 and 3 and described in Supplementary Methods.

We further characterized caQTL variants to assess their regulatory potential. Lead caQTL variants were enriched strongly near the peaks they were predicted to regulate, consistent with a *cis*-regulatory mechanism (Extended Data Fig. 3j). We next overlapped our pseudobulk caQTLs and GTEx bulk arterial tissue eQTLs by cell type and found several hundred shared QTL variants in SMCs, endothelial cells, myeloid cells and fibroblasts (Extended Data Fig. 3i). Furthermore, we found a strong correlation in effect size between our SMC caQTLs and bulk coronary artery eQTLs from GTEx (Extended Data Fig. 3k). Finally, in SMCs we trained a convolutional neural net model (seq2PRINT[22]) that incorporates chromatin accessibility and ATAC footprinting to predict transcription factor binding probability at a base-pair resolution. We found that caQTL variants within accessible peaks had significantly higher predicted transcription factor binding probabilities compared with all other non-caQTL SNP sites

within accessible peaks ($P = 8.2 \times 10^{-40}$, Wilcoxon rank sum) (Extended Data Fig. 3l).

To identify disease-associated caQTLs, we overlapped sc-caQTLs for each cell type with CAD GWAS variants (defined as lead variants associated with CAD at 1% FDR[1] and variants in linkage disequilibrium with these at $R^2 \geq 0.8$ in the European population from 1000 Genomes Project; Supplementary Table 27), which yielded numerous disease-relevant cell-type-specific caQTLs (Extended Data Fig. 3m). For example, we found that rs7182567 was an SMC-specific caQTL that was also a GWAS variant ($P = 5.4 \times 10^{-26}$) and a bulk GTEx eQTL for *ADAMTS7* in tibial artery (Extended Data Fig. 4a). Similarly, we found that rs8017642 is an endothelial cell-specific caQTL in the *NEK9* locus (Extended Data Fig. 4b), whereas rs2337154 and rs1443120 are fibroblast- and myeloid-specific caQTLs, respectively (Extended Data Fig. 4c,d).

Next, we leveraged the paired RNA data to sub-cluster SMCs into transcriptionally distinct cell states, forming a continuum from 'healthy' contractile SMC states (SMC1–3) to de-differentiated 'disease-associated' SMC states including fibromyocytes (FMCs) and chondromyocytes (CMCs) (Extended Data Fig. 4e). Notably, we found that *ADAMTS7* was expressed most highly in FMCs (Extended Data Fig. 4e), indicating a role in disease regulation. To identify functionally distinct endothelial cell states, we mapped our data to a high-resolution scRNA-seq atherosclerosis atlas[23] and identified four transcriptionally distinct endothelial subsets (Extended Data Fig. 4f). Notably, *NEK9* was expressed in endothelial cells enriched with endothelium-to-mesenchymal transition (EMT) genes such as *COL1A1* and *FN1* (refs. 24,25) (Extended Data Fig. 4f,g). Pathway enrichment in the EMT population showed increased EMT and stress-response signals (Extended Data Fig. 4h). Next, we created a gene set score comprised of 41 endothelial cell-specific genes thought to be associated causally with CAD[26] that showed maximal enrichment in intimal endothelial cells (Extended Data Fig. 4i). Similarly, we created a gene signature of a gene programme[26] identified previously as driving CAD pathogenesis in endothelial cells (Programme 8) and found an enrichment of this programme in angiogenic and/or vasa vasorum endothelial cells (Extended Data Fig. 4j).

The CAD disease-associated variant rs658956 illustrates how a cell-type-specific caQTL can impart functional transcriptional changes. The variant is an SMC-specific and fibroblast-specific caQTL and regulates a putative enhancer element (chr. 1: 59169569–59170069), which is linked to the gene *FGGY-DT* in an SMC-specific manner by scE2G (Fig. 2c,d). Consistently, we found that rs658956 is also a coronary artery eQTL for *FGGY-DT* expression ($P = 3.4 \times 10^{-12}$) wherein the alternative allele (T) is associated with decreased expression relative to the reference allele (G) (Fig. 2e) and is also predicted to overlie a high probability transcription factor binding site (Fig. 2f). Motif-based transcription factor binding site analysis found that rs658956 is predicted to disrupt a core AP-1 (JUN/FOS) transcription factor binding motif (5′-TGAGTCA-3′) (Fig. 2g). Electrophoretic mobility shift assay (EMSA) demonstrated a binding shift in the reference allele sequence (containing G) that was not present in the alternative allele sequence (containing T) when these alleles were exposed to nuclear lysate from SMCs (Fig. 2h). Antibody supershift EMSA confirmed that the shifted AP-1 complex band consists of JUN and FOS transcription factors (Extended Data Fig. 4k). Together, these results suggest that the disease-associated variant rs658956 impacts transcription factor binding, thereby decreasing the chromatin accessibility of the enhancer and subsequently reducing *FGGY-DT* expression in a SMC-specific manner.

Dynamic caQTLs across SMC states

In atherosclerosis, SMCs undergo a well-characterized phenotypic switch from a contractile state to a fibroblast-like FMC phenotype[16], driven by disease-associated transcriptional programmes. To characterize this transition, individual nuclei were assigned an FMC score based on FMC state marker genes for downstream modelling (Fig. 2i and Extended Data Fig. 5a–d). Notably, this FMC gene score is dynamic with disease across species, disease burden, pseudotime trajectory, in vitro HCASMC stress and transcriptional regulation (Extended Data Fig. 5e–k). To identify state-dependent caQTL effects, we used a Poisson mixed-effects (PME) single-cell model (Extended Data Fig. 6a–c; Methods). Applying this PME model to the lead variants of SMC-specific pseudobulk caQTLs, we identified 712 caQTLs whose effects were dynamic with respect to the FMC score (FDR ≤ 5%) (Supplementary Tables 29 and 30). Notably, eight of these dynamic caQTLs were also disease-associated variants. For example, rs658956 exhibited a significant interaction with FMC score (FDR-adjusted likelihood ratio test (LRT) P value = 2.36×10^{-4}) (Fig. 2j,k). A negative interaction term indicates that, as the FMC score increased, the negative association between rs658956 and local chromatin accessibility was amplified (Fig. 2j,k). Using the scE2G map to link dynamic caQTL-related enhancer elements (caPeaks) to target genes, we found that extracellular matrix (ECM) remodelling was the most enriched pathway among linked genes (Extended Data Fig. 6d,e and Supplementary Table 31). Notably, several of these genes linked to dynamic caPeaks were enriched in FMCs (Extended Data Fig. 6f and Supplementary Table 32). Similar state-dependent effects were observed at other CAD loci such as dynamic caQTL variant rs11838776 in the *COL4A1* locus, suggesting that *COL4A1* may play a role during transition toward an FMC state (Extended Data Fig. 6g–k). Collectively, these dynamic analyses demonstrate that disease-associated variants can exert cell-state-dependent regulatory effects on chromatin accessibility.

Single-cell CAD variant-to-gene map

To map CAD disease-associated variants to cell types and genes in an unbiased fashion, we built a disease-relevant single-cell variant-to-enhancer-to-gene map (scV2E2G) integrating cell-type-specific regulatory elements, CAD GWAS variants, scE2G links and caQTLs[21,27] (Fig. 3a; Methods).

(1) ATAC–seq peaks containing a CAD GWAS variant (n = 1,360).
(2) Cell-type-specific enhancers containing a CAD variant: SMCs (883), endothelial cells (729), fibroblasts (784) and myeloid cells (760).
(3) scE2G-linked genes: SMCs (325), endothelial cells (295), fibroblasts (295) and myeloid cells (325).
(4) Links also supported by caQTLs: SMCs (56 genes), endothelial cells (8), fibroblasts (13) and myeloid cells (27).

Overlapping scE2G with sc-caQTLs provides a high-fidelity set of predicted variants that may be driving disease by directly regulating intermediate molecular traits in a cell-type-specific manner.

SMC CAD enhancer–gene map

Focusing on the 325 genes linked to SMC enhancers harbouring CAD GWAS variants, we performed several enrichment analyses. These analyses found that these linked genes were over-represented in nearest lead genes from CAD and myocardial infarction GWAS, identified GATA2 as the most enriched transcription factor and highlighted collagen and ECM remodelling as key biological processes (Extended Data Fig. 7a and Supplementary Tables 37–39). Correlation analysis between gene expression and FMC score identified *FN1*, *COL4A1*, *COL4A2*, *MFGE8* and *MYO9B*—all known CAD-associated genes implicated in ECM remodelling—among the most highly correlated transcripts (Extended Data Fig. 7b and Supplementary Table 40). Notably, the enhancer harbouring variant rs7246865 emerged as potentially important in disease, acting as both a cell-state dynamic caPeak and pseudobulk caPeak in SMCs with a high-confidence scE2G link (Fig. 3b). We confirmed this predicted link experimentally using CRISPRi in HCASMCs. Repressing the rs7246865-containing enhancer significantly reduced *MYO9B* expression relative to a non-targeting control (Fig. 3c), consistent with

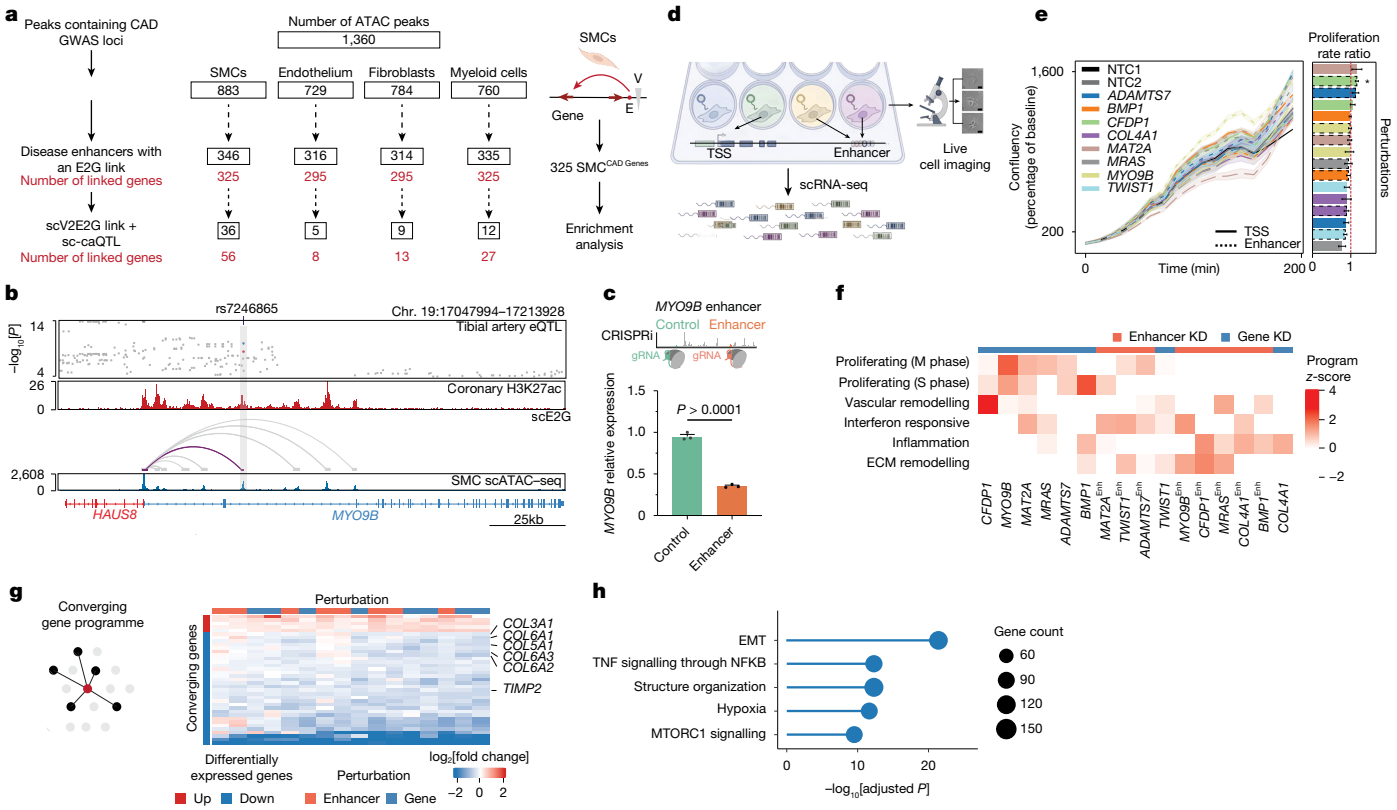


Fig. 3 | scV2E2G map of CAD. a, Number of CAD GWAS variants mapping to snATAC–seq peaks and linked genes across cell types with overlapping caQTLs. **b**, ENCODE coronary artery H3K27ac and snATAC–seq SMC tracks at rs7246865 SNP with scE2G link to *MYO9B* and *HAUS8*. The *P* value was calculated using a two-sided *t*-test of the GTEx eQTL linear regression coefficient against zero. **c**, CRISPRi in HCASMCs against a non-targeting control and enhancer containing rs7246865 with quantitative PCR (qPCR) showing reduced *MYO9B* expression; $n = 3$ biologically independent samples were used. Data represent mean ± s.e.m.; *P* values were calculated using two-sided unpaired *t*-tests or one-way analysis of variance with Dunnett's correction for multiple comparisons, as appropriate. **d**, Arrayed Perturb–seq study design. **e**, HCASMC proliferation curves (left) and proliferation rates normalized to the mean of non-targeting controls (right). Shading, ±1 s.e.m.; error bars, s.d. of 5,000 bootstrap resamples. Data are mean ± s.e.m. from $n = 4$ independent wells per condition, with four images per well at each time point. *P* values were calculated using two-sided Welch's *t*-tests versus the combined NTC1 or NTC2 control; CFDP1 enhancer, $P = 0.035$; $^*P < 0.05$. **f**, Consensus non-negative matrix factorization (cNMF) gene programme *z* score heatmap by perturbation; $k = 6$ components were retained for cNMF analysis. **g**, Converging genes across all perturbations which increase (red) or decrease (blue) relative to non-targeting control heatmap grouped by perturbation. Notable ECM remodelling genes are annotated. **h**, Converging programme pathway enrichment lollipop graph using genes down-regulated across all perturbations relative to non-targeting control. MTORC1, mammalian target of rapamycin complex 1; NFKB, nuclear factor κB; TNF, tumour necrosis factor. Illustrations in **d** created in BioRender; Lavine, K. https://biorender.com/diaxji8 (2026).

previous functional validation at this locus[1]. Furthermore, CRISPRi validated scE2G predictions in SMCs for three additional loci near genes *LOXL1*, *NAV1* and *PALLD*; enhancer repression altered gene expression significantly at each of these loci relative to non-targeting controls (Extended Data Fig. 7c–e).

Perturb–seq of CAD programmes

To further validate scE2G links, we performed arrayed CRISPRi Perturb–seq in HCASMCs, delivering individual sgRNAs to target either the transcription start site (TSS) or the nominated enhancer at eight CAD GWAS loci (Fig. 3d, Supplementary Fig. 4 and Supplementary Table 41). Alongside transcriptional profiling, we measured SMC proliferation by live-cell imaging, revealing that several perturbations (including *ADAMTS7*, *CFDP1* and *COL4A1*) and significantly reduced proliferation rate, with enhancer knockdowns generally phenocopying their cognate TSS perturbations (Fig. 3e, Extended Data Fig. 8a and Supplementary Table 42). Following quality control, we retained 76,179 cells and confirmed robust, perturbation-specific transcriptional responses, with selective target-gene knockdown, stronger effects for TSS than enhancer perturbations and separation from non-targeting controls (Extended Data Fig. 8b–j).

Gene set enrichment analysis (GSEA) revealed locus-specific pathway enrichments: *BMP1* and *ADAMTS7* knockdowns strongly enriched MYC and E2F target-gene sets, whereas *CFDP1* and *COL4A1* knockdowns suppressed interferon response pathways, indicating distinct downstream transcriptional consequences across CAD loci (Extended Data Fig. 9a). Consensus non-negative matrix factorization identified six biologically interpretable programmes: inflammation, ECM remodelling, vascular remodelling, interferon response and two proliferative programmes corresponding to S and M phase (Extended Data Fig. 9b,c and Supplementary Tables 43 and 44). Notably, the ECM remodelling programme showed the strongest positive correlation with the FMC module score, with enhancer knockdowns of CFDP1, MYO9B and BMP1 exhibiting the highest mean FMC scores (Extended Data Fig. 9d–h). Programme-level analysis further revealed that *COL4A1* and its enhancer were the strongest drivers of the inflammation programme, whereas the *MRAS* and *CFDP1* enhancer perturbations modulated both inflammation and ECM remodelling (Fig. 3f).

Finally, robust rank aggregation (RRA) across pseudobulk DESeq2 profiles revealed a convergent gene programme dominated by downregulation of several collagen genes (*COL3A1*, *COL4A1*, *COL5A1*, *COL6A1–A3*) and *TIMP2*, alongside the upregulation of metabolic and stress-response genes (Fig. 3g, Extended Data Fig. 9i and Supplementary

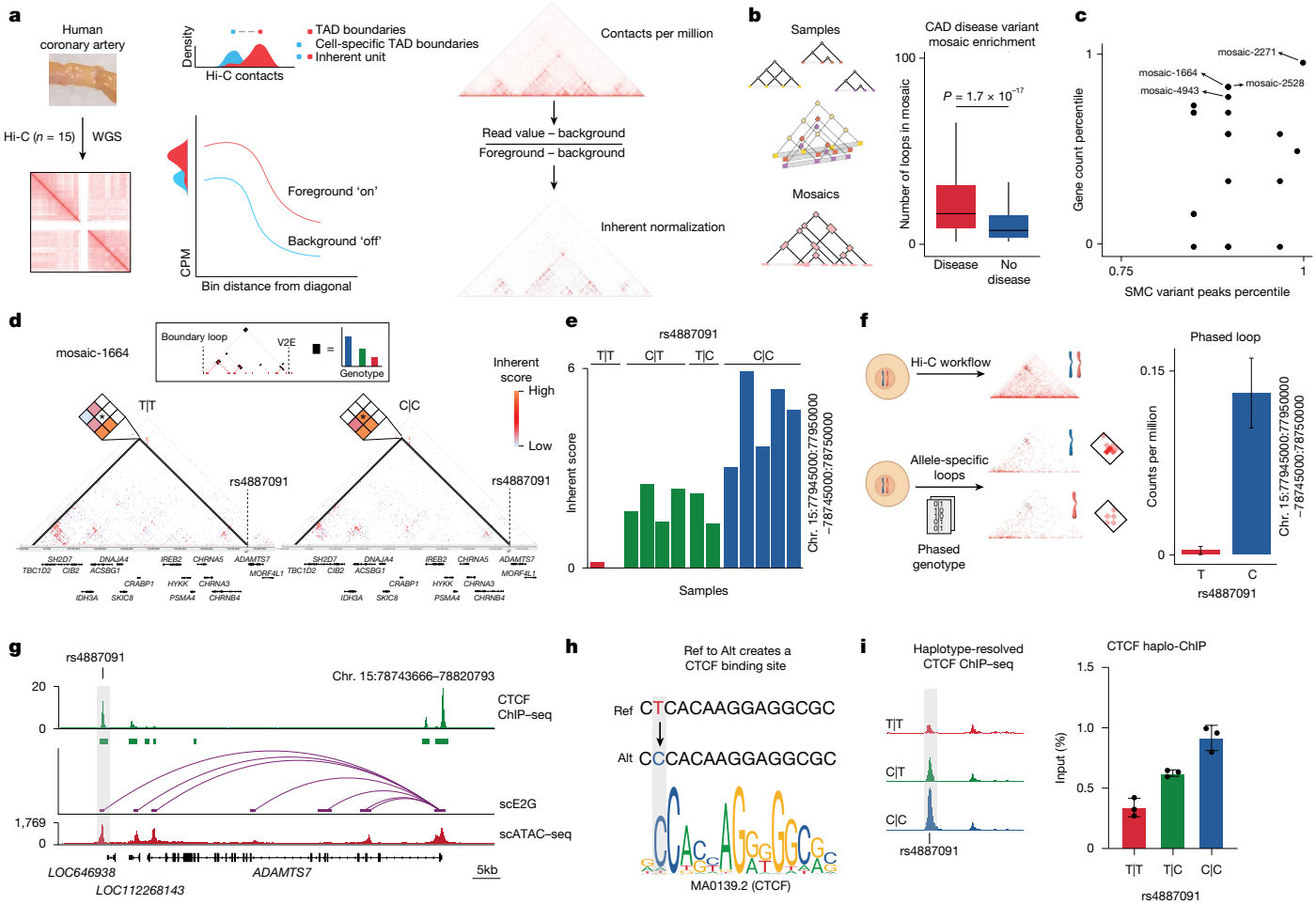


Fig. 4 | Disease chromatin network. a, Study design and computational framework with inherent normalization for looping and TAD analysis from tissue Hi-C data. **b**, Number of loops in mosaics with and without a CAD variant shows significantly greater number of Hi-C loops in mosaics that contain at least one CAD variant. Data represent medians ± interquartile range. **c**, Mosaic prioritization by number of genes and SMC enhancers containing a CAD variant. **d**, Boundary loop in mosaic-1664 originating at rs4887091; heatmaps: left, participant homozygous for the reference allele T|T; right, participant homozygous for the alternate allele C|C. The enlarged area shows loop contact coloured by inherent score. **e**, Inherent score of the boundary loop from d in all samples grouped by genotypes at rs4887091, where each bar represents a single person. **f**, Allele-specific loops from phased genotype framework (left) and boundary loop inherent score resolved from phased genotypes showing increased looping strength in the reference allele (C) at rs4887091. Data represent means ± s.e.m. **g**, CTCF ChIP–seq in HCASMCs overlaid with SMC snATAC–seq and scE2G at the rs4887091 locus. **h**, rs4887091 ref (T) to alt is predicted to create a CTCF binding site. **i**, Haplotype-resolved CTCF ChIP–seq in HCASMCs from different cells that are homozygous (T|T), heterozygous (T|C) and homozygous (C|C) showing a dosage-dependent binding by genotype at rs4887091; $n = 3$ biologically independent samples were used. Data represent means ± s.e.m. In **b** and **f**, $n = 15$ biologically independent samples were used to construct the mosaics and derive allele-specific loops; data represent means ± s.e.m.; P values were calculated using the Mann–Whitney U test.

Table 45). Pathway enrichment of this convergent programme identified EMT, tumour necrosis factor signalling through nuclear factor κB, and mammalian target of rapamycin complex 1 signalling as the top enriched terms (Fig. 3h). Collectively, these findings implicate a shared regulatory framework that governs SMC phenotypic states across independent, disease-associated loci.

CAD variant chromatin networks

To investigate how disease-associated variants may influence broader gene networks through chromatin looping, we performed genome-wide Hi-C in 15 human coronary arteries to construct large-scale chromatin networks (Fig. 4a, Extended Data Fig. 10a–f and Supplementary Figs. 5 and 6). To further complement Hi-C in bulk tissue, we performed H3K27Ac Hi-ChIP–seq in cultured human coronary artery smooth muscle cells and endothelial cells. Comparing the number of enhancer–gene connection overlap by method, we found that scE2G$^{\text{Multi-ome}}$ and Hi-ChIP yielded the highest number of enhancer–gene connections, with a smaller fraction from bulk coronary tissue Hi-C (Supplementary Fig. 7). We then identified groups of genomic loci that interact preferentially with one another in specific chromatin networks termed 'mosaics' (Fig. 4a; Methods). We identified 5,130 mosaics containing 11,494 protein-coding genes. Notably, mosaics harbouring a disease-associated variant exhibited significantly higher structural complexity, containing an average of 25 loops compared with 14 loops in non-disease-associated mosaics ($P = 1.7 \times 10^{-17}$) (Fig. 4b). This approach provides a framework for mapping disease-associated variants to a catalogue of genes regulated distally through interconnected chromatin networks (Supplementary Tables 46–49).

Next, we performed CTCF ChIP–seq in HCASMC to prioritize disease-associated SMC enhancers that are bound by CTCF. Of all CAD disease-associated variants, 498 overlapped CTCF peaks, including 200 within accessible chromatin regions. CAD-associated variants were enriched significantly at CTCF-bound regions above permutation-based expectation (1.25-fold, $P = 0.02$; Extended Data Fig. 10g).

rs4887091 tunes chromatin looping

Among the top-ranking regulatory mosaics, we identified mosaic-1664, which contains the CAD-associated variant rs4887091 located within an SMC enhancer (Fig. 4c). In this mosaic, we found an 800-kb boundary loop exhibiting a strong rs4887091 genotype-dependent increase in contact strength for the alternate allele (C) relative to the reference allele (T) (Fig. 4d). Specifically, the loop strength increased in a dose-dependent manner when quantified across all samples, where there was minimal contact strength in homozygotes (T|T), moderate strength in heterozygotes (T|C, C|T) and highest strength in homozygotes (C|C) (Fig. 4e). Phased genotypes confirmed increased looping strength on the C allele ($P = 3.1 \times 10^{-3}$) (Fig. 4f).

To characterize the regulatory architecture of rs4887091 within mosaic-1664, we identified all genes linked to the SMC enhancer containing rs4887091: Hi-C identified a boundary loop containing *ADAMTS7*, *MORF4L1*, *CHRNA3*, *IDH3A*, *ACSBG1*, *CIB2*, *SH2D7* and *TBC1D2* promoters (Fig. 4d), whereas scE2G identified *ADAMTS7* (the nearest gene) (Fig. 4g). Consistent with *ADAMTS7* being a functional target of this regulatory element, CRISPRi Perturb–seq targeting the rs4887091-containing enhancer reduced *ADAMTS7* expression in HCASMCs (Extended Data Fig. 8f). CTCF ChIP–seq in HCASMCs demonstrated that rs4887091 overlaps a CTCF binding site (Fig. 4g). Furthermore, motif-based transcription factor binding site analysis found that rs4887091, which changes the reference (T) to alternate (C) allele, is predicted to create a core CTCF binding motif (Fig. 4h). Consistent with this prediction, haplotype-resolved CTCF ChIP–seq demonstrated allelic dosage-dependent increases in CTCF occupancy on the C allele ($P = 0.0003$) (Fig. 4i). Validation at control loci and orthogonal TaqMan genotyping confirmed the accuracy of the haplotype-resolved CTCF occupancy measurements (Extended Data Fig. 10h–j). Together, these data support a model in which rs4887091 creates an allele-specific CTCF binding site associated with increased local chromatin loop strength, although the effect of this loop alteration on gene expression remains to be established.

Distal mapping of CAD variants

Although our scE2G predictions are limited to enhancer–gene interactions within a *cis*-window, leveraging Hi-C data can link CAD GWAS variants to distal genes. For the loops present in all 15 samples, we catalogued genes linked to a CAD GWAS variant through direct chromatin contact between putative enhancers and promoters and identified 11 such genes: *ANTXR1*, *ARMC12*, *CNNM2*, *FKBP5*, *CAPG*, *SH2D6*, *AMOTL2*, *NUAK2*, *RGS1*, *PLCL1* and *FOXC2* (Fig. 5a). Among these, *ANTXR1*, *CAPG*, *SH2D6*, *AMOTL2*, *NUAK2* and *RGS1* are not the nearest genes. As an example, the CAD-associated variant rs9876658 was linked to *AMOTL2* through a direct Hi-C contact. rs9876658 falls within a snATAC–seq peak in SMCs, endothelial cells and fibroblasts (but not myeloid cells) highlighting stroma-specific regulation (Fig. 5b). H3K27ac ChIP–seq showed that rs9876658 overlapped a H3K27ac peak in SMCs but not in endothelial cells (Fig. 5b). Hi-ChIP–seq further identified a direct loop between the rs9876658-containing enhancer and the *AMOTL2* promoter in SMCs but not in endothelial cells (Fig. 5b). Collectively, these analyses predict that the enhancer containing rs9876658 functionally regulates expression of *AMOTL2* in SMCs but not in endothelial cells.

rs9876658 regulates *AMOTL2* to control SMC proliferation

We found that targeting the *AMOTL2* TSS led to a complete knockdown of *AMOTL2* expression in HCASMCs and human coronary artery endothelial cells (HCAECs) (Fig. 5c). Importantly, targeting the rs9876658-containing enhancer reduced *AMOTL2* expression in HCASMCs but not HCAECs, bolstering our prediction (Fig. 5c). Collectively, these analyses nominate rs9876658 as a causal regulator of *AMOTL2* expression in SMCs but not endothelial cells and implicate this gene as a new candidate gene for atherosclerosis through an effect in SMCs. To dissect the functional impact of targeting *AMOTL2* in HCASMCs, we compared non-targeting control and *AMOTL2* knockdown HCASMCs. We observed disordered cellular morphology as well as decreased calponin (CNN1) and alpha-smooth muscle actin (ACTA2) expression by immunofluorescence in *AMOTL2* knockdown cells, suggesting a shift away from the quiescent contractile state (Fig. 5d). We performed bulk RNA sequencing in stressed non-targeting control and *AMOTL2* knockdown cells (Fig. 5e). Gene ontology pathway analysis revealed an upregulation of cell-cycle-related genes and a downregulation of cell-motility-associated pathways (Fig. 5f). Consistent with our immunofluorescence staining (Fig. 5d), we saw a reduction in contractile SMC gene expression (*ACTA2*, *CNN1*, *TAGLN* and *LMOD1*) in *AMOTL2* knockdown cells relative to control (Fig. 5g). Finally, *AMOTL2* knockdown cells had a significantly greater proliferation rate (Fig. 5h). Taken together, these findings establish a functional role for *AMOTL2* in SMC proliferation and migration—pathways integral in the pathogenesis of atherosclerosis.

Discussion

Although GWAS have identified hundreds of CAD-associated variants[1,5,28,29], translation to therapies remains limited by incomplete understanding of the causal cell types and gene programmes[30,31]. The emergence of multi-omic single-cell technologies has created a unique opportunity to profile human tissue at unprecedented resolution[14,32]. Numerous studies have leveraged these technologies to uncover cellular heterogeneity in human tissue across many diseases, contributing to a human cell atlas[17,18,24,33]. Few studies have mapped common disease variants to causal cell types and target genes directly in human tissue, with previous work limited by unimodal profiling or cell-based systems that lack the native tissue context.

Our study used a new technique—single-nucleus multi-ome (RNA+ATAC)—in human coronary arteries from 44 patients with paired genotyping for unbiased characterization of the transcriptome and epigenome at single-cell resolution. Capturing paired gene expression and chromatin accessibility from the same nucleus allows for more robust peak-gene linkage[21] and represents an important milestone for regulatory biology of the human vasculature. We also used a new computational framework—scE2G[21]—to build an unbiased genome-wide scE2G map in human CAD. Notably, our scV2E2G framework identified a substantial set of CAD-associated enhancer–gene links in SMCs, including 325 linked genes, providing a basis for downstream functional interrogation of SMC regulatory programmes. This systematic approach can be applied broadly across diseases and allows for an unbiased framework to prioritize pathological genes and cell types driving common heritable human diseases.

To nominate functional regulatory variants in a broader cohort, we integrated our data with published human coronary artery snATAC–seq[19] to identify 11,269 sc-caQTLs, enabling QTL discovery in cell types that are obscured in bulk analyses[34,35]. Mechanistic dissection of rs658956 illustrates how non-coding CAD variants can act through cell-type-specific disruption of transcription factor binding and enhancer activity[36]. The disruption of an AP-1 (JUN/FOS) motif by this variant is particularly notable given the well-established role of AP-1 in SMC phenotypic switching and the recent evidence that AP-1 family members act as pioneer factors governing vascular cell chromatin landscapes[37] and carry epigenetic memory of inflammation[38]. Furthermore, SMC-specific regulation of the lncRNA FGGY-DT highlights the potential contribution of cell-restricted non-coding RNA programmes to cardiovascular disease[39–41], although its downstream role remains to be defined.

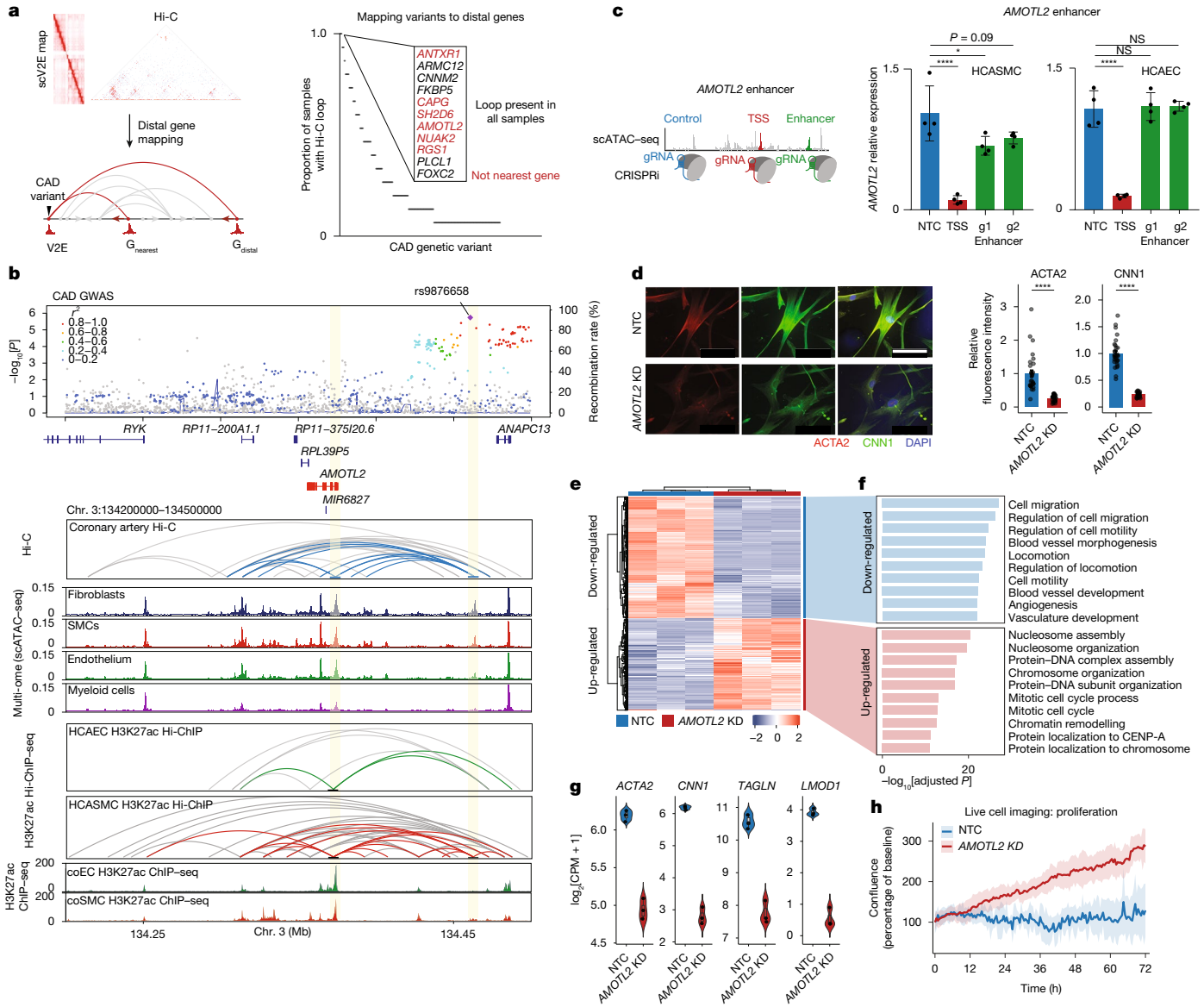


Fig. 5 | CAD-associated enhancer regulates *AMOTL2* in SMCs. a, Framework linking CAD variants to distal genes using Hi-C (left) and proportion of 15 samples containing each loop (right). Boxed genes are linked to a CAD variant by a direct Hi-C loop in all samples; red genes are non-nearest. **b**, LocusZoom plot for rs9876658 overlaid with human coronary Hi-C, coronary snATAC–seq and H3K27ac Hi-ChIP–seq/ChIP–seq from primary HCASMCs and HCAECs, showing an SMC-specific enhancer linked to the *AMOTL2* TSS in SMCs but not in endothelial cells. *P* values are two-sided CAD GWAS association *P* values from fixed-effect inverse-variance weighted meta-analysis. **c**, CRISPRi of the rs9876658-containing enhancer in HCASMCs and HCAECs regulates *AMOTL2* in SMCs but not in endothelial cells; non-targeting and TSS controls are shown; $n = 4$ biologically independent samples. Data are mean ± s.e.m.; *P* values were calculated using two-sided unpaired *t*-tests; $^{*}P < 0.05$, $^{****}P < 0.0001$. **d**, Immunofluorescence (4′,6-diamidino-2-phenylindole (DAPI), CNN1, ACTA2) in HCASMCs with non-targeting control or *AMOTL2* knockdown. Quantification of ACTA2 and CNN1 fluorescence per DAPI^{+} nucleus in *AMOTL2* NTC and knockdown cells across independent fields of view ($n = 30$). Data are mean ± s.e.m.; *P* values were calculated using two-sided unpaired *t*-tests. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$. **e**, Heatmap of top differentially expressed genes in non-targeting control versus *AMOTL2* knockdown in stressed HCASMCs. **f**, Gene ontology biological process enrichment following *AMOTL2* knockdown in stressed HCASMCs. *P* values were calculated using one-sided Fisher's exact tests with Benjamini–Hochberg correction. **g**, Violin plots of contractile genes in *AMOTL2* knockdown versus control from bulk RNA-seq; $n = 3$ biologically independent samples. **h**, Cell proliferation by confluency (percentage of baseline) in control versus *AMOTL2* knockdown over 72 h; $n = 8$ biologically independent samples; shading, mean ± s.e.m. Scale bar, 100 µm.

Traditional pseudobulk QTL discovery lacks the ability to discriminate between molecular traits that are influenced by disease state[3]. Here we used single-cell genomics to define cell-state dynamics in one of the most fluid cell types in the atherosclerotic plaque. In atherosclerosis, contractile SMCs undergo de-differentiation and acquire an FMC phenotype with distinct transcriptional signatures[16]. We applied a continuous score reflecting SMC to FMC transition to map dynamic caQTLs that are dependent on cell-state. Using this framework, we uncovered rs658956 as a caQTL for *FGGY-DT* enhancer that is associated with altered activity as SMCs transition to FMCs. This is an application of single-cell-state-dependent QTL modelling in human cardiovascular disease and expands our understanding of gene regulation in heterogeneous cell states.

We applied arrayed Perturb–seq[42] to GWAS-nominated enhancers and their SMC-linked target genes, enabling transcriptome-wide functional validation of non-coding regulatory elements and their downstream gene programmes. Critically, the arrayed format preserves non-cell-autonomous transcriptional effects that can be obscured in

pooled screens where shared culture conditions confound intrinsic regulatory responses through paracrine signalling, as in a recent CAD variant-to-gene-to-programme framework[26]. The convergence of several enhancer perturbations on inflammatory and ECM-remodelling programmes is consistent with the established role of said processes in atherosclerosis, including induction of matrix metalloproteinases and collagens[43]. Together, these data support a model in which CAD variants act through cell-type-specific enhancers to co-ordinately dysregulate inflammation, ECM homeostasis and proliferation during transition from contractile SMCs to pathological FMCs[18,44].

Given that the genome is organized as a dynamic 3D structure[45], chromatin looping through topologically associating domains (TADs) represents a higher-order regulatory layer governing gene expression[45,46]. Although Hi-C has been used in cell-based systems and other tissues to map chromatin architecture and connect disease variants to distal genes[47,48], large-scale chromatin structures are highly tissue- and cell-type-specific[27,48,49] and remain unexplored in diseased human arterial tissue. Here, genome-wide Hi-C in 15 human coronary arteries enabled tissue-specific mapping of chromatin networks and distal variant–gene interactions. Notably, we discovered that rs4887091 creates an allele-specific CTCF binding site with dosage-dependent effects on both CTCF occupancy and loop strength, defining a mechanistically distinct non-coding regulatory mechanism involving 3D genome architecture[48]. Haplotype-resolved analysis confirmed the allele-specific increase in looping strength[50,51], consistent with recent demonstrations that phased Hi-C can resolve allele-specific chromatin structure at disease loci[52] and that variants within CTCF sites can alter looping and gene regulation in a dosage-sensitive manner[50,51]. The linked gene, *ADAMTS7*, is causally implicated in CAD, with Adamts7 deficiency reducing atherosclerosis and its catalytic activity regulating SMC migration[53,54]. Although our data do not establish altered looping as the proximal mechanism controlling ADAMTS7 expression, allele-dependent changes to chromatin architecture may represent a previously unrecognized regulatory mechanism at this locus.

Furthermore, our tissue Hi-C framework enabled systematic genome-wide nomination of distal causal genes. We identified 11 genes linked to CAD variants through chromatin contacts, most of which, for example *AMOTL2*, were not the gene nearest to the disease-associated variant. CRISPRi targeting of the enhancer linked distally to *AMOTL2* reduced *AMOTL2* expression in HCASMCs but not HCAECs, and *AMOTL2* knockdown in HCASMCs increased cellular proliferation and suppressed contractile marker expression, consistent with a shift away from a quiescent state. This phenotype aligns with the known function of AMOTL2 as a scaffold protein[55,56] that retains YAP in the cytoplasm to suppress Hippo-driven proliferative transcription, such that its loss de-represses YAP nuclear activity and promotes cell-cycle progression[57]. The connection between a CAD-associated variant and a Hippo pathway regulator with established roles in vascular cell proliferation and mechanosensing[58,59] points to dysregulation of YAP and/or TAZ signalling in SMCs as a potentially underappreciated axis in atherosclerotic disease.

Our study has several limitations. First, nuclear RNA provides more limited cell-state resolution than whole-cell RNA, so we leveraged high-resolution published single-cell data for reference mapping and state imputation. Second, caQTL discovery remains constrained by sample size, cell abundance and allele frequency, limiting power in rarer cell types and variants and complicating interpretation of differences in caQTLs across cell types. Residual population stratification in our multi-ancestry cohort may have also resulted in false positives, although our summary statistics can support future meta-analyses in larger cohorts. Third, multi-ome profiling does not capture post-transcriptional regulation such as alternative splicing. Finally, bulk coronary Hi-C precludes genome-wide cell-type-specific looping inference; we therefore used targeted Hi-ChIP–seq for cell-type resolution at selected loci. Although CRISPRi Perturb–seq supports ADAMTS7 as a responsive target of the rs4887091-containing enhancer, direct allele editing or CTCF motif perturbation will be required to establish whether allele-dependent looping causally regulates ADAMTS7 expression.

In conclusion, we built a multi-omic gene-expression and chromatin-accessibility map of human CAD and provide a comprehensive framework to map CAD GWAS variants to causal genes and affected cell types. Furthermore, we use pseudobulk and single-cell QTL modelling to characterize state-dependent and independent regulatory mechanism of disease variants in human tissue. We used tissue Hi-C to build large-scale chromatin networks and uncover how disease variants impact distal gene regulation. Finally, we integrated tissue -omics with in vitro epigenetic profiling to prioritize and functionally test candidate genes using Perturb–seq and CRISPRi. Collectively, we provide a disease agnostic framework to apply human genetic findings to nominate pathologic cell states and genes driving disease, constructing a comprehensive GWAS variant-to-gene map with genetic- and tissue-level convergence for future mechanistic and therapeutic studies.

Online content

Any methods, additional references, Nature Portfolio 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/s41586-026-11275-z.

1. Aragam, K. G. et al. Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nat. Genet.* **54**, 1803–1815 (2022).
2. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.* **18**, 331–344 (2017).
3. Nathan, A. et al. Single-cell eQTL models reveal dynamic T cell state dependence of disease loci. *Nature* **606**, 120–128 (2022).
4. Boyle, E. A., Li, Y. I. & Pritchard, J. K. An expanded view of complex traits: from polygenic to omnigenic. *Cell* **169**, 1177–1186 (2017).
5. Tcheandjieu, C. et al. Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat. Med.* **28**, 1679–1692 (2022).
6. Musunuru, K. & Kathiresan, S. Genetics of common, complex coronary artery disease. *Cell* **177**, 132–145 (2019).
7. Shapiro, M. D., Tavori, H. & Fazio, S. PCSK9: from basic science discoveries to clinical trials. *Circ. Res.* **122**, 1420 (2018).
8. Abifadel, M. et al. Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
9. Ridker, P. M. et al. Inflammation, cholesterol, lipoprotein(a), and 30-year cardiovascular outcomes in women. *New Engl. J. Med.* **391**, 2087–2097 (2024).
10. Bayturan, O. et al. Clinical predictors of plaque progression despite very low levels of low-density lipoprotein cholesterol. *J. Am. Coll. Cardiol.* **55**, 2736–2742 (2010).
11. Kamb, A., Harper, S. & Stefansson, K. Human genetics as a foundation for innovative drug development. *Nat. Biotechnol.* **31**, 975–978 (2013).
12. Young, E. P. & Stitziel, N. O. Capitalizing on insights from human genetics to identify novel therapeutic targets for coronary artery disease. *Annu. Rev. Med.* **70**, 19–32 (2019).
13. Cano-Gamez, E. & Trynka, G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. *Front. Genet.* **11**, 505357 (2020).
14. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The technology and biology of single-cell RNA sequencing. *Mol. Cell* **58**, 610–620 (2015).
15. Chaffin, M. et al. Single-nucleus profiling of human dilated and hypertrophic cardiomyopathy. *Nature* https://doi.org/10.1038/S41586-022-04817-8 (2022).
16. Wirka, 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).
17. Amrute, J. M. et al. Targeting immune–fibroblast cell communication in heart failure. *Nature* https://doi.org/10.1038/s41586-024-08008-5 (2024).
18. Amrute, J. M. et al. Targeting modulated vascular smooth muscle cells in atherosclerosis via FAP-directed immunotherapy. *Science* https://doi.org/10.1126/science.adx1736 (2026).
19. Turner, A. W. et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.* **54**, 804–816 (2022).
20. Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
21. Sheth, M. U. et al. Mapping enhancer-gene regulatory interactions from single-cell data. *Nat. Genet.* **58**, 1941–1952 (2026).
22. Hu, Y. et al. Multiscale footprints reveal the organization of cis-regulatory elements. *Nature* **638**, 779–786 (2025).

23. Mosquera, J. V. et al. Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. *Cell Rep.* **42**, 113380 https://doi.org/10.1016/j.celrep.2023.113380 (2023).
24. Litviňuková, M. et al. Cells of the adult human heart. *Nature* **588**, 466–472 (2020).
25. Reichart, D. et al. Pathogenic variants damage cell composition and single-cell transcription in cardiomyopathies. *Science* **377**, eabo1984 (2022).
26. Schnitzler, G. R. et al. Convergence of coronary artery disease genes onto endothelial cell programs. *Nature* **626**, 799–807 (2024).
27. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
28. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
29. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
30. Hartmann, K., Seweryn, M. & Sadee, W. Interpreting coronary artery disease GWAS results: a functional genomics approach assessing biological significance. *PLoS ONE* **17**, e0244904 (2022).
31. Visscher, P. M. et al. 10 Years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
32. Elmentaite, R., Domínguez Conde, C., Yang, L. & Teichmann, S. A. Single-cell atlases: shared and tissue-specific cell types across human organs. *Nat. Rev. Genet.* **23**, 395–410 (2022).
33. Kuppe, C. et al. Spatial multi-omic map of human myocardial infarction. *Nature* https://doi.org/10.1038/s41586-022-05060-x (2022).
34. Cuomo, A. S. E., Nathan, A., Raychaudhuri, S., MacArthur, D. G. & Powell, J. E. Single-cell genomics meets human genetics. *Nat. Rev. Genet.* **24**, 535 (2023).
35. Aguet, F. et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
36. Zhao, Q. et al. Molecular mechanisms of coronary disease revealed using quantitative trait loci for TCF21 binding, chromatin accessibility, and chromosomal looping. *Genome Biol.* **21**, 135 (2020).
37. Li, S., Tan, Y., Tong, J. & Xu, X. The role of AP-1 in distinct cardiac cell types: pathological repair and maladaptive remodeling. *Biochim. Biophys. Acta* **1872**, 168136 (2026).
38. Nagaraja, S. et al. Epigenetic memory of colitis promotes tumour growth. *Nature* https://doi.org/10.1038/S41586-026-10258-4 (2026).
39. Uchida, S. & Dimmeler, S. Long noncoding RNAs in cardiovascular diseases. *Circ. Res.* **116**, 737–750 (2015).
40. Dong, K. et al. CARMN is an evolutionarily conserved smooth muscle cell-specific LncRNA that maintains contractile phenotype by binding myocardin. *Circulation* **144**, 1856–1875 (2021).
41. Shi, H. et al. Discovery of transacting long noncoding RNAs that regulate smooth muscle cell phenotype. *Circ. Res.* **132**, 795–811 (2023).
42. Dixit, A. et al. Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866 (2016).
43. Bennett, M. R., Sinha, S. & Owens, G. K. Vascular smooth muscle cells in atherosclerosis. *Circ. Res.* **118**, 692 (2016).
44. Lambert, J. et al. Network-based prioritization and validation of regulators of vascular smooth muscle cell proliferation in disease. *Nat. Cardiovasc. Res.* **3**, 714–733 (2024).
45. Schleif, R. DNA looping. *Annu. Rev. Biochem.* **61**, 199–223 (1992).
46. Grubert, F. et al. Landscape of cohesin-mediated chromatin loops in the human genome. *Nature* **583**, 737–743 (2020).
47. Grant, Z. L. et al. Dose-dependent sensitivity of human three-dimensional chromatin to a heart disease–linked transcription factor. *Science* **393**, eadv5434 (2026).
48. Xie, Y. et al. Single-cell multiomics and chromatin structure reveal gene-regulatory dynamics in heart failure. *Science* **393**, eady6893 (2026).
49. Heffel, M. G. et al. Temporally distinct 3D multi-omic dynamics in the developing human brain. *Nature* **635**, 481–489 (2024).
50. Bhattacharyya, S. & Ay, F. Identifying genetic variants associated with chromatin looping and genome function. *Nat. Commun.* **15**, 8174 (2024).
51. Saint Just Ribeiro, M., Tripathi, P., Namjou, B., Harley, J. B. & Chepelev, I. Haplotype-specific chromatin looping reveals genetic interactions of regulatory regions modulating gene expression in 8p23.1. *Front. Genet.* **13**, 1008582 (2022).
52. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
53. Mizoguchi, T. et al. Coronary disease association with ADAMTS7 is due to protease activity. *Circ. Res.* **129**, 458–470 (2021).
54. Bauer, R. C. et al. Knockout of Adamts7, a novel CAD locus in humans, reduces atherosclerosis in mice. *Circulation* **131**, 1202 (2015).
55. Wang, Y. et al. Angiomotin-like2 gene (amotl2) is required for migration and proliferation of endothelial cells during angiogenesis. *J. Biol. Chem.* **286**, 41095 (2011).
56. Wang, H., Ye, M. & Jin, X. Role of angiomotin family members in human diseases (Review). *Exp. Ther. Med.* **27**, 258 (2024).
57. Mannion, A. J. et al. The junctional mechanosensor AmotL2 regulates YAP promotor accessibility. Preprint at *bioRxiv* https://doi.org/10.1101/2023.01.13.523596 (2023).
58. Wang, Y. et al. Arterial wall stress induces phenotypic switching of arterial smooth muscle cells in vascular remodeling by activating the YAP/TAZ signaling pathway. *Cell. Physiol. Biochem.* **51**, 842–853 (2018).
59. Daoud, F. et al. YAP/TAZ deletion in vascular smooth muscle cells mirrors atherosclerosis-associated transcriptional programs. *J. Mol. Cell. Cardiol. Plus* **14**, 10048-7 (2025).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2026

[1]Department of Medicine, Center for Cardiovascular Research, Division of Cardiology, Washington University School of Medicine, Saint Louis, MO, USA. [2]Amgen Global Research, South San Francisco, CA, USA. [3]Helmholtz-Institute for Translational AngioCardioScience (HI-TAC) of the Max Delbrück Center for Molecular Medicine in the Helmholtz Association (MDC) at Heidelberg University, Heidelberg, Germany. [4]Cardiovascular Metabolic Disease Translational Research Programme, National University Health System, Centre for Translational Medicine, Singapore, Singapore. [5]Institute of Molecular and Cell Biology, Singapore, Singapore. [6]Axiotl Inc., Cleveland, OH, USA. [7]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [8]Basic Sciences and Engineering Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA, USA. [9]Department of Bioengineering, Stanford University, Stanford, CA, USA. [10]Department of Medicine, Division of Cardiovascular Medicine, Stanford University, Stanford, CA, USA. [11]Montreal Heart Institute, Montreal, Quebec, Canada. [12]Department of Medicine, Université de Montréal, Montreal, Quebec, Canada. [13]Amgen R&D Postdoctoral Fellows Program, South San Francisco, CA, USA. [14]Department of Biology, University of Copenhagen, Copenhagen, Denmark. [15]The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute, Cambridge, MA, USA. [16]Department of Pharmacy and Pharmaceutical Sciences, Faculty of Science, National University of Singapore, Singapore, Singapore. [17]Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [18]Agency for Science, Technology and Research (A*STAR), Singapore, Singapore. [19]Varios Bio, Singapore, Singapore. [20]Institute of Experimental Medicine and Systems Biology, Medical Faculty, RWTH Aachen University, Aachen, Germany. [21]Department of Medicine, Cardiovascular Research Institute, Division of Cardiology, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [22]The Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [23]Department of Pathology and Immunology, Washington University School of Medicine, Saint Louis, MO, USA. [24]Department of Surgery, Division of Cardiothoracic Surgery, Washington University School of Medicine, Saint Louis, MO, USA. [25]Department of Developmental Biology, Washington University School of Medicine, Saint Louis, MO, USA. [26]Department of Nephrology, Medical Faculty, RWTH Aachen, Aachen, Germany. [27]Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA. [28]Center for Genomic Health, Yale University, New Haven, CT, USA. [29]Department of Genetics, Yale University, New Haven, CT, USA. [30]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [31]Department of Medicine, Divisions of Genetics and Cardiovascular Medicine, Brigham and Women's Hospital, Boston, MA, USA. [32]Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA, USA. [33]Department of Internal Medicine, Nephrology and Transplantation Erasmus Medical Center, Rotterdam, Netherlands. [34]deCODE genetics, Reykjavik, Iceland. [35]Department of Public Health Sciences, Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA. [36]Agency for Science, Genome Institute of Singapore (GIS), Technology and Research (A*STAR), Singapore, Singapore. [37]Precision Medicine Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore. [38]Yong Loo Lin School of Medicine, NUS Centre for Cancer Research, National University of Singapore, Singapore, Singapore. [39]Gene Regulation Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [40]Department of Genetics, Washington University School of Medicine, Saint Louis, MO, USA. [41]John Cochran Veterans Affairs Medical Center, Saint Louis, MO, USA. [42]These authors contributed equally: Junedh M. Amrute, Paul C. Lee, Ittai Eres, Tracy Yamawaki. ✉e-mail: junedh.amrute@mdc-berlin.de; nstitziel@wustl.edu

Methods

Ethical approval for human specimens

The study is compliant with all relevant ethical regulations and has been approved by the Washington University School of Medicine Institutional Review Board (no. 201104172). Informed consent was obtained from each patient before tissue collection by Washington University School of Medicine; no compensation was provided in exchange for participation in the study, and participants consented to publication of research data. Written consent was obtained. Demographic data have been de-identified and are provided in Supplementary Table 1. Patients included in this study span diverse race, age and sex to provide an inclusive trans-ethnic study population.

Inclusion criteria

Before tissue collection, specific inclusion criteria were used to ensure well controlled study groups. Any patients with human immunodeficiency virus or hepatitis and known genetic cardiomyopathies were excluded from this study. The left anterior descending coronary artery was isolated, and flash frozen from donor hearts (patients with stable ejection fractions who experienced non-cardiac cause of death) and from patients with chronic heart failure (ischaemic and non-ischaemic cardiomyopathy) at the time of heart transplantation. For all samples, the proximal left anterior descending coronary artery was used. All demographic and clinical data have been de-identified and are provided in Supplementary Table 1. Recruitment was therefore determined by donor and/or transplant tissue availability and clinical indication rather than population-based sampling, which may limit generalizability to the broader CAD population.

Nuclei isolation for multi-ome sequencing

The left anterior descending coronary artery was dissected from explanted hearts; epicardial fat was removed, and arteries were flash frozen using liquid nitrogen. Identical regions from the proximal left anterior descending artery were used from all patients. Single-nuclei suspensions were generated as described previously[17]. Nuclei were isolated according to the 10x Genomics protocol (Nuclei Isolation Complex Sample for ATAC GEX Sequencing RevB, CG000375) and flow cytometry for 7-AAD (Sigma, SML1633-1ML) positive nuclei was used for sorting using a BD FACS Melody (BD Biosciences) with a 100-µm nozzle. Protocol CG000338 from 10x Genomics was used for Chromium Next GEM Single Cell Multi-ome ATAC + Gene Expression. Briefly, following nuclei isolation, permeabilization was performed, followed by transposition, GEM generation and barcoding using ChipJ (10x Genomics, PN1000234), post-GEM clean up, pre-amplification PCR, cDNA amplification, library construction and sequencing. Gene-expression and ATAC libraries were sequenced to a read depth of 50,000 and 25,000, respectively, using a NovaSeq 6000 platform (Illumina) at the McDonnell Genome Institute. Data were acquired and sorted using BD FACSChorus v.1.1.

Multi-ome data processing

Raw fastq files were aligned to the human GRCh38 reference genome using CellRanger ARC (10x Genomics, v.6.1), with ATAC data processed in ArchR (v.1.0.3)[60] and RNA data in Seurat (v.5.5.1). After quality control, doublet removal, SCTransform normalization (sctransform v.0.4.3)[61], Harmony integration (v.2.0.5)[62], clustering and RNA-based cell-type annotation, RNA annotations were transferred to the paired ATAC profiles. Detailed quality control thresholds and clustering procedures are provided in Supplementary Fig. 1.

scE2G model predictions

We used the scE2GMulti-ome model[21] to predict enhancer–gene connections in 11 cell types from 10x Genomics multi-ome data. In each cell type, scE2G scores every candidate element–gene pair (where candidate elements are ATAC peaks within 5 Mb of the TSS of the gene) by integrating several features. Regulatory enhancer–gene interactions were defined as element–gene pairs with a score greater than 0.171.

The scE2G$^{multi\text{-}ome}$ model is a supervised classifier adapted for multi-omic data from ENCODE-rE2G[63]. It integrates eight features, including (1) a pseudobulk activity-by-contact score, where 3D contact is estimated by an inverse function of genomic distance; (2) the Kendall correlation across single cells between element accessibility and gene expression; (3) whether the gene is 'expressed ubiquitously' and (4) several other measures of genomic distance and chromatin accessibility around the element and promoter. The score threshold of 0.171 was determined as the score yielding 70% recall when evaluating predictions in K562 cells against CRISPRi-validated enhancer–gene pairs.

scE2G predictions for all cell types can be found in Supplementary Tables 3–14. To construct the disease-relevant scV2E2G map, we selected ATAC–seq peaks from the four main cell types that contained a CAD GWAS variant, defined cell-type-specific enhancers by overlap with coronary artery H3K27ac peaks, linked these enhancers to genes using scE2G and finally overlapped scE2G peaks with sc-caQTLs to nominate links supported by a molecular trait.

LDSC and GWAS variant overlap

To test for enrichment of disease risk in the four main cell types, stratified LDSC[64] was used to partition heritability in cell-type marker peaks. First, cell-type specific marker peaks (FDR < 0.01, $\log_2$[fold change] > 1) were obtained using ArchR (v.1.0.3) and converted to GRCh37 coordinates using liftover[65]. GWAS summary statistics for CAD[1], diabetes[66], abdominal aortic aneurysms[67], DBP and SBP[68], and stroke [69] were obtained, and their formats were standardized using MungeSumstats (v.1.20.0)[70]. Stratified LD score analysis (LDSC v.1.0.1) was then carried out with --h2-cts flag for cell-type specific analyses. *P* values obtained from stratified LDSC were considered significant for a condition if they were lower than Bonferroni-corrected *P* value threshold of 8.33×10^{-3} (for six conditions tested). To examine overlap between CAD GWAS variants and ATAC peaks, lead variants from 241 genome-wide significant ($P < 5 \times 10^{-8}$) loci and 897 conditionally independent variants meeting FDR cutoff (FDR < 0.01) were obtained from Aragam and colleagues[1]. Additional variants that are in high linkage disequilibrium ($r^2 > 0.80$ within 250 kb) with these lead variants were obtained from the 1000 Genomes European panel using PLINK v.1.9 --tag-r2 flag. Furthermore, these variants were combined with the functional fine-mapped credible set variants from Aragam and colleagues[1] to create a superset of GWAS variants to be tested for overlap. These variants were overlapped with ATAC peak regions using the findOverlaps function in the IRanges R package (v.2.46.0).

Genotyping

To obtain high-quality genotypes from patients, DNA from peripheral leukocytes of 39 participants was genotyped using Illumina GSA-24-V3 SNP array. Following initial processing using Illumina GenomeStudio software v.2.0.5, variants with both (1) minor allele frequency (MAF) >5% and call rate <95% or (2) MAF <5% and call rate <99% were excluded. No patients were excluded by the call rate exclusion filter (<95%). Additional genotypes were further imputed from the TOPMED panel on the Michigan Imputation Server v.1.7.1 using minimac v.4-1.0.2 (ref. 71) and phasing with Eagle v.2.4.1 (ref. 72). Following imputation, variants with imputation $R^2 > 0.3$ and MAF > 0.05 were further kept for caQTL analysis for a final dataset containing 7,250,405 variants. To conduct combined caQTL mapping, the variant call format (VCF) files from Turner and colleagues[19] were merged with the imputed genotype dataset, and shared SNPs in only two datasets were kept for downstream analysis. Turner and colleagues carried out low-coverage whole-genome sequencing; further details are available in the original publication[19]. Following the merge, 5,229,397 variants were available for QTL analysis. Local ancestries were inferred using RFMix v.2.03-r0 with default settings[73].

Yoruba in Ibadan, Nigeria (YRI, $n = 186$) and Utah residents with ancestry from northern and western Europe (CEU, $n = 183$) from the 1000 Genomes Project[74] were used as African and/or European reference populations, respectively.

snATAC–seq reference mapping

To achieve uniform cell-type annotations across datasets, snATAC–seq data from Turner and colleagues[19] were reference mapped onto the multi-ome dataset rather than integrated. Raw snATAC–seq data from Turner and colleagues were downloaded and reprocessed using identical quality control thresholds (minimum fragment count and TSS enrichment cut-offs) as applied to the multi-ome snATAC–seq, ensuring consistent data processing upstream of annotation. Gene activity scores were computed from the snATAC–seq data from Turner and colleagues using ArchR (v.1.0.3)[60] and used as the query, with the snRNA-seq component of the multi-ome data serving as the reference for label transfer through Seurat's anchor-based framework (v.5.5.1). This approach leveraged the higher-quality transcriptomic annotations from the multi-ome dataset to annotate the external snATAC–seq data without merging the two datasets. Annotation concordance between transferred labels and author-provided annotations from Turner and colleagues[19] was assessed using Jaccard Index across the four main cell populations used for QTL mapping.

Pseudobulk caQTL discovery

Similar to previous studies for sc-caQTL discovery[19,75], we used RASQUAL[76] to identify pseudobulk caQTLs in four main cell types (SMCs, endothelial cells, myeloid cells and fibroblasts). RASQUAL v.1.1 was chosen as our mapping method due to its ability to simultaneously model allelic imbalance and count-based QTL effect to improve fine-mapping of putative causal variants. RASQUAL maximizes power for caQTL discovery by simultaneously modelling allelic imbalance and total read counts in each locus. To generate fragment counts within each ATAC peak regions, we loaded each fragment object and used Signac (v.1.17.1) FeatureMatrix function to generate a cell-by-peak count matrix for all cells within each cell type. This matrix was subsequently aggregated at a sample pseudobulk level using rollup function from the R slam package (v.0.1-56). For RASQUAL's allele-specific modelling, createASVCF.sh script in paired-read mode was used to generate allele-specific fragment counts at each site with additional BAM flag of -F 1280 to exclude secondary or PCR duplicate reads. For each cell type, only samples with at least 20 cells of that cell type and only peaks with at least five reads on average across all samples were included for the final analysis.

For the caQTL association analysis, all variants within a ±10-kb *cis*-window of the peaks were tested for association. Total library size for each sample was included as an offset. Age, sex, sequencing site and first four principal components of genotype data (generated from plink --pca) were included as covariates. For multiple testing correction, we used the two-step correction procedure recommended by RASQUAL[76], including locus-level correction followed by a genome-wide correction using an empirical null distribution generated using genotype permutations. The caQTL variants were queried subsequently in bulk eQTL datasets using Qtlizer (v.1.22.0)[77]. Power, MAF-stratified validation, model comparison and genotype-PC sensitivity analyses are described in Supplementary Fig. 2. For locus-level multiple testing, a Q value corresponding to SNP-level FDR was calculated using the Benjamini–Hochberg method, and the SNP with the lowest Q value was selected for each locus. For genome-wide correction, an empirical null distribution was calculated from four RASQUAL runs using randomly permuted genotypes; the averaged null distribution was used to obtain Q value cut-offs corresponding to 1%, 5% and 10% genome-wide FDR.

We compared the myeloid cell caQTLs with previously published monocyte caQTLs discovered in peripheral blood mononuclear cells from 15 healthy donors. We found that 26.7% (767 of 2,868) of the lead macrophage caQTL variants were also the lead variant among the previously identified 3,528 monocyte caQTLs. Although only a minority, this difference may reflect the profoundly distinct biological environments between peripheral blood mononuclear cells and myeloid cells in atherosclerotic lesions[75].

Pseudobulk caQTLs for the main cell types can be found in Supplementary Tables 15–26. To assess transcription factor binding potential at caQTLs, we trained seq2PRINT[22] in SMCs using chromatin accessibility and ATAC footprinting to predict transcription factor binding probability at base-pair resolution and compared caQTL variants with non-caQTL *cis*-SNPs within accessible peaks.

FMC module score

By mapping our multi-ome dataset onto a recently published coronary artery CITE-seq dataset[18], we demonstrated concordant mapping scores across SMC cell states, and projecting the FMC module score into this independent dataset confirmed the score was most enriched in FMCs (Extended Data Fig. 5b–d). In human spatial transcriptomics data, the FMC score showed greater enrichment in diseased relative to healthy coronary vessels (Extended Data Fig. 5e), and in carotid plaques the FMC score correlated positively with percent stenosis (Spearman $\rho = 0.58$, $P = 0.019$), linking FMC identity directly to disease burden (Extended Data Fig. 5f). Pseudotime trajectory analysis demonstrated progressive enrichment of the FMC score as SMCs differentiate, recapitulated by both our module and an independently published gene set (Extended Data Fig. 5g,h). Using an in vitro serum starvation and replenishment model of phenotypic switching in HCASMCs followed by scRNA-seq, we confirmed that the FMC score is elevated significantly in the stressed state (Extended Data Fig. 5i). Consistent with this, published time-course murine atherosclerosis data in SMC lineage-traced cells found progressive expansion of the FMC module with disease progression (Extended Data Fig. 5j). Finally, the FMC score was correlated positively with ChromVAR-predicted activity[78] of established FMC transcriptional regulators including the MEF2 family, linking chromatin remodelling to transcriptional dynamics along this differentiation axis (Extended Data Fig. 5k). Collectively, these analyses establish the reproducibility of the FMC module across independent human datasets, its association with clinical disease burden, and its relevance to both in vitro and in vivo models of SMC phenotypic switching.

sc-caQTL discovery

In addition to the pseudobulk caQTL model, we used a PME single-cell model to map caQTLs in SMCs. This approach models cell-level fragment counts with random effect covariates to account for relatedness shared by cells from the same sample and batch. Such attempts to model discrete count distributions at a single-cell level have been shown to generally boost power in single-cell QTL discovery compared with pseudobulk and/or linear regression frameworks[79–81]. Our PME model was adapted from the PME model used by Nathan and colleagues[3] for single-cell eQTL mapping as follows:

$$\begin{aligned}\log[A] = {} & \beta_G X_G + \beta_{\text{age}} X_{\text{age}} + \beta_{\text{sex}} X_{\text{sex}} + \beta_{n\text{Frags}} \log[X_{n\text{Frags}}] \\ & + \beta_{\text{TSSEnrich}} X_{\text{TSSEnrich}} + \sum_{k=1}^{4} \beta_{\text{accPCk}} X_{\text{accPCk}} \\ & + \sum_{k=1}^{4} \beta_{\text{genoPCk}} X_{\text{genoPCk}} + (\varphi_d | d) + (\epsilon_s | s) + \theta\end{aligned}$$

where ATAC fragment counts (A) was modelled with fixed-effect covariates (G, genotype; nFrags, number of fragments per cell, TSSEnrich, TSS enrichment per cell; accPC, principal components of single-nucleus ATAC counts; genoPC, principal components of genotype data), random effect covariates (d, donor ID; s, sequencing site) and intercept (θ). All quantitative covariates were scaled and centred before the regression analysis. Significance of the genotype effect term (β_G) was tested by LRT of the full model containing the genotype term and a null model

without the genotype term. Due to the high computational load of running Poisson regression analysis on all cells, this analysis was restricted to lead caQTL variants from the pseudobulk RASQUAL analysis, which models the fragment counts using a negative binomial model. All Poisson models were fit using the glmer function from the lme4 R package (v.2.0-6). We confirmed that our data is generally not over-dispersed by plotting the mean variance relationship between a random subset of peaks. Similarly, using a zero-inflated Poisson regression model or a negative binomial model on a subset of our hits revealed little differences in overall regression estimates. Given that a small proportion of our pseudobulk hits probably represent null results, we did not conduct FDR correction due to our inability to accurately estimate the proportion of results from the null distribution. Instead, we compared *z* values between our pseudobulk and PME models.

A key benefit of modelling individual cells instead of sample-level pseudobulk is our ability to simultaneously model cell-state covariates (for example, differentiation trajectories, gene and/or protein expression) as well as potential interaction between genotypes and cell-state covariates. We defined FMC identity score by calculating the gene activity of the top 100 FMC (modulated SMC) marker genes defined by Wirka and colleagues[16] from our integrated snATAC–seq meta-map using ArchR (v.1.0.3) addGeneScoreMatrix() and Seurat (v.5.5.1) AddModuleScore(). To test for 'dynamic' caQTL where the genotype effect is modified by SMC cell state, we then added the FMC score and a genotype-FMC score interaction term as covariates in the following model:

$$\begin{aligned}\log[A] = {} & \beta_G X_G + \beta_{\text{age}} X_{\text{age}} + \beta_{\text{sex}} X_{\text{sex}} + \beta_{n\text{Frags}} \log[X_{n\text{Frags}}] \\ & + \beta_{\text{TSSEnrich}} X_{\text{TSSEnrich}} + \sum_{k=1}^{4} \beta_{\text{accPCk}} X_{\text{accPCk}} \\ & + \sum_{k=1}^{4} \beta_{\text{genoPCk}} X_{\text{genoPCk}} + (\varphi_d|\mathrm{d}) + (\epsilon_s|s) \\ & + \beta_{\text{FMC}} X_{\text{FMC}} + \beta_{\text{FMCxG}} X_{\text{FMCxG}} + \theta\end{aligned}$$

Again, this analysis was carried out just on the significant (FDR < 5%) lead variants from the pseudobulk caQTL analyses, although it is entirely possible that there are non-significant pseudobulk caQTLs (that is, genotype effect is near zero for average FMC score) that may have significant interaction terms. We obtained *P* values for the genotype-FMC score interaction term using an LRT. We subsequently applied Storey *Q* value correction to LRT *P* values of the interaction term across all tested variants for multiple-testing correction. Dynamic caQTLs can be found in Supplementary Tables 29 and 30.

Electrophoretic mobility shift assay

For EMSA, reference allele probe (5′-atagttcagaaacctgactcagccccttctt-3′) and alternate allele probe (5′- atagttcagaaaccttactcagccccttctt-3′) for rs658956 were synthesized and labelled with IRDye700 (Integrated DNA Technologies). The labelled (100 nM) and competitor (5 μM) probes were incubated with 5–10 μg of SMC nuclear extract in binding buffer (final concentration 8% Ficoll, 160 mM KCl, 20 mM HEPES, 10 mM $MgCl_2$, 2.5 mM dithiothreitol, 0.25% Tween-20) and 1 μg of Poly dI*dC competitor at 4 °C for 1 h. For supershift assays, the protein complex was also incubated with anti-cJun (Santa Cruz Biotechnology, sc-74543), anti-cFos (Santa Cruz Biotechnology, sc-166940) or anti-JunB (Santa Cruz Biotechnology, sc-8051) before addition of the probe. Following incubation, the reaction mixture was run on a 5% TBE native gel in 4 °C for 30 min to 1 h. The gel was subsequently imaged using a Bio-Rad ChemiDoc Gel Imager.

CRISPRi for key CAD loci

The most effective gRNA sequence targeting an enhancer was cloned individually into the pBA904 vector and transfected into CRISPRi-HCASMC. After 5 days, RNA was collected, extracted and converted to cDNA. The effect of enhancer targeting was then assessed by comparing the expression levels of the target genes *LOXL1*, *MYO9B*, *NAV1* and *PALLD* with that of the control. RNA was isolated using RNeasy plus micro kit (Qiagen, 74034) and total cDNA was prepared using High-capacity RNA-to-cDNA kit (Life Technologies, 4388950).

HCASMC preparation for scRNA-seq

Parental and dCas9-Zim3 SMCs were cultured for 3 days in full medium (SmGM-2 Smooth Muscle Cell Growth Medium; Lonza, CC-3182) followed by 3 days serum starvation (base medium only) and 3 days recovery with full medium. dCas9-Zim3 SMCs were also cultured under two additional conditions for profiling: (1) 3 days full medium and (2) 3 days full medium followed by 3 days serum starvation. Cells were then processed using the Single Cell 3′ v.4 OCM kit (10x Genomics, PN-1000779) according to the manufacturer's instructions. The library was sequenced on an Illumina Nextseq2000 with read lengths (Read 1: 28, Read 2: 90, Index 1: 10, Index 2: 10). Sequencing data were demultiplexed and aligned to the human genome (GRCh38-2024-A from 10x Genomics) using CellRanger (v.9.0.1).

In vitro HCASMC scRNA-seq analysis

Library preparation was conducted with Chromium GEM-X Single Cell 3′ Reagent Kits v.4 (10x Genomics). Libraries were sequenced on NovaSeq X Plus. The initial 10x scRNA-seq dataset was processed using the CellRanger pipeline (v.8.0.0). The workflow included demultiplexing the bcl files with CellRanger mkfastq, followed by running CellRanger count using the pre-built CellRanger reference GRCh38-2020-A from 10x Genomics.

Arrayed Perturb–seq pipeline

Lentivirus carrying gRNAs were produced by VectorBuilder by modifying a CROPseq vector to include a T7–U6 hybrid promoter as described previously[82]. Four guides were designed per TSS, three guides per enhancer region and two non-targeting controls. gRNA lentiviruses were pooled for each enhancer or TSS at a 1:1 ratio based on titre for knockdown. Cells were transfected at a multiplicity of infection of 5 and left to recover for 3 days. Puromycin (0.5 μg ml^{-1}) was used to select for transfected cells for 4 days. Cells then underwent the 3-day starvation and 3-day recovery period as described above during which growth was monitored in an Incucyte S3 incubator and images analysed using Incucyte 2020B software. Cells were then used for single-cell transcriptomic analysis as described below.

A total of 100,000 cells from each condition were fixed with the Low Input Cell Fixation v.3 kit (Parse Biosciences, ECLC3303) according to the manufacturer's instructions. Cells were processed with the Evercode WT v.3 kit (Parse Biosciences, ECWT3300). Eight sub-libraries of 12,500 cells were generated according to manufacturer's instructions and sequenced on an Illumina Nextseq2000 with read lengths (Read 1, 64; Read 2, 58; Index 1: 8; Index 2, 8). Sequencing data were demultiplexed and aligned to the human genome (GRCh38 Parse Biosciences) using split-pipe (v.1.6.3).

Analysis of Perturb–seq data

Data loading and quality control. Raw count matrices generated by Parse Biosciences split-pipe demultiplexing were imported into R using the ReadMtx function from Seurat v.4, loading the count matrix, cell metadata and gene annotation files from the DGE_filtered directory. A Seurat object was created requiring genes detected in at least three cells and a minimum of 200 features per cell. Mitochondrial content was calculated using PercentageFeatureSet. Cells were filtered to retain those with 200–7,500 detected genes and less than 25% mitochondrial reads. Untransduced cells and two genomic coordinate-labelled perturbation conditions were excluded. Perturbation identities were mapped from sample barcodes to gene or enhancer labels, and cells were classified

as non-targeting or perturbed. Gene-expression data were normalized using Seurat NormalizeData, variable features identified with FindVariableFeatures and expression values scaled with ScaleData. Principal component analysis (PCA) was performed on the RNA assay, followed by UMAP embedding using the top 30 principal components.

Perturbation signature calculation, Mixscape perturbation classification and linear discriminant analysis. Local perturbation signatures were calculated using CalcPerturbSig from Seurat, subtracting the average expression of the $k = 20$ nearest non-targeting neighbour cells in PCA space (30 dimensions) to remove technical variation, producing a perturbation assay. The perturbation assay was centred, projected into PCA space using variable features from the RNA assay and visualized by UMAP. Mixscape[83] was run on the scaled perturbation assay using RunMixscape with a minimum of five differentially expressed genes, ten iterations and perturbation type set to knockdown. This classified cells as knockdown, escaped or non-targeting based on posterior probability of perturbation. Escaped cells were excluded from all downstream analyses. Knockdown classification rates were computed per perturbation as the proportion of cells assigned knockdown status. Linkage disequilibrium analysis was performed on knockdown and non-targeting control cells using MixscapeLDA on the PRTB assay (ten principal components, $\log_2$[fold change] threshold 0.25) and a UMAP was computed on the top linear discriminant analysis dimensions for visualization of perturbation-specific clustering.

Pseudobulk differential expression. Differential expression between KD and non-targeting control cells was performed per perturbation using DESeq2 (v.1.52.0). Cells were assigned randomly to three pseudo-replicates within each perturbation and non-targeting control group, and raw counts were aggregated. Genes with fewer than ten total counts across pseudo-replicates were excluded. Differential expression was tested using a negative binomial model with perturbation condition as the independent variable. Results were filtered at FDR < 0.05.

Gene set enrichment analysis. Pre-ranked GSEA was performed per perturbation using fgsea (v.1.38.0) with genes ranked by $\log_2$[fold change] from DESeq2. Gene sets from MSigDB Hallmark and Gene Ontology Biological Process collections were obtained using the msigdbr (v.26.1.1) package (*Homo sapiens*). Minimum and maximum gene set sizes were set to 10 and 500, respectively, with 1,000 permutations. *P* values were adjusted using the Benjamini–Hochberg method across perturbations.

Cross-perturbation transcriptional similarity. Pairwise cosine similarity was computed between the $\log_2$[fold change] vectors of all perturbations across the union of tested genes. Missing values were replaced with zero. Perturbations were clustered hierarchically on the resulting similarity matrix and visualized as a heatmap using pheatmap (v.1.0.13).

Consensus non-negative matrix factorization. cNMF[26,84] was applied to the raw count matrix of Mixscape-classified knockdown and non-targeting control cells. Counts were exported as an AnnData h5ad file and processed using the cNMF Python package (v.1.7.1). Genes expressed in fewer than five cells and all-zero genes were removed before factorization. Factorization was performed over $k = 5$–10 components with 100 iterations per k value and random seed 42. The optimal $k = 8$ was selected based on stability and reconstruction error, with a local-density threshold of 0.01 applied during consensus clustering; 99 of 800 (12%) outlier spectra were removed. Six of eight programmes were retained based on coherent gene set enrichment; two programmes dominated by unannotated loci and lacking enrichment were excluded. cNMF usage scores were imported back into R and added to the Seurat object metadata. Programme identity was assigned by over-representation analysis of the top 100 programme-specific genes against MSigDB Hallmark and Gene Ontology Biological Process gene sets using enricher from clusterProfiler (v.4.20.0).

Transcription factor motif enrichment and transcription factor activity estimation. Transcription factor motif enrichment per cNMF programme was assessed by over-representation analysis of top 100 programme genes against Gene Transcription Regulation Database transcription factor target-gene sets from MSigDB C3. Transcription factor activity was estimated using the univariate linear model method implemented in decoupleR (v.2.17.0), with the DoRothEA regulon[85] (confidence levels A–C) as the prior network. cNMF gene specificity score vectors (genes × programmes) were used as input. Transcription factors with the highest variance across programmes were retained for visualization.

FMC and stimulation module scores. FMC module scores were computed using AddModuleScore in Seurat on a curated list of FMC marker genes as described above. Stimulated SMC up- and down-regulated gene sets were derived from FindAllMarkers differential expression between stimulated (Zim3_3d3d) and control (Zim3_ctrl) conditions, filtered at FDR < 0.05 and $|\log_2$[fold change]$| > 0.25$. A signed stimulation score was computed as the difference between up- and down-regulated module scores. Spearman correlations between module scores and cNMF programme usage were computed using cor.test. Per perturbation mean scores were compared using the Wilcoxon rank sum test with Benjamini–Hochberg correction. The Spearman correlation between each perturbation's full $\log_2$[fold change] profile and the signed stimulated SMC $\log_2$[fold change] vector was computed across shared genes.

Convergent gene programme identification. RRA[86] was applied across all perturbation DESeq2 gene rankings using the RobustRankAggreg (v.1.2.1) R package, separately for upregulated and down-regulated genes ranked by adjusted *P* value. Genes passing FDR < 0.05 after Benjamini–Hochberg correction of RRA scores were retained as the convergent programme. Each gene was assigned a direction based on the RRA result with the lower score. The top 30 convergent genes by number of perturbations in which they reached significance were selected for visualization. A convergent programme score was computed per cell as the difference between up and down module scores using AddModuleScore. Pathway enrichment of convergent up and down-regulated genes was performed using enricher against MSigDB Hallmark and GO Biological Process gene sets.

Coronary artery Hi-C processing

Bulk in situ Hi-C libraries for 15 different coronary artery samples across a multi-ethnic cohort were generated with the Arima Hi-C kit, according to the manufacturer's protocols. Specifically, one Hi-C library was created from each biological sample, with an input of approximately 800,000 cells each. Restriction enzymes cutting at ^GATC and G^ANTC were used in library construction, and approximately 600 million reads were sequenced on an Illumina HiSeq 4000 per library.

Hi-C reads were processed with the nf-core Hi-C pipeline (v.2.0.0)[87] and downstream maps were analysed at 5-kb resolution. Resolution assessment, library quality metrics, loop-caller benchmarking, inherent normalization and visualization procedures[88,89] are described in Supplementary Figs. 5–7. Phased contact matrices using sample-specific genotypes were processed using Axiotl's Turnkey platform (tinker.axiotl.com), which leverages Hi-C-Pro v.3.1.0 and cloud compute.

Hi-C mosaic analysis

We called cohort-specific 3D structures using the AQuA Tools framework[90] for 3D genomics. Sample-wise chromatin structures were called in BEDPE format using extract_bedpe. We used union_bedpe to stack and merge BEDPE coordinates coming from various samples into one BEDPE file that contained the union set coordinates of all possible Hi-C

connections in the cohort. This allowed for resolving overlapping interactions and outputting a common scaffold. To assign membership (as unique identifiers to column 7) to the stacked loops based on shared overlap of BEDPE anchors, we used cluster_bedpe. This resulting file became the mosaic for the cohort. To assess their functional and disease relevance, we intersected two-dimensional mosaic coordinates with one-dimensional genomic features including genes and snATAC peaks harbouring CAD GWAS variants. Against the mosaic BEDPE, contact values for each sample were queried in inherent and counts per million (CPM) numeric spaces using query_bedpe. All downstream analyses were performed in R.

CTCF ChIP experiments

All experiments were performed on primary HCASMC between passages 4 and 7. Antibodies used for ChIP–qPCR were all pre-validated according to ChIP–seq guidelines. Purified rabbit monoclonal antibody against human CTCF (Abcam, ab128873; clone EPR7314(B)). Cell lines with various genotypes of rs4887091 were used. Approximately 2×10^6 HCASMC for each line were fixed with 1% formaldehyde and quenched by glycine. The cells were washed three times with PBS and then harvested in ChIP lysis buffer (50 mM Tris-HCl, pH 8, 5 mM EDTA, 0.5% SDS). Crosslinked chromatin was sheared for 2 × 1 min by sonication (Branson SFX250 Sonifier) before extensive centrifugation. Four volumes of ChIP dilution buffer (20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) was added to the supernatant. The resulting lysate was then incubated with Dynabeads Protein G (Thermo Scientific, 10009D) and antibodies at 4 °C overnight. Beads were washed once with buffer 1 (20 mM Tris pH 8, 2 mM EDTA, 150 mM NaCl, 1% Triton X-100, 0.1% SDS), once with buffer 2 (10 mM Tris pH 8, 1 mM EDTA, 500 mM NaCl, 1% Triton X-100, 0.1% SDS), once with buffer 3 (10 mM Tris pH 8, 1 mM EDTA, 250 mM LiCl, 1% NP40, 1% sodium deoxycholate monohydrate) and twice with TE buffer. DNA was eluted by ChIP elution buffer (0.1 M $NaHCO_3$, 1% SDS, 20 μg ml^{-1} proteinase K). The elution was incubated at 65 °C overnight, and DNA was extracted with a DNA purification kit (Zymo, D4013). The purified DNA from various lines were analysed by qPCR using the ABI ViiA 7, with the same primer applied to each SNP. Assays were repeated at least three times. Data are average values ± s.d. of representative experiments.

Enrichment of CAD variants in CTCF binding sites

CAD-associated variants were represented as 1-bp genomic intervals (hg38) and intersected with HCASMC CTCF ChIP–seq peak intervals. Observed overlap was defined as the number of CAD variants overlapping at least one CTCF peak. To estimate overlap expected by chance while preserving genomic structure, we performed chromosome-wise circular permutation of CTCF peak coordinates. For each chromosome, peak start coordinates were shifted by a random offset uniformly sampled from $[1, L-1]$, where L is chromosome length, with wraparound at chromosome boundaries. Peak widths and chromosome assignments were preserved. For each of B = 1,000 permutations, we recalculated the number of CAD variants overlapping permuted CTCF peaks to generate an empirical null distribution.

Fold enrichment was calculated as

$$\text{Fold} = \frac{\text{Observed overlap fraction}}{\text{Null mean overlap fraction}}$$

Empirical one-sided P values were computed as

$$P = \frac{(\text{null} \geq \text{observed}) + 1}{B + 1}$$

For parametric approximation, we set p_0 to the null mean overlap probability (mean(null overlaps)/N) and computed the exact one-sided binomial probability of observing ≥k overlaps among N variants.

CTCF allele-specific binding quantification

Enriched and input DNA were quantified by SYBR Green qPCR on ViiA 7. In addition to SYBR Green primers, we also used TaqMan genotyping probe (ThermoFisher, AN7EEGX, lot no. P260615-009) to directly amplify the different alleles of the DNA in heterozygous lines. Relative CTCF enrichment for target (rs4887091) and control loci was calculated as percent input. Target allele-specific binding effects were confirmed by benchmarking enrichment profiles against the expected quantitative gradients of the in-house validated allelic controls.

pos_rs2493274_F3: CCAAGCACCCACCATTCTG
pos_rs2493274_R3: ATCTGGGAATGCTGGCAGT
pos_rs56140443_F2: TGTCCTCCCTTATCCAGCTT
pos_rs56140443_R2: TAATTGCTGTGTGGGTGAGC
pos_rs78444094_F3: CCCACATCCATCTCCTTTGC
pos_rs78444094_R3: AATGAAGGTGGAGGTGCTGA
neg_CTCF+rs9850507_F3: GGGAATCAAGCTGTGTTTGC
neg_CTCF+rs9850507_R3: CAGGGCAGAGAAGTTGTTGG
neg_CTCF+rs6825404_F1: ACATTACTTGGGCAGGTGGA
neg_CTCF+rs6825404_R1: CCCCTAAACATCCTCGTTGAT
neg_CTCF+rs566243_F3: TGGGACCGGATAGTTCTTTGT
neg_CTCF+rs566243_R3: GACTAGGTAGAGGCAGGGGT.

TaqMan genotyping probe:
GGCCGACCTGGCCCCTCCCACACTGCTGGGCCTTGCATGGTCCTGCAGCGCC[T/C]CCTTGTGAGCGGCTCGCGGTACGGTCTCCCCTGCAGCTTGCTTTAGAAAGCGG.

CTCF binding controls

In silico screening of allelic control SNPs. To identify endogenous allelic controls across the four cell lines, CTCF ChIP–seq peaks were intersected with corresponding whole-genome sequencing VCF files. SNPs exhibiting complete genotypic gradients (homozygous reference, heterozygous, homozygous alternative) were scored for CTCF motif perturbation using the JASPAR matrix (MA0139.2). Allelic positive controls were defined by severe motif disruption (ΔScore ≥ 6.0, max score >8.0), whereas negative controls exhibited preserved motif integrity (ΔScore ≤ 0.2, both scores >8.0).

Primer design and polymorphism masking. To prevent cross-cell-line amplification bias, SYBR Green qPCR primers (75–150 bp amplicons) were designed using Primer3 v.2.6.1. An in silico masking pipeline strictly discarded any candidate primer pairs whose binding sites overlapped with endogenous variants in the merged four-cell-line VCF, ensuring 100% conserved annealing kinetics across all biological samples.

H3K27ac HCASMC and HCAEC Hi-ChIP

Hi-ChIP was adapted from previously published protocols with some modifications[91,92]. Briefly, at least 1 million healthy HCASMCs (Thermo Scientific, C-017-5C, lot no. 1689414) and healthy HCAECs (ATCC, PCS-100-020, lot no. 62382179) were crosslinked with 1% formaldehyde. The initial Hi-C portion followed the Arima Hi-C protocol described in the Arima High Coverage Hi-C Kit (Arima Genomics, A410110). For the immunoprecipitation step, sheared chromatin after the Hi-C portion of the protocol was immunoprecipitated with 5 μg of H3K27ac antibody (Abcam, ab4729) with 50 μl of protein G Dynabeads (Invitrogen 10004D) and allowed to incubate overnight at 4 °C. The protein G Dynabeads were washed and the immunoprecipitated DNA was de-crosslinked at 67 °C for 2 h and eluted (50 mM sodium bicarbonate, 1% SDS). Subsequently, Dynabeads MyOne Streptavidin C1 beads (Invitrogen, 65002) were used to enrich ligated junctions as per Hi-C protocol. Library preparation was performed using the NEB Ultra II library preparation kit (NEB, E7645L), according to the manufacturer's protocol. Twelve PCR cycles were performed using indexed primers and subsequently DNA fragments between 300 bp and 500 bp were size selected using

the Omega MagBind NGS clean up Magnetic beads (Omega, M1378). The libraries were sequenced by paired-end sequencing with at least 300 million read pairs at 2 × 151-bp read length on Illumina HiSeq 4000 platforms. Raw paired-end reads in Fastq format for respective samples were processed using the Hi-CuP pipeline[93] (v.0.7.4). The hg38 human genome was used as reference to generate the Hi-CuP Digest file using two restriction motifs GATC and GANTC using the hicup_digester command for running the Hi-CuP pipeline. Paired contacts were extracted from aligned filtered bam files using Samtools (v.1.13) and used as input for the pre function in the Juicer package to generate hic files[94]. Paired interaction counts at 5 kb resolution were extracted from the matrix, and high-confidence chromatin loops were identified using the FitHiChIP tool (v.11.0) as described previously using default parameters[95]. Using HCASMC and HCAEC ChIP–seq peak files as input, and initial paired interaction files, high-confidence peak-to-all interactions with a 5-kb bin size were processed using the bash FitHiChIP_HiCPro.sh script (https://ay-lab.github.io/FitHiChIP/). High-confidence chromatin interactions were sorted and indexed in bed format and converted to long-range format for visualization on the WashU browser. The overlay of Hi-ChIP, ChIP–seq and GWAS data was plotted using the locuszoomr package (v.0.3.5) and R (v.4.3.1).

H3K27ac ChIP–seq

ChIP–seq was performed as described previously with minor optimizations[96]. Briefly, 1 million HCASMCs and HCAECs from the same lot used for Hi-ChIP experiments were crosslinked with 1% formaldehyde for 10 min at room temperature and quenched with glycine (125 mM) for 5 min. The cells were rinsed twice with cold 1× PBS and sonicated in nuclei lysis buffer (50 mM HEPES–KOH, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X, 0.1% sodium deoxycholate, 1% SDS, Takara, 1× protease inhibitor) using the Bioruptor sonicator to obtain chromatin fragments between 200 bp and 500 bp. The fragmented chromatin was immunoprecipitated overnight at 4 °C, with 5 µg of H3K27ac antibody (Abcam, ab4729) and 50 µl of protein G beads (Invitrogen, 10004D). Subsequently, the beads were washed and immunoprecipitated DNA was de-crosslinked and eluted in elution buffer (50 mM Tris-HCl, pH 7.5, and 10 mM EDTA). Eluted ChIP fragments were isolated using phenol–chloroform extraction method, and DNA was purified by ethanol precipitation. Library preparation was performed on the eluted ChIP DNA using the NEB Ultra II library preparation kit, according to the manufacturer's protocol. Ten PCR cycles were performed using indexed primers and the library of DNA fragments with sizes between 300 bp and 500 bp was selected using the Omega MagBind NGS clean up Magnetic beads (Omega, M1378). The libraries were sent for paired-end sequencing, with at least 50 million read pairs at 2 × 151-bp read length on the Illumina HiSeq 4000 platform. Paired-end sequencing reads were aligned to human genome (hg38) using BWA mem v.0.7.5 (ref. 97), and reads were deduplicated using Picard v.3.4.0 MarkDuplicates. Aligned mapped reads in BAM file were used as input for peak-calling using Dfilter[98] to identify significant H3K27ac peaks using the following setting: -ks = 60, -bs = 100, -lpvalue = 8. Bigwig tracks were normalized for sequencing depths and the sum of normalized binned tag-count represents the peak height.

Culture of HCAEC and HCASMC primary cells

Healthy HCASMCs (Thermo Scientific, C-017-5C, lot no. 1689414) were cultured in human vascular smooth muscle basal medium (Gibco, M231500), in Smooth Muscle Growth Supplement (Gibco, S00725) and 1% penicillin–streptomycin. Healthy HCAECs (ATCC, PCS-100-020, lot no. 62382179) were cultured in Endothelial Cell Growth medium MV2 (PromoCell, C-22022).

Construction of lentiviral plasmids and virus production

The Lentiguide-Puro plasmid was obtained as a gift from F. Zhang (Addgene, 52963, LentiGuide-Puro) in which the non-targeting control sgRNAs and sgRNAs targeting the TSS and regulatory enhancer were annealed and ligated into the *Esp*3I cut-site of the Lentiguide-Puro vector using T4 Ligase (NEB, M0202) as per the manufacturer's instructions. Ligation products were transformed into One Shot Stbl3 Chemically competent *Escherichia coli* (Invitrogen, C737303) and plasmids of positive clones were isolated using FavorPrep Plasmid DNA Extract Kit (Favorgen, FAPDE001-1). The CRISPR interference expression vector for Lenti-dCas9-KRAB-blast (Addgene, 89567) was a gift from G. Hon. For lentivirus production, HEK293T cells were cultured in DMEM + 10% fetal bovine serum (FBS) until 70% confluency before transfection. Then, 10 µg of individual plasmids, 7.5 µg of pMDLg/RRE, 2.5 µg of pRSV-Rev and 2.5 µg of pMD2.G lentivirus packaging plasmids were co-transfected with 50 µl of polyethylenimine diluted in 3 ml of Opti-MEM I Reduced Serum Medium (ThermoFisher Scientific, 31985070). Medium was refreshed overnight with reduced 5% FBS in DMEM medium, and the expended medium was collected twice after 24 h and 48 h, respectively. Pooled supernatant was filtered through a polyethersulfone 0.45 µm filter and viral particles were concentrated using Viro-PEG Lentivirus Concentrator (Ozbiosciences, LVG100) as per the manufacturer's instructions.

The sgRNAs were as follows:

Non-targeting control_sgRNA_F: 5′-CACCGAACGTGCTGACGATGCGGGC-3′

Non-targeting control_sgRNA_R: 5′-AAACGCCCGCATCGTCAGCACGTTC-3′

AMOTL2_TSS_sgRNA_F: 5′-CACCGGCGCGAACAGCCAGAGCGT-3′

AMOTL2_TSS_sgRNA_R: 5′- AAACACGCTCTGGCTGTTCGCGCC-3′

AMOTL2_Enhancer_sgRNA1_F: 5′- CACCGTATTCATAGACATCACTAA-3′

AMOTL2_Enhancer_sgRNA1_R: 5′- AAACTTAGTGATGTCTATGAATAC-3′

AMOTL2_Enhancer_sgRNA2_F: 5′- CACCGATCCCTATGGAATCCTTGG-3′

AMOTL2_Enhancer_sgRNA2_R: 5′- AAACCCAAGGATTCCATAGGGATC-3′.

Primary cell transduction for CRISPR interference

At least two independent sgRNAs targeting a single regulatory element were tested against the non-targeting control. sgRNA targeting the TSS was used as positive control. Healthy HCASMCs (Thermo Scientific, C-017-5C, lot no. 1689414) and HCAECs (ATCC, PCS-100-020, lot no. 62382179) were both co-transduced in 5 µg ml^{-1} polybrene with lentiviral particles packaged from Lenti-dCas9-KRAB-blast (Addgene, 89567)[99], and individual sgRNAs were cloned into Lentiguide-Puro backbone (Addgene, 52963)[100] at a multiplicity of infection of 3. Respective HCASMC and HCAEC growth media were refreshed after 24 h of infection, and co-selection with 10 µg ml^{-1} blasticidin (Gibco, A1113903) and 1 µg ml^{-1} puromycin (Gibco, A1113803) was performed 72 h following infection. Selected cells were allowed to recover and expanded for 2 weeks before RNA isolation using Trizol Reagent (Gibco, 15596026) and Direct-Zol RNA miniprep (Zymo Research, R2052) as per the manufacturer's instructions. cDNA conversion was performed with at least 200 ng RNA using HiScript III RT SuperMix (Vazyme Biotech, R323-01), and reverse transcription-PCR was performed using gene-specific primers to assess for gene expression.

The following reverse transcription-PCR primers were used:

AMOTL2_F: 5′-AGTGAGCGACAAACAGCAGACG-3′

AMOTL2_R: 5′-ATCTCTGCTCCCGTGTTTGGCA-3′

GAPDH_F: 5′-GTGGACCTGACCTGCCGTCT-3′

GAPDH_R: 5′-GGAGGAGTGGGTGTCGCTGT-3′

PPIA_F: 5′-CACCGTGTTCTTCGACATTG-3′

PPIA_R: 5′-TTCTGCTGTCTTTGGGACCT-3′.

Cell kinetic proliferation assay

Cell kinetic proliferation live-cell imaging was performed using the HoloMonitor M4 system (Phase Holographic Imaging, PHI AB) using

the Cell Kinetic Proliferation assay protocol as per the manufacturer's instructions. Non-targeting controls and *AMOTL2* dCas9-KRAB CRISPRi knockdown HCASMCs were seeded and cultured in vascular smooth muscle cell medium, in smooth muscle growth supplement, at a confluence of 15,000 cells cm^{-2} in a 24-well vessel (VWR, 734-2325) ($n = 8$). The 24-well HoloLid (PHI AB, 71130) was used to enhance the visualization and tracking of cells. At least five capture areas were taken per well, and time-lapse imaging at every 30-min time interval was taken across wells over 72 h. Cellular segmentation was performed using the HoloMonitor App Suite (v.3.5.2.446). Cellular confluence was captured and normalized to baseline at 0 h.

Bulk RNA-seq of HCASMCs

Total RNA was extracted using Trizol Reagent (ThermoFisher, 15596026) and Direct-Zol RNA miniprep Kit (Zymo Research, R2050). At least three independent replicates were used for RNA library preparations. RNA quality and yield was assessed using Agilent RNA 6000 pico kit (Agilent, 5067-1513) for quality. Bulk RNA-seq library preparations were prepared using VAHTS Universal V10 RNA-seq Library Prep Kit for Illumina (Vazyme, NR606-01) with Ribodepletion, according to the manufacturer's instructions. cDNA libraries were prepared using Illumina RNA Truseq indexes and quantified using Agilent Tapestation. Sequencing was performed on the NovaSeq X Plus (PE150) by NovogeneAIT Singapore, yielding 50 million raw pair-end reads. Raw Fastq files were trimmed using the trim-galore tool (v.0.65) from Babraham Bioinformatics to sequence length of 50. Paired-end reads were mapped and aligned to the human reference genome GRCh38–hg38 using STAR aligner (v.2.7.0) as per default settings and using gene annotations from GENCODE v.38. Aligned gene counts were mapped using HTSeq-count v.2.1.2 and normalized to TPM by sequencing depth and by gene length. Differential gene-expression analysis was performed using EdgeR (v.3.42.4). Contrasts were compared between non-targeting control and AMOTL2 TSS knockdown cells. A negative binomial generalized log-linear model was used to model gene counts using the glmFit function. |log[fold change]| >0.5 and FDR < 0.05 cut-offs were used to identify significantly differentially expressed genes following multiple testing correction. Figures were generated with R v.3.5.1.

Immunofluorescence staining and imaging

HCASMCs were fixed in 4% paraformaldehyde (ThermoFisher, 5090495) for 15 min at room temperature, then washed with 3× PBS for 5 min at room temperature. Cells were permeabilized with PBS with 0.1% Triton X-100 for 15 min, followed by blocking in 5% normal goat serum in PBS (blocking buffer) for 1 h at room temperature. Primary anti-Calponin (1:100) (ProteinTech, 13938-1-AP), and primary anti-α-smooth muscle actin antibody (1:100) (Sigma-Aldrich, A5228) were diluted in blocking buffer and added to permeabilized cells overnight at 4 °C. After washing with PBS three times for 5 min, cells were incubated with Alexa fluor-conjugated secondary antibodies (1:1,000) (Goat anti-mouse 594, ThermoFisher, A-11032; goat anti-rabbit 488, ThermoFisher, A-11008) and 1 μg ml^{-1} DAPI for 1 h at room temperature. Stained cells were washed three times for 5 min each with PBS, and images were acquired under the same laser exposure and magnification using a fluorescence microscope (EVOS M7000).

Reporting summary

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

Data availability

Raw and processed sequencing files can be found as follows: Human Coronary Multi-ome: GSE328800. Human Coronary Hi-C: GSE328970. Genotyping for Human Samples: dbGaP accession ID phs004118.v1.p1. HCASMC Perturb–seq: GSE347104. CTCF ChIP–seq in HCASMCs: GSE319778. H3K27ac/Hi-ChIP in HCASMC and HCAEC: GSE282556 and GSE282557. Bulk RNA-seq AMOTL2: GSE329169. Public and/or reference resources used in this study include GTEx v.8, the 1000 Genomes Project Phase 3 reference panel and the GRCh38–hg38 human reference genome. Previously published datasets used for validation or reference mapping include GSE131780, GSE175621 and GSE188422.

Code availability

Scripts used for analysis in this manuscript can be found at GitHub (https://github.com/jamrute/CAD-scV2E2G).

60. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Genet.* **53**, 403–411 (2021).
61. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
62. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
63. Gschwind, A. R. An encyclopedia of human enhancer–gene regulatory interactions. *Nature* **657**, 179–190 https://doi.org/10.1038/s41586-026-10781-4 (2026).
64. Finucane, H. K. et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).
65. Hinrichs, A. S. et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**, D590–D598 (2006).
66. Suzuki, K. et al. Genetic drivers of heterogeneity in type 2 diabetes pathophysiology. *Nature* **627**, 347–357 (2024).
67. Roychowdhury, T. et al. Genome-wide association meta-analysis identifies risk loci for abdominal aortic aneurysm and highlights PCSK9 as a therapeutic target. *Nat. Genet.* **55**, 1831–1842 (2023).
68. Keaton, J. M. et al. Genome-wide analysis in over 1 million individuals of European ancestry yields improved polygenic risk scores for blood pressure traits. *Nat. Genet.* **56**, 778–791 (2024).
69. Malik, R. et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat. Genet.* **50**, 524–537 (2018).
70. Murphy, A. E., Schilder, B. M. & Skene, N. G. MungeSumstats: a Bioconductor package for the standardization and quality control of many GWAS summary statistics. *Bioinformatics* **37**, 4593–4596 (2021).
71. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
72. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
73. Maples, B. K., Gravel, S., Kenny, E. E. & Bustamante, C. D. RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288 (2013).
74. Auton, A. et al. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
75. Benaglio, P. et al. Mapping genetic effects on cell type-specific chromatin accessibility and annotating complex immune trait variants using single nucleus ATAC-seq in peripheral blood. *PLoS Genet.* **19**, e1010759 (2023).
76. Kumasaka, N., Knights, A. J. & Gaffney, D. J. Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213 (2016).
77. Munz, M. et al. Qtlizer: comprehensive QTL annotation of GWAS results. *Sci Rep.* **10**, 20417 (2020).
78. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
79. Fitzgerald, T., Jones, A. & Engelhardt, B. E. A Poisson reduced-rank regression model for association mapping in sequencing data. *BMC Bioinf.* **23**, 529 (2022).
80. Zhou, W. et al. Efficient and accurate mixed model association tool for single-cell eQTL analysis. Preprint at *medRxiv* https://doi.org/10.1101/2024.05.15.24307317 (2024).
81. Alquicira-Hernandez, J. et al. Efficient genome-wide mapping of reproducible, context-dependent eQTLs at single-cell resolution. Preprint at *bioRxiv* https://doi.org/10.64898/2026.08.25.747138 (2026).
82. Kudo, T. et al. Multiplexed, image-based pooled screens in primary cells and tissues with PerturbView. *Nat. Biotechnol.* **43**, 1091–1100 (2025).
83. Papalexi, E. et al. Characterizing the molecular regulation of inhibitory immune checkpoints with multi-modal single-cell screens. *Nat. Genet.* **53**, 322 (2021).
84. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *eLife* **8**, e43803 (2019).
85. Garcia-Alonso, L., Holland, C. H., Ibrahim, M. M., Turei, D. & Saez-Rodriguez, J. Benchmark and integration of resources for the estimation of human transcription factor activities. *Genome Res.* **29**, 1363–1375 (2019).
86. Kolde, R., Laur, S., Adler, P. & Vilo, J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **28**, 573 (2012).
87. Ewels, P. A. et al. The nf-core framework for community-curated bioinformatics pipelines. *Nat. Biotechnol.* **38**, 276–278 (2020).
88. McArthur, E. & Capra, J. A. Topologically associating domain boundaries that are stable across diverse cell types are evolutionarily constrained and enriched for heritability. *Am. J. Hum. Genet.* **108**, 269–283 (2021).
89. An, L. et al. OnTAD: hierarchical domain structure reveals the divergence of activity among TADs and boundaries. *Genome Biol.* **20**, 282 (2019).

90. Chakraborty, M. et al. AQuA Tools: clear and reliable BEDPE operations for 3D genomics. *Bioinformatics* **41**, btaf510 (2025).
91. Mumbach, M. R. et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat. Methods* **13**, 919–922 (2016).
92. Anene-Nzelu, C. G. et al. Assigning distal genomic enhancers to cardiac disease-causing genes. *Circulation* **142**, 910–912 (2020).
93. Wingett, S. et al. HiCUP: Pipeline for mapping and processing Hi-C data. *F1000Res.* **4**, 1310 (2015).
94. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
95. Bhattacharyya, S., Chandra, V., Vijayanand, P. & Ay, F. Identification of significant chromatin contacts from HiChIP data by FitHiChIP. *Nat. Commun.* **10**, 4221 (2019).
96. Tan, W. L. W. et al. Epigenomes of human hearts reveal new genetic variants relevant for cardiac disease and phenotype. *Circ. Res.* **127**, 761–777 (2020).
97. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
98. Kumar, V. et al. Uniform, optimal signal processing of mapped deep-sequencing data. *Nat. Biotechnol.* **31**, 615–622 (2013).
99. Xie, S., Duan, J., Li, B., Zhou, P. & Hon, G. C. Multiplexed engineering and analysis of combinatorial enhancer activity in single cells. *Mol. Cell* **66**, 285–299 (2017).
100. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).

Acknowledgements We thank M. Ackers-Johnson from NUS Cardiovascular Research Institute for the provision of the primary HCAEC and HCASMC cell lines. We thank the Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine for help with genomic analysis.

Author contributions N.O.S. and I.M.H. conceived the study. J.M.A., T.Y. and A.B. isolated nuclei and performed 10x Multi-ome experiments on human coronary arteries. P.C.L., X.L., Y.Z., C.W.-R., C.L.M. and B.L. performed QTL analyses. M.U.S., W.-L.Q., A.T., R.A. and J.M.E. performed enhancer–gene linkage analyses. J.M.A., I.E., T.Y., M.C., R.G.G., K.J.L., Y.-H.H., R.S., C.-M.L., B.A. and N.O.S. designed and contributed to analysis of the Hi-C data. C.J.M.L., R.G., C.A.-N., W.Q.S. and R.S.-Y.F. performed in vitro Hi-ChIP, H3K27ac and CRISPR interference experiments. Q.Z., T.N., D.Y.L., M.R. and T.Q. performed in vitro CRISPR interference validation and CTCF experiments. P.C.L., H.C.P.P., Y.W., J.O., R.E.W. and A.A. performed EMSA experiments. J.M.A., T.Y., C.J.M.L., E.Y., W.-L.L., J.C., D.R.L., S.H., S.M. and S.J. designed, performed and contributed to the Perturb–seq experiments. J.M.A., P.C.L., C.J.M.L., R.S.W., S.K., D.K., C.K., P.C., I.M.H., R.M.G., S.M.H., R.K., P.S., H.Y.C., B.L., S.M., S.J., T.Q. and N.O.S. provided guidance on GWAS mapping, data interpretation and cell-specific analyses. C.J.K. analysed genotyping data. J.M.A., S.J. and N.O.S. supervised the work. J.M.A., P.C.L., T.Q. and N.O.S. drafted the manuscript with assistance from all authors. N.O.S. is responsible for all aspects of this manuscript including experimental design, data analysis and manuscript production. All authors edited and approved the final version of the manuscript.

Funding J.M.A. was supported by the American Heart Association Predoctoral Fellowship (826325) and by Leducq Foundation Network Seed Grant (grant no. 20CVD02). J.M.A. and P.C.L. were also supported by the Washington University School of Medicine Medical Scientist Training Program. This work was supported by National Institutes of Health (NIH) grant nos. UM1HG008853 (I.M.H. and N.O.S.), R01HG013371 (I.M.H. and N.O.S.), R01HL159171 (N.O.S.), R01HL171045 (T.Q.), R01HL134817 (T.Q.), R01HL139478 (T.Q.), R01HL156846 (T.Q.), R01HL151535 (T.Q.), R01HL158525 (T.Q.), UM1HG011972 (J.M.E. and T.Q.), U01HG011762 (T.Q.), R01HL159176 (J.M.E.), R01HL164811 (R.G. and J.M.E.), F30HL179947 (P.C.L.), Leducq Foundation 26CVD02 (T.Q. and N.O.S.), American Heart Association 25PRE137062 (P.C.L.), 23SCISA1144703 (P.C.), 24SCEFIA1248386 (P.C.), 69520CDA35310303 (P.C.) and the Novo Nordisk Foundation Center for Genomic Mechanisms of Disease (grant no. NNF21SA0072102). M.U.S. acknowledges the support of a National Science Foundation Graduate Research Fellowship (DGE-1656518) and a graduate fellowship award from Knight-Hennessy Scholars at Stanford University. N.O.S. was also supported in part by the Foundation for Barnes-Jewish Hospital. R.S.-Y.F. is funded by Individual Research Grants from the National Medical Research Council of Singapore (MOH-001480-00) and MOE Academic Research Fund (AcRF) Tier 3 (MOE-000333-00). B.L. is funded by MOE Academic Research Fund Tier 2 (MOE-T2EP30123-0015; MOE-T2EP30225-0016), by the National Research Foundation (NRF) Fellowship (NRF-NRFF18-2026-0006), by NRF AI-for-Science (AI4S) Challenge (NRF-AI4SCH-2025-0002), and by A*STAR under the Nucleic Acid Therapeutics Initiative (H24J5a0066). The study was partially supported by an Amgen sponsored research agreement. The Genome Technology Access Center at the McDonnell Genome Institute at Washington University School of Medicine is partially supported by NCI Cancer Center Support grant no. P30 CA91842 to the Siteman Cancer Center.

Competing Interests J.M.A., I.E., T.Y., E.Y., W.-L.L., J.C., R.W., D.R.L., S.M.H., P.S., H.Y.C., S.M., S.J., C.-M.L. and B.A. were or are employed by Amgen. J.M.A., W.Q.S, and R.S.Y-.F. are founders and shareholders of Viventia Therapeutics and Varios Bio. M.C., R.G.G. and R.S. were or are employed by Axiotl. S.M.H. was or is employed by Genentech. B.A. was or is employed by Merck. W.Q.S. was or is employed by Varios Bio and Viventia Therapeutics. All other authors declare no competing interests.

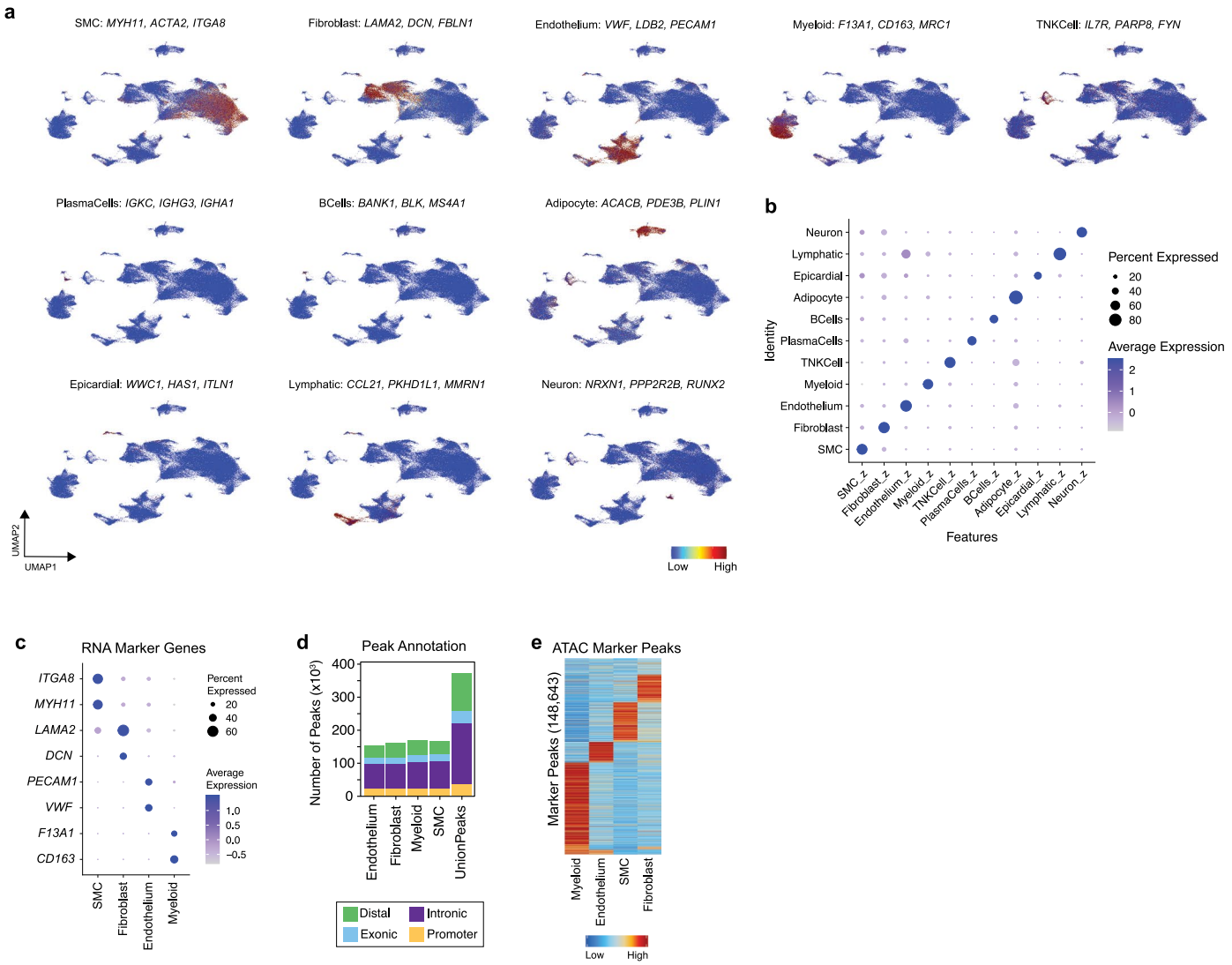
Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-026-11275-z.

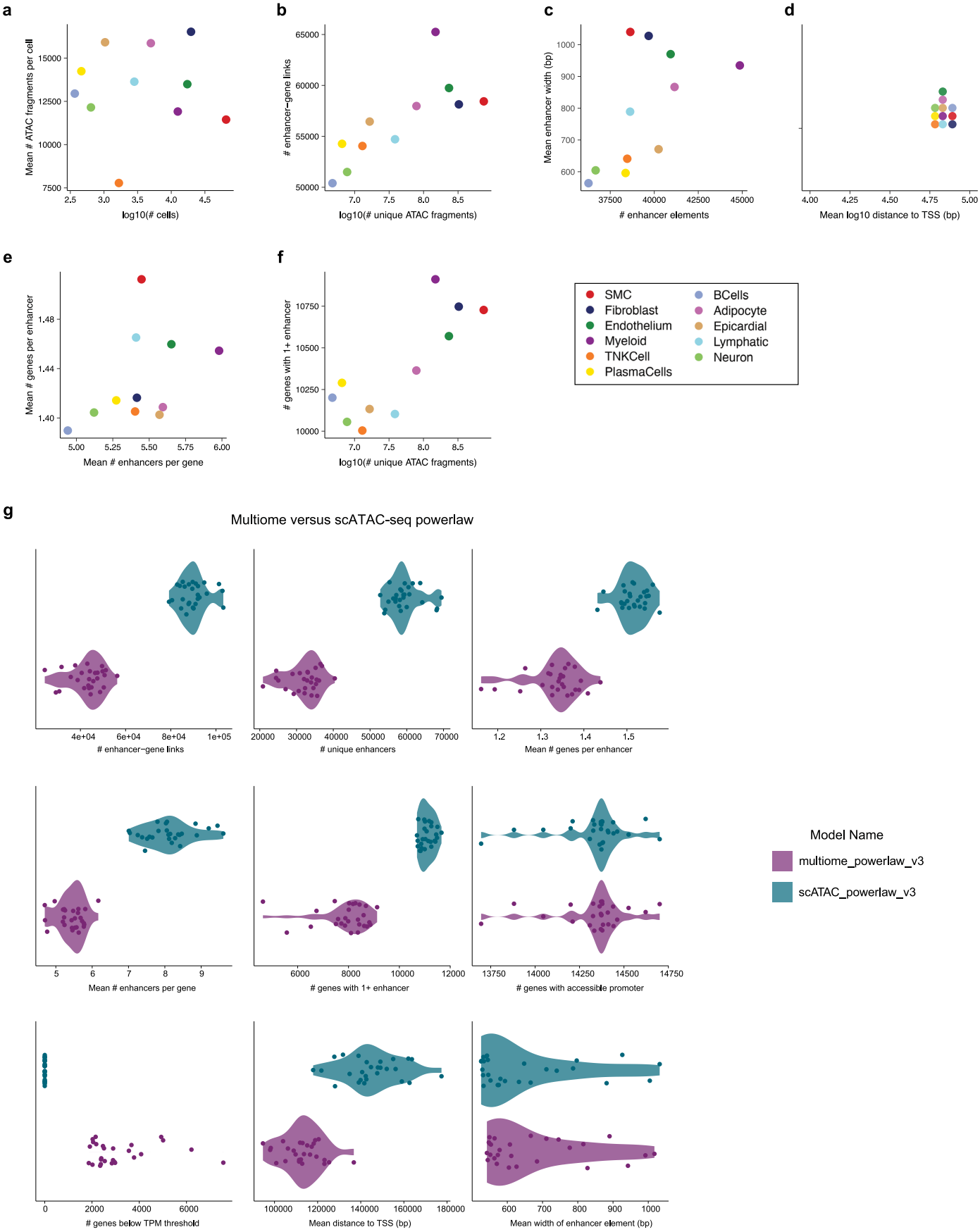
Correspondence and requests for materials should be addressed to Junedh M. Amrute or Nathan O. Stitziel.

Peer review information *Nature* thanks Joseph Powell, Carlos Talavera-López and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at http://www.nature.com/reprints.

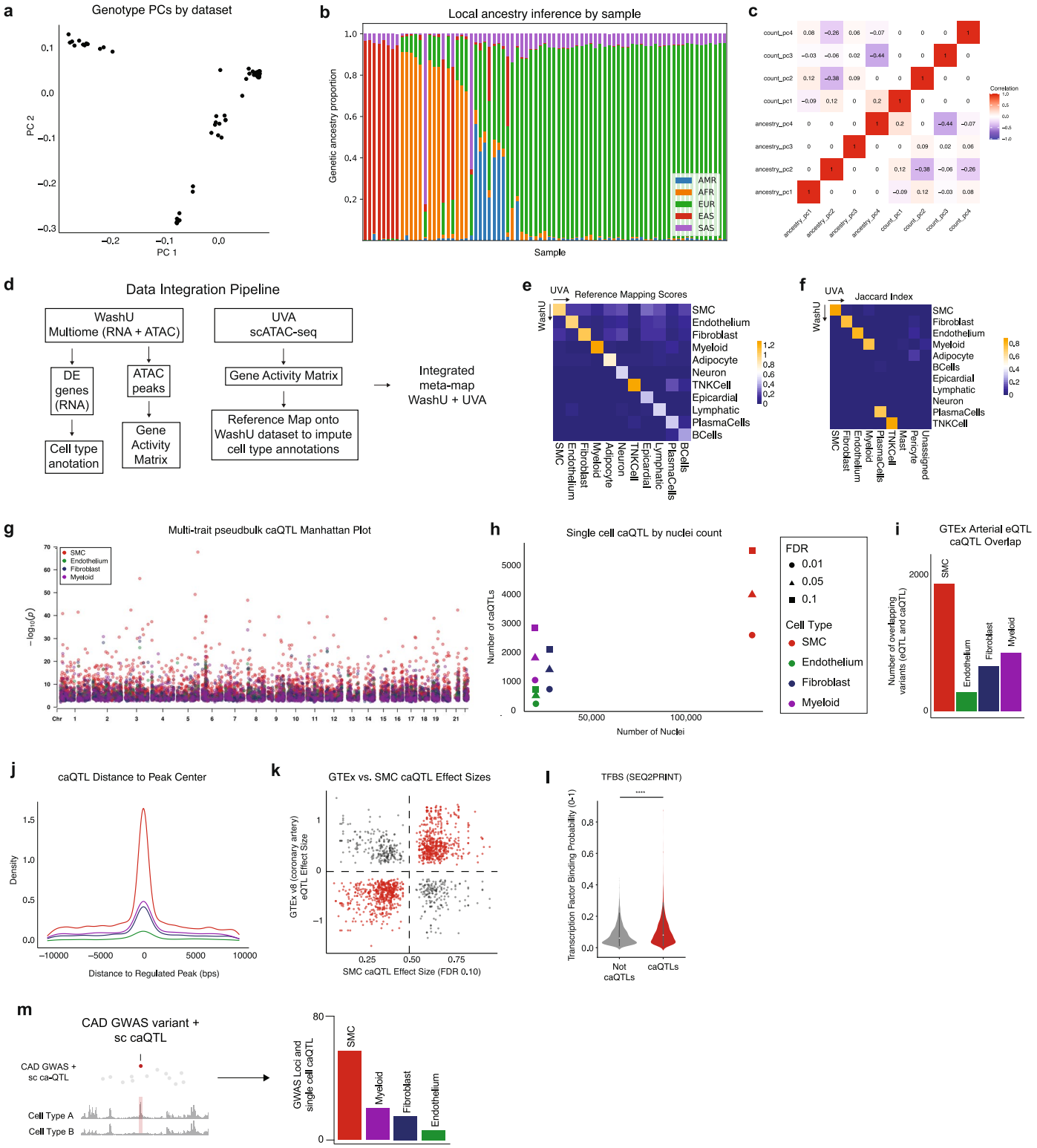


Extended Data Fig. 1 | Multimodal clustering of Multiome data. (a) Gene set scores for marker genes (adjusted P value < 0.05 and $\log_2$FC > 0.58) in each cell type. P values were calculated using one-sided Fisher's exact tests with Benjamini–Hochberg correction. (b) Dot plot of marker gene set score from (a) grouped by cell type. (c) Dot plot of canonical RNA marker genes by cell type, colored by average expression; dot size indicates percent of cells in the cluster which express the gene. (d) Multiome peaks grouped by distal, exonic, intronic, and promoter from macs2 peak calls split by cell type. (e) Differentially accessible marker peaks heatmap by cell type; statistically significant peaks (FDR < 0.1 and $\log_2$FC > 0.5).



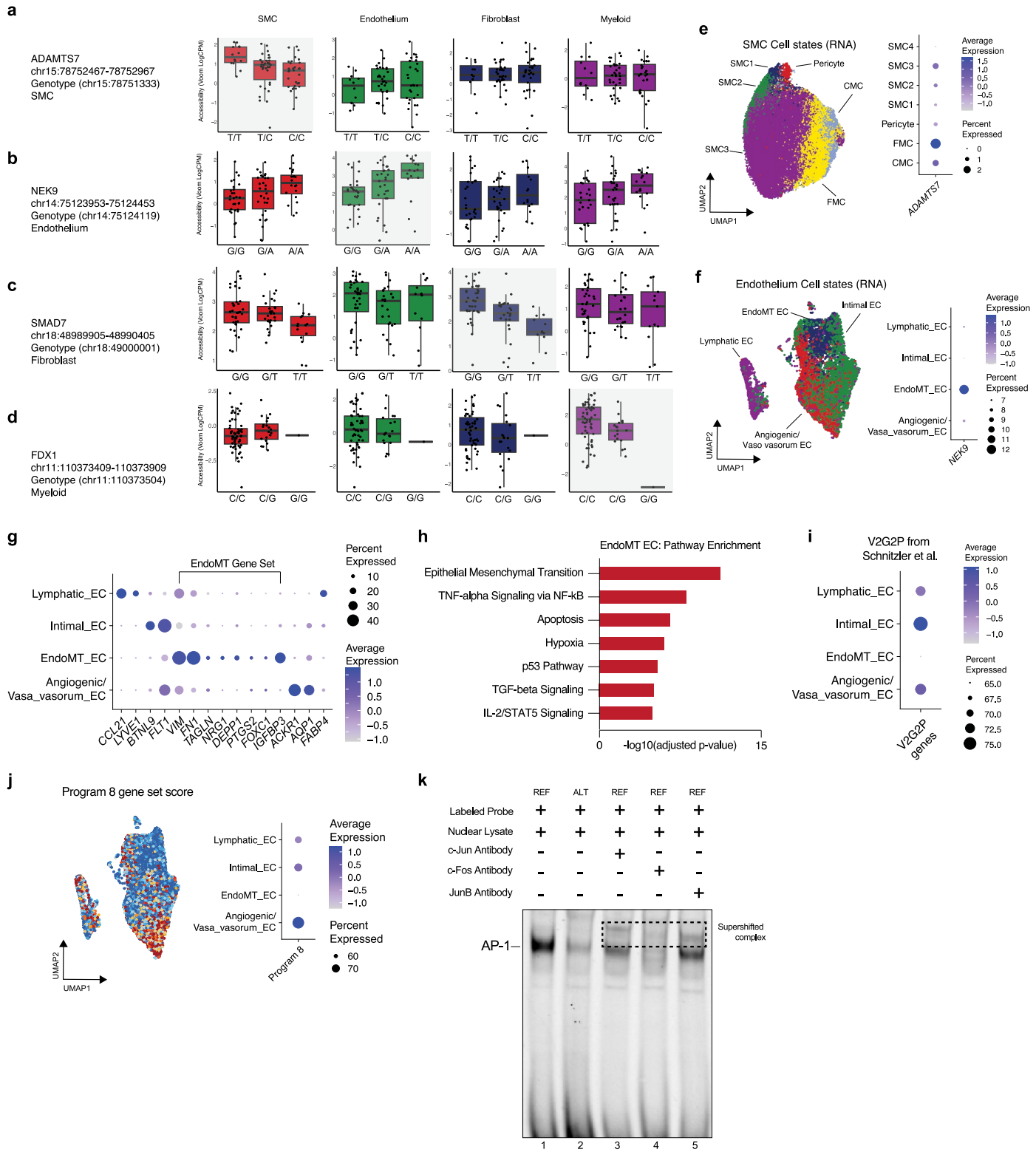
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | scE2G^{Multiome} enhancer gene predictions. Each point represents enhancer-gene predictions by scE2G^{Multiome} in one cell cluster (N = 11). All properties were computed on binarized predictions omitting promoter elements. (a) Average number of unique ATAC fragments per cell by number of cells per cluster. (b) Number of enhancer-gene links by number of unique ATAC fragments per cluster. (c) Mean enhancer width by number of unique enhancer elements regulating at least one distal gene. (d) Mean distance between enhancer element and target gene transcription start site. (e) Mean number of enhancers per gene by mean number of genes per enhancer. (f) Number of genes with at least one non-promoter enhancer by number of unique ATAC fragments. (g) Metrics comparing scE2G^{Multiome} model with scE2G^{scATACseq} model.

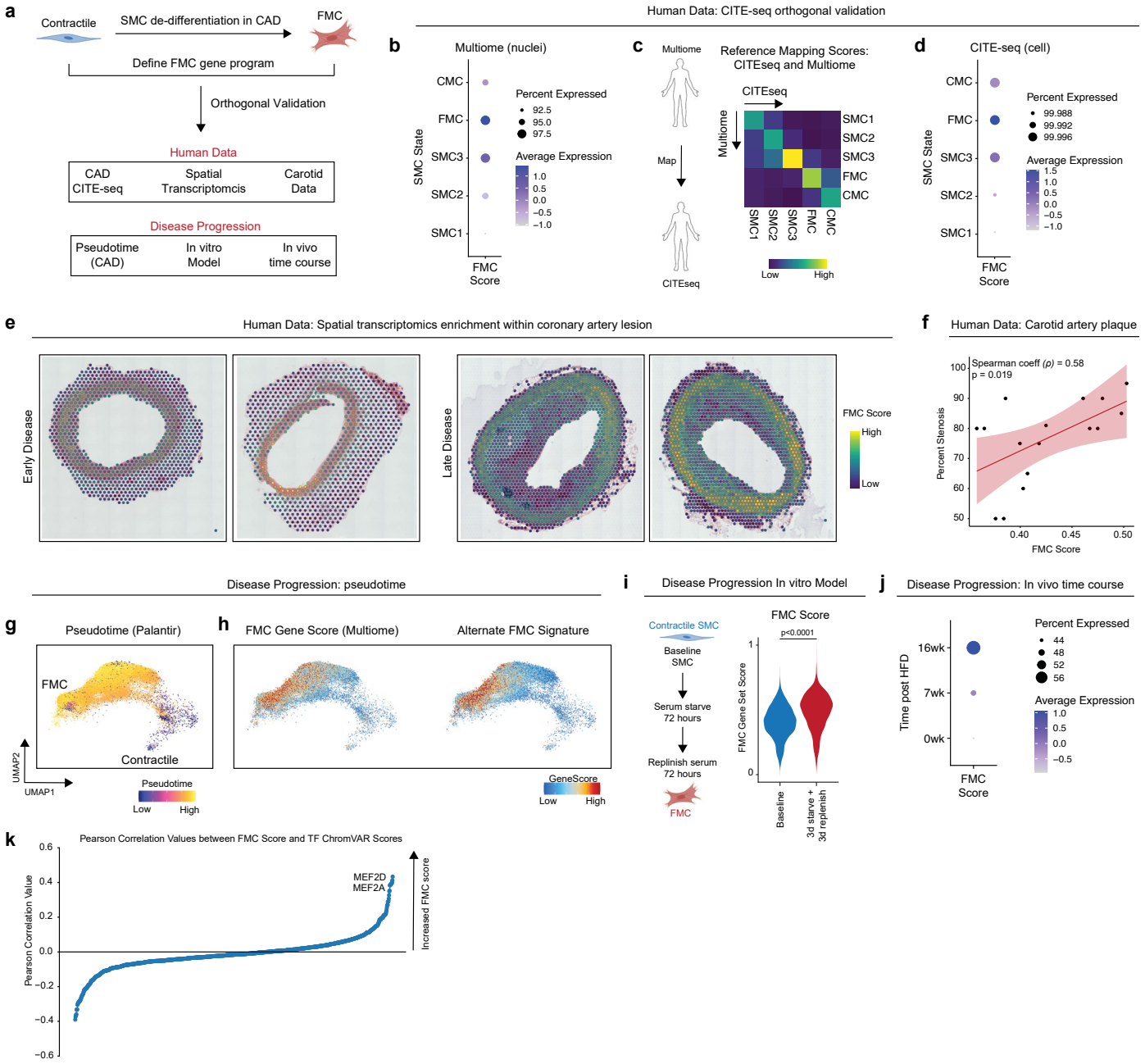


Extended Data Fig. 3 | See next page for caption.

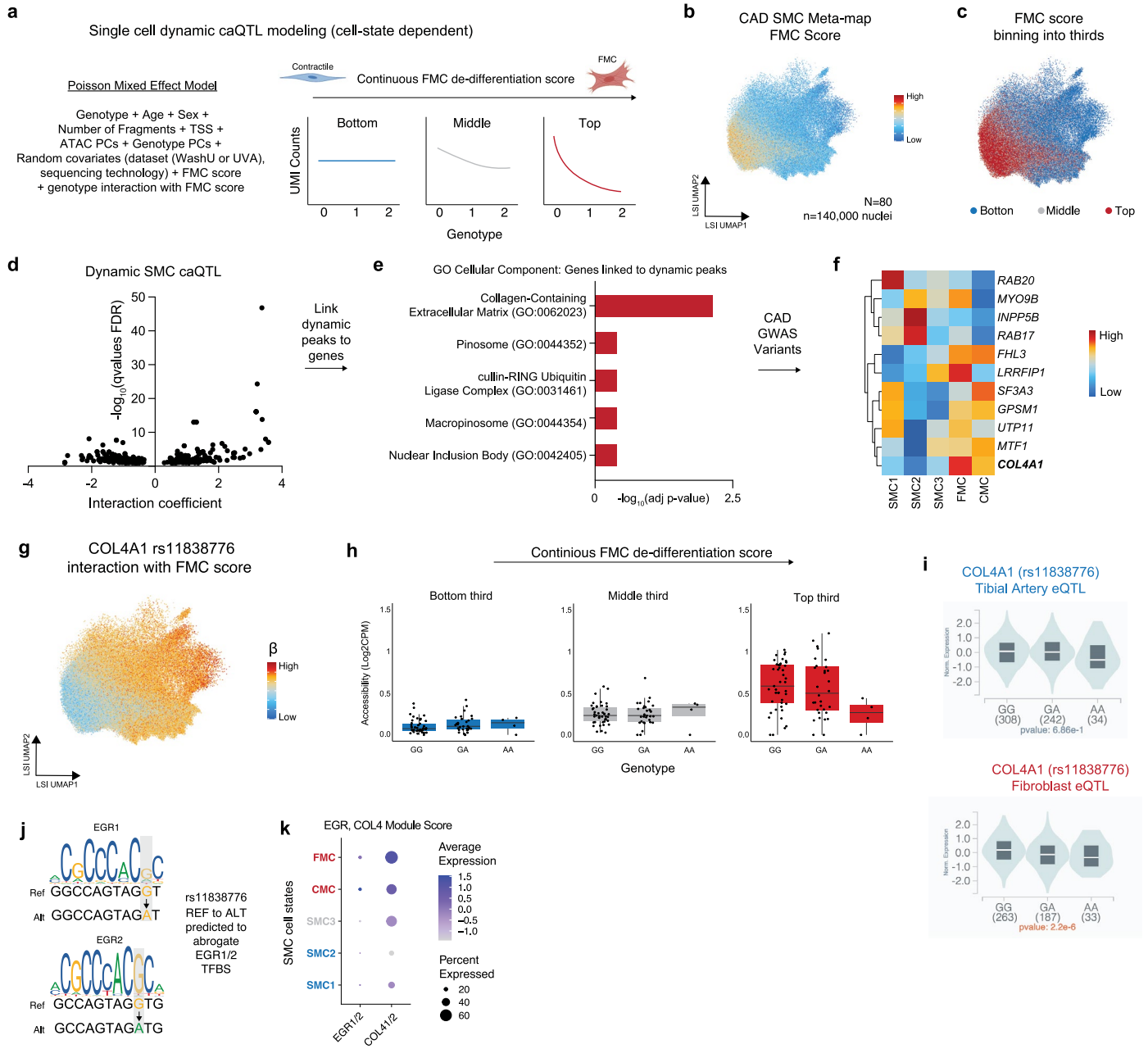
Extended Data Fig. 3 | Genotyping and pseudobulk caQTL discovery. (a) PC1 and 2 for genotyping data (post imputation from the TOPMED panel and filtering for variants with imputation $R^2 < 0.3$ and MAF > 0.05) for each sample. (b) Local ancestry inference; YRI (n = 186) and CEU (n = 183) from the 1000 Genomes Project[74] were used as AFR / EUR reference populations, respectively. Ancestry proportion stack plot by sample. (c) Correlation heatmap for ancestry PCs and ATAC count PCs. (d) Data integration pipeline for WashU and UVA cohorts. (e) Reference mapping score heatmap for UVA snATAC-seq data onto WashU Multiome data. (f) Jaccard Index for UVA cell annotations based on WashU Multiome mapping versus author designated annotations from UVA. (g) Multi-trait pseudobulk Manhattan plot for caQTL colored by cell type genome-wide. P values were calculated using a likelihood-ratio test in RASQUAL (χ^2 distribution, 1 degree of freedom). (h) Number of pseudobulk caQTL by number of nuclei colored by cell type at various FDR cut-offs. (i) sc-caQTL overlap with GTEx arterial eQTL by cell type. (j) Distance from caQTL to the regulated peak colored by cell type. (k) Correlation of effect sizes between bulk arterial eQTL variants and pseudobulk caQTL variants. Red indicates a positive correlation while black indicates a negative correlation. Note that pseudobulk caQTL variant effect sizes are estimated by RASQUAL's pi parameter, which is centered at 0.5. Pearson r = 0.398, $P < 2.2 \times 10^{-16}$ (two-sided Pearson correlation test). (l) Transcription factor binding scores (TFBS) for caQTL and non-caQTL variants overlapping snATAC-seq peaks. TFBS were compared using a two-sided Wilcoxon rank-sum test ($P = 8.2 \times 10^{-40}$). caQTLs were identified from n = 88 biologically independent samples. (m) Overlap of caQTLs with CAD GWAS variants by cell type.



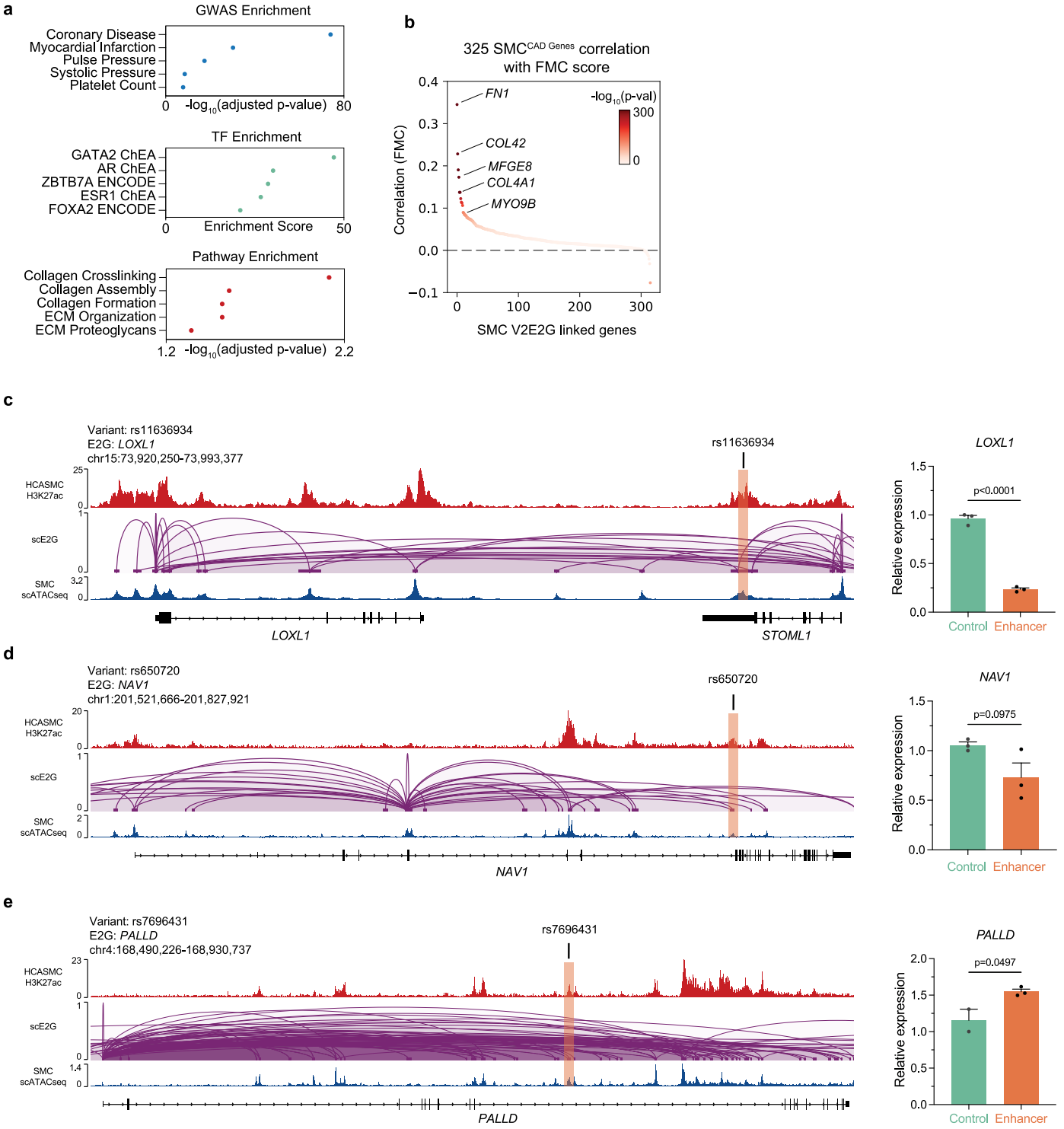
Extended Data Fig. 4 | Cell type and state-specific regulation of molecular traits. Pseudobulk accessibility by genotype plots for (a) *ADAMTS7*, (b) *NEK9*, (c) *SMAD7*, and (d) *FDX1* loci enhancers are cell type specific caQTLs for SMCs, endothelium, fibroblasts and myeloid cells respectively. n = 88 biologically independent samples were used for the caQTLs. (e) UMAP embedding plot of SMC cell state annotations (left) and dotplot of *ADAMTS7* expression by SMC cell state. (f) UMAP embedding plot of endothelial cell state annotations (left) and dotplot of *NEK9* expression by endothelial cell state. (g) DotPlot of canonical marker genes for endothelium cell biology grouped by mapped cell state annotation. (h) Pathway enrichment for statistically significant genes upregulated in EndoMT (adjusted P value < 0.05 and log2FC > 0.58). P values were calculated using one-sided Fisher's exact tests with Benjamini–Hochberg correction. (i) Endothelium acting loci from V2G2P in Schnitzler et al.[26] gene set score grouped by EC state. (j) Program 8 gene set score from Schnitzler et al. in UMAP embedding plot (left) and dot plot grouped by EC state (right). (k) Antibody supershift EMSA of AP-1 binding to the reference (REF) and alternate (ALT) probes. Nuclear lysate was incubated with REF or ALT probe alone (lanes 1–2) or REF probe with antibodies against c-Jun, c-Fos, or JunB (lanes 3–5, respectively). Supershifted complexes are indicated by the dashed box. Supershifts with c-Jun and c-Fos antibodies confirmed that the AP-1 complex contains c-Jun and c-Fos. This experiment was repeated 3 times; representative experiment shown.



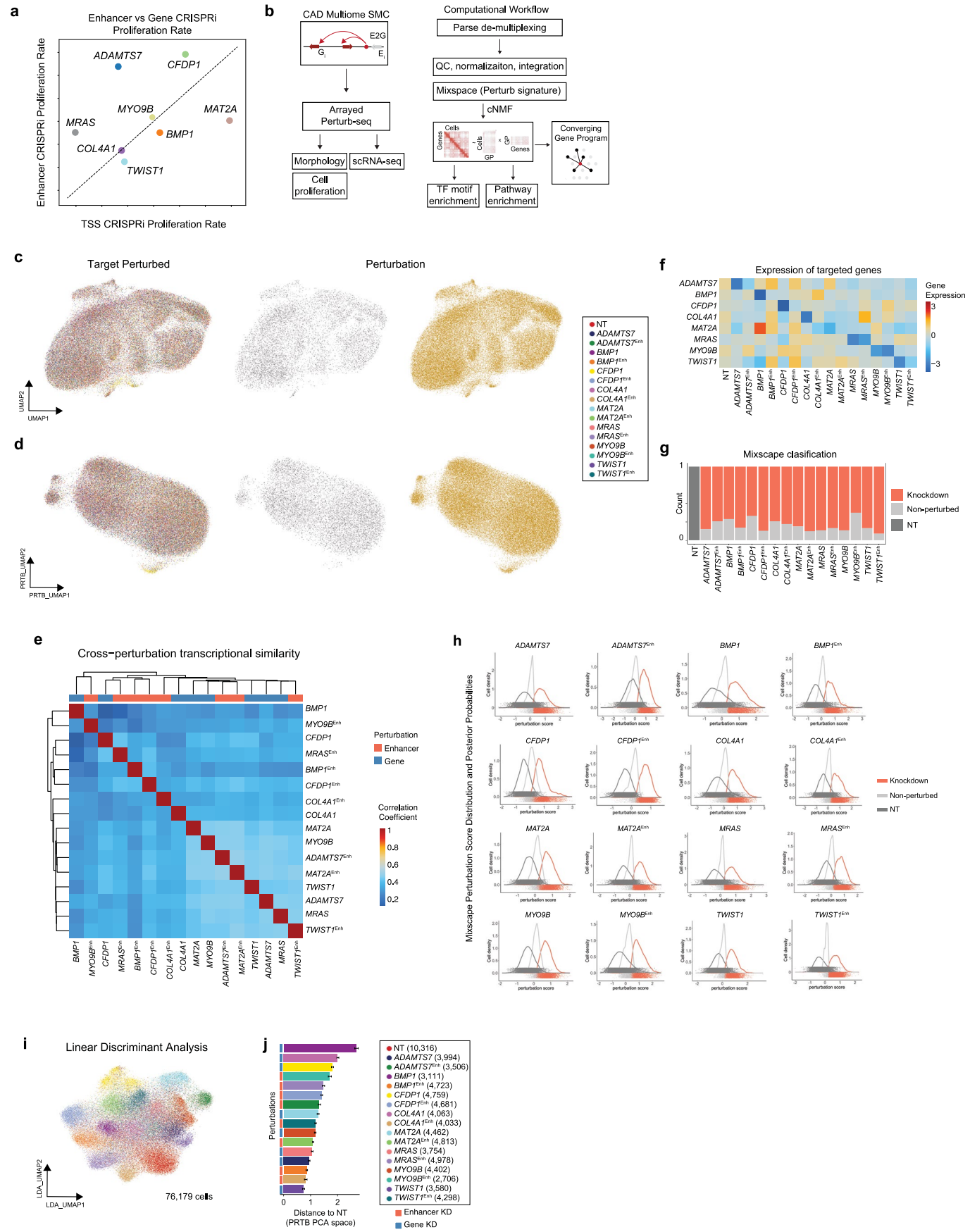
Extended Data Fig. 5 | FMC cell state characterization. (a) Schematic for FMC characterization across datasets. (b) FMC module score dot plot in Multiome SMC states. (c) Reference mapping score heatmap of Multiome SMC states onto published CITE-seq data[18]. (d) FMC module score dot plot in CITE-seq SMC states. (e) FMC module score in spatial transcriptomics data in early and late disease coronary lesions. (f) Regression plot for percent stenosis versus FMC module score from carotid artery scRNA-seq data. Spearman ρ = 0.58, P = 0.019. The shaded area is the 95% confidence interval around the fitted linear regression line. (g) Pseudotime calculated using Palantir in SMC states with contractile SMC states as the starting state. (h) FMC gene set score from Wirka et al.[16] (left) and Amrute et al.[18] (right) overlaid on SMC states Multiome data. (i) FMC module score in baseline and stressed (72 hr serum starve followed by 48 hr serum replenishment) HCASMCs. P values were calculated using two-sided Wilcoxon rank-sum tests. (j) FMC module score dot plot in time course SMC lineage traced data grouped by time post-HFD feeding. (k) Pearson correlation score between FMC score and TF ChromVAR score. Illustrations in **a**, **c** and **i** created in BioRender: **a** and **i**, Lavine, K. https://biorender.com/m3r2gap (2026); **c**, Lavine, K. https://biorender.com/cwi31gk (2026).



Extended Data Fig. 6 | Dynamic caQTL modeling. (a) Dynamic single cell caQTL modeling strategy using the PME model for cell-state dependent single nuclei caQTL analysis. Covariate adjustment and interaction between a continuous FMC cell state score (binned into low, medium and high) and genotype (see Methods). (b) FMC gene set score (top 100 genes from Wirka et al. for FMC marker gene) in integrated SMC meta-map. (c) Nuclei binned into tertiles based on FMC score. (d) SMC single cell dynamic caQTLs volcano plot showing -$\log_{10}$(adjusted P value) by interaction coefficient ($\beta_{total} = \beta_G + \text{FMC score} \times \beta_{FMCxG}$). RASQUAL q values were determined using Benjamini–Hochberg correction followed by permutation-based genome-wide FDR calibration. (e) Gene ontology cellular component pathway analysis for genes linked to dynamic caQTLs. P values were calculated using one-sided Fisher's exact tests with Benjamini–Hochberg correction. (f) Heatmap of genes linked to dynamic caQTLs which are also CAD GWAS variants. (g) Interaction coefficient of the rs11838776 caQTL for *COL4A1* in SMCs with FMC score. (h) Box plots show the caQTL effect for SMCs in the bottom (left), middle (center), and top (right) thirds of FMC scores. n = 79 biologically independent samples were used for the caQTLs. (i) rs11838776 eQTL in tibial artery (left) and cultured fibroblasts (right) for *COL4A1* showing no correlation in tibial arteries but a negative correlation in fibroblasts by genotype. n = 268 biologically independent samples were used; P values were calculated using two-sided t-tests of the GTEx eQTL linear regression coefficient against zero. (j) rs11838776 ref (G) to alt (A) predicted to remove an EGR1/2 binding site. (k) *EGR1 + 2* and *COL4A1 + 2* gene set score dotplot by SMC cell state showing greatest expression in CMC/FMC cell states. Illustrations in **a** created in BioRender: Lavine, K. https://biorender.com/m3r2gap (2026).

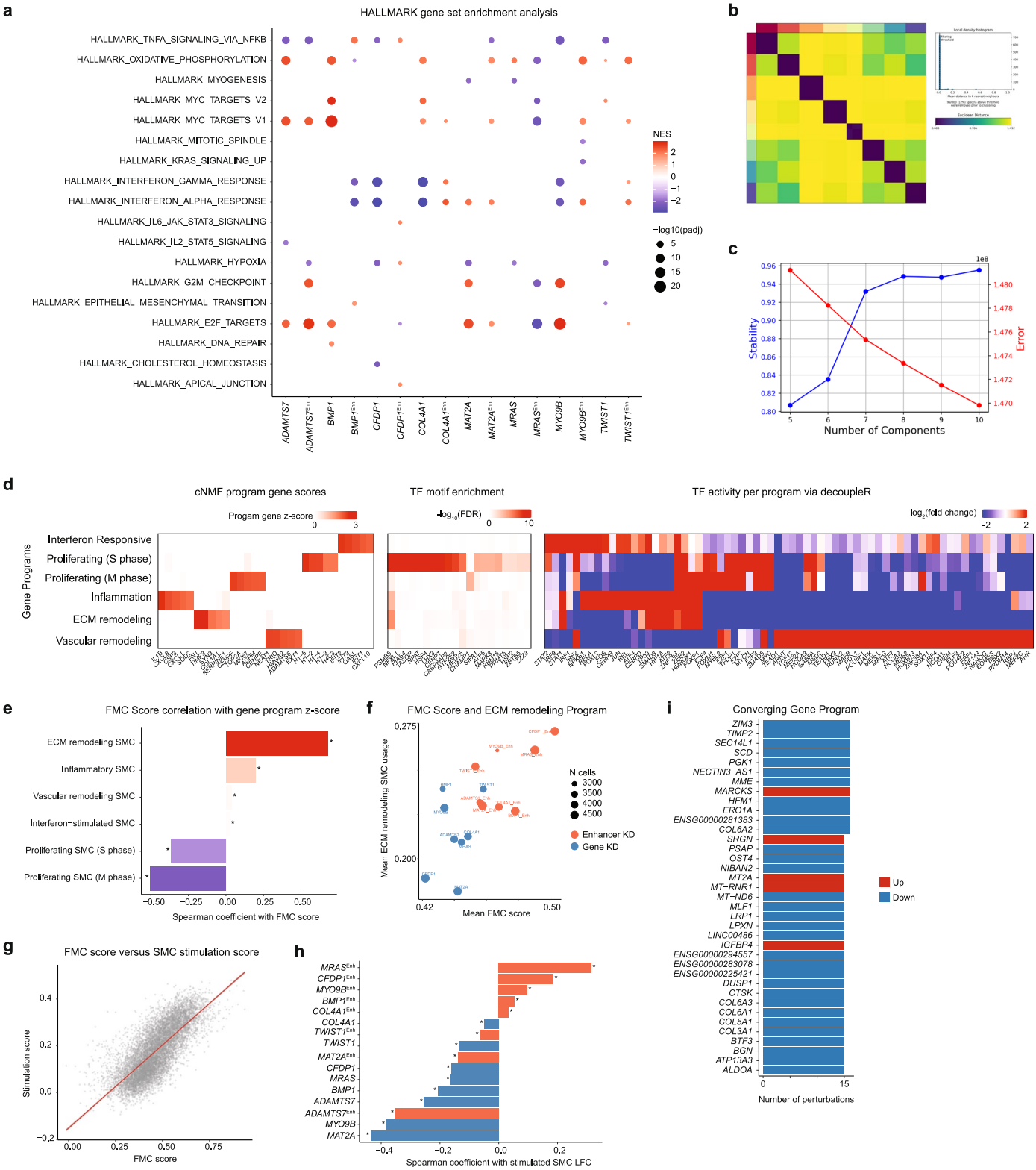


Extended Data Fig. 7 | CRISPRi validation of CAD loci in SMCs. (a) Enrichment analysis for 325 genes linked to SMC enhancers with a CAD disease variant: GWAS enrichment (top), TF enrichment (middle), and pathway enrichment (bottom). P values were calculated using one-sided Fisher's exact tests with Benjamini–Hochberg correction. (b) Correlation between FMC module score and expression of the 325 SMC genes identified in Fig. 3a colored by P value. P values were calculated using two-sided Pearson correlation, with Bonferroni correction for 325 comparisons ($P < 1.54 \times 10^{-4}$). (c) scE2G plot showing HCASMC H3K27ac, scE2G^{Multiome}, and scATAC-seq in SMCs at rs11636934 locus (left) with CRISPRi perturbation of enhancer containing rs11636934 and non-targeting control with *LOXL1* qPCR. (d) scE2G plot showing HCASMC H3K27ac, scE2G^{Multiome}, and scATAC-seq in SMCs at rs650720 locus (left) with CRISPRi perturbation of enhancer containing rs650720 and non-targeting control with *NAV1* qPCR. (e) scE2G plot showing HCASMC H3K27ac, scE2G^{Multiome}, and scATAC-seq in SMCs at rs7696431 locus (left) with CRISPRi perturbation of enhancer containing rs7696431 and non-targeting control with *PALLD* qPCR. In (c-e) n = 3 biologically independent samples were used; data represent the mean ± SEM; P values were calculated using two-sided unpaired t-tests or one-way ANOVA with Dunnett's correction for multiple comparisons, as appropriate.



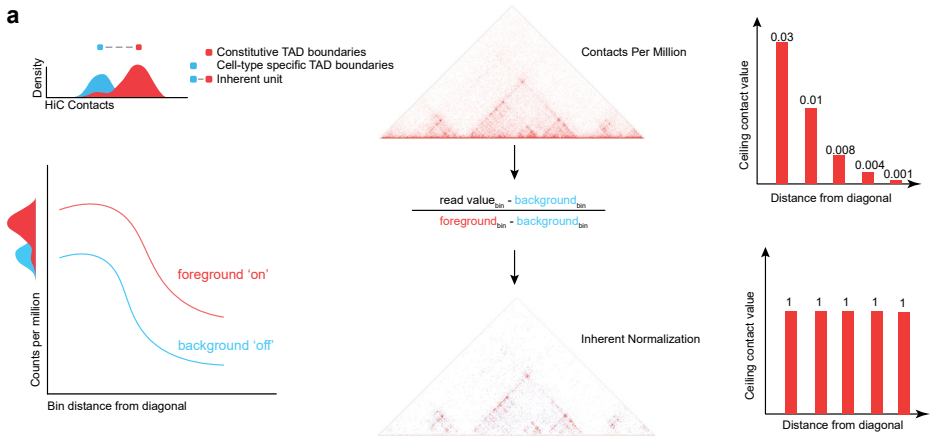
Extended Data Fig. 8 | See next page for caption.

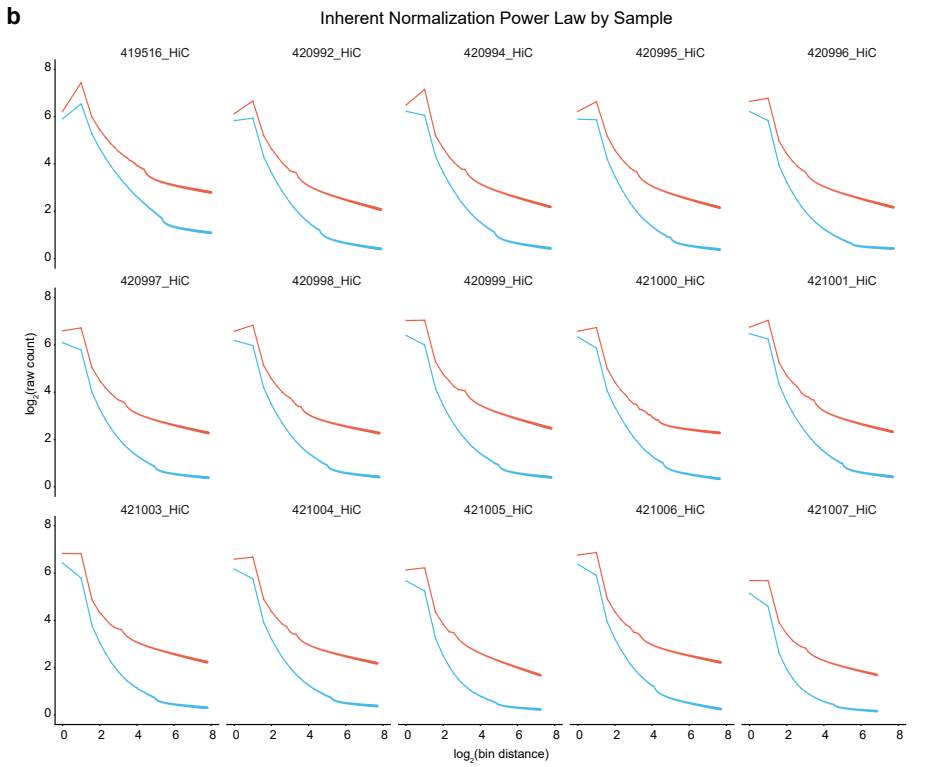
Extended Data Fig. 8 | Arrayed perturb-seq analysis pipeline. (a) Scatter plot of proliferation rate of HCASMCs with TSS repression versus enhancer repression; dotted line indicates parity. (b) Computational workflow for analysis of arrayed perturb-seq scRNA-seq data. Integrated data across all perturbations grouped by perturbation (left) and NT control versus perturbed cell (right) in (c) UMAP space and (d) perturbation UMAP space. Perturbation signatures were calculated using k = 20 non-targeting nearest neighbors in PCA space to remove technical variation. Dimensionality reduction on the PRTB assay revealed perturbation-specific clustering independent of technical confounders. (e) Pairwise cosine similarity between log2 fold-change profiles of all perturbations versus NT cells, derived from pseudobulk DESeq2 analysis. Perturbations are hierarchically clustered and annotated by type (gene knockdown, blue; enhancer knockdown, salmon). (f) Heatmap of gene expression of targets by perturbation. Diagonal indicates repression led to reduced target gene expression. (g) Cell composition stack plot for Mixscape classification of cells into NT control, knocked down, and non-perturbed. Across all perturbations ~0.8 were knocked down. (h) Mixscape perturbation score distribution and posterior probabilities for all targets. (i) Linear discrimination analysis with UMAP visualization colored by perturbation. (j) Distance between perturbed cells and NT control in perturbation PCA space. Data represent the mean ± 95% bootstrap confidence intervals derived from 1,000 resamples of KD-classified cells for each perturbation. Numbers on bars indicate the total number of Mixscape-classified KD cells per condition.

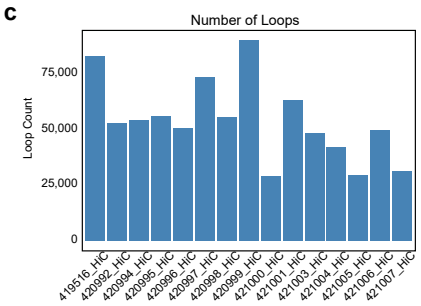


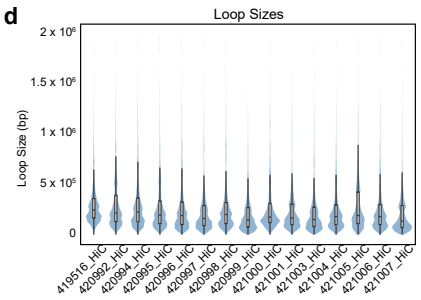
Extended Data Fig. 9 | See next page for caption.

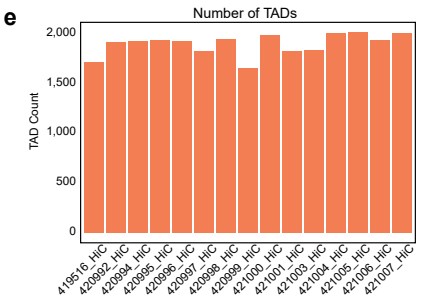
Extended Data Fig. 9 | Perturb-seq gene program analysis with cNMF. (a) GSEA dotplot of MSigDB Hallmark gene sets per perturbation, ranked by normalized enrichment score (NES). Dot size reflects -log10(adjusted P value). P values were determined by permutation testing in GSEA, with FDR correction for multiple comparisons. (b) cNMF consensus clustering heatmap at k = 8 showing pairwise Euclidean distances between NMF spectra across 100 iterations. Well-separated diagonal blocks indicate reproducible programs. Local density histogram (inset) shows distribution of mean distances to k-nearest neighbor spectra; 99/800 (12%) outlier spectra exceeding the density threshold (local-density-threshold=0.01) were removed prior to consensus averaging. (c) cNMF component selection plot showing stability (blue) and reconstruction error (red) across k = 5–10; k = 8 was selected. (d) Three-panel characterization of cNMF gene programs (k = 8, 6 biologically interpretable programs retained): program gene specificity z-scores (left), TF motif enrichment (-log10 FDR, center), and TF activity scores estimated by decoupleR ULM using the DoRothEA regulon (log2 fold change, right). (e) Spearman correlation of each cNMF program with FMC module score across KD-classified cells. Asterisks indicate FDR < 0.05. (f) Per-perturbation mean FMC score versus mean ECM remodeling program usage. Dot size reflects number of KD-classified cells. (g) Cell-level scatter of FMC score versus SMC stimulation score (n = 10,000 downsampled cells). Red line indicates linear fit. (h) Spearman correlation of each perturbation's log2 fold-change profile with the stimulated SMC differential expression signature. Asterisks indicate FDR < 0.05. (i) Convergent gene program identified by robust rank aggregation (RRA) across all perturbation DE profiles. Genes are ranked by number of perturbations in which they reach significance (FDR < 0.05), colored by direction of effect.

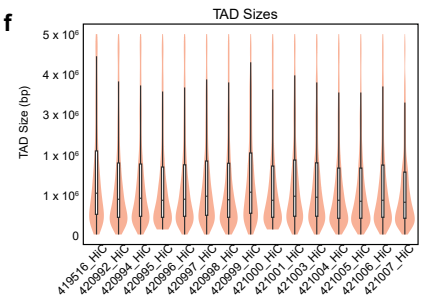


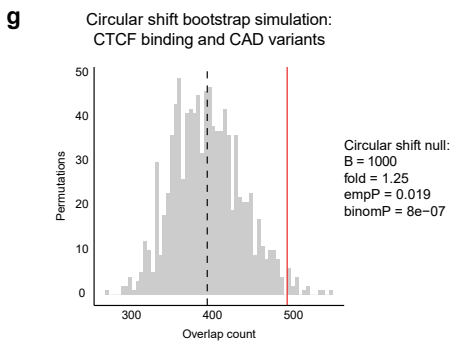


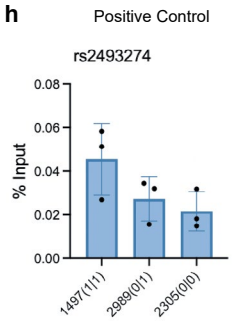


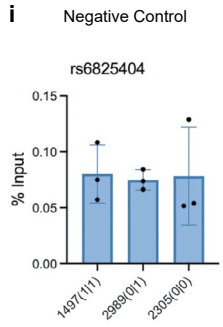


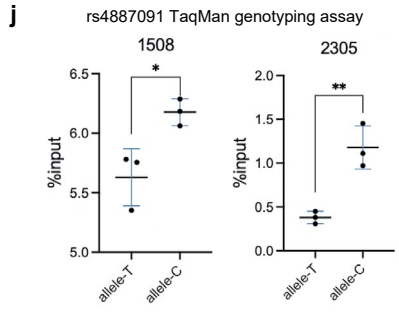












Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Hi-C quality metrics, TAD boundary classification, inherent normalization, and CTCF enrichment at CAD variants. (a) Schematic of inherent normalization for Hi-C contact matrices, comparing CPM (top) and inherent normalization (bottom) with ceiling contact value versus distance from the diagonal. (b) Power-law decay curves for 15 Hi-C samples, plotting $\log_2$(raw contact count) against $\log_2$(bin distance). Red and blue curves represent interactions from constitutive and cell-type-specific TAD boundaries, respectively; their separation defines the distance-dependent ruler used to transform contacts to a 0–1 scale. (c) Number of chromatin loops per donor. (d) Loop-size distributions per donor. (e) Number of TADs per donor. (f) TAD-size distributions per donor. (g) Circular-shift bootstrap assessing enrichment of CAD-associated variants at HCASMC CTCF ChIP-seq peaks. The histogram shows the null distribution; the dashed line indicates the observed overlap. Fold enrichment was calculated as observed/null mean overlap. Empirical one-sided P values were calculated as (# null permutations ≥ observed + 1)/(B + 1), with an exact one-sided binomial test as a parametric approximation (fold = 1.25; empirical P = 0.019; binomial P = 8×10^{-7}). CTCF ChIP–qPCR at the positive-control locus rs2493274 (h) and negative-control locus rs6825404 (i) across individuals with the indicated genotypes. (j) Orthogonal allele-specific TaqMan quantification of CTCF occupancy at rs4887091 in individuals 1508 and 2305. In (d, f), n = 15 and (h–j) n = 3 biologically independent samples; data represent the mean ± SEM; P values were calculated using two-sided paired t-tests (1508: P = 0.0224; 2305: P = 0.0281).

nature portfolio

Corresponding author(s): Junedh Amrute
Nathan Stitziel

Last updated by author(s): Sep 13, 2026

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	10x Genomics Cell Ranger ARC v6.1; Cell Ranger v8.0.0 and v9.0.1; Parse Biosciences split-pipe v1.6.3; BD FACSChorus v1.1; Illumina GenomeStudio v2.0.5; Incucyte 2020B; HoloMonitor App Suite v3.5.2.446.
Data analysis	R v4.3.1 and v3.5.1; ArchR v1.0.3; Seurat v5.5.1 and v4; sctransform v0.4.3; Harmony v2.0.5; scrublet v0.2.3; MACS2 v2.2.9.1; MungeSumstats v1.20.0; LDSC v1.0.1; PLINK v1.9; IRanges v2.46.0; Michigan Imputation Server v1.7.1; minimac4 v1.0.2; Eagle v2.4.1; RFMix v2.03-r0; RASQUAL v1.1; Signac v1.17.1; slam v0.1-56; Qtlizer v1.22.0; chromVAR v1.34.1; lme4 v2.0-6; QTLTools v1.3.1; DESeq2 v1.52.0; fgsea v1.38.0; msigdbr v26.1.1; pheatmap v1.0.13; cNMF v1.7.1; clusterProfiler v4.20.0; decoupleR v2.17.0; RobustRankAggreg v1.2.1; nf-core Hi-C v2.0.0; HiCRes v2.0; HiC-Pro v3.1.0; FitHiC v2.0.8; OnTAD v1.4; cooltools v0.7.1; AQuA Tools (2025 release); HiCuP v0.7.4; SAMtools v1.13; Juicer; FitHiChIP v11.0; locuszoomr v0.3.5; BWA-MEM v0.7.5; Picard v3.4.0; Dfilter; trim-galore v0.65; STAR v2.7.0; HTSeq v2.1.2; edgeR v3.42.4; Primer3 v2.6.1.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

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

Raw/processed study data: Human Coronary Multiome GSE328800; Human Coronary Hi-C GSE328970; genotyping dbGaP phs004118.v1.p1; HCASMC Perturb-seq GEO accession pending; CTCF ChIP-seq GSE319778; H3K27ac ChIP-seq/HiChIP GSE282556/GSE282557; AMOTL2 bulk RNA-seq GSE329169. Public/reference resources: GTEx v8; 1000 Genomes Project Phase 3; GRCh38/hg38 reference genome; GSE131780, GSE175621 and GSE188422; GWAS summary statistics as cited in Methods. Study analysis scripts: https://github.com/jamrute

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Biological sex and other demographics are reported in Supplementary Table 1; sex was included as a covariate in caQTL models.
Reporting on race, ethnicity, or other socially relevant groupings	Race/ethnicity/ancestry are reported in Supplementary Table 1; genotype principal components and local ancestry were used to account for genetic ancestry in relevant analyses.
Population characteristics	Human coronary artery tissue was obtained from 44 Multiome donors; the integrated caQTL meta-map comprised 88 samples. Clinical and demographic characteristics are provided in Supplementary Table 1.
Recruitment	Coronary arteries were obtained from explanted non-failing donor hearts (stable ejection fraction, non-cardiac cause of death) and chronic heart-failure transplant recipients. Tissue availability and clinical indication determined sampling; HIV, hepatitis and known genetic cardiomyopathies were excluded. This donor/transplant-based, non-population sampling may limit generalizability to the broader CAD population.
Ethics oversight	Approved by the Washington University School of Medicine IRB (IRB #201104172). Written informed consent was obtained from each patient before tissue collection; no compensation was provided and participants consented to publication. Data were de-identified. [AUTHOR CONFIRMATION REQUIRED: state whether any participants were minors and, if applicable, how consent from legally authorized representatives was obtained.]

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

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-flat.pdf

Life sciences study design

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

Sample size	No formal statistical power calculation was performed for human tissue studies. Human sample size was governed by tissue availability and the ability to recover sufficient nuclei; assay-specific replicate numbers are reported in Methods and figure legends.
Data exclusions	Human exclusion criteria were prespecified: HIV, hepatitis and known genetic cardiomyopathies. Sequencing/cell-level QC thresholds were prespecified in Methods and Supplementary Methods; no outcomes-based sample exclusions were performed.
Replication	Reproducibility was assessed using independent human donors and independent primary-cell replicate wells/sgRNAs as specified in Methods and figure legends. CTCF ChIP-qPCR assays were repeated at least three times; bulk RNA-seq used at least three independent replicates; other replicate numbers are reported per experiment.
Randomization	Human clinical group and genotype assignment were not randomized because they reflected donor/transplant status or inherited genotype. In vitro perturbations were assigned by prespecified sgRNA/condition and processed under identical protocols; computational subsampling and permutation analyses used random sampling where described.
Blinding	Investigators were blinded to clinical groups during initial single-cell clustering/annotation. Blinding was not applicable for analyses requiring known genotype, perturbation or condition labels; automated/prespecified QC and analysis procedures were used to limit subjective bias.

Behavioural & social sciences study design

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

Study description	*Briefly describe the study type including whether data are quantitative, qualitative, or mixed-methods (e.g. qualitative cross-sectional, quantitative experimental, mixed-methods case study).*
Research sample	*State the research sample (e.g. Harvard university undergraduates, villagers in rural India) and provide relevant demographic information (e.g. age, sex) and indicate whether the sample is representative. Provide a rationale for the study sample chosen. For studies involving existing datasets, please describe the dataset and source.*
Sampling strategy	*Describe the sampling procedure (e.g. random, snowball, stratified, convenience). Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient. For qualitative data, please indicate whether data saturation was considered, and what criteria were used to decide that no further sampling was needed.*
Data collection	*Provide details about the data collection procedure, including the instruments or devices used to record the data (e.g. pen and paper, computer, eye tracker, video or audio equipment) whether anyone was present besides the participant(s) and the researcher, and whether the researcher was blind to experimental condition and/or the study hypothesis during data collection.*
Timing	*Indicate the start and stop dates of data collection. If there is a gap between collection periods, state the dates for each sample cohort.*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, provide the exact number of exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Non-participation	*State how many participants dropped out/declined participation and the reason(s) given OR provide response rate OR state that no participants dropped out/declined participation.*
Randomization	*If participants were not allocated into experimental groups, state so OR describe how participants were allocated to groups, and if allocation was not random, describe how covariates were controlled.*

Ecological, evolutionary & environmental sciences study design

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

Study description	*Briefly describe the study. For quantitative data include treatment factors and interactions, design structure (e.g. factorial, nested, hierarchical), nature and number of experimental units and replicates.*
Research sample	*Describe the research sample (e.g. a group of tagged Passer domesticus, all Stenocereus thurberi within Organ Pipe Cactus National Monument), and provide a rationale for the sample choice. When relevant, describe the organism taxa, source, sex, age range and any manipulations. State what population the sample is meant to represent when applicable. For studies involving existing datasets, describe the data and its source.*
Sampling strategy	*Note the sampling procedure. Describe the statistical methods that were used to predetermine sample size OR if no sample-size calculation was performed, describe how sample sizes were chosen and provide a rationale for why these sample sizes are sufficient.*
Data collection	*Describe the data collection procedure, including who recorded the data and how.*
Timing and spatial scale	*Indicate the start and stop dates of data collection, noting the frequency and periodicity of sampling and providing a rationale for these choices. If there is a gap between collection periods, state the dates for each sample cohort. Specify the spatial scale from which the data are taken*
Data exclusions	*If no data were excluded from the analyses, state so OR if data were excluded, describe the exclusions and the rationale behind them, indicating whether exclusion criteria were pre-established.*
Reproducibility	*Describe the measures taken to verify the reproducibility of experimental findings. For each experiment, note whether any attempts to repeat the experiment failed OR state that all attempts to repeat the experiment were successful.*
Randomization	*Describe how samples/organisms/participants were allocated into groups. If allocation was not random, describe how covariates were controlled. If this is not relevant to your study, explain why.*
Blinding	*Describe the extent of blinding used during data acquisition and analysis. If blinding was not possible, describe why OR explain why blinding was not relevant to your study.*

Did the study involve field work? ☐ Yes ☐ No

Field work, collection and transport

Field conditions	*Describe the study conditions for field work, providing relevant parameters (e.g. temperature, rainfall).*
Location	*State the location of the sampling or experiment, providing relevant parameters (e.g. latitude and longitude, elevation, water depth).*
Access & import/export	*Describe the efforts you have made to access habitats and to collect and import/export your samples in a responsible manner and in compliance with local, national and international laws, noting any permits that were obtained (give the name of the issuing authority, the date of issue, and any identifying information).*
Disturbance	*Describe any disturbance caused by the study and how it was minimized.*

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

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	Flow/sorting: 7-AAD (Sigma, SML1633-1ML). ChIP: rabbit recombinant monoclonal anti-CTCF, clone EPR7314(B), Abcam ab128873 (amount/dilution to be confirmed); rabbit polyclonal anti-H3K27ac, Abcam ab4729, 5 ug/IP. IF: anti-Calponin, ProteinTech 13938-1-AP, 1:100; anti-alpha-smooth muscle actin, Sigma A5228, clone 1A4, 1:100; Goat anti-Mouse Alexa Fluor 594, Thermo Fisher A-11032, 1:1000; Goat anti-Rabbit Alexa Fluor 488, Thermo Fisher A-11008, 1:1000. EMSA supershift: anti-c-Jun Santa Cruz sc-74543, anti-c-Fos sc-166940, anti-JunB sc-8051 (amount/dilution to be confirmed).
Validation	CTCF and H3K27ac antibodies were ChIP-grade/prevalidated by the suppliers and experimental specificity was assessed using genotype/allelic controls. Calponin and alpha-SMA antibodies were used in vendor-validated immunofluorescence applications. AP-1 supershift antibodies were supported by the observed c-Jun/c-Fos supershift pattern. Manufacturer validation details are available from Abcam, Sigma, ProteinTech, Thermo Fisher and Santa Cruz.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Human healthy coronary artery smooth muscle cells (HCASMC; Thermo Scientific, C-017-5C, lot 1689414) and human healthy coronary artery endothelial cells (HCAEC; ATCC, PCS-100-020, lot 62382179).
Authentication	Primary HCASMC and HCAEC were obtained from commercial suppliers with lot identifiers; supplier identity/authentication procedures apply.
Mycoplasma contamination	All cell cultures tested negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No cell lines used in this study are listed as commonly misidentified in the ICLAC register.

Palaeontology and Archaeology

Specimen provenance	*Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information). Permits should encompass collection and, where applicable, export.*

Specimen deposition	*Indicate where the specimens have been deposited to permit free access by other researchers.*
Dating methods	*If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.*

☐ Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	*For laboratory animals, report species, strain and age OR state that the study did not involve laboratory animals.*
Wild animals	*Provide details on animals observed in or captured in the field; report species and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Reporting on sex	*Indicate if findings apply to only one sex; describe whether sex was considered in study design, methods used for assigning sex. Provide data disaggregated for sex where this information has been collected in the source data as appropriate; provide overall numbers in this Reporting Summary. Please state if this information has not been collected. Report sex-based analyses where performed, justify reasons for lack of sex-based analysis.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	*Identify the organization(s) that approved or provided guidance on the study protocol, OR state that no ethical approval or guidance was required and explain why not.*

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

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	*Provide the trial registration number from ClinicalTrials.gov or an equivalent agency.*
Study protocol	*Note where the full trial protocol can be accessed OR if not available, explain why.*
Data collection	*Describe the settings and locales of data collection, noting the time periods of recruitment and data collection.*
Outcomes	*Describe how you pre-defined primary and secondary outcome measures and how you assessed these measures.*

Dual use research of concern

Policy information about dual use research of concern

Hazards

Could the accidental, deliberate or reckless misuse of agents or technologies generated in the work, or the application of information presented in the manuscript, pose a threat to:

No	Yes	
☐	☐	Public health
☐	☐	National security
☐	☐	Crops and/or livestock
☐	☐	Ecosystems
☐	☐	Any other significant area

Experiments of concern

Does the work involve any of these experiments of concern:

No	Yes	
☐	☐	Demonstrate how to render a vaccine ineffective
☐	☐	Confer resistance to therapeutically useful antibiotics or antiviral agents
☐	☐	Enhance the virulence of a pathogen or render a nonpathogen virulent
☐	☐	Increase transmissibility of a pathogen
☐	☐	Alter the host range of a pathogen
☐	☐	Enable evasion of diagnostic/detection modalities
☐	☐	Enable the weaponization of a biological agent or toxin
☐	☐	Any other potentially harmful combination of experiments and agents

Plants

Seed stocks

Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.

Novel plant genotypes

Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.

Authentication

Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

CTCF ChIP-seq: GSE319778; H3K27ac ChIP-seq: GSE282556; H3K27ac HiChIP: GSE282557. Raw data are in GEO/SRA and processed peak/track/interaction files are deposited with the GEO series.

Files in database submission

GSE282556: raw SRA reads, peak-call BED files and normalized bigWig tracks for H3K27ac ChIP samples (inputs raw where applicable). GSE282557: raw HiChIP sequencing and processed interaction/loop files as deposited. GSE319778: raw and processed CTCF ChIP-seq files as deposited in GEO.

Genome browser session
(e.g. UCSC)

No longer applicable (final submission).

Methodology

Replicates

H3K27ac ChIP-seq: two HCASMC ChIP replicates and one HCAEC ChIP sample with matched cell-type inputs (GSE282556); replicate HCASMC peak/track files are deposited. CTCF ChIP-qPCR assays were repeated at least three times; genome-wide CTCF ChIP-seq is deposited under GSE319778.

Sequencing depth

H3K27ac ChIP-seq libraries: at least 50 million paired-end read pairs/sample, 2 x 151 bp, Illumina HiSeq 4000. H3K27ac HiChIP: at least 300 million paired-end read pairs/sample, 2 x 151 bp. CTCF ChIP-seq sequencing details are provided in GSE319778.

Antibodies

CTCF: rabbit recombinant monoclonal anti-human CTCF, Abcam ab128873, clone EPR7314(B) (amount/dilution to be confirmed). H3K27ac: rabbit polyclonal anti-H3K27ac, Abcam ab4729, 5 ug per immunoprecipitation.

Peak calling parameters

H3K27ac reads were aligned to hg38 with BWA-MEM v0.7.5, duplicates removed with Picard MarkDuplicates v3.4.0, and peaks called with Dfilter using -ks=60, -bs=100, -lpvalue=8; bigWig tracks were depth-normalized. CTCF ChIP-seq processed files and peaks are deposited in GSE319778.

Data quality

Chromatin was sonicated to 200-500 bp; H3K27ac libraries were size-selected to 300-500 bp and sequenced at at least 50 million read pairs. Read deduplication, depth normalization and deposited peak calls/tracks were used for QC. Numbers of peaks meeting FDR 5% and >5-fold enrichment were not used as prespecified QC criteria in the manuscript.

Software

BWA-MEM v0.7.5; Picard v3.4.0; Dfilter; HiCuP v0.7.4; SAMtools v1.13; Juicer; FitHiChIP v11.0; locuszoomr v0.3.5; R v4.3.1.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Nuclei were isolated using 10x Genomics protocol CG00375 and sorted for 7-AAD-positive intact nuclei before Chromium Next GEM Single Cell Multiome ATAC + Gene Expression (CG000338).
Instrument	BD FACSMelody (BD Biosciences), 100-um nozzle.
Software	BD FACSChorus v1.1.
Cell population abundance	Target approximately 10,000 nuclei/cells per preparation; final post-QC nuclei numbers are reported in the manuscript and Supplementary Fig. 1.
Gating strategy	7-AAD-positive intact nuclei; gating strategy shown in Supplementary Fig. 1.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

Magnetic resonance imaging

Experimental design

Design type	*Indicate task or resting state; event-related or block design.*
Design specifications	*Specify the number of blocks, trials or experimental units per session and/or subject, and specify the length of each trial or block (if trials are blocked) and interval between trials.*
Behavioral performance measures	*State number and/or type of variables recorded (e.g. correct button press, response time) and what statistics were used to establish that the subjects were performing the task as expected (e.g. mean, range, and/or standard deviation across subjects).*

Acquisition

Imaging type(s)	*Specify: functional, structural, diffusion, perfusion.*
Field strength	*Specify in Tesla*
Sequence & imaging parameters	*Specify the pulse sequence type (gradient echo, spin echo, etc.), imaging type (EPI, spiral, etc.), field of view, matrix size, slice thickness, orientation and TE/TR/flip angle.*
Area of acquisition	*State whether a whole brain scan was used OR define the area of acquisition, describing how the region was determined.*

Diffusion MRI ☐ Used ☐ Not used

Preprocessing

Preprocessing software	*Provide detail on software version and revision number and on specific parameters (model/functions, brain extraction, segmentation, smoothing kernel size, etc.).*
Normalization	*If data were normalized/standardized, describe the approach(es): specify linear or non-linear and define image types used for transformation OR indicate that data were not normalized and explain rationale for lack of normalization.*
Normalization template	*Describe the template used for normalization/transformation, specifying subject space or group standardized space (e.g. original Talairach, MNI305, ICBM152) OR indicate that the data were not normalized.*
Noise and artifact removal	*Describe your procedure(s) for artifact and structured noise removal, specifying motion parameters, tissue signals and physiological signals (heart rate, respiration).*

Volume censoring

Define your software and/or method and criteria for volume censoring, and state the extent of such censoring.

Statistical modeling & inference

Model type and settings

Specify type (mass univariate, multivariate, RSA, predictive, etc.) and describe essential details of the model at the first and second levels (e.g. fixed, random or mixed effects; drift or auto-correlation).

Effect(s) tested

Define precise effect in terms of the task or stimulus conditions instead of psychological concepts and indicate whether ANOVA or factorial designs were used.

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference

Specify voxel-wise or cluster-wise and report all relevant parameters for cluster-wise methods.

(See Eklund et al. 2016)

Correction

Describe the type of correction and how it is obtained for multiple comparisons (e.g. FWE, FDR, permutation or Monte Carlo).

Models & analysis

n/a	Involved in the study
☐	☐ Functional and/or effective connectivity
☐	☐ Graph analysis
☐	☐ Multivariate modeling or predictive analysis

Functional and/or effective connectivity

Report the measures of dependence used and the model details (e.g. Pearson correlation, partial correlation, mutual information).

Graph analysis

Report the dependent variable and connectivity measure, specifying weighted graph or binarized graph, subject- or group-level, and the global and/or node summaries used (e.g. clustering coefficient, efficiency, etc.).

Multivariate modeling and predictive analysis

Specify independent variables, features extraction and dimension reduction, model, training and evaluation metrics.