Title

Whole-genome sequencing uncovers two loci for coronary artery calcification and identifies ARSE as a regulator of vascular calcification.

Journal

Nature Cardiovascular Research, 2(12)

Authors

de Vries, Paul
Conomos, Matthew
Singh, Kuldeep
et al.

Publication Date

2023-12-01

DOI

10.1038/s44161-023-00375-y

Peer reviewed

HHS Public Access

Author manuscript

Nat Cardiovasc Res. Author manuscript; available in PMC 2024 May 30.

Published in final edited form as:

Nat Cardiovasc Res. 2023 December ; 2(12): 1159–1172. doi:10.1038/s44161-023-00375-y.

Whole-genome sequencing uncovers two loci for coronary artery calcification and identifies ARSE as a regulator of vascular calcification

A full list of authors and affiliations appears at the end of the article.

Abstract

Coronary artery calcification (CAC) is a measure of atherosclerosis and a well-established predictor of coronary artery disease (CAD) events. Here we describe a genome-wide association study (GWAS) of CAC in 22,400 participants from multiple ancestral groups. We confirmed associations with four known loci and identified two additional loci associated with CAC (*ARSE* and *MMP16*), with evidence of significant associations in replication analyses for both novel loci. Functional assays of *ARSE* and *MMP16* in human vascular smooth muscle cells (VSMCs) demonstrate that *ARSE* is a promoter of VSMC calcification and VSMC phenotype switching from a contractile to a calcifying or osteogenic phenotype. Furthermore, we show that the association of variants near *ARSE* with reduced CAC is likely explained by reduced *ARSE* expression with the G allele of enhancer variant rs5982944. Our study highlights ARSE as an important contributor to atherosclerotic vascular calcification, and a potential drug target for vascular calcific disease.

Main

Coronary artery calcification (CAC) is a measure of atherosclerosis and a strong independent predictor of coronary artery disease (CAD) events.[1,2] Several studies have demonstrated that the amount of CAC is associated with risk for CAD events.[3] Cardiovascular risk prediction is enhanced when calcium score is added to cardiovascular risk factors including body mass index, diabetes mellitus, hypertension, total cholesterol, and smoking history.[4] Coronary calcification is not only a manifestation of atherosclerotic plaque but is also thought to contribute directly to plaque rupture when present as

Corresponding authors Paul S. de Vries (paul.s.devries@uth.tmc.edu), The University of Texas Health Science Center at Houston, 1200 Pressler St, Houston TX, 77030; Rajeev Malhotra (rmalhotra@mgh.harvard.edu), Massachusetts General Hospital, 55 Fruit Street, Yawkey 5700, Boston, MA 02114.

*These authors contributed equally: Paul S. de Vries, Matthew P. Conomos, Kuldeep Singh, Christopher J. Nicholson

ΔThese authors jointly supervised this work: Patricia A. Peyser, Clint L. Miller, and Rajeev Malhotra

Drs. Nickerson and Zapol are deceased.

Author Contributions

PSdV, MPC, KS, CJN, DJ, SML, PAP, CLM, and RMa contributed to the study design and conception, and to the drafting of the manuscript.

PdV, KS, CJN, DJ, JCB, LFB, SLC, JGT, NDP, LRY, NHC, XS, AMS, AK, ABN, BIF, BGK, CPM, DRJ Jr, DAN, EB, FFW, GH, GLK, HVD, JB, JDC, JGW, JAS, JDS, KAVM, KAY, LAL, LSE, MF, QW, RAM, RAG, SCN, SMD, SLRK, VN, YIM, TNK, CJOD, ACM, DWB, LCB, AC, BDM, BMP, JJC, TLA, NOS, CAL, RSV, WSP, PAP, and RMa contributed to data collection and processing.

PSdV, MPC, KS, CJN, DJ, NRH, DW, XG, JY, EPY, CT, ATH, MRB, SM, SLC, JW, MSS, XS, GJ, JVM, MY, PN, RMe, VF, VN, WZ, AKM, GP, TNK, NOS, CLM, RSV, WSP, and RMa contributed to data analysis or interpretation.

KS, CJN, WJ, SL, CLLC, RHL, MSM, WT, WMZ, and RMa conducted and interpreted the *in vitro* and histologic experiments.

All authors critically reviewed and approved the manuscript.

microcalcifications whereas more extensive sheets of calcification are associated with greater plaque stability.[5]

Vascular calcification is an actively regulated process that is characterized by the switch of vascular smooth muscle cells from a contractile phenotype to an osteogenic phenotype with expression of bone markers that include Runt-related transcription factor 2 (RUNX2), a master regulator of the phenotype switch, in addition to other markers such as bone gamma-carboxyglutamate protein (BGLAP*)* and alkaline phosphatase (ALPL).[6,7] There is a strong genetic component to CAC and CAD events. Over 200 loci for CAD have been uncovered by previous GWAS,[8–14] but only four loci have been associated with CAC (9p21, *PHACTR1/EDN1*, *APOE*, and *APOB*).[15–18] These four loci are associated with CAD events and their effects appear to be primarily related to the progression of atherosclerosis,[19–22] although *PHACTR1*/*EDN1* may also affect calcification directly.[23] The identification of additional CAC loci could provide insights into novel pathogenic mechanisms underlying atherosclerotic cardiovascular disease. To identify new CAC loci, we used whole genome sequencing (WGS) data from the Trans-Omics for Precision Medicine (TOPMed) program.

Results

Traditional single variant analyses restricted to variants with minor allele count ≥50 were performed, as well as aggregate analyses of rare variants with minor allele frequency <1%. A total of 22,400 participants from ten different studies were included in our discovery analyses (Supplementary Table 1). These participants were stratified into five population groups based on their reported race/ethnicity: 596 Amish, 603 Asian, 8,606 Black, 1,010 Hispanic/Latino, and 11,585 White. The mean age was 58, and 53% of participants were male. Detectable CAC was present in 50% of the population (25% had CAC 1-100, 14% had CAC 101-400, and 11% had CAC>400).

A total of 28,457,765 variants with minor allele count ≥50 in our pooled study population were tested for association with log(CAC+1), including 2,139,838 X chromosome variants (Manhattan plot: Fig. 1; QQ plot: Supplementary Data Fig. 1). Genetic variants exhibiting statistically significant association ($P<5\times10^{-9}$) were identified in regions defined by 9p21, *APOE* (19q13.32), *PHACTR1/EDN1* (6p24.1), *ARSE* (also known as *ARSL*, Xp22.33), *APOB* (2p24.1), and *MMP16* (8q21.3) (Table 1, Supplementary Data Fig. 2–7). We created 95% credible sets of variants for each locus, corresponding to the smallest set of variants at a locus that has a 95% posterior probability of containing the causal variant (Supplementary Table 2).[24] In stratified analyses, the direction of the effects of variants in each locus were consistent between men and women (Supplementary Table 3) and generally consistent across the five race/ethnic populations (Supplementary Table 4). Only the associations of rs7412 (*APOE*) and rs5982944 (*ARSE*) showed differences in the effect direction among population groups. In both cases the effect direction in the Hispanic/Latino population group was discordant with the remaining groups. The six loci were also associated with the presence of CAC as a dichotomous trait using a variety of CAC thresholds (Supplementary Table 5). We performed conditional analyses to identify independent secondary signals at the six associated loci, leading to the identification of conditionally independent associations (FDR < 0.05) at two loci: 9p21 and *APOE* (Supplementary Table 6 and Supplementary

Data Fig. 8). The primary and secondary variants at *APOE* (rs7412 and rs429358) make up the classical *APOE* haplotypes e2, e3, and e4. Analyses by *APOE* haplotype pair using e3/e3 as the reference group are shown in Supplementary Table 7, demonstrating decreased CAC with the e2 haplotype and increased CAC with the e4 haplotype. We also examined whether 241 variants associated with CAD by Aragam et al.[14] were associated with CAC in our study. Of these variants, 233 were included in our analysis. Besides 9p21, *PHACTR1*, and *APOE*, we found suggestive associations (P<0.00021) at four additional loci, including *LOC283033*/*CXCL12*, *COL4A1*/*COL4A2*, *ADAMTS7*, and *LDLR* (Supplementary Table 8).

Rare variants (minor allele frequency < 1%) were tested for association with coronary artery calcification (CAC) using two gene-based variant aggregation and filtering strategies, resulting in a total of 41,868 groupings of rare variants. Both of the aggregation strategies resulted in a genome-wide significant ($P<1.2\times10^{-6}$) aggregation unit for CAC that mapped to the *APOB* gene, but these associations were no longer significant (P>0.05) after adjusting for rs5742904, the index variant identified from the single variant analysis at this locus. Supplementary Table 9 shows the variants that were included in the aggregate tests, as well as the single variant analysis P-value for those variants with minor allele count ≥ 5.

Of the six genome-wide significant loci, variants in *ARSE* and *MMP16* have not been reported previously to be associated with CAC or CAD at a genome-wide significance level. The inverse association of the G allele of index variant rs5982944 at *ARSE* with CAC appeared to follow a recessive mode of inheritance (Extended Data Fig. 1), and was more statistically significant in a recessive model (Beta=−0.43; $P=2.8\times10^{-17}$) than in an additive model (Beta=−0.20; $P=9.1\times10^{-14}$). Because the G allele of rs5982944 is largely restricted to individuals with African ancestry, we replicated the associations with the index variants at the two loci (rs5982944 at *ARSE* and rs13268080 at *MMP16*) from TOPMed cohorts in independent samples from two race/ethnicity groups with African ancestry: Black and Brazilian (Supplementary Table 10 and Supplementary Table 1). The Brazilian population group was ancestrally heterogeneous and included individuals who self-reported as White, Black, and Brown (admixed).[25] Using a Bonferroni adjusted P-value threshold of P<0.0083, the *ARSE* index variant was significantly associated with CAC in the Black population group in a recessive model (Beta=−0.41; P=0.0053), but not in the Brazilian population group. Furthermore, the *MMP16* index variant was significantly associated with CAC in the Brazilian population group (Beta=0.18; P=0.0021), but not in the Black population group. All effect directions were consistent with those from the discovery analysis.

The *ARSE* index variant was also associated with thoracic ascending and descending aorta calcification in the COPDGene study[26] according to both the additive (Beta=-0.34; $P=1.7\times10^{-10}$) and recessive models (Beta=-0.65; $P=8.0\times10^{-10}$). The phenotypic correlation between coronary artery and thoracic aortic calcification was r=0.53 ($P<2.2\times10^{-16}$). The *ARSE* index variant was also associated with high-density lipoprotein (P=0.0014), as well as other phenotypes at nominal significance, including carotid plaque (P=0.0048), systolic blood pressure (P=0.042), and low-density lipoprotein cholesterol (P=0.042). . The *MMP16* variant was not significantly associated with any of the other atherosclerotic phenotypes and was only nominally associated with CAD events (Supplementary Tables 11 and 12).

Credible set analysis indicated that index variant rs5982944 at the *ARSE* locus had a 85% posterior probability of being the causal variant (Supplementary Table 2). This index variant is located within an enhancer region with accessible chromatin, and was significantly associated with gene expression levels of *ARSE* (*cis*-eQTL) in multiple cells and tissues, including cultured fibroblasts and aorta (Supplementary Table 2). A high degree of colocalization was observed between *ARSE* genetic variants associated with CAC and *ARSE* gene expression in cultured fibroblasts (Extended Data Fig. 2), supporting a regulatory effect at the locus. The rs5982944-G allele that was associated with lower CAC was associated with decreased expression levels of *ARSE* in cultured fibroblasts (normalized effect size=-0.90, $P=3.5\times10^{-18}$), indicating that higher levels of ARSE may stimulate arterial calcification.

The index variant at *MMP16* was associated with *MMP16* expression (Supplementary Table 2). Specifically, in aorta the G allele of the *MMP16* index variant, which was associated with higher CAC levels, was associated with decreased mRNA expression levels of *MMP16* (GTEx v8, normalized effect size=-0.22, $P=2.9\times10^{-7}$). Co-localization analyses for *MMP16* suggest that higher *MMP16* expression may inhibit calcification (Extended Data Fig. 3). While the index variant was located 387 kb from *MMP16*, we show that the variant is likely to interact with the promoter region of *MMP16* in the 3D configuration of the chromatin. The 95% credible set of SNPs at the *MMP16* locus were intersected with pooled HiC-based pairwise chromatin interactions as well as and H3K27ac HiChIP pairwise enhancer-promoter interactions, both generated in human coronary artery VSMCs.[27,28] These chromatin loops were visualized at *MMP16* and *ARSE* loci on the latest Roadmap Epigenomics browser. Using Hi-C data from human coronary artery VSMCs a chromatin loop is observed that suggests a 3D interaction between variants in the credible set and the promoter region of *MMP16* (Supplementary Data Fig. 9).

To determine the vascular cell type-specific expression pattern for *ARSE* and *MMP16*, we queried these genes in a recent meta-analysis of 4 public human atherosclerosis datasets in coronary and carotid artery samples (n=119,578 cells). From these datasets we observed *MMP16* expression in vascular smooth muscle cells, fibroblasts and endothelial cells (Extended Data Fig. 4). While *ARSE* was expressed in fewer cells overall, we did observe expression in vascular smooth muscle cells, endothelial cells (particularly those annotated as endothelial-mesenchymal transition ECs marked by fibronectin expression) as well as some fibroblasts. Together these results provide additional support for the role of *ARSE* and *MMP16* in the main cell types underlying coronary calcification.

Atherosclerotic vascular calcification is characterized by the switch of VSMCs from a contractile phenotype to an osteogenic phenotype. The osteogenic phenotype of VSMCs is characterized by expression of Runt-related transcription factor 2 (RUNX2), a master regulator of the phenotype switch, in addition to other markers such as bone gamma-carboxyglutamate protein (BGLAP) and alkaline phosphatase (ALPL), as well as a decrease in contractile markers such as calponin (CNN1).[29,30] Therefore, we performed a series of functional perturbation studies in human coronary artery VSMCs to investigate whether *MMP16* and *ARSE* regulate osteogenic phenotype switch and calcification in VSMCs.

MMP16 is expressed in coronary artery VSMCs, and MMP16 expression decreased by ~75% when VSMCs were grown in osteogenic media compared to normal media (Extended Data Fig. 5a). Silencing *MMP16* did not affect VSMC osteogenic phenotype or calcification (Extended Data Fig. 5a–b**)**.

ARSE is endogenously expressed in coronary artery VSMCs (Fig. 2a). Compared with cells grown in normal media, mRNA and protein expression of *ARSE* in cells grown in osteogenic media was 5-fold and 1.7-fold higher, respectively (Fig. 2a–b). We next sought to determine whether specific knockdown of *ARSE* mRNA inhibited the switch of coronary artery VSMCs from the contractile to the osteogenic phenotype. Compared with control siRNA, siARSE decreased *ARSE* mRNA levels by >95% (Fig. 2a) without affecting expression of *ARSD*, the closest homologue of *ARSE* (Supplementary Data Fig. 10). *ARSE* silencing reduced the expression of key markers of the osteogenic phenotype switch (Fig. 2c), namely *RUNX2* (>40%), *BGLAP* (>80%), and *ALPL* expression (50%), and also increased the mRNA expression of the contractile protein calponin (*CNN1*) by > 2-fold in osteogenic media (Fig. 2c). The fact that silencing ARSE was associated with a greater than 5-fold increase in *CNN1* expression in cells grown in normal growth media suggests an important role for *ARSE* in maintaining the contractile phenotype of VSMCs. Further, we observed that VSMCs treated with siARSE had increased expression of other important markers of VSMC contractility[31] (Extended Data Fig. 6). In immunoblot analyses, knockdown of *ARSE* was also associated with a ~50% decrease in RUNX2 and a 35% increase in CNN1 protein expression (Fig. 2d). *ARSE* silencing resulted not only in a >60% decrease in calcium deposition by cells treated with osteogenic media for ten days (**Fig. 2e)**, but also in a >3-fold increase in coronary artery VSMC contractility, as measured using a collagen gel contraction assay (Fig. 2f).

Since knockdown of *ARSE* reduced *RUNX2* expression and inhibited VSMC calcification, we predicted that increasing *ARSE* levels with adenovirus (Ad.ARSE) would promote an osteogenic phenotype and calcification of VSMCs. In cells grown in normal media, Ad.ARSE treatment increased RUNX2 protein levels by >7-fold and decreased CNN1 protein levels by ~70% (Fig. 3a). Ad.ARSE treatment resulted in a >10-fold increase in coronary VSMC calcification (Fig. 3b) and led to a >70% decrease in coronary artery VSMC contractility (Fig. 3c). Considering the consistent association of the *ARSE* locus with both coronary and aortic calcification, the same silencing and overexpression experiments were also performed in human aortic VSMCs. The findings in aortic VSMCs resembled those in coronary artery VSMCs (Extended data Fig. 7). Together these results highlight an important role for ARSE in regulating VSMC phenotype switch to osteogenic cells, as well as VSMC calcification underlying the pathogenesis of atherosclerosis and CAC .

To study ARSE expression in human tissue, we used human coronary arteries from control patients and patients with ischemic coronary artery disease (Extended Data Fig. 8a–b). Immunofluorescence analysis showed that ARSE is expressed at a low level in the control coronary arteries. However, ARSE expression is significantly higher in ischemic coronary arteries as compared to control arteries. Colocalization analysis showed that ARSE expression modestly correlates with α-smooth muscle actin expression (a marker of smooth muscle cells). Alizarin red staining was significantly higher in ischemic diseased arteries

as compared to control arteries, corresponding to the increased ARSE expression. We also found that RUNX2 (a marker of calcification) expression was higher in calcified diseased arteries and colocalized with ARSE. Given the findings of ARSE as a regulator of VSMC calcification, we sought to investigate the functional role of the index variant. To validate whether SNP rs5982944 influences gene expression, a luciferase reporter assay was performed. The rs5982944-A allele and rs5982944-G allele were cloned into Luciferase-pcDNA3 vectors and co-transfected with the renilla luciferase vector (internal control reporter) into human coronary artery smooth muscle cells,human aortic smooth muscle cells, and HEK-293 cells. Cells transfected with the rs5982944-G allele showed significantly less luciferase activity as compared to the rs5982944-A allele in all three cell types (Extended Data Fig. 9). This result validates our *in-silico* analyses showing that SNP rs5982944 affects ARSE gene expression. Together these results provide support for the role of increased expression of *ARSE* in the pathogenesis of atherosclerotic coronary calcification.

Discussion

In summary, we performed a GWAS in 22,400 diverse individuals and identified six genetic loci associated with CAC, a strong and independent risk factor for future cardiovascular disease events.[1,2] Of these loci, variants in the *ARSE* and *MMP16* loci have not been previously reported to be associated with CAC.[15–17] *ARSE* is located on the X chromosome, which was excluded from previous GWAS of CAC. Furthermore, the index variant at *ARSE* was largely restricted to population groups with African ancestry. Additionally, by performing functional experiments of *ARSE* in human coronary artery and aortic VSMCs, we implicate *ARSE* as a major regulator of the osteogenic phenotype switch and calcification of VSMCs (Extended Data Fig. 10). Rare variants in the *ARSE* gene that cause congenital ARSE deficiency result in X-linked chondrodysplasia punctata 1 (CDPX1), a disorder characterized by abnormal cartilage and bone development.[32] Specific characteristics of CDPX1 include speckled calcifications at the end of bones, short stature, short fingers, and a depressed nasal bridge. Complications can include extensive calcifications within the tracheal and bronchial cartilage, but not atherosclerotic vascular calcification. The underlying mechanisms by which ARSE deficiency leads to CDPX1 have not yet been elucidated, but disturbed regulation of calcification appears to be a central feature.

We performed a series of functional studies that point towards an important role of *ARSE* in atherosclerotic vascular calcification as a promoter of osteogenic phenotype switch in both coronary artery and aortic VSMCs (Extended Data Fig. 10). Silencing of *ARSE* resulted in decreased levels of markers of the osteogenic phenotype, decreased calcification, and increased VSMC contractility, while overexpression of *ARSE* resulted in increased levels of markers of the osteogenic phenotype, increased calcification, and decreased contractility. Notably, we observed that, in the absence of osteogenic media, *ARSE* not only upregulates production of RUNX2, a master regulator of bone formation and atherosclerotic vascular calcification, but also decreases the expression of the contractile protein calponin.[33,34] Using a luciferase reported assay, we also show that the association of variants near *ARSE* is likely explained by enhancer variant rs5982944, with the G allele reducing *ARSE* expression. Overall, our findings implicate ARSE in the regulation of VSMC phenotype. Because mice

do not carry the *ARSE* gene, further exploration of the mechanism through which *ARSE* affects atherosclerotic vascular calcification will require the use of other animal models or *in vitro* approaches.

In silico approaches pointed towards *MMP16* as a candidate causal gene for CAC. Given previous reports of *MMP16* being expressed in VSMCs in atherosclerotic plaques,[35] we used gene silencing to test whether MMP-16 regulates the phenotype switch from contractile to osteogenic VSMCs. However, we found no evidence for this in our *in vitro* experiments, which may reflect the possibility that MMP-16 may not be working primarily in VSMCs to affect atherosclerotic plaque formation. Instead, it may be that MMP-16 is functioning primarily in endothelial cells, inflammatory cells such as macrophages, or in the extracellular matrix. The MMP-16 protein cleaves several important extracellular matrix proteins, including but not limited to fibronectin, collagen type III, gelatin, laminin-1, vitronectin, and fibrin.[36] These substrates may each have unique roles in the pathogenesis of atherosclerosis and/or calcification. Fibronectin promotes plaque formation in mice, but decreases vulnerability to plaque rupture by promoting the formation of a thick fibrous cap.[37] Importantly, fibronectin may promote atherosclerotic vascular calcification.[38,39] Thus, degradation of fibronectin through cleavage by MMP-16 may lead to decreased calcification: both directly and indirectly through the inhibition of atherosclerosis. In contrast, MMP-16 also activates MMP-2,[36] which promotes atherosclerotic plaque formation in mice,[40,41] and may promote atherosclerotic vascular calcification.[42–44]

In conclusion, we discovered and replicated associations of two loci with CAC that harbor *ARSE* and *MMP16*. By performing functional experiments of *ARSE* in human coronary artery and aortic VSMCs, we implicate *ARSE* as a major regulator of the phenotype switch from contractile to osteogenic VSMCs, and highlight enhancer variant rs5982944 as the putative causal variant. Thus, *ARSE* represents a novel locus associated with CAC that exerts direct effects on atherosclerotic vascular calcification *in vitro*. Our findings highlight *ARSE* as a potential target of therapy for vascular calcific disease.

Methods

Ethics statement

All human research was approved by the relevant institutional review boards for each study and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Study Population

A total of 22,400 participants with TOPMed freeze 6 WGS and harmonized CAC data were included in our discovery analysis. This included participants from ten studies: Genetics of Cardiometabolic Health in the Amish (Amish; n = 596), Coronary Artery Risk Development in Young Adults (CARDIA; n = 2,781), Cardiovascular Health Study (CHS; n = 356), Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene; n = 8,665), Diabetes Heart Study (DHS; n = 381), Framingham Heart Study (FHS; n = 1,557), Genetic Study of Atherosclerosis Risk (GeneSTAR; n = 631), Genetic Epidemiology Network of

Arteriopathy (GENOA; n = 493), Jackson Heart Study (JHS; n = 1,605), and the Multi-Ethnic Study of Atherosclerosis (MESA; n = 5,335). Summaries of the study design of each of these included studies can be found in the Supplementary Methods. The study population was diverse, with reported membership of race/ethnicity population groups as follows: Amish (n = 596), Asian (n = 601), Black (n = 8,434), Central American (n = 64), Cuban (n = 36), Dominican (n = 145), Mexican (n = 541), Puerto Rican (n = 141), South American (n = 81), and White (n = 11,392). Due to their small sample size, the Central American, Cuban, Dominican, Mexican, Puerto Rican, and South American population groups were combined into a single Hispanic/Latino population group (n = 1,010) for association analyses. Additionally, 369 participants had either unreported or non-specific (e.g. 'Multiple' or 'Other') race/ethnicity membership. In order to include these 369 participants in all analyses, their most likely race/ethnicity membership was imputed based on their genetic ancestry using the Harmonized Ancestry and Race/Ethnicity (HARE) method.[45] Further details are provided in the Supplementary Methods and Supplementary Table 13. This research was approved by the institutional review board at the University of Texas Health Science Center at Houston. All included studies were approved by their respective institutional review committees and all included participants gave written informed consent.

Whole genome sequencing

Within TOPMed, WGS was conducted at a mean depth of >30X using Illumina HiSeq X Ten instruments at five sequencing centers. Variant discovery and genotype calling for freeze 6 were conducted jointly across the ten discovery studies, as well as additional studies not included in our current analysis, using the GotCloud pipeline by the TOPMed Informatics Research Center. This procedure resulted in a single genotype call set encompassing all TOPMed studies. Quality control of genetic variants performed by the TOPMed Information Resource Center consisted of the removal of variants failing the support vector machine filter, with excess heterozygosity or Mendelian inconsistencies, or overlapping centromeric or other low complexity regions.[46] Quality control of samples performed by the TOPMed Data Coordinating Center consisted of the removal of duplicate samples pertaining to the same individual, samples with discrepancies between genetic and reported sex, samples with discrepancies between genetically inferred and reported pedigrees, and samples with poor quality based on concordance of WGS and genotyping array data. On the X chromosome, dosages of variants were coded as [0,2] in men and [0,1,2] in women.

Quantification of Coronary Artery Calcification

CAC was determined in a standardized manner from computed tomography images using the Agatston score.[47,48] The primary phenotype used for genetic association analysis was the continuous trait log(CAC+1). As secondary phenotypes, we also analyzed presence of CAC for index variants at significant loci, comparing participants with 1) CAC > 0 to those with CAC = 0, 2) CAC ≥ 100 to those with CAC < 100, and 3) CAC ≥ 400 to those with CAC < 400. Phenotype harmonization was performed by the TOPMed Data Coordinating Center.[49] Further detail regarding the measurement of CAC in the included studies can be found in the Supplementary Methods.

Single Variant Association Tests

Genome-wide tests for single variant association with log(CAC+1) were performed using linear mixed models. The first step of this procedure was to fit the 'null model' under the null hypothesis of no individual genetic variant associations (i.e. without any individual genotype terms in the model). Fixed effect covariates in the null model included age at CAC measurement, sex, study, and the first eleven PC-AiR[50] principal components (PCs) of genetic ancestry. Sex by PC interactions were included because it has previously been reported that the association between race and CAC varies markedly by sex.[51] A 4th degree sparse empirical kinship matrix computed with PC-Relate[52] was included to account for genetic relatedness among participants. Additional details on the computation of the ancestry PCs and the sparse kinship matrix are provided in the Supplementary Methods. We also allowed for heterogeneous residual variances across combined study by race/ethnicity groups (e.g. COPDGene - White), as it has previously been shown that this can improve control of genomic inflation.[53] To improve power and control of false positives with a non-normally distributed phenotype, we implemented a fully-adjusted two-stage procedure for rank-normalization[54] when fitting the null model:

1. **Stage 1:** We fitted a linear mixed model using log(CAC+1) as the outcome, with the fixed effect covariates, sparse kinship matrix, and heterogeneous residual variance model as described above. We performed a rank-based inverse-normal transformation of the marginal residuals, and subsequently re-scaled by their original variance. This re-scaling allows for clearer interpretation of estimated genotype effect sizes from the association tests.
2. **Stage 2:** We fitted a second linear mixed model using the rank-normalized and re-scaled residuals as the outcome, with the same fixed effect covariates, sparse kinship matrix, and heterogeneous residual variance model as in Stage 1.

The output of the Stage 2 null model was then used to perform genome-wide score tests of genetic association for all individual genetic variants with minor allele count ≥50. Genome-wide significance was determined at the $P < 5\times10^{-9}$ level. Sex- and population-specific analyses are described in the Supplementary Methods. All association tests were performed using the GENESIS software, and the primary analyses assumed an additive genetic model.[55]

In addition to our hypothesis-free genome-wide approach, we performed a focused analysis of 241 variants associated with CAD by Aragam et al.[14] A Bonferroni correction for the number of tested variants was performed to determine the appropriate significance threshold.

Conditional Analyses

Conditional single variant association tests were then performed separately by locus to assess the number of independent association signals. For each locus with one or more genome-wide significant ($P < 5\times10^{-9}$) variants, a conditional association analysis was performed by including the variant with the most significant P-value from the locus as a fixed effect covariate in the null model and then testing all other variants within ±250 kb. The same fully-adjusted two-stage linear mixed model association testing procedure

as described for the main analysis was used for the conditional analyses. The Benjamini-Hochberg procedure was applied, and secondary signals with a false discovery rate (FDR) < 0.05 were considered statistically significant. This conditioning process was repeated iteratively, adding the variant with the most significant P-value after each round of conditioning as another fixed effect covariate in the null model, until no variants at the locus had an FDR < 0.05.

Secondary Analysis of Presence of CAC

The association testing procedure for presence of CAC using thresholds of CAC > 0, CAC ≥ 100, and CAC ≥ 400 was very similar to that for log(CAC+1), but with a few differences. Rather than fitting a linear mixed model, we fitted a generalized linear mixed model with binomial family and logit link via the penalized quasi-likelihood approach of GMMAT.[56] Because the variance model of a generalized linear mixed model is specified by the family and link function, the heterogeneous residual variance groups did not apply and were not used. Additionally, the two-stage rank-normalization procedure did not apply to this model. Therefore, the null model was fitted using a single stage generalized linear mixed model with the binary outcome and the same fixed effects and sparse kinship matrix as the main analysis. The null model output was then used to perform genome-wide single variant score tests with saddle point approximation of the P-values.[57,58] Saddle point approximation provides better calibration of P-values than the traditional score P-value when fitting a mixed model for a binary trait in a sample with an imbalanced case to control ratio, particularly for low frequency and rare variants.

Gene-based Aggregate Rare Variant Association Tests

Multi-variant association tests were also performed genome-wide to assess the cumulative association of rare variants in gene-centric aggregation units based on the GENCODE V24 gene models[59] with log(CAC+1). Two annotation-based aggregation and filtering strategies were implemented. The first strategy focused primarily on variants in protein-coding regions and included variants that were high-confidence loss-of-function variants according to the Loss-of-Function Transcript Effect Estimator (https://github.com/konradjk/loftee), missense variants with MetaSVM_score>0,[60] inframe insertions, inframe deletions, or synonymous variants with fathmm_XF_coding_score>0.5. The prediction of the consequences of variants were obtained from the Ensembl variant effect predictor.[61] The second strategy included variants from the first strategy as well as regulatory variants either proximal to a gene or predicted to be regulating the gene. Specifically, we retained variants within the upstream 5 Kb region (putative promoters) of a gene or in GeneHancers[62] (putative enhancers) that were labelled by Ensembl regulatory build annotation[63] as promoters, promoter flanking regions, enhancers, CTCF binding sites, transcription factor binding sites, or open chromatin regions. Only regulatory variants with a Fathmm-XF score>0.5 were retained.[64] The annotation based variant filtering and gene-centric aggregation was performed using a local MySQL database built from annotations generated by the Whole Genome Sequence Annotator version v0.75[65] and formatted using WGSAParsr version 5.0.9. After performing annotation-based aggregation and filtering, variants were further filtered to those that were non-monomorphic with MAF < 1% among study participants, and had less than 10% of samples with sequencing read depth < 10 at that particular variant. The aggregate

association testing was performed using the Efficient Variant-Set Mixed Model Association Test (SMMAT).[66] The SMMAT test used the same null model as was fitted for the single variant association tests. For each aggregation unit, SMMAT efficiently combines a burden test P-value with an asymptotically independent adjusted SKAT test P-value using Fisher's method. This testing approach is more powerful than either a burden or SKAT test alone, and is computationally more efficient than the SKAT-O test.[67] Wu weights[68] based on the variant MAF were used to 'upweight' rarer variants in the aggregation units. Only aggregation units with a cumulative minor allele count≥5 across all included variants were tested for association. Statistical significance was determined using a Bonferroni threshold of 1.2×10^{-6}, corresponding to the 41,868 aggregation units tested across both strategies.

Replication

We replicated the association of CAC with the index variants at the two newly-identified loci in independent samples from two race/ethnicity groups: Black and Brazilian. The Black population group consisted of a meta-analysis of 787 BioImage participants and 583 participants from MESA who did not have WGS data as a part of TOPMed, while the ancestrally heterogeneous Brazilian population group corresponded to 1758 participants of the ELSA study (Supplementary Methods). Association analyses for the *ARSE* index variant were performed using both additive and recessive models. A Bonferroni-adjusted significance threshold of P<0.0083 was used to adjust for six statistical tests. Additional details about these analyses can be found in the Supplementary Methods.

Association of novel CAC loci with other atherosclerosis phenotypes and risk factors

We examined the association of index variants at the two loci with a range of other atherosclerosis phenotypes and outcomes, including carotid intima media thickness, carotid plaque, thoracic aortic calcification, and CAD events. Association analyses for the *ARSE* index variant were performed using both additive and recessive models. We used WGS data from nine TOPMed studies to examine associations with carotid intima media thickness and from seven TOPMed studies to examine associations with carotid plaque (Supplementary Methods). We used WGS data from COPDGene to examine associations with thoracic aortic calcification. For the *MMP16* index variant we used a trans-ethnic meta-analysis of the Million Veterans Program (MVP) to determine the association with CAD events. For the *ARSE* index variant, we restricted association analyses for CAD events to Black MVP participants. We used results from published WGS-based GWAS in TOPMed to examine the association of the ARSE and MMP16 index variants with type 2 diabetes, systolic blood pressure, diastolic blood pressure, high-density lipoprotein, low-density lipoprotein, and natural log transformed triglycerides.[69–72] A Bonferroni-adjusted significance threshold of P<0.0021 was used to adjust for 24 statistical tests. Additional details about these analyses can be found in the Supplementary Methods.

Fine-mapping studies

Causal variant credible sets—We used previously described Bayesian methods to define 95% credible sets of variants for each locus, corresponding to the smallest set of variants at a locus that has a 95% posterior probability of containing the causal variant.[24]

Gene expression analyses—We integrated CAC GWAS summary level data with aortic artery gene expression quantitative trait locus (eQTL) summary data for *ARSE* and *MMP16* from GTEx v8 (n=387). To account for linkage disequilibrium (LD) at each locus, we used the cosmopolitan 1000 Genomes phase3 reference panel to generate the LD matrix for *MMP16* and the African 1000 Genomes phase 3 reference panel for *ARSE*.[73] Given the strong association in GTeX v7 of the *ARSE* top variant with *ARSE* expression in cultured fibroblasts, we repeated the analyses for *ARSE* using GTEx v8 data from cultured fibroblasts (n=483).

We used the R package "coloc" to identify colocalizing signals between our CAC GWAS data and cis-eQTLs.[74] Coloc is a Bayesian statistical approach that calculates the posterior probability that the GWAS and eQTL associations share a common signal. For each locus, we calculated the posterior probability of a shared signal, assuming 1 causal variant. Locuscompare plots (https://github.com/boxiangliu/locuscomparer) were used to visualize top colocalizing signals using either European or African 1000G ph3 reference populations to visualize the LD.

Chromatin interaction analysis—The 95% credible set of CAC GWAS SNPs were intersected with pooled HiC-based pairwise chromatin interactions as well as H3K27ac HiChIP pairwise enhancer-promoter interactions, both generated in human VSMCs (Mumbach et al., 2017; Zhao et al., 2020). These chromatin loops were visualized at the *MMP16* locus on the latest Roadmap Epigenomics browser. Results shown are for GRCh38 human genome build.

Functional Studies

Cell lines—Human coronary artery (Cat# 350-05a) and aortic (Cat# 355-75a) VSMCs were obtained from Cell Applications. HEK-293 cells were obtained from ATCC (Cat #CRL-1573).

Plasmid construction, transfection and luciferase reporter assay—The rs5982944-A allele and rs5982944-G allele constructs (+/− 50 base pairs) were synthesized and subcloned into Luciferase-pcDNA3 plasmid (GenScript USA Inc.) Luciferase-pcDNA3 plasmid was a gift from Dr. William Kaelin (Addgene plasmid # 18964; http://n2t.net/addgene:18964; RRID: Addgene_18964). The rs5982944-A allele and rs5982944-G allele constructs were inserted in between the CMV promoter and coding sequence of the luciferase gene. These constructs were transfected into human coronary smooth muscle cells (HCSMCs), human aortic smooth muscle cells (HASMCs) and HEK-293 cells according to manufacturer protocol (Invitrogen). Briefly, HCSMCs, HASMCs, and HEK-293 cells were transfected with 5μg of rs5982944-A allele and rs5982944-G allele constructs using lipofectamine 3000 (#L3000015, Invitrogen). Renilla plasmid (pRL-CMV, #E2261, Promega) was co-transfected with the luciferase constructs (1:100). Cells were harvested 72 hr post-transfection and luciferase activity was measured using the dual-luciferase reporter assay (Promega) according to the manufacturer's instructions. Firefly luciferase activity was normalized to renilla luciferase activity.

siRNA-mediated silencing of *ARSE* and *MMP16*—siRNA directed against *ARSE* (siARSE, # L-008578-00-0005), *MMP16* (siMMP16, # L-005957-00-0005), and scrambled control siRNA (siCTRL, # D-001810-10-05) were obtained from Horizon discovery (Dharmacon). Cells were transfected with siRNA using Lipofectamine RNAiMAX reagent (ThermoFisher Scientific, #13778150), according to the manufacturer's instructions .

Adenovirus-mediated over-expression of *ARSE*—Recombinant adenoviruses expressing either human wild-type ARSE (NM_00047) with Enhanced Green Fluorescent Protein (eGFP) as a reporter under its own promoter or expressing eGFP alone were obtained from Vector Biolabs (human adenovirus type 5 [dE1/E3], promoter: cytomegalovirus, Catalog #1768, Malvern, PA). Human coronary artery and aortic VSMCs were transduced with the adenovirus vectors in regular growth medium. After 24hrs, cells expressing eGFP were detected using fluorescent microscopy. Protein levels were analyzed by immunoblot five days after transduction.

Measurement of gene expression by quantitative RT-PCR—Total RNA from aortas and cultured cells was extracted by the phenol/guanidinium method.[75] Reverse transcription was performed using the High-Capacity cDNA Reverse Transcription Kit (#4368814, Applied Biosystems, Foster City, CA, USA). A Mastercycler ep Realplex (Eppendorf, Hamburg, Germany) was used for real-time amplification and quantification of transcripts. Relative expression of target transcripts was normalized to levels of 18S ribosomal RNA (#4331182).[76] Taqman® gene expression assays were used to quantify mRNA levels of *ARSE* (#4331182), *MMP16* (#4331182), *RUNX2* (#4331182), *BGLAP* (#4331182), *ALPL* (#4331182), *CNN1(*#4331182*), ACTA2* (#4331182), *TAGLN* (#4331182), and *ARSD* (#4331182). All of the Taqman primers for qPCR assays were procured from ThermoFisher Scientific, USA.

Immunoblot techniques—Human VSMCs were homogenized in RIPA buffer containing protease and phosphatase inhibitors (Sigma). Lysates (20 μg/lane) were mixed with denaturing buffer (1× Laemmli loading buffer with 10% of β-mercaptoethanol) and analyzed by SDS–PAGE/Western. Rabbit polyclonal anti-ARSE antibodies were used to detect the ~70 kDa isoform of human ARSE in human VSMCs (Abcam, ab238485, 1:1000). A rabbit polyclonal antibody directed against the human RUNX2 (Santa Cruz, sc-10758, 1:1000) and a rabbit monoclonal antibody directed against CNN1 (Abcam, ab46794, 1:1000) were used to detect RUNX2 and calponin protein, respectively. A rabbit polyclonal and mouse monoclonal antibody directed against glyceraldehyde 3-phosphate dehydrogenase (GAPDH, Cell Signaling #2118, 1:5000) and vinculin (VCL, Santa Cruz, sc-25336, 1:1000) were used to detect GAPDH and VCL protein, respectively. Blots were incubated with fluorescent-dye labeled anti-rabbit IgG IRDye 800CW (LI-COR, Lincoln, NE, #926-32211, 1:10,000) and protein bands were imaged using a LI-COR Odyssey detection system (LI-COR, Lincoln, NE).

Calcification of human VSMCs—To stimulate calcification, cells were treated with osteogenic media: Dulbecco's modified eagle medium supplemented with 10% fetal calf serum, 10 mM β-glycerophosphate disodium, 50 μg/mL L-ascorbic acid, and 10 nM

dexamethasone, as previously described.[6,77] Osteogenic media was replaced every 48 hours for 7-10 days. Cells were either harvested for expression studies (see below) or fixed in 10% formalin and incubated with Alizarin Red (Sigma) to detect calcification. Cells were treated with a 2% Alizarin Red solution (pH 4.1-4.3) for 40 minutes, followed by multiple washes with distilled water.

Collagen matrix cell contraction assay—VSMC contraction was measured by the extent of deformation of collagen lattices, as previously described.[78] Cells were treated with siRNA or adenovirus for 24 hours prior to being embedded in collagen matrices, per manufacturer's protocol (Cell Contraction Assay #CBA-201, Cell Biolabs, Inc). After 48 hours, the collagen lattice was released from the culture dish. Upon releasing the collagen lattice, the embedded cells are free to contract the deformable collagen lattice, resulting in a reduction of the lattice surface area. After detachment of the collagen gel lattice from the dish, changes in the gel surface area were quantified using ImageJ software.

Human coronary artery tissue procurement

All research described herein complies with ethical guidelines for human subjects research under approved institutional review board protocols at Stanford University (no. 4237 and no. 11925) and the University of Virginia (no. 20008), for the procurement and use of human tissues and information, respectively. Freshly explanted hearts from orthotopic heart transplant recipients were obtained at Stanford University under approved institutional review board protocols with written, informed consents. Hearts were arrested in cardioplegic solution and rapidly transported from the operating room to the adjacent laboratory on ice. The proximal 5-6 cm of three major coronary arteries (LAD, LCX, RCA) were dissected from the epicardium, trimmed of surrounding adipose, rinsed in cold PBS and snap-frozen in liquid nitrogen. Human coronary artery tissue biospecimens were also obtained at Stanford University from non-diseased donor hearts rejected for orthotopic heart transplantation and processed following the same protocol as hearts for transplant. Reasons for rejected hearts included size incompatibility, risk for cardiotoxicity or comorbidities. Tissues were de-identified and clinical and histopathology information was used to classify ischemic, non-ischemic hearts and calcific and non-calcific arterial segments. All normal arteries originated from hearts with left ventricular ejection fraction (LVEF) greater than 50%. Frozen tissues were transferred to the University of Virginia through a material transfer agreement and IRB approved protocols.

Coronary and carotid artery atherosclerosis single-cell meta-analysis

QC and normalization of scRNA-seq sequencing libraries: Raw count matrices from each library across the 4 studies were downloaded from GEO and Zenodo. The 22 sequencing libraries were processed in a standardized manner using Seurat v4 and first pass clustering with SCTransform normalization without removing low-quality cells. Doublets were removed using scDblFinder and ambient RNA removed using DecontX. The decontaminated raw count matrices were loaded into each Seurat object and the following quality filters were used to retain cells that had 1) >= 200 and <= 4000 uniquely expressed genes 2) >= 200 and <= 20000 UMIs 3) <= 10% of reads mapped to the mitochondrial genome. To avoid clustering results confounded by cell cycle state, cell cycle variance

was regressed out during SCTransform normalization. We then carried out dimensionality reduction of the normalized count matrix using PCA. The first 30 principal components (PCs) were used as input for clustering in Seurat, which relies on a Shared-Nearest-Neighbors (SNN) and Louvain community detection approach. We then applied Uniform Manifold Approximation and Projection (UMAP) non-linear dimensionality reduction using the first 30 PCs. UMAP embeddings were used for visualization of Louvain clustering results. Processed matrices were then stored as seurat objects for batch-correction.

Integration, cell type annotation and subclustering: We used reciprocal PCA (rPCA) method in Seurat to harmonize the processing sequencing libraries based on several benchmarking metrics, including running time, clustering purity and biological conservation scores. Upon integration we performed dimensionality reduction of the data using PCA and a shared nearest neighbors (SNN) graph was constructed using 50 nearest neighbors and the first PCs as input. Gene markers for each cluster were identified in Seurat v4 using PrepSCTMarkers() and FindAllMarkers() functions. Differentially expressed genes by cluster were considered for genes expressed in >25% of the clusters being compared (one versus all) at log Fold change =0.25 and B-H adjusted p-value < 0.05. Clusters were annotated using a combination of manual curation and comparison with lineage training datasets, as well as automated label transfer from the SCTransform processed Tabula Sapiens vasculature reference. Subclustering analysis was performed by subsetting the integrated object to SMCs, and then identifying the variable features, repeating the dimensionality reduction and embedding, SCTransform normalization and then repeating the PCA and UMAP embedding. Feature plots were generated for candidate genes *ARSE* and *MMP16.*

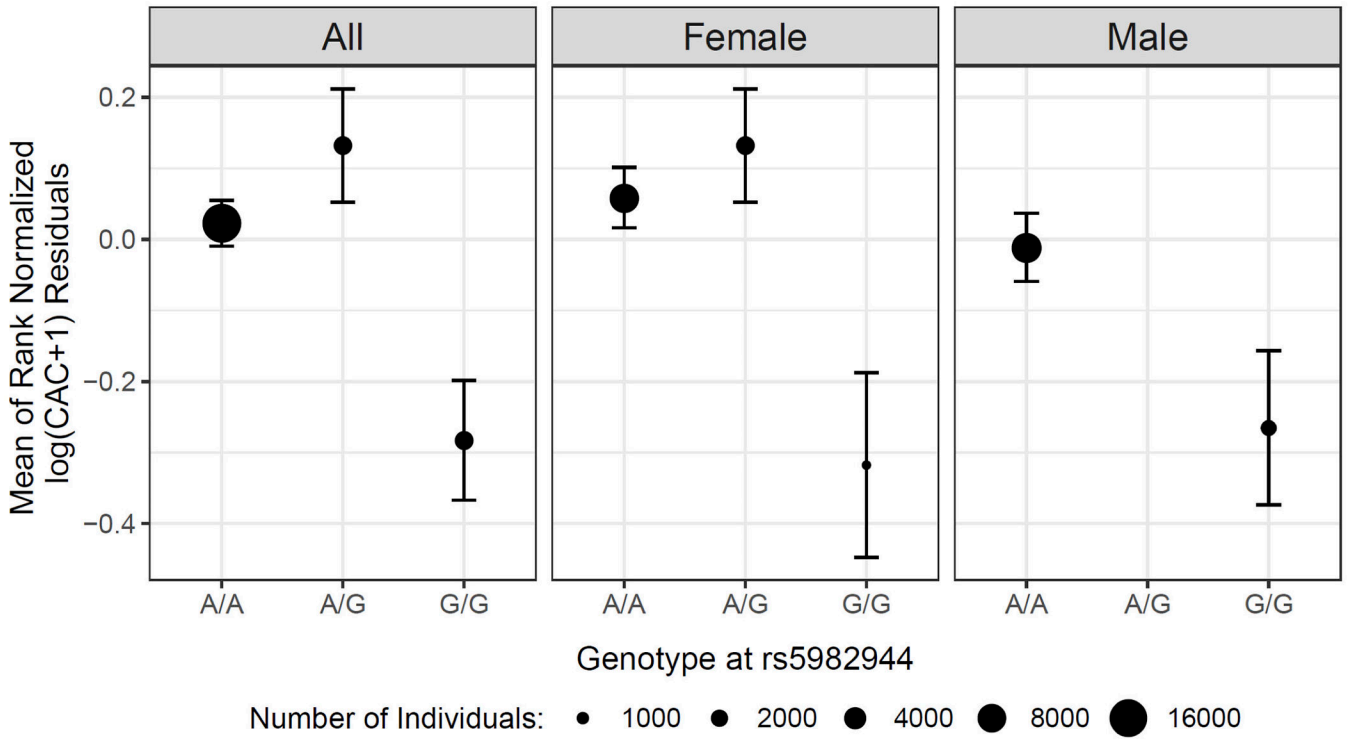
Immunofluorescence—Ischemic and normal explanted human coronary artery tissue biospecimens were obtained from diseased heart transplant donors consenting for research studies as previously described.[79] Coronary arteries were embedded in optimal cutting temperature (OCT) and cryosectioned at 8mm**.** Frozen slides were washed with sterile PBS twice for 2 min, followed by fixation with formaldehyde at 4% for 10 min. Then slides were washed with PBS twice for 2 min, and tissue sections were permeabilized with Triton-x-100 at 0.05% for 10 min. Coronary artery sections were blocked with 10% donkey serum for 1 hour, followed by incubation overnight at 4°C with anti-ARSE antibody (Abcam #ab238485), FITC-conjugated anti-α-smooth muscle actin (Sigma #F3777), and anti-RUNX2 (Santa Cruz Biotechnology, #sc-390715) at 1:100 dilution. Slides were washed with 0.1% PBS-tween three times for 3 min each, followed by incubation with secondary antibodies (CyTM3 AffiniPure Donkey anti-Rabbit IgG (H+L), Jackson immunoresearch #711-165-152; ThermoFisher Scientific, USA, #A32787) at 1:400 dilution for 1 hr at room temperature. Then, slides were washed with PBS-tween at 0.1%, 4 times for 3 min each, and slides were mounted with a diamond mounting medium containing DAPI. Immunofluorescence images were captured using Nikon SoRa spinning disk confocal microscope and Leica TCS SP8 confocal microscope. For calcification, coronary arteries were fixed with 4% paraformaldehyde, followed by washing with distilled water. Coronary artery sections were stained with 1% Alizarin red S solution (Sigma-Aldrich #A5533).

Statistical analysis—Statistical analysis of the functional studies was performed using Graph Pad Prism 8.0 (GraphPad Software, La Jolla, CA) and Stata 13.0 (StataCorp LLC). The Shapiro-Wilk test was used to determine the normality of each continuous variable, and all such variables were found to be normally distributed. Data are reported as mean ± standard error unless otherwise indicated. Two group comparisons of continuous variables were performed using the two-tailed Student t test. For more than two group comparisons of continuous variables, two-tailed 1-way analysis of variance (ANOVA) with Sidak post-hoc testing was employed. All *in vitro* experiments were performed at least in duplicate. A two-sided $p<0.05$ was considered to indicate statistical significance.

Data Availability

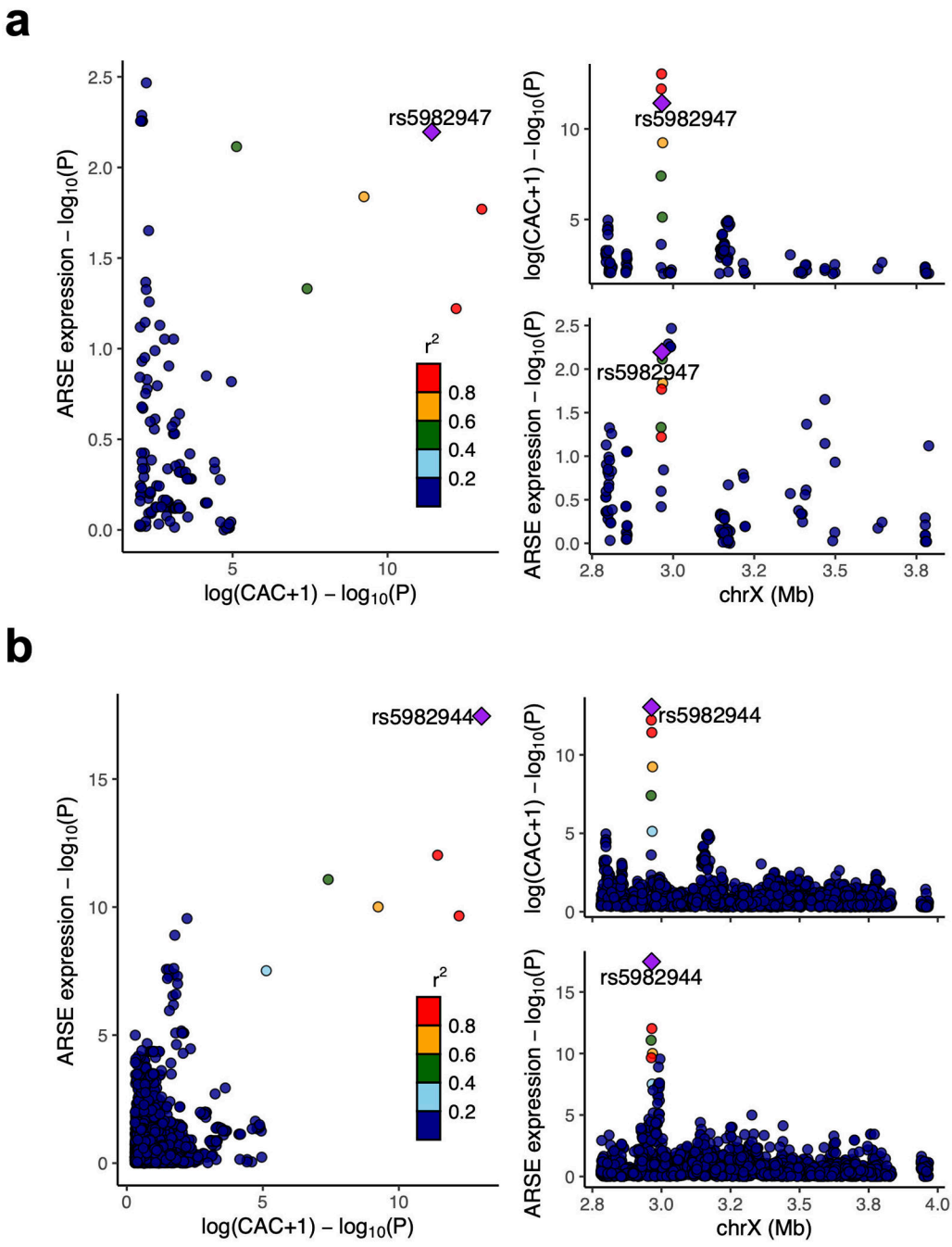
All data supporting the findings in this study are included in the main article and associated files. An overview of the TOPMed participant consents and data access procedures is provided at https://topmed.nhlbi.nih.gov/topmed-data-access-scientific-community. Participant-level genotype and phenotype data are available to approved investigators via dbGaP. The dbGaP accession numbers for all TOPMed studies referenced in this paper are listed in Supplementary Table 14. Additionally, genomic summary results pertaining to the pooled GWAS of CAC score are available at phs001974, as detailed at https://topmed.nhlbi.nih.gov/topmed-genomic-summary-results-public.

Extended Data



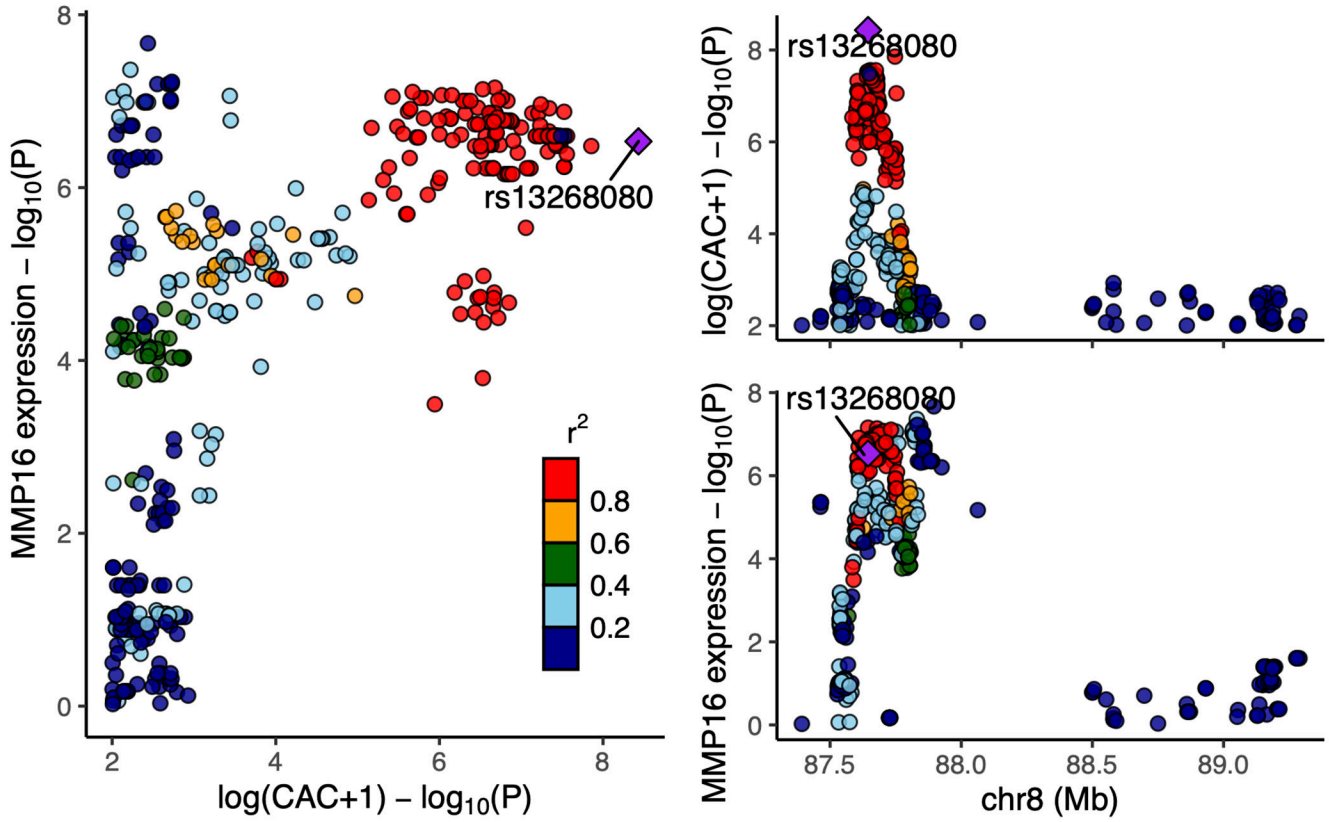
Extended Data Fig. 1: Coronary artery calcification levels across *ARSE* index variant genotypes suggest a recessive mode of inheritance.

Mean log (coronary artery calcification + 1) across genotypes for index variants at ARSE. Error bars indicate the 95% confidence interval for the mean (n=22,400).



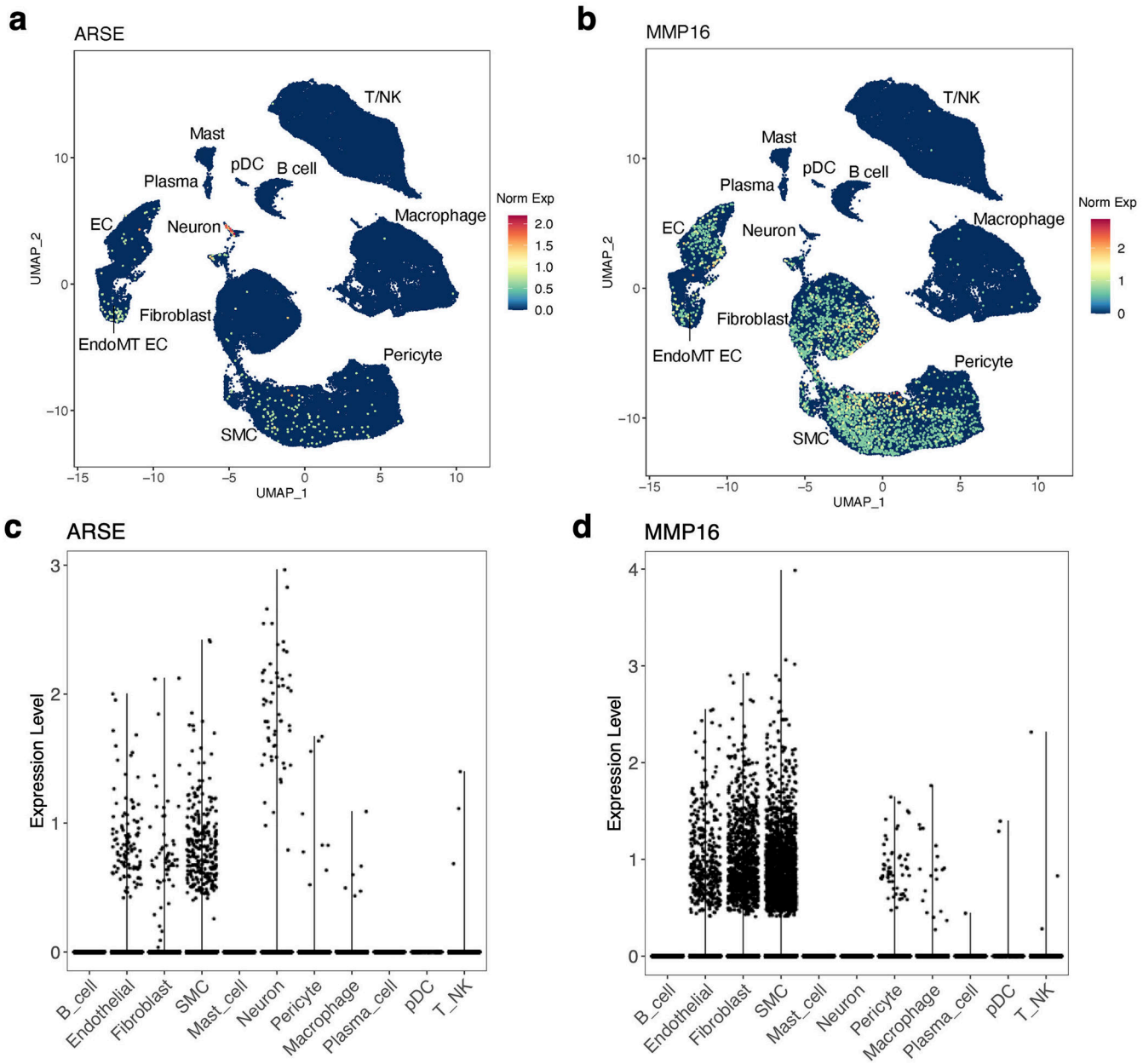
Extended Data Fig. 2: Co-localization of genetic associations with coronary artery calcification and genetic associations with *ARSE* expression in a) cultured fibroblasts and b) aorta.

These plots were created with Locuscomparer (https://github.com/boxiangliu/locuscomparer) using the African 1000G ph3 reference population to calculate the linkage disequilibrium r^2. Unadjusted two-sided P values are provided. The posterior probability for causal variant sharing was 99.9% in cultured fibroblasts and 8.9% in aorta.



Extended Data Fig. 3: Co-localization of genetic associations with coronary artery calcification and genetic associations with *MMP16* expression in aorta.

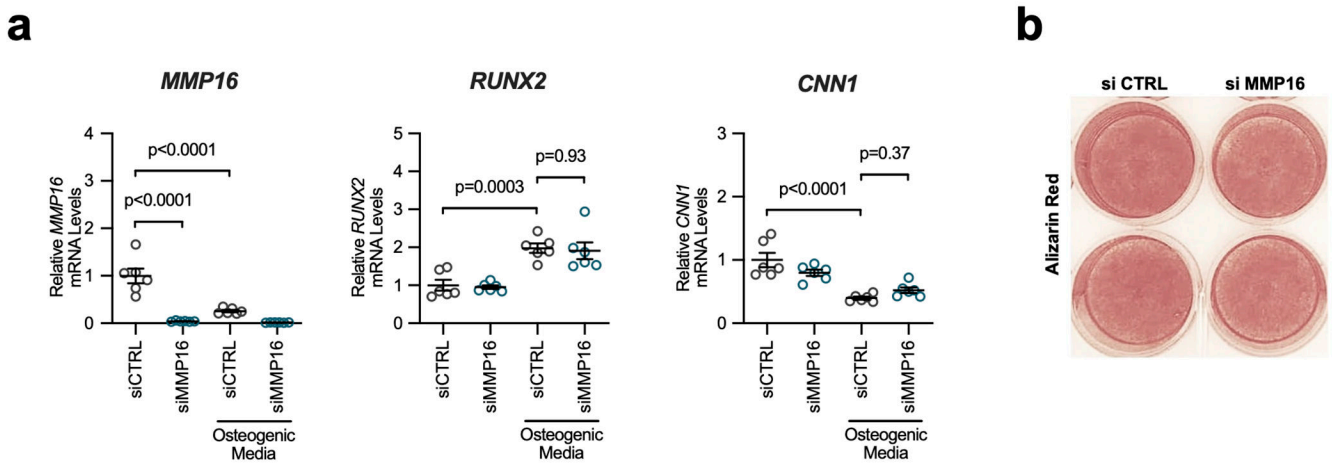
This plot was created with Locuscomparer (https://github.com/boxiangliu/locuscomparer) using the European 1000G ph3 reference population to calculate the linkage disequilibrium r^2. Unadjusted two-sided P values are provided. The posterior probability for causal variant sharing was 81.9%.



Extended Data Fig. 4: Cell type-specific gene expression of *ARSE* and *MMP16* in an integrated human atherosclerosis reference dataset.

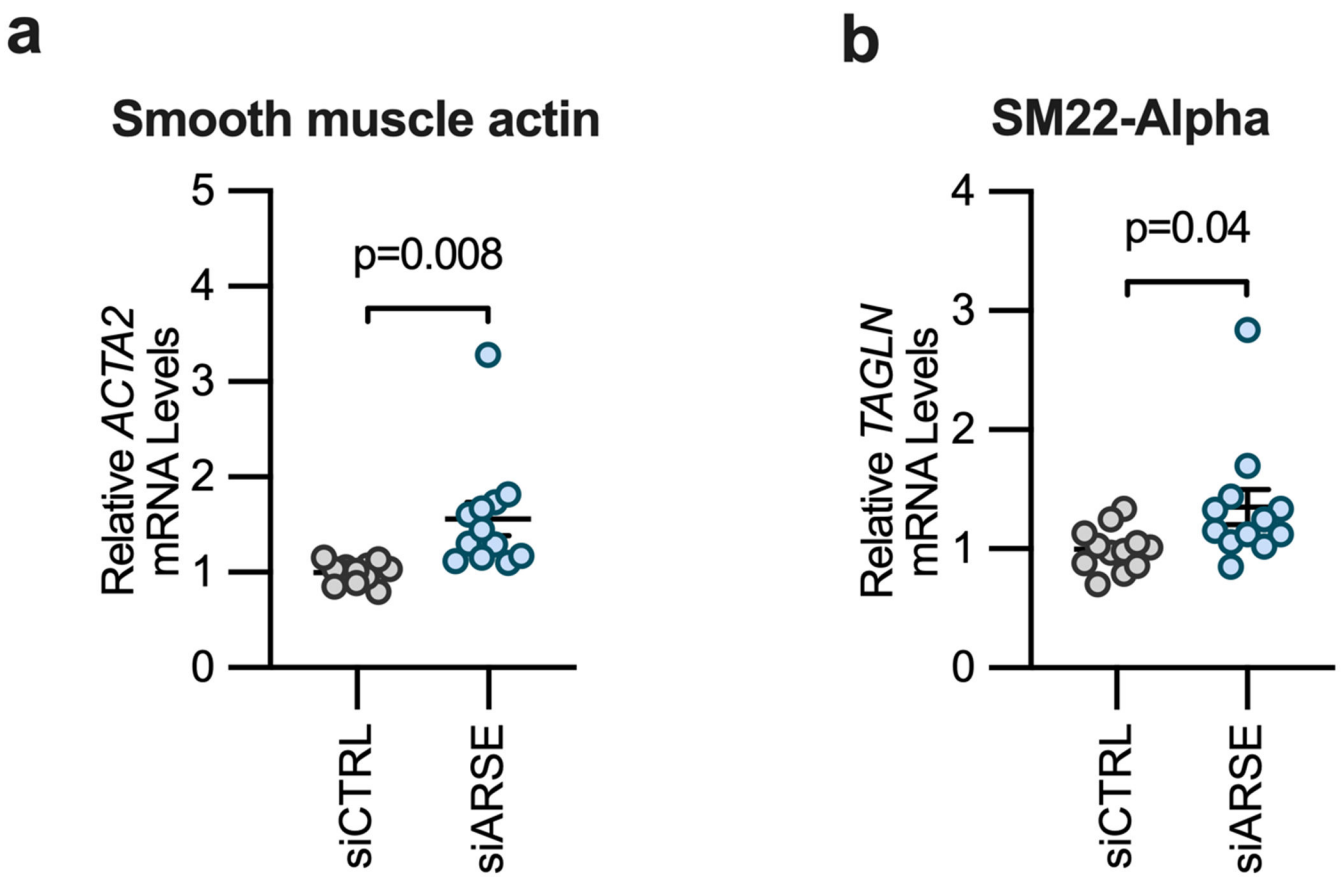
(a-b) Uniform Manifold Approximation and Projection (UMAP) embeddings from an integrated human carotid and coronary artery atherosclerosis single-cell RNA-seq reference dataset (Methods), showing **(a)** *ARSE* and **(b)** *MMP16*, normalized gene expression depicted by the heatmap from SCTransform normalized read counts. Individual sequencing libraries across four studies were harmonized after QC and batch correction with reciprocal PCA (rPCA). Clusters were annotated with level 1 cell type labels using transfer learning with cell labels from the Tabula Sapiens vasculature subset. Level 2 cell type label for endothelial-mesenchymal transition (EndoMT) endothelial cells expressing ARSE and MMP16 are also highlighted. **(c-d)** Scatter plots showing the normalized expression level

of **(c)** *ARSE* and **(d)** *MMP16*, across the level 1 cell types. EC: Endothelial cells; SMC: Smooth muscle cells; T/NK: T cells and Natural Killer cells; pDC: plasmacytoid dendritic cells.



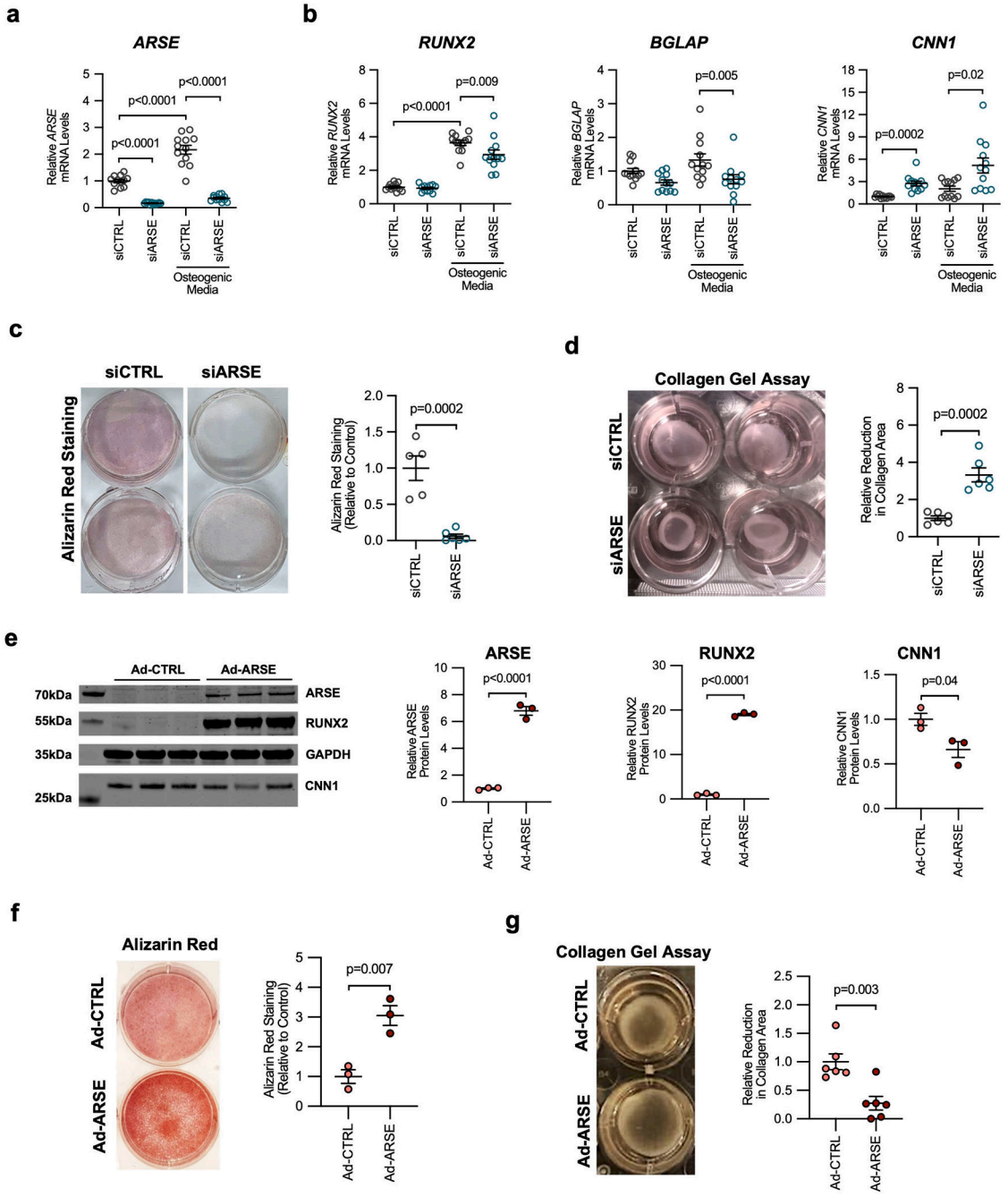
Extended Data Fig. 5: Silencing *MMP16* expression has no effect on osteogenic phenotypic switching in human coronary artery vascular smooth muscle cells.

a) Treatment of human coronary artery vascular smooth muscle cells (n = 6 biologically independent samples in each group) with osteogenic media decreased *MMP16* mRNA expression by ~74% (left panel). Treatment of cells grown in osteogenic media with siMMP16 (resulting in >90% knockdown of *MMP16* mRNA) had no effect on *RUNX2* (middle panel) or *CNN1* (right panel) mRNA levels. Statistical comparisons were made using a two-tailed one-way ANOVA with Sidak's post-hoc comparison testing. The mean ± SEM is depicted in plots. **b)** Treatment of human coronary artery vascular smooth muscle cells grown in osteogenic media with siMMP16 had no effect on calcification, as evidenced by decreased Alizarin Red S staining.



Extended Data Fig. 6: Silencing *ARSE* expression increases contractile gene expression in human aortic vascular smooth muscle cells.

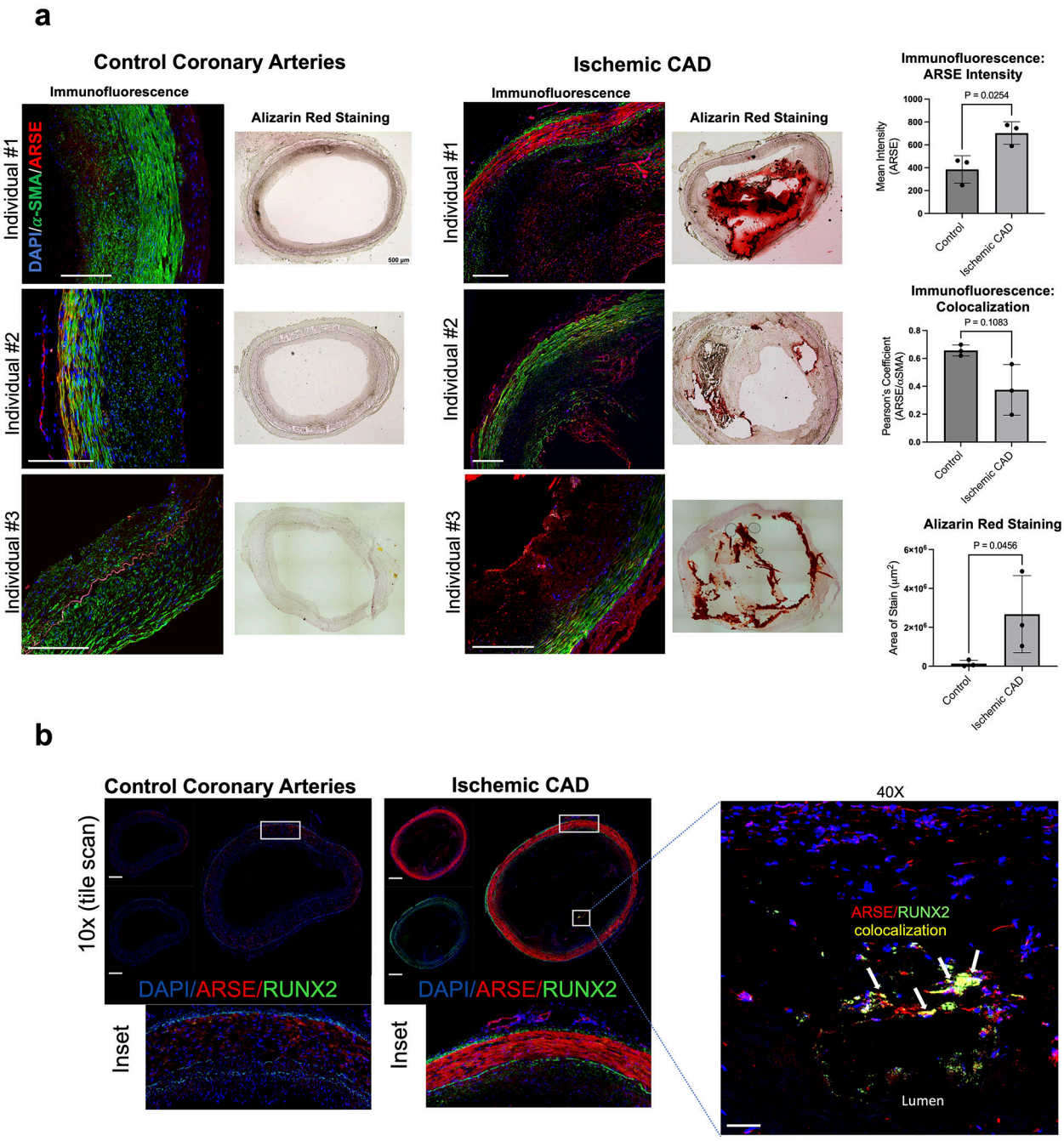
Silencing ARSE in cells (n = 12 biologically independent samples in each group) grown in normal media increased **a)** *ACTA2* and **b)** *TAGLN* mRNA levels by ~ 56% and 35%, respectively. Statistical comparisons were made using a two-tailed Student t test. The mean is depicted in plots, with the error bars representing the standard error of the mean.



Extended Data Fig. 7: Human aortic vascular smooth muscle cell calcification, bone and contractile marker expression, and contractility are affected by changes in *ARSE* expression.

a) Treatment of human aortic vascular smooth muscle cells (n = 12 biologically independent samples in each group) with osteogenic media increased *ARSE* mRNA expression > 2-fold. **b**) Treatment of cells grown in osteogenic media with siARSE (resulting in >90% knockdown of *ARSE* mRNA) decreased *RUNX2* (left panel), and *BGLAP* (middle panel) mRNA levels by ~20% and ~43% respectively, and increased CNN1 mRNA levels by > 150% (right panel). Silencing ARSE in cells grown in normal media increased CNN1

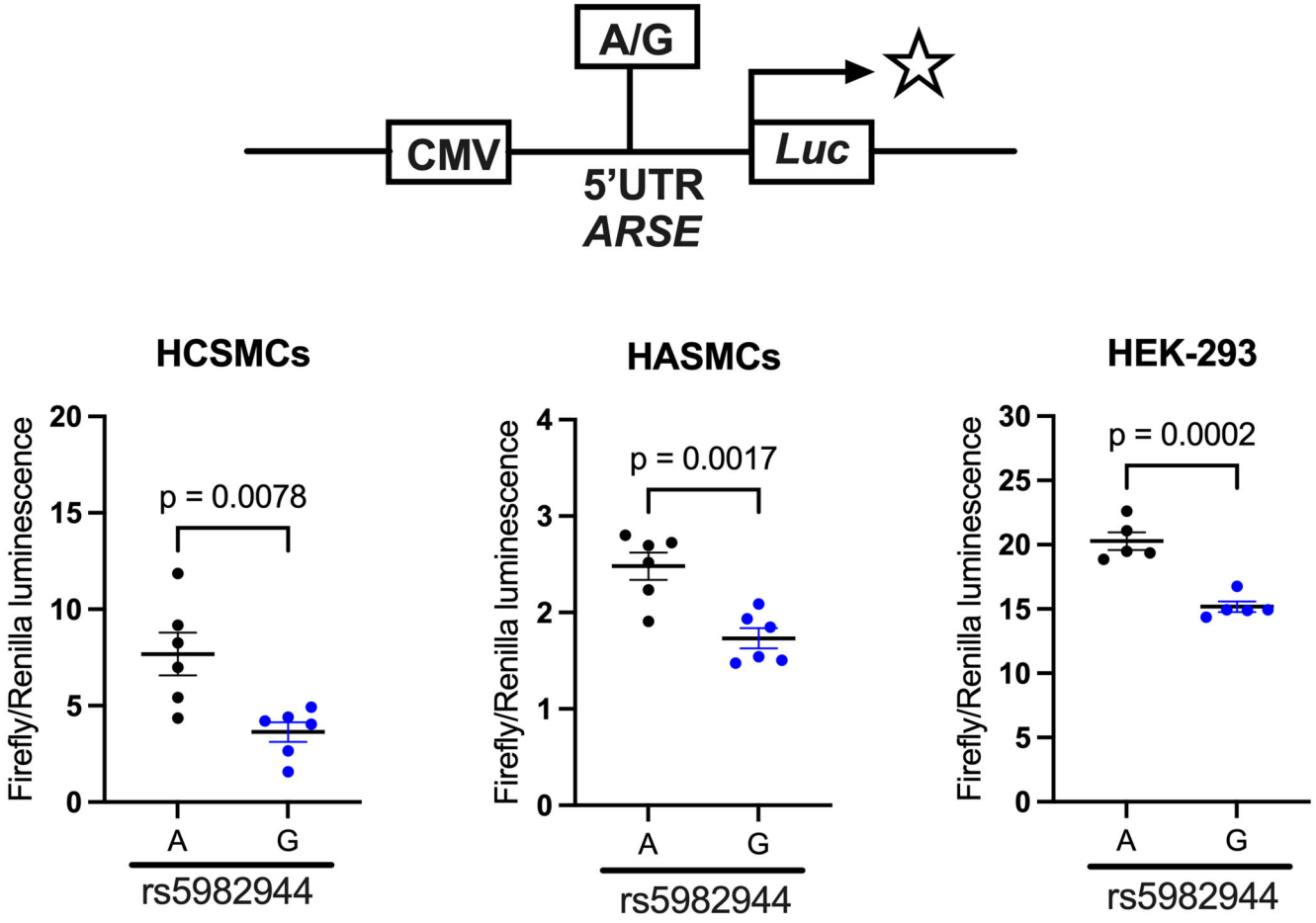
mRNA levels by > 2.5-fold. **c**) Treatment of cells grown in osteogenic media with siARSE reduced calcification, as evidenced by decreased Alizarin Red S staining (n=5 biologically independent samples in each group). **d**) Reduced ARSE expression in cells grown in collagen discs (left panel) resulted in a >3-fold increase in contraction (right panel, n=6 biologically independent samples in each group). **e)** Protein expression of ARSE, RUNX2 and CNN1 were confirmed by Western blot using antibodies directed against ARSE, RUNX2, CNN1 and GAPDH (for a loading control). Adenoviral expression of the 70-kDa isoform of ARSE in human aortic vascular smooth muscle cells was associated with a >15-fold increase in RUNX2 protein levels and an approximately 34% decrease in CNN1 protein levels, when cells were harvested 5 days after viral transduction (n=3 biologically independent samples in each group). **f**) As shown by Alizarin Red staining, increased ARSE expression resulted in augmented calcification in human aortic vascular smooth muscle cells (n=3 biologically independent samples in each group). **g**) Increased ARSE expression also caused a >70% decrease (right panel, n=6 biologically independent samples in each group) in contraction of human aortic vascular smooth muscle cells grown in collagen discs (left panel). Statistical comparisons were made using either a two-tailed one-way ANOVA with Sidak's post-hoc comparison testing or a two-tailed Student t test. The mean is depicted in plots, with the error bars representing the standard error of the mean.



Extended Data Fig. 8: ARSE expression and calcification in normal and ischemic human coronary arteries.

a) Cross sections of human coronary arteries from control subjects and patients with ischemic coronary artery disease (n=3 individuals in each group with 2 sections stained for each individual) were stained for ARSE (red), α-smooth muscle actin (green) and DNA (blue, DAPI). Immunofluorescence analysis shows a higher expression of ARSE in diseased arteries. Alizarin red staining for calcification was high in the coronary arteries of diseased patients with no significant stain observed in the control group. Scale bars, 200 μm for

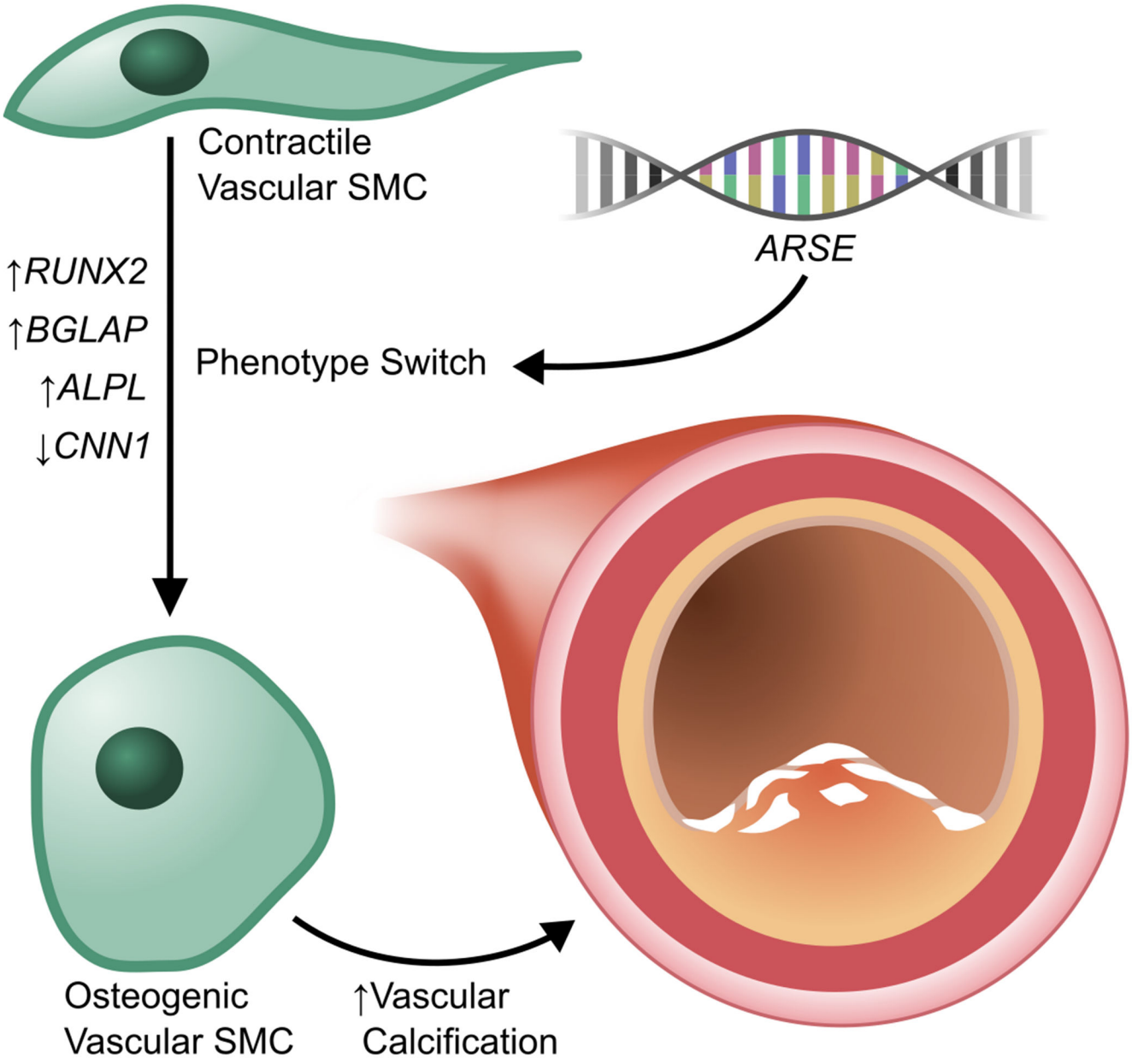
each immunofluorescence image; 500 μm for each Alizarin red staining image. Statistical comparisons were made using a two-tailed Student's t test. The mean is depicted in plots, with the error bars representing the standard error of the mean. **b)** Cross sections of human coronary arteries (n=1 each for control and ischemic patient) were stained for ARSE (red), RUNX2 (green, VSMC calcification marker) and DNA (blue, DAPI). Immunofluorescence analysis shows a higher expression of ARSE in calcified diseased arteries that colocalized with increased RUNX2 expression. Scale bar 500 μm.



Extended Data Fig. 9: Luciferase reporter assay to analyze the functional impact of SNP rs5982944 (A/G).

The rs5982944-A allele and rs5982944-G allele firefly luciferase constructs were co-transfected with renilla luciferase plasmid into human coronary smooth muscle cells (HCSMCs), human aortic smooth muscle cells (HASMCs) and HEK-293 cells (n=6 biologically independent samples for HCSMCs and HASMCs and n=5 biologically independent samples for HEK-293). Firefly luciferase activity and renilla luciferase activity (internal control reporter vector) were measured sequentially in cell lysates. Firefly luciferase activity in cell lysates transfected with either the rs5982944-A construct or the rs5982944-G construct was normalized to renilla luciferase activity. The mean is depicted in

plots, with the error bars representing the standard error of the mean. Statistical comparisons were made using the two-tailed unpaired t-test.



Extended Data Fig. 10: Model of ARSE-induced phenotype switch from contractile to osteogenic vascular smooth muscle cells.

Atherosclerotic vascular calcification is characterized by the phenotype switch of vascular smooth muscle cells (VSMCs) from a contractile phenotype to a proliferative, osteogenic phenotype. The osteogenic phenotype of VSMCs is characterized by decreased expression of contractile proteins such as calponin (CNN1), but increased expression of Runt-related transcription factor 2 (RUNX2), a master regulator of the phenotype switch, in addition to other markers of calcification such as bone gamma-carboxyglutamate protein (BGLAP) and

alkaline phosphatase (ALPL). We identified ARSE as a major regulator of the phenotype switch.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Authors

Paul S. de Vries[1,*], Matthew P. Conomos[2,*], Kuldeep Singh[3,*], Christopher J. Nicholson[3,*], Deepti Jain[2], Natalie R. Hasbani[1], Wanlin Jiang[3], Sujin Lee[3], Christian L Lino Cardenas[3], Sharon M. Lutz[4,5], Doris Wong[6], Xiuqing Guo[7], Jie Yao[7], Erica P. Young[8], Catherine Tcheandjieu[9,10], Austin T. Hilliard[9,11], Joshua C. Bis[12], Lawrence F. Bielak[13], Michael R. Brown[1], Shaila Musharoff[9,14], Shoa L. Clarke[9,10], James G. Terry[15], Nicholette D. Palmer[16], Lisa R. Yanek[17], Huichun Xu[18], Nancy Heard-Costa[19,20], Jennifer Wessel[21,22], Margaret Sunitha Selvaraj[23,24,25], Rebecca H. Li[3], Xiao Sun[26,27], Adam W. Turner[6], Adrienne M. Stilp[2], Alyna Khan[2], Anne B. Newman[28], Asif Rasheed[29], Barry I Freedman[30], Brian G. Kral[31], Caitlin P. McHugh[2], Chani Hodonsky[6], Danish Saleheen[29,32,33], David M. Herrington[34], David R. Jacobs Jr[35], Deborah A. Nickerson[36,37], Eric Boerwinkle[1,38], Fei Fei Wang[2], Gerardo Heiss[39], Goo Jun[1], Greg L. Kinney[40], Haakon H. Sigurslid[3], HarshaVardhan Doddapaneni[38], Ira M. Hall[41], Isabela M. Bensenor[42], Jai Broome[2], James D. Crapo[43], James G. Wilson[44], Jennifer A. Smith[13,45], John Blangero[46,47], Jose D. Vargas[48], Jose Verdezoto Mosquera[6], Joshua D. Smith[36,37], Karine A. Viaud-Martinez[49], Kathleen A. Ryan[18], Kendra A. Young[40], Kent D. Taylor[7], Leslie A. Lange[50], Leslie S. Emery[2], Marcio S. Bittencourt[42], Matthew J. Budoff[51], May E. Montasser[18], Miao Yu[13], Michael C. Mahaney[46,47], Mohammed S Mahamdeh[3], Myriam Fornage[1,52], Nora Franceschini[53], Paulo A. Lotufo[42], Pradeep Natarajan[23,24,25], Quenna Wong[2], Rasika A. Mathias[17,54], Richard A. Gibbs[38,55], Ron Do[56,57], Roxana Mehran[58], Russell P. Tracy[59], Ryan W. Kim[60], Sarah C. Nelson[2], Scott M. Damrauer[61,62], Sharon L.R. Kardia[13], Stephen S. Rich[6], Valentin Fuster[63,64], Valerio Napolioni[65], Wei Zhao[13], Wenjie Tian[3], Xianyong Yin[66], Yuan-I Min[67], Alisa K. Manning[68,69], Gina Peloso[70], Tanika N. Kelly[27], Christopher J. O'Donnell[71,72], Alanna C. Morrison[1], Joanne E. Curran[46,47], Warren M. Zapol[73], Donald W. Bowden[16], Lewis C. Becker[31], Adolfo Correa[67,74], Braxton D. Mitchell[18,75], Bruce M. Psaty[12,76,77], John Jeffrey Carr[15], Alexandre C. Pereira[78,79], Themistocles L. Assimes[9,10], Nathan O. Stitziel[8,80,81], John E. Hokanson[40], Cecelia A. Laurie[2], Jerome I. Rotter[7], Ramachandran S. Vasan[20,82,83], Wendy S. Post[31], Patricia A. Peyser[13,Δ], Clint L. Miller[6,Δ], Rajeev Malhotra[3,Δ]

Affiliations

[1]Human Genetics Center, Department of Epidemiology, Human Genetics, and Environmental Sciences, School of Public Health, The University of Texas Health Science Center at Houston, Houston, TX, USA

[2]Genetic Analysis Center, Department of Biostatistics, School of Public Health, University of Washington, Seattle, WA, USA

[3]Cardiovascular Research Center, Division of Cardiology, Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

[4]PRecisiOn Medicine Translational Research (PROMoTeR) Center, Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA, USA

[5]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA

[6]Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, VA, USA

[7]The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

[8]Cardiovascular Division, Department of Internal Medicine, Washington University School of Medicine, St. Louis, MO, USA

[9]VA Palo Alto Healthcare System, Palo Alto, CA, USA

[10]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

[11]Palo Alto Veterans Institute for Research, Palo Alto, CA, USA

[12]Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, WA, USA

[13]School of Public Health, Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA

[14]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[15]Department of Radiology, Vanderbilt Translational and Clinical Cardiovascular Research Center, Vanderbilt University Medical Center, Nashville, TN, USA

[16]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA

[17]Division of General Internal Medicine, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[18]Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

[19]Boston University School of Medicine, Boston, MA, USA

[20]Boston University and National Heart, Lung, and Blood Institute’s Framingham Heart Study, Framingham, MA, USA

[21]Department of Epidemiology, Fairbanks School of Public Health, Indiana University, Indianapolis, IN, USA

[22]Diabetes Translational Research Center, Indiana University, Indianapolis, IN, USA

[23]Cardiovascular Research Center and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA

[24]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA

[25]Department of Medicine, Harvard Medical School, Boston, MA, USA

[26]School of Public Health and Tropical Medicine, Department of Epidemiology, Tulane University, New Orleans, LA, USA

[27]College of Medicine, Department of Medicine, Division of Nephrology, University of Illinois Chicago, Chicago, IL, USA

[28]Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

[29]Center For Non-Communicable Diseases, Karachi, Pakistan

[30]Section on Nephrology, Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[31]Division of Cardiology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[32]Department of Medicine, Columbia University Irving Medical Center, New York, NY, USA

[33]Department of Cardiology, Columbia University Irving Medical Center, New York, NY, USA

[34]Department of Internal Medicine, Section of Cardiovascular Medicine, Wake Forest School of Medicine, Winston-Salem, NC, USA

[35]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, MN, USA

[36]Department of Genome Sciences, University of Washington, Seattle, WA, USA

[37]Northwest Genomics Center, University of Washington, Seattle, WA, USA

[38]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA

[39]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA

[40]Department of Epidemiology, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, CO, USA

[41]Yale Center for Genomic Health, Yale School of Medicine, New Haven, CT, USA

[42]Center for Clinical and Epidemiological Research, University Hospital, University of Sao Paulo Medical School, São Paulo, Brazil

[43]Department of Medicine, National Jewish Health, Denver, CO, USA

[44]Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, MA, USA

[45]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA

[46]Department of Human Genetics, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA

[47]South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, Brownsville, TX, USA

[48]Medstar Heart and Vascular Institute, Medstar Georgetown University Hospital, Washington, DC, USA

[49]Illumina Laboratory Services, Illumina Inc., San Diego, CA, USA

[50]Department of Medicine, University of Colorado Denver, Anschutz Medical Campus, Aurora, CO, USA

[51]Department of Medicine, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, CA, USA

[52]Institute of Molecular Medicine, McGovern Medical School, The University of Texas Health Science Center at Houston, Houston, TX, USA

[53]Department of Epidemiology, Gillings School of Global Public health, University of North Carolina, Chapel Hill, NC, USA

[54]Division of Allergy and Clinical Immunology, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[55]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX, USA

[56]The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[57]Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[58]Icahn School of Medicine at Mount Sinai, New York, NY, USA

[59]Department of Pathology and Laboratory Medicine, Robert Larner, M.D. College of Medicine, University of Vermont, Burlington, VT, USA

[60]Psomagen Inc., Rockville, MD, USA

[61]Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA, USA

[62]Department of Surgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[63]Centro Nacional de Investigaciones Cardiovasculares Carlos III, Madrid, Spain

[64]Mount Sinai Heart Center, New York, NY, USA

[65]Genomic And Molecular Epidemiology (GAME) Lab, School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

[66]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA

[67]Jackson Heart Study, Department of Medicine, University of Mississippi Medical Center, Jackson, MS, USA

[68]Clinical and Translation Epidemiology Unit, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[69]Programs in Metabolism and Medical and Population Genetics, Broad Institute, Cambridge, MA, USA

[70]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

[71]VA Boston Healthcare System, Boston, MA, USA

[72]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA

[73]Department of Anesthesia, Critical Care and Pain Medicine at Massachusetts General Hospital, Boston, MA, USA

[74]Department of Population Health Science, University of Mississippi Medical Center, Jackson, MS, USA

[75]Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD, USA

[76]Department of Epidemiology, University of Washington, Seattle, WA, USA

[77]Department of Health Services, University of Washington, Seattle, WA, USA

[78]Department of Genetics, Harvard Medical School, Boston, MA, USA

[79]Laboratory of Genetics and Molecular Cardiology, Heart Institute, University of São Paulo, São Paulo, Brazil

[80]Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

[81]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA

[82]Department of Medicine, Boston University School of Medicine, Boston, MA, USA

[83]Department of Epidemiology, Boston University School of Public Health, Boston, MA, USA

Acknowledgements

This work was funded by National Heart, Lung and Blood Institute (NHLBI) grant number R01HL146860. Paul S. de Vries and Natalie R. Hasbani were additionally supported by American Heart Association grant number 18CDA34110116. Matthew P. Conomos was supported by NHLBI grant U01HL137162. Rajeev Malhotra was supported by the NHLBI (R01HL142809, R01HL159514, and R01HL162928), the American Heart Association (18TPA34230025), and the Wild Family Foundation. Paul S. de Vries and Rajeev Malhotra were supported by NHLBI grant number R01 HL162928. Tanika N. Kelly was supported by NHLBI grant number R01HL120393. Gina Peloso, Margaret Sunitha Selvaraj and Pradeep Natarajan were supported by NHLBI grant number R01HL142711. Clint L. Miller was supported by funding from NHLBI (R01HL148239 and R01HL164577), Fondation Leducq 'PlaqOmics' (18CVD02), and the Chan Zuckerberg Initiative, LLC and Silicon Valley Community Foundation.

Molecular data for the Trans-Omics in Precision Medicine (TOPMed) program was supported by the NHLBI. See the TOPMed Omics Support Table (Supplementary Table 14) for study specific omics support information. Core support including centralized genomic read mapping and genotype calling, along with variant quality metrics and filtering were provided by the TOPMed Informatics Research Center (3R01HL-117626-02S1; contract HHSN2682018000002I). Core support including phenotype harmonization, data management, sample-identity QC, and general program coordination were provided by the TOPMed Data Coordinating Center (R01HL-120393; U01HL-120393; contract HHSN2682018000001I). Study-specific funding is detailed in the Supplementary Methods. We thank L. Adrienne Cupples for her helpful comments that improved the manuscript. We gratefully acknowledge the studies and participants who provided biological samples and data for TOPMed.

Competing Interests

Bruce M. Psaty serves on the Steering Committee of the Yale Open Data Access Project funded by Johnson & Johnson. Leslie S. Emery is now an employee of Celgene/Bristol Myers Squibb. Celgene/Bristol Myers Squibb had no role in the funding, design, conduct, or interpretation of this study. Nathan O. Stitziel has received research funding from Regeneron Pharmaceuticals, unrelated to this work. Karine A. Viaud-Martinez is an employee at Illumina Inc. Ryan W. Kim is an employee at Psomagen Inc. Roxona Mehran reports institutional research grants from Abbott Laboratories, Abiomed, Applied Therapeutics, AstraZeneca, Bayer, Beth Israel Deaconess, Bristol-Myers Squibb, CERC, Chiesi, Concept Medical, CSL Behring, DSI, Medtronic, Novartis Pharmaceuticals, OrbusNeich; consultant fees from Abbott Laboratories, Boston Scientific, CardiaWave, Chiesi, Janssen Scientific Affairs, Medscape/WebMD, Medtelligence (Janssen Scientific Affairs), Roivant Sciences, Sanofi, Siemens Medical Solutions; consultant fees paid to the institution from Abbott Laboratories, Bristol-Myers Squibb; advisory board, funding paid to the institution from Spectranetics/Philips/Volcano Corp; consultant (spouse) from Abiomed, The Medicines Company, Merck; Equity <1% from Claret Medical, Elixir Medical; DSMB Membership fees paid to the institution from Watermark Research Partners; consulting (no fee) from Idorsia Pharmaceuticals Ltd., Regeneron Pharmaceuticals; Associate Editor for ACC, AMA. Rajeev Malhotra is a consultant for MyoKardia (now owned by BMS), Epizon Pharma, Renovacor, and Third Pole, a co-founder of Patch and Angea Biotherapeutics, and has received research funding from Angea Biotherapeutics, Bayer Pharmaceuticals, and Amgen. Adrienne M. Stilp receives funding from Seven Bridges Genomics to develop tools for the NHLBI BioData Catalyst consortium. Pradeep Natarajan reports grants from Amgen, Apple, Boston Scientific, AstraZeneca, Allelica, Novartis, and Genentech, consulting income from GV, Blackstone Life Sciences, Foresite Labs, Apple, AstraZeneca, Allelica, Novartis, HeartFlow, and Genentech, is a scientific advisor to Esperion Therapeutics, Preciseli, and TenSixteen Bio, is a scientific co-founder of TenSixteen Bio, and spousal employment at Vertex, all unrelated to the present work. Clint L. Miller received a research grant from AstraZeneca for an unrelated project. The remaining authors declare no competing interests.

References

1. Kavousi M et al. Evaluation of newer risk markers for coronary heart disease risk classification: a cohort study. Ann Intern Med 156, 438–44 (2012). [PubMed: 22431676]
2. Budoff MJ et al. Ten-year association of coronary artery calcium with atherosclerotic cardiovascular disease (ASCVD) events: the multi-ethnic study of atherosclerosis (MESA). Eur Heart J 39, 2401–2408 (2018). [PubMed: 29688297]
3. Greenland P, Blaha MJ, Budoff MJ, Erbel R & Watson KE Coronary Calcium Score and Cardiovascular Risk. J Am Coll Cardiol 72, 434–447 (2018). [PubMed: 30025580]
4. Polonsky TS et al. Coronary artery calcium score and risk classification for coronary heart disease prediction. JAMA 303, 1610–6 (2010). [PubMed: 20424251]
5. Shi X et al. Calcification in Atherosclerotic Plaque Vulnerability: Friend or Foe? Front Physiol 11, 56 (2020). [PubMed: 32116766]

6. Malhotra R et al. HDAC9 is implicated in atherosclerotic aortic calcification and affects vascular smooth muscle cell phenotype. Nat Genet 51, 1580–1587 (2019). [PubMed: 31659325]
7. Sutton NR et al. Molecular Mechanisms of Vascular Health: Insights From Vascular Aging and Calcification. Arterioscler Thromb Vasc Biol 43, 15–29 (2023). [PubMed: 36412195]
8. Nikpay M et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. Nat Genet 47, 1121–1130 (2015). [PubMed: 26343387]
9. Klarin D et al. Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. Nat Genet 49, 1392–1397 (2017). [PubMed: 28714974]
10. Verweij N, Eppinga RN, Hagemeijer Y & van der Harst P Identification of 15 novel risk loci for coronary artery disease and genetic risk of recurrent events, atrial fibrillation and heart failure. Sci Rep 7, 2761 (2017). [PubMed: 28584231]
11. Howson JMM et al. Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms. Nat Genet 49, 1113–1119 (2017). [PubMed: 28530674]
12. Koyama S et al. Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. Nat Genet 52, 1169–1177 (2020). [PubMed: 33020668]
13. Matsunaga H et al. Transethnic Meta-Analysis of Genome-Wide Association Studies Identifies Three New Loci and Characterizes Population-Specific Differences for Coronary Artery Disease. Circ Genom Precis Med 13, e002670 (2020). [PubMed: 32469254]
14. Aragam KG 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). [PubMed: 36474045]
15. Natarajan P et al. Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. Circ Cardiovasc Genet 9, 511–520 (2016). [PubMed: 27872105]
16. Wojczynski MK et al. Genetics of coronary artery calcification among African Americans, a meta-analysis. BMC Med Genet 14, 75 (2013). [PubMed: 23870195]
17. O'Donnell CJ et al. Genome-wide association study for coronary artery calcification with follow-up in myocardial infarction. Circulation 124, 2855–64 (2011). [PubMed: 22144573]
18. Bielak LF & Peyser PA Genetics of Subclinical Coronary Atherosclerosis. Curr Genet Med Rep 6, 116–123 (2018). [PubMed: 30854262]
19. Lo Sardo V et al. Unveiling the Role of the Most Impactful Cardiovascular Risk Locus through Haplotype Editing. Cell 175, 1796–1810 e20 (2018). [PubMed: 30528432]
20. Gupta RM et al. A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression. Cell 170, 522–533 e15 (2017). [PubMed: 28753427]
21. Innerarity TL et al. Familial defective apolipoprotein B-100: low density lipoproteins with abnormal receptor binding. Proc Natl Acad Sci U S A 84, 6919–23 (1987). [PubMed: 3477815]
22. Mahley RW Apolipoprotein E: cholesterol transport protein with expanding role in cell biology. Science 240, 622–30 (1988). [PubMed: 3283935]
23. Aherrahrou R, Aherrahrou Z, Schunkert H & Erdmann J Coronary artery disease associated gene Phactr1 modulates severity of vascular calcification in vitro. Biochem Biophys Res Commun 491, 396–402 (2017). [PubMed: 28720499]
24. Wellcome Trust Case Control Consortium et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat Genet 44, 1294–301 (2012). [PubMed: 23104008]
25. Pereira AC et al. Age, Gender, and Race-Based Coronary Artery Calcium Score Percentiles in the Brazilian Longitudinal Study of Adult Health (ELSA-Brasil). Clin Cardiol 39, 352–9 (2016). [PubMed: 27082165]
26. Budoff MJ et al. Coronary artery and thoracic calcium on noncontrast thoracic CT scans: comparison of ungated and gated examinations in patients from the COPD Gene cohort. J Cardiovasc Comput Tomogr 5, 113–8 (2011). [PubMed: 21167806]
27. Mumbach MR et al. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat Genet 49, 1602–1612 (2017). [PubMed: 28945252]

28. 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). [PubMed: 32513244]
29. Durham AL, Speer MY, Scatena M, Giachelli CM & Shanahan CM Role of smooth muscle cells in vascular calcification: implications in atherosclerosis and arterial stiffness. Cardiovasc Res 114, 590–600 (2018). [PubMed: 29514202]
30. Leopold JA Vascular calcification: Mechanisms of vascular smooth muscle cell calcification. Trends Cardiovasc Med 25, 267–74 (2015). [PubMed: 25435520]
31. Nicholson CJ et al. Reversal of Aging-Induced Increases in Aortic Stiffness by Targeting Cytoskeletal Protein-Protein Interfaces. J Am Heart Assoc 7(2018).
32. Brunetti-Pierri N et al. X-linked recessive chondrodysplasia punctata: spectrum of arylsulfatase E gene mutations and expanded clinical variability. Am J Med Genet A 117A, 164–8 (2003). [PubMed: 12567415]
33. Lin ME, Chen T, Leaf EM, Speer MY & Giachelli CM Runx2 Expression in Smooth Muscle Cells Is Required for Arterial Medial Calcification in Mice. Am J Pathol 185, 1958–69 (2015). [PubMed: 25987250]
34. Sun Y et al. Smooth muscle cell-specific runx2 deficiency inhibits vascular calcification. Circ Res 111, 543–52 (2012). [PubMed: 22773442]
35. Uzui H et al. Increased expression of membrane type 3-matrix metalloproteinase in human atherosclerotic plaque: role of activated macrophages and inflammatory cytokines. Circulation 106, 3024–30 (2002). [PubMed: 12473546]
36. Itoh Y Membrane-type matrix metalloproteinases: Their functions and regulations. Matrix Biol 44-46, 207–23 (2015). [PubMed: 25794647]
37. Rohwedder I et al. Plasma fibronectin deficiency impedes atherosclerosis progression and fibrous cap formation. EMBO Mol Med 4, 564–76 (2012). [PubMed: 22514136]
38. Ding HT, Wang CG, Zhang TL & Wang K Fibronectin enhances in vitro vascular calcification by promoting osteoblastic differentiation of vascular smooth muscle cells via ERK pathway. J Cell Biochem 99, 1343–52 (2006). [PubMed: 16795048]
39. Watson KE, Parhami F, Shin V & Demer LL Fibronectin and collagen I matrixes promote calcification of vascular cells in vitro, whereas collagen IV matrix is inhibitory. Arterioscler Thromb Vasc Biol 18, 1964–71 (1998). [PubMed: 9848891]
40. Kuzuya M et al. Effect of MMP-2 deficiency on atherosclerotic lesion formation in apoE-deficient mice. Arterioscler Thromb Vasc Biol 26, 1120–5 (2006). [PubMed: 16556856]
41. Wagsater D, Zhu C, Bjorkegren J, Skogsberg J & Eriksson P MMP-2 and MMP-9 are prominent matrix metalloproteinases during atherosclerosis development in the Ldlr(-/-)Apob(100/100) mouse. Int J Mol Med 28, 247–53 (2011). [PubMed: 21567073]
42. Hecht E et al. The matrix metalloproteinases 2 and 9 initiate uraemic vascular calcifications. Nephrol Dial Transplant 31, 789–97 (2016). [PubMed: 26333546]
43. Jiang L et al. Calpain-1 regulation of matrix metalloproteinase 2 activity in vascular smooth muscle cells facilitates age-associated aortic wall calcification and fibrosis. Hypertension 60, 1192–9 (2012). [PubMed: 23006733]
44. Bailey M et al. Involvement of matrix metalloproteinases and tenascin-C in elastin calcification. Cardiovasc Pathol 13, 146–55 (2004). [PubMed: 15081471]

METHODS-ONLY REFERENCES

45. Fang H et al. Harmonizing Genetic Ancestry and Self-identified Race/Ethnicity in Genome-wide Association Studies. Am J Hum Genet 105, 763–772 (2019). [PubMed: 31564439]
46. Taliun D et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. bioRxiv (2019).
47. Agatston AS et al. Quantification of coronary artery calcium using ultrafast computed tomography. J Am Coll Cardiol 15, 827–32 (1990). [PubMed: 2407762]
48. Carr JJ et al. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary

Artery Risk Development in Young Adults (CARDIA) study. Radiology 234, 35–43 (2005). [PubMed: 15618373]

49. Stilp AM et al. A system for phenotype harmonization in the NHLBI Trans-Omics for Precision Medicine (TOPMed) Program. bioRxiv (2020).
50. Conomos MP, Miller MB & Thornton TA Robust inference of population structure for ancestry prediction and correction of stratification in the presence of relatedness. Genet Epidemiol 39, 276–93 (2015). [PubMed: 25810074]
51. Bild DE et al. Ethnic differences in coronary calcification: the Multi-Ethnic Study of Atherosclerosis (MESA). Circulation 111, 1313–20 (2005). [PubMed: 15769774]
52. Conomos MP, Reiner AP, Weir BS & Thornton TA Model-free Estimation of Recent Genetic Relatedness. Am J Hum Genet 98, 127–48 (2016). [PubMed: 26748516]
53. Conomos MP et al. Genetic Diversity and Association Studies in US Hispanic/Latino Populations: Applications in the Hispanic Community Health Study/Study of Latinos. Am J Hum Genet 98, 165–84 (2016). [PubMed: 26748518]
54. Sofer T et al. A fully adjusted two-stage procedure for rank-normalization in genetic association studies. Genet Epidemiol 43, 263–275 (2019). [PubMed: 30653739]
55. Gogarten SM et al. Genetic association testing using the GENESIS R/Bioconductor package. Bioinformatics 35, 5346–5348 (2019). [PubMed: 31329242]
56. Chen H et al. Control for Population Structure and Relatedness for Binary Traits in Genetic Association Studies via Logistic Mixed Models. Am J Hum Genet 98, 653–66 (2016). [PubMed: 27018471]
57. Dey R, Schmidt EM, Abecasis GR & Lee S A Fast and Accurate Algorithm to Test for Binary Phenotypes and Its Application to PheWAS. Am J Hum Genet 101, 37–49 (2017). [PubMed: 28602423]
58. Zhou W et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. Nat Genet 50, 1335–1341 (2018). [PubMed: 30104761]
59. Frankish A et al. GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Res 47, D766–D773 (2019). [PubMed: 30357393]
60. Dong C et al. Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. Hum Mol Genet 24, 2125–37 (2015). [PubMed: 25552646]
61. McLaren W et al. The Ensembl Variant Effect Predictor. Genome Biol 17, 122 (2016). [PubMed: 27268795]
62. Fishilevich S et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. Database (Oxford) 2017(2017).
63. Zerbino DR, Wilder SP, Johnson N, Juettemann T & Flicek PR The ensembl regulatory build. Genome Biol 16, 56 (2015). [PubMed: 25887522]
64. Rogers MF et al. FATHMM-XF: accurate prediction of pathogenic point mutations via extended features. Bioinformatics 34, 511–513 (2018). [PubMed: 28968714]
65. Liu X et al. WGSA: an annotation pipeline for human genome sequencing studies. J Med Genet 53, 111–2 (2016). [PubMed: 26395054]
66. Chen H et al. Efficient Variant Set Mixed Model Association Tests for Continuous and Binary Traits in Large-Scale Whole-Genome Sequencing Studies. Am J Hum Genet 104, 260–274 (2019). [PubMed: 30639324]
67. Lee S et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 91, 224–37 (2012). [PubMed: 22863193]
68. Wu MC et al. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89, 82–93 (2011). [PubMed: 21737059]
69. Kelly TN et al. Insights From a Large-Scale Whole-Genome Sequencing Study of Systolic Blood Pressure, Diastolic Blood Pressure, and Hypertension. Hypertension 79, 1656–1667 (2022). [PubMed: 35652341]

70. Selvaraj MS et al. Whole genome sequence analysis of blood lipid levels in >66,000 individuals. Nat Commun 13, 5995 (2022). [PubMed: 36220816]

71. Natarajan P et al. Chromosome Xq23 is associated with lower atherogenic lipid concentrations and favorable cardiometabolic indices. Nat Commun 12, 2182 (2021). [PubMed: 33846329]

72. Wessel J et al. Rare Non-coding Variation Identified by Large Scale Whole Genome Sequencing Reveals Unexplained Heritability of Type 2 Diabetes. medRxiv, 2020.11.13.20221812 (2020).

73. 1000 Genomes Project Consortium et al. A global reference for human genetic variation. Nature 526, 68–74 (2015). [PubMed: 26432245]

74. Giambartolomei C et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 10, e1004383 (2014). [PubMed: 24830394]

75. Chomczynski P & Sacchi N Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162, 156–9 (1987). [PubMed: 2440339]

76. Livak KJ & Schmittgen TD Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods 25, 402–8 (2001). [PubMed: 11846609]

77. O'Rourke C et al. Calcification of Vascular Smooth Muscle Cells and Imaging of Aortic Calcification and Inflammation. J Vis Exp (2016).

78. Kang H et al. Bone morphogenetic protein 4 promotes vascular smooth muscle contractility by activating microRNA-21 (miR-21), which down-regulates expression of family of dedicator of cytokinesis (DOCK) proteins. J Biol Chem 287, 3976–86 (2012). [PubMed: 22158624]

79. Turner AW et al. Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. Nat Genet 54, 804–816 (2022). [PubMed: 35590109]

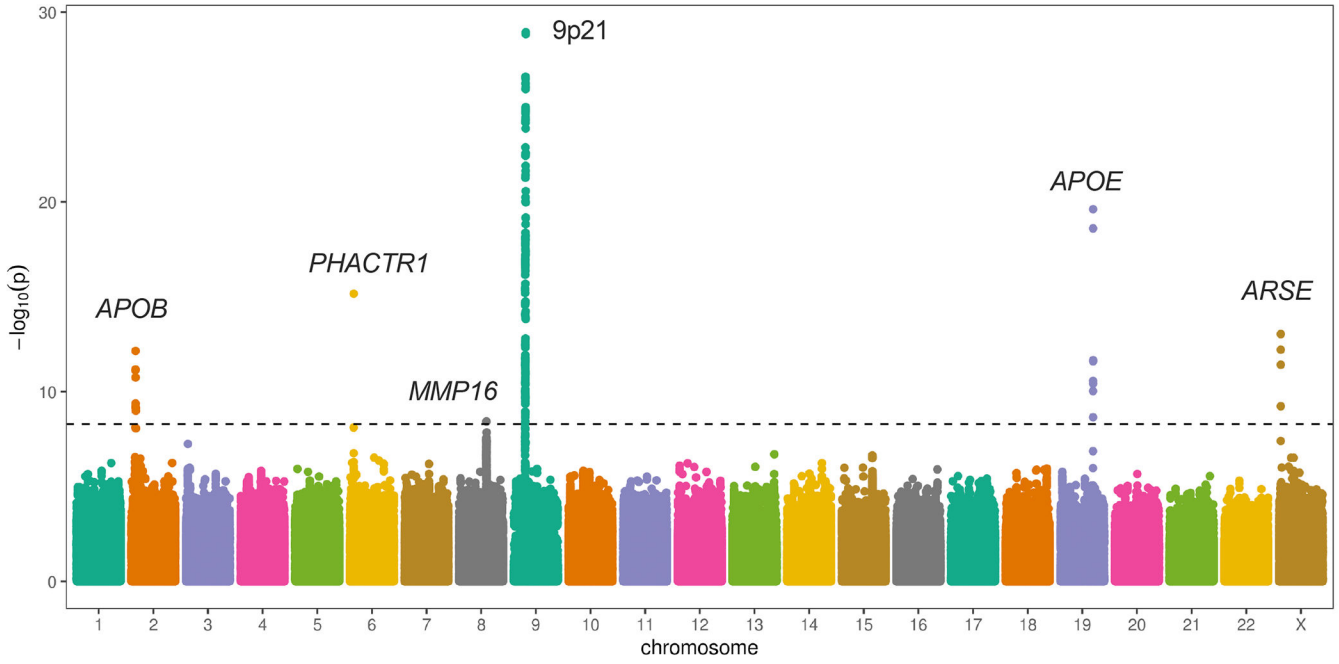


Fig. 1: Manhattan plot for the genome-wide association study of log(CAC+1).

In the Manhattan plot, each genetic variant is inluded as a dot, with the position on the x axis corresponding to their genomic position and the position on the y axis corresponding to the significance the association, denoted by $-\log_{10}$ transformed two-sided P-values. The plot shows 6 genetic loci (including 2 novel and 4 known) associated with coronary artery calcification score at a significance level of $P < 5\times10^{-9}$ (dotted line) in the pooled analysis of 22,400 individuals from 10 studies.

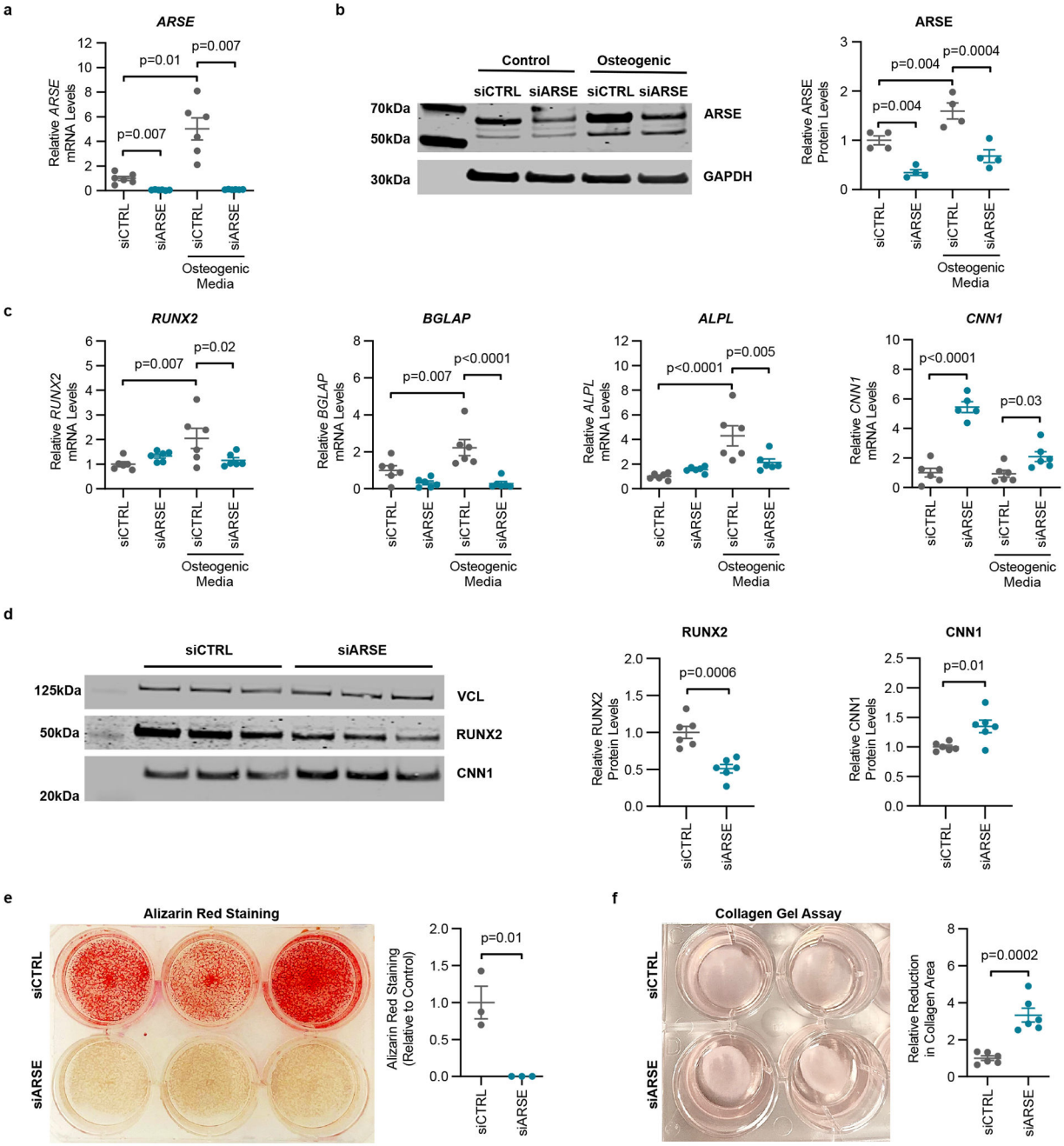


Fig. 2: Silencing *ARSE* expression inhibits osteogenic phenotype switch in human coronary artery vascular smooth muscle cells.

a) Treatment of human coronary artery vascular smooth muscle cells (n = 6 biologically independent samples in each group) with osteogenic media for 3 days increased *ARSE* mRNA expression approximately 5-fold. Treatment of cells grown in osteogenic media with siARSE resulted in >90% knockdown of *ARSE* mRNA. **b**) Protein expression of ARSE was measured by immunoblot (left panel) using antibodies directed against ARSE and GAPDH (for a loading control). Treatment of cells with osteogenic media increased

ARSE protein levels by 1.5-fold (right panel). Treatment of cells grown in osteogenic media with siARSE resulted in >70% reduction of *ARSE* protein (n=4 biologically independent samples in each group). **c)** Treatment of cells grown in osteogenic media with siARSE ameliorated osteogenic phenotype switch as evidenced by decreased *RUNX2*, *BGLAP*, and *ALPL* mRNA levels, and increased *CNN1* mRNA levels ~2-fold. Of note, silencing *ARSE* in cells grown in normal media increased *CNN1* mRNA levels by > 5-fold (n=6 biologically independent samples in each group except n=5 for siARSE CNN1 data). **d)** Reduced *ARSE* expression was also associated with an approximately 50% decrease in RUNX2 and >30% increase in CNN1 protein levels assessed by immunoblot using antibodies directed against RUNX2, CNN1 and VCL (for a loading control) (n=6 biologically independent samples in each group). **e**) Treatment of cells grown in osteogenic media with siARSE reduced calcification by approximately 60% (right panel, n = 3 biologically independent samples in each group), as evidenced by decreased Alizarin Red staining (left panel). **f**) Reduced ARSE expression with siARSE treatment in human coronary artery vascular smooth muscle cells grown in collagen discs (left panel) resulted in a >3-fold increase in contraction (right panel, n=6 biologically independent samples in each group). Statistical comparisons were made using either a two-tailed one-way ANOVA with Sidak's post-hoc comparison testing (for more than two groups) or a two-tailed Student t test (for two groups). The mean is depicted in plots, with the error bars representing the mean ± the standard error of the mean.

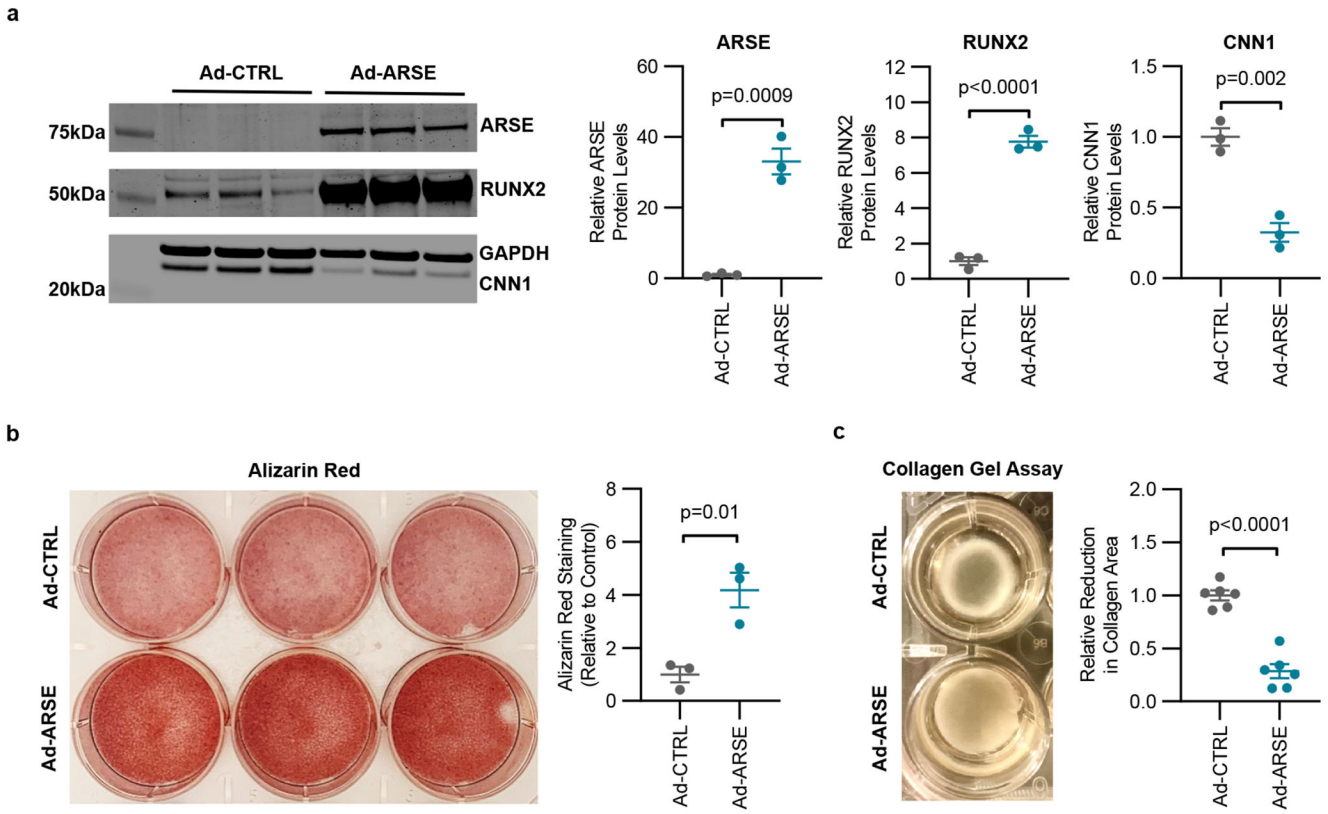


Fig. 3: Overexpression of *ARSE* induces calcification in human coronary artery vascular smooth muscle cells.

a) Adenoviral expression of ARSE in human coronary artery vascular smooth muscle cells was associated with an 8-fold increase in RUNX2 protein levels and an approximately 70% decrease in CNN1 protein levels, when cells were harvested 5 days after viral transduction (n=3 biologically independent samples in each group). Protein expression was determined by immunoblot (left panel) using antibodies directed against ARSE, RUNX2, CNN1 and GAPDH (for a loading control) with quantification shown in the right panel. **b**) As shown by Alizarin Red staining (left panel), increased ARSE expression resulted in augmented calcification in human coronary artery vascular smooth muscle cells (right panel, n = 3 biologically independent samples in each group). Two independent experiments were performed with representative images shown. **c**) Increased ARSE expression also caused a >70% decrease (right panel, n=6 biologically independent samples in each group) in contraction of human coronary artery vascular smooth muscle cells grown in collagen discs (left panel). Statistical comparisons were made using a two-tailed Student t test. The mean is depicted in plots, with the error bars representing the mean ± the standard error of the mean.

Table 1:

Index variants at genome-wide significant loci for log(CAC+1).

Variant	Chr:pos	Locus/Gene	Alleles	Annotation	Overall Freq (%)	White Freq (%)	Black Freq (%)	Beta (SE)log(CAC+1)	Plog(CAC+1)
rs4977575	*9:22124745*	9p21	G/C	downstream	66.4	50.8	89.4	0.27 (0.02)	1.1×10^{-29}
rs7412	*19:44908822*	*APOE*	T/C	missense	8.7	8.0	10.7	−0.33 (0.04)	2.4×10^{-20}
rs9349379	*6:12903725*	*PHACTR1/ EDN1*	G/A	intronic	28.3	40.3	9.2	0.20 (0.02)	6.8×10^{-16}
rs5982944	*X:2964339*	*ARSE*	G/A	intronic	17.1	0.3	41.8	−0.20 (0.03)	9.1×10^{-14}
rs5742904	*2:21006288*	*APOB*	T/C	missense	0.2	0.1	0	1.83 (0.25)	7.1×10^{-13}
rs13268080	*8:87644802*	*MMP16*	G/A	intergenic	78.7	77.5	83.2	0.15 (0.02)	3.7×10^{-09}

CAC refers to coronary artery calcification score; Chr:pos shows the chromosome number and position in build 38; Locus/Gene shows the candidate effector gene(s) at each locus, or the cytogenetic band if this is unclear; Alleles shows the coded/noncoded alleles; Frequency (Freq), Beta, and standard error (SE) all apply to the coded allele.