Supplementary Note

Multi-ancestry genome-wide analysis identifies effector genes and druggable pathways for coronary artery calcification

Supplementary Methods

Conditional and joint analyses. We utilized COJO to evaluate evidence for independent associations within loci that were identified as significant in the combined study population[1]. Due to reliance on LD and the lack of an appropriate reference population for African Americans, we ran COJO only for the European ancestry meta-analyzed summary statistics. We used relevant 1000G Phase 3 populations to generate LD references including all EUR populations aside from FIN. We used Plink v1.90b6.16 to generate bed files for the 1000G EUR population and restricted the analyses to the variants that were not monomorphic in the combined study cohorts[2]. COJO was subsequently run on the 1000G bed files using a maximum p-value threshold of $1x10^{-5}$ and a window size of 1Mb.

Credible set analyses. 95% credible sets were generated for each significantly associated locus separately in the meta-analyzed summary statistics from European and African ancestry-only summary statistics using a script adapted from [3] (available on Github: https://github.com/MillerLab-CPHG/Fine_mapping/tree/main/scripts/R). In short, Bayesian posterior probability values were generated for each variant within the respective locus, after which variants were selected until a cumulative 95% posterior probability was achieved. For improved interpretability, we present the total number of variants with a posterior probability >1% to represent the size of the credible set, and report the variant with the highest posterior probability by ancestry group for associations within each locus.

Functional annotation and gene mapping. We first annotated the function of variants using the Ensembl Variant Effect Predictor (v92)[4]. We used FUMA (v1.3.6a, http://fuma.ctglab.nl/)[5] for functional annotation of the GWAS results based on 1000G phase 3 (version 5 based on ALL populations) and Ensembl 92 (only protein-coding genes were considered, n = 19,177). Variants were filtered while performing annotation based on positional gene mapping (maximum distance 10kb), eQTL-based gene mapping, and 3D chromatin interaction mapping using a minimum CADD score ≥ 12.37 (considered to be suggestively deleterious), a maximum RegulomeDB score of 7, a maximum 15-core chromatin state of 7 using all available tissues. For eQTL mapping aorta, coronary and tibial artery derived data (GTEx v8) were used and a FDR < 0.05 was applied. For the 3D chromatin interaction analysis an FDR < 10^{-6} threshold was applied using all the available data from Roadmap. We defined the promoter region in a window of 250 Kb upstream and 500 Kb downstream of the TSS to overlap genes with significantly interacting regions, and only variants overlapping enhancers and genes whose promoters overlap epigenomic regions were used. Variants were annotated using ANNOVAR version 2017-07-17[6] as implemented in FUMA.

GTEx query. Candidate CAC genes were queried in the Genotype Tissue Expression (GTEx) database (version 8) via the online portal (https://gtexportal.org/home/). This release includes 54 unique human tissues from 948 donors. Bulk RNA-seq processing and analysis details are

available on the portal website. Queries of normalized mean gene expression levels per tissue were limited to the most relevant tissues and clustered using the Multi-Gene Query Tool.

Activity-by contact and enhancer linking. We used the combined EA and AA CAC variants filtered to suggestive associations ($P < 1 \times 10^{-5}$) to intersect with genome-wide enhancer-gene predictions calculated in the activity-by-contact models[7]. We used both the ENCODE coronary artery tissue and human coronary artery smooth muscle cell based annotations from H3K27ac histone modification ChIP-seq and chromatin accessibility datasets, and we considered significant enhancer-gene predictions with ABC scores > 0.02, as previously described[8]. For enhancer-gene linking annotations we again used the above CAC summary associations at $P < 1 \times 10^{-5}$ and intersected these variants with human aortic enhancer states (6,7, and 12) calculated from the Epigenomics Roadmap consortium[9].

Athero-Express Biobank Study

Patient population

Atherosclerotic plaques were obtained from patients undergoing a carotid endarterectomy (CEA) procedure and included in the Athero-Express Biobank Study (AE, www.atheroexpress.nl), an ongoing biobank study at the University Medical Centre Utrecht (Utrecht, The Netherlands) and the St. Antonius Hospital (Nieuwegein, The Netherlands)[10]. This study complies with the Declaration of Helsinki, and all participants provided informed consent. The medical ethical committees of the respective hospitals approved this study which was registered under number TME/C-01.18. The study design of the AE was described before[10], but in brief: during surgery blood and plaques are obtained, stored at -80°C and plaque material is routinely used for histological analysis[10,11].

Histological phenotyping

We described the standardized (immuno)histochemical analysis protocols used in the AE before[10,11]. In short, 10-micron cross-sections of the paraffin-embedded segments were cut using a microtome and examined under a microscope. We quantitatively scored the number of macrophages (CD68) and smooth muscle cells (SMCs, α-actin), as percentage of the microscopy field area by computerized analysis usingAnalySIS 3.2 software (Soft Imaging Systems GmbH, Münster, Germany). Intraplaque vessel density (CD34) was assessed as the average number per 3 hotspots. Intraplaque hemorrhage (IPH) was scored as no/yes using a hematoxylin and eosin staining (HE). Intraplaque fat was defined as less or more than 40% fat per total plaque area using HE. The amount of calcification (using HE) and collagen (picrosirius red) were binary scored as no/minor vs. moderate/heavy staining.

Assessment of overall plaque vulnerability was performed as previously described by Verhoeven et al.[12]. Briefly, macrophages and smooth muscle cells were semi-quantitatively defined as no/minor or moderate/heavy. Each plaque characteristic that defines a stable plaque

(i.e., no/minor macrophages, moderate/heavy collagen, moderate/heavy smooth muscle cells and <10% fat) was given a score of 0, while each plaque characteristic that defines a vulnerable plaque (i.e., moderate/heavy macrophages, no/minor collagen, no/minor smooth muscle cells and ≥10% fat) was given a score of 1. The score of each plaque characteristic was summed resulting in a final plaque score ranging from 0 (most stable plaque) to 4 (most vulnerable plaque). All histological observations were performed by the same dedicated technician and interobserver analyses have been reported previously[13].

DNA isolation, genotyping, and imputation

DNA isolation and genotyping

We genotyped the AE in three separate, but consecutive experiments. In short, DNA was extracted from EDTA blood or (when no blood was available) plaque samples of 1,858 consecutive patients from the Athero-Express Biobank Study and genotyped in 3 batches. For the Athero-Express Genomics Study 1 (AEGS1) 891 patients (602 males, 262 females, 27 unknown sex), included between 2002 and 2007, were genotyped (440,763 markers) using the Affymetrix Genome-Wide Human SNP Array 5.0 (SNP5) chip (Affymetrix Inc., Santa Clara, CA, USA) at Eurofins Genomics (https://www.eurofinsgenomics.eu/, formerly known as AROS). For the Athero-Express Genomics Study 2 (AEGS2) 954 patients (640 makes, 313 females, 1 unknown sex), included between 2002 and 2013, were genotyped (587,351 markers) using the Affymetrix Axiom® GW CEU 1 Array (AxM) at the Genome Analysis Center (https://www.helmholtz-muenchen.de). The two first batches, AEGS1 and AEGS2, were described before[14]. For the Athero-Express Genomics Study 3 (AEGS3) 658 patients (448 males, 203 females, 5 unknown sex), included between 2002 and 2016, were genotyped (693,931 markers) using the Illumina GSA MD v1 BeadArray (GSA) at Human Genomics Facility, HUGE-F (http://glimdna.org/index.html). All experiments were carried out according to OECD standards. We used the genotyping calling algorithms as advised by Affymetrix (AEGS1 and AEGS2) and Illumina (AEGS3): BRLMM-P, AxiomGT1, and Illumina GenomeStudio respectively.

Quality control after genotyping

After genotype calling, we adhered to community standard quality control and assurance (QCA) procedures of the genotype data from AEGS1, AEGS2, and AEGS3[14,15]. Samples with low average genotype calling and sex discrepancies (compared to the clinical data available) were excluded. The data was further filtered on 1) individual (sample) call rate > 97%, 2) SNP call rate > 97%, 3) minor allele frequencies (MAF) > 3%, 4) average heterozygosity rate ± 3.0 s.d., 5) relatedness (pi-hat > 0.20), 6) Hardy–Weinberg Equilibrium (HWE $p < 1.0\times10-3$), and 7) Monomorphic SNPs ($< 1.0\times10-6$). After QCA 2,493 samples remained, 108 of non-European descent/ancestry, and 156 related pairs. These comprise 890 samples and 407,712 SNPs in AEGS1, 869 samples and 534,508 SNPs in AEGS2, and 649,954 samples and 534,508 SNPs in AEGS3 remained.

Imputation

Before phasing using SHAPEIT2, data was lifted to genome build b37 using the liftOver tool from UCSC (https://genome.ucsc.edu/cgi-bin/hgLiftOver). Finally, data was imputed with 1000G phase 3, version 5 and HRC release 1.1 as a reference using the Michigan Imputation Server (https://imputationserver.sph.umich.edu/)[16]. These results were further integrated using QCTOOL v2, where HRC imputed variants are given precedence over 1000G phase 3 imputed variants.

Plaque phenotype association

For the association of the 11 independent loci with the plaque vulnerability and plaque morphological characteristics, we queried 1000G phase 3 using FUMA and included 853 variants ±250kb from the lead variants, with LD r^2 values (1000G EUR) provided for each variant. Next, we performed regression analyses adjusted for age, sex and principal components, thus: phenotype ~ age + sex + chip-used + PC1 + PC2 + year-of-surgery. Continuous variables were inverse-rank normal transformed. Codes and raw results from these analyses are found here: https://github.com/CirculatoryHealth/CHARGE_1000G_CAC; input genetic and phenotype data are available through.

RNA isolation and (single-cell) RNA sequencing

RNA isolation and library preparation

A total of 700 segments were selected from patients who were included in the study between 2002 and 2016. The RNA isolated from the archived advanced atherosclerotic lesion is fragmented. We have ultimately employed the CEL-seq2 method[17]. CEL-seq2 yielded the highest mappability reads to the annotated genes compared to other library preparation protocols. The methodology captures 3'-end of polyadenylated RNA species and includes unique molecular identifiers (UMIs), which allow direct counting of unique RNA molecules in each sample.

Sequencing read mapping and quality filtering

Libraries were sequenced on the Illumina Nextseq500 platform; a high output paired-end run of 2 × 75 bp was performed (Utrecht Sequencing Facility). The reads were demultiplexed and aligned to human cDNA reference (Ensembl 84) using the BWA (0.7.13). Multiple reads mapping to the same gene with the same unique molecular identifier (UMI, 6bp long) were counted as a single read. The raw read counts were corrected for UMI sampling (corrected_count=-4096*(ln(1-(raw_count/4096)))), normalized for sequencing depth and quantile normalized (core scripts can be found in https://github.com/mmokry/bulkCEL-seq2 and https://github.com/mmokry/seurat_meets_bulk_AE). We have detected a median of 19.501 (SD = 5.874) genes per sample with at least one unique read and discarded samples (n=46) with less than 9000 detected genes from further analysis. For all the subsequent analyses, we

have excluded all the ribosomal genes and used only the protein-coding genes with annotated HGCN names.

Single-cell RNA sequencing

Atherosclerotic plaques were collected from 35 individuals and processed for single-cell RNA sequencing as described before[18]. In short, time between surgical removal and plaque processing did not exceed 10 minutes; note that the inclusion of a small medial layer in the dissected tissue could not be excluded during the surgical procedure. The remainder of the plaque was washed in RPMI and minced into small pieces with a razor blade. The tissue was then digested in RPMI 1640 containing 2.5 mg/mL Collagenase IV (ThermoFisher Scientific), 0.25 mg/mL DNAse I (Sigma), 2.5 mg/mL Human Albumin Fraction V (MP Biomedicals) and 1 mM Flavopiridol (Selleckchem) at 37°C for 30 minutes. Subsequently, the plaque cell suspension was filtered through a 70 µm cell strainer and washed with RPMI 1640. Cells were kept in RPMI 1640 with 1% Fetal Calf Serum until subsequent staining for fluorescence-activated cell sorting. Remaining, unstained cells were cryostored in liquid nitrogen.

We adapted a CELseq2-protocol for single-cell RNA sequencing and processed the data as we described before[18]. Analyses were performed using Seurat (version 3.xx). Prior to processing, reads were filtered for mitochondrial and ribosomal genes, *MALAT1*, *KCNQ1OT1*, *UGDH-AS1*, and *EEF1A*. In order to omit doublets and low-quality cells, only cells expressing between 500 and 10.000 genes and genes expressed in at least 3 cells were used for further analysis. Data was log-normalized and scaled with the exclusion of unique molecular identifiers (UMIs). Top variable genes for all samples were used to combine samples into one object using Seurat function RunMultiCCA(), after which samples were aligned using AlignSubspace() with reduction.type=CCA and grouping.var="plate". Subsequently, canonical correlation analysis (CCA) reduction was performed with a resolution of 1.2 for 15 dimensions to identify clusters and to perform t-distributed stochastic neighbor embedding (tSNE). Cell types were assigned to cell clusters by evaluating gene expression of individual cell clusters using differential gene expression (Wilcoxon rank sum test) and analysis with SingleR8 against BLUEPRINT reference data.

Coronary artery single-cell RNA-seq analysis

Human coronary artery scRNA data read count matrix was retrieved from the Gene Expression Omnibus (GEO) using GSE131780 and loaded into R 4.1.1, and was preprocessed using standard parameters of the R packages 'Seurat' v.4, and 'Monocle3' as required, read count matrix was read into R and converted to a 'Seurat object' using the CreateSeuratObject() function. Then, the object underwent removal of mitochondrial and low-quality reads, followed by normalization and variable feature selection using the NormalizeData() and FindVariableFeature() functions, respectively. Uniform manifold approximation projections (UMAPs) were then calculated using the runUMAP() function using the first 30 dimensions.

Custom scripting was created to export UMAP of the clusters from 'Seurat' into 'Monocle3' before pseudotemporal analysis.

PheWAS. For our PheWAS analysis, we queried the following CAC genes based on their multiple lines of functional evidence and specific association with CAC: *ADK*, *ARID5B*, *IGFBP3*, *ENPP1*/*ENPP3*, and *FGF23.* Gene names were queried in the GWAS ATLAS resource (https://atlas.ctglab.nl)[19] and the UKBiobank TOPMed-imputed PheWeb (https://pheweb.org/UKB-TOPMed/about).

Immunofluorescence analysis in coronary arteries. Freshly isolated coronary arteries were obtained from consented heart transplant recipients or heart donors as described[20,21]. Briefly, hearts were arrested in cardioplegic solution, transferred on ice, and left anterior descending (LAD), left circumflex artery and right coronary arteries were dissected from the epicardium, with surrounding adipose and myocardial tissue carefully removed. Coronary artery segments were grossly scored for presence of lesions and embedded in OCT, snap frozen in liquid nitrogen and stored at -80°C until analysis. OCT blocks were cryosectioned at 8mm. Histological analysis using hematoxylin and eosin (H&E), Oil Red O, and picrosirius red staining was performed on adjacent slides to characterize subclinical (non-calcified lesions) and advanced (calcified lesions) stages of atherosclerosis. Frozen slides were washed with sterile PBS twice for 2min followed by fixation with 4% formaldehyde for 10 min. Slides were then washed with PBS twice for 2min and tissue was permeabilized with triton at 0.05% for 10 min. Coronary artery tissues were blocked with donkey serum at 10% for 1 hour followed by incubation overnight at 4°C with primary antibody as follows: mouse anti-ENPP1 (SC-166649, Santa Cruz Biotechnology, Dallas, TX), rabbit anti-IGFBP3 (#10189-2-AP, Proteintech, Rosemont, IL), rabbit anti-ARID5B (#HPA015037, Atlas Antibodies, Stockholm, SWE), or mouse anti-ADK (SC-514588, Santa Cruz Biotechnology) at 1:100 dilution and mouse anti-α-SMA (SC-53142, Santa Cruz Biotechnology) at 1:100 dilution. Slides were washed with PBS-tween at 0.1% 3 times for 3 min each followed by incubation with appropriate secondary antibodies: Alexa 488 anti-goat (#A11055, ThermoFisher Scientific), Alexa 555 anti-mouse (#A31570), ThermoFisher Scientific), Alexa 647 anti-rabbit (#A31573, ThermoFisher Scientific), Alexa 647 anti-mouse (#A31571, ThermoFisher Scientific), Alexa 555 anti-goat (#A21432, ThermoFisher Scientific), and Alexa 488 anti-mouse (#A21202, ThermoFisher Scientific) all at 1:400 dilution for 1h at room temperature. Slides were then washed with PBS-tween at 0.1%, 4 times for 3 min each and then coverslipped with a diamond mounting medium containing DAPI. Slides were visualized with the Leica TCS SP8 confocal microscopy station and micrographs were digitized with the Leica Application Suite X software.

Quantitative Real-Time PCR. HCASMCs were obtained from Cell Applications (#350-05a). To assess the effects of ENPP1, IGFBP3, ARID5B, and ADK on gene expression, siRNA directed against each of these genes (siENPP1, siIGFBP3, siARID5B, siADK) or control siRNA (siCTRL) was obtained from Horizon Discovery (Waterbeach, UK). Cells were transfected with siRNA

(20nM final concentration) using Lipofectamine RNAiMAX reagent (Life Technologies, Carlsbad, CA), according to the manufacturer protocol. Cells were treated with siRNA for 24h and then grown in osteogenic conditions for 48h. Osteogenic conditions consisted of serum-free smooth muscle cell media (#311-500, Cell Applications) supplemented with 10% FBS, 10mM β-glycerophosphate disodium, 10 nM dexamethasone, and 50 ug/mL L-ascorbic acid, as described previously[22,23]. Total cellular RNA was extracted using Trizol (ThermoFisher, Waltham, MA). cDNA was obtained via reverse transcription, using MultiScribe Reverse Transcriptase (Applied Biosystems, Foster City, CA). Quantitative PCR was performed using TaqMan Fast Advance Master Mix (Applied Biosystems). TaqMan assays were used for the genes of interest using the following assay IDs: *18S* (Hs99999901_s1), *ADK* (Hs00417073_m1), *ARID5B* (Hs01382781_m1), *CNN1* (Hs00959434_m1), *ENPP1* (Hs01054040_m1), *IGFBP3* (Hs00181211_m1), and *RUNX2* (Hs01047973_m1). Real-time amplification and quantification of transcripts was performed using a QuantStudio 3 (Applied Biosystems, Foster City, CA). The relative C_T method was used to quantify relative expression of target genes; values were normalized to levels of 18S ribosomal RNA.

Immunoblotting. Cells were treated on day 1 and day 6 with siRNA and grown in osteogenic conditions, as described above. After 9 days, cells were homogenized in RIPA buffer containing protease and phosphatase inhibitor (Sigma-Aldrich, Saint Louis, MO). Protein concentration was then determined via BCA; lysates were then diluted in RIPA, sample reducing agent (Invitrogen, Waltham, MA) and LDS buffer (Invitrogen). Samples were run on NuPAGE 4-12% BIS-TRIS gels (Invitrogen) alongside PageRuler Plus prestained protein ladder (Thermo Fisher). Following gel electrophoresis, separated proteins were transferred to PVDF membrane using the iBlot 2 Dry Blotting System (Invitrogen). Membranes were blocked in 5% bovine serum albumin (BSA) in Tris-buffered saline with 0.05% Tween (TBST) for 1 hour and then incubated in primary antibody. Rabbit polyclonal antibodies (1:1000 dilution) directed against RUNX2 (#12556S, Cell Signaling Technology, Danvers, MA) were used to detect RUNX2 protein. Rabbit polyclonal antibodies (1:10000 dilution) directed against glyceraldehyde 3-phosphate dehydrogenase (GAPDH, #TA890003, OriGene, Rockville, MD) were used to detect GAPDH protein. Blots were then incubated in goat anti-rabbit secondary antibody (#31460, ThermoFisher) at 1:1000 concentration and imaged using a ChemiDoc Touch imaging system (Bio-Rad Laboratories, Hercules, CA).

Study-specific acknowledgements

Coronary Artery Risk Development in Young Adult (CARDIA): The Coronary Artery Risk Development in Young Adults Study (CARDIA) is conducted and supported by the NHLBI in collaboration with the University of Alabama at Birmingham (HHSN268201800005I & HHSN268201800007I), Northwestern University (HHSN268201800003I), University of Minnesota (HHSN268201800006I), and Kaiser Foundation Research Institute

(HHSN268201800004I). The Y25 cardiac CT scans were supported by an NHLBI award to Vanderbilt University Medical Center (R01-HL098445).

Genetic Epidemiology of Chronic Obstructive Pulmonary Disease (COPDGene): The COPDGene project described was supported by Award Number U01 HL089897 and Award Number U01 HL089856 from the NHLBI. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NHLBI or the NIH. The COPDGene project is also supported by the COPD Foundation through contributions made to an Industry Advisory Board composed of AstraZeneca, Boehringer Ingelheim, GlaxoSmithKline, Novartis, Pfizer, Siemens and Sunovion. A full listing of COPDGene investigators can be found at: http://www.copdgene.org/directory

Diabetes Heart Study (DHS): This work was supported by R01 HL92301, R01 HL67348, R01 NS058700, R01 AR48797, R01 DK071891, R01 AG058921, the General Clinical Research Center of the Wake Forest University School of Medicine (M01 RR07122, F32 HL085989), the American Diabetes Association, and a pilot grant from the Claude Pepper Older Americans Independence Center of Wake Forest University Health Sciences (P60 AG10484).

Framingham Heart Study (FHS): This project has been funded in whole or in part with Federal funds from the National Heart Lung and Blood Institute, National Institutes of Health, Department of Health and Human Services, under Contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 and grant supplement R01 HL092577-06S1. We also acknowledge the dedication of the FHS study participants without whom this research would not be possible.

Genetic Study of Atherosclerosis Risk (GeneSTAR): We would like to thank our participants and staff for their valuable contributions. GeneSTAR was supported by grants from the NIH/NHLBI (HL72518, HL087698, HL099747, K23HL105897, HL112064) and by a grant from the National Center for Research Resources and the National Center for Advancing Translational Sciences, NIH (UL1 RR 025005) to the Johns Hopkins Institute for Clinical & Translational Research.

Genetic Epidemiology Network of Arteriopathy (GENOA): Support for GENOA was provided by the NHLBI (HL054457, HL054464, HL054481, HL119443, HL087660, and HL085571) of the NIH. We would like to thank the Mayo Clinic Genotyping Core. We would also like to thank the GENOA families.

Jackson Heart Study (JHS): The Jackson Heart Study (JHS) is supported and conducted in collaboration with Jackson State University (HHSN268201800013I), Tougaloo College (HHSN268201800014I), the Mississippi State Department of Health (HHSN268201800015I) and the University of Mississippi Medical Center (HHSN268201800010I, HHSN268201800011I and HHSN268201800012I) contracts from the NHLBI and the National Institute for Minority Health

and Health Disparities (NIMHD). The authors also wish to thank the staff and participants of the JHS.

Leiden Longevity Study (LLS): The Leiden Longevity Study has received funding from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement number 259679. This study was financially supported by the Innovation-Oriented Research Program on Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), and by BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007 and 184.033.111).

Multi-Ethnic Study of Atherosclerosis (MESA): The MESA projects are conducted and supported by the National Heart, Lung, and Blood Institute (NHLBI) in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, and UL1-TR-001420, UL1TR001881, DK063491, R01HL105756, R01HL122801, and by the National Center for Research Resources, Grant UL1RR033176. A full list of participating MESA investigators and institutes can be found at http://www.mesa-nhlbi.org.

The Young Finns Study (YFS): The Young Finns Study (YFS) has been financially supported by the Academy of Finland: grants 322098, 286284, 134309 (Eye), 126925, 121584, 124282, 129378 (Salve), 117797 (Gendi), and 141071 (Skidi); the Social Insurance Institution of Finland; Competitive State Research Financing of the Expert Responsibility area of Kuopio, Tampere and Turku University Hospitals (grant X51001); Juho Vainio Foundation; Paavo Nurmi Foundation; Finnish Foundation for Cardiovascular Research ; Finnish Cultural Foundation; The Sigrid Juselius Foundation; Tampere Tuberculosis Foundation; Emil Aaltonen Foundation; Yrjö Jahnsson Foundation; Signe and Ane Gyllenberg Foundation; Diabetes Research Foundation of Finnish Diabetes Association; EU Horizon 2020 (grant 755320 for TAXINOMISIS and grant 848146 for To Aition); European Research Council (grant 742927 for MULTIEPIGEN project); Tampere University Hospital Supporting Foundation, Finnish Society of Clinical Chemistry and the Cancer Foundation Finland.

Supplementary tables

Supplemental tables are available as Excel spreadsheets. Below is a list of each table.

Supplementary Table 1. Characteristics of the included cohorts in the meta-analysis of CAC quantity

Supplementary Table 2. Genomic risk loci and lead independent SNPs in GWAS of CAC quantity

Supplementary Table 3. Sex-specific GWAS and SNP-sex interaction tests for novel and known independent lead CAC SNPs in a subset of cohorts

Supplementary Table 4. Genetic signals identified by conditional analyses

Supplementary Table 5. Summary of top ranked SNPs and 95% credible set using combined and ancestry-specific summary statistics

Supplementary Table 6. Genes for CAC quantity mapped by positional mapping, eQTL mapping or chromatin interaction mapping in FUMA

Supplementary Table 7. Genes associated with CAC identified through a genome-wide gene-based association study

Supplementary Table 8. Polygenic priority score (PoPS) gene prioritization analysis

Supplementary Table 9. Summary-level Mendelian Randomization prioritization of CAC loci (qval< 0.10)

Supplementary Table 10. Colocalization results for CAC genes prioritized using coloc (PP4 > 0.80)

Supplementary Table 11a. Epigenome based fine-mapping of CAC loci using activity-by-contact mapping in cardiometabolic cells/tissues

Supplementary Table 11b. Epigenome based fine-mapping of CAC loci using enhancer gene linking in cardiometabolic cells/tissues

Supplementary Table 12. Cell type-specific heritability enrichment of CAC associated variants in open chromatin regions

Supplementary Table 13. Pathway enrichment analysis of CAC and CAD specific and shared associated genes

Supplementary Table 14. Associations of CAC associated loci with atherosclerotic plaque phenotypes

Supplementary Table 15. Genetic correlations between CAC and atherosclerosis traits using LD score and meta-GWAS

Supplementary Table 16a. Mendelian Randomization analysis to test for the causal relationship between risk factors or CAD and CAC

Supplementary Table 16b. Mendelian Randomization analysis to test for the causal relationship between CAC and CAD using all CAC loci

Supplementary Table 16c. Mendelian Randomization analysis to test for the causal relationship between CAD and CAC using CAC-specific loci

Supplementary Table 17. Association of CAC variants with CAD

Supplementary Table 18. Phenome-wide association study analysis of CAC variants

Supplementary Table 19a. Druggability of CAC candidate genes with multiple lines of functional evidence

Supplementary Table 19b. Druggability of CAC candidate genes with one line of functional evidence

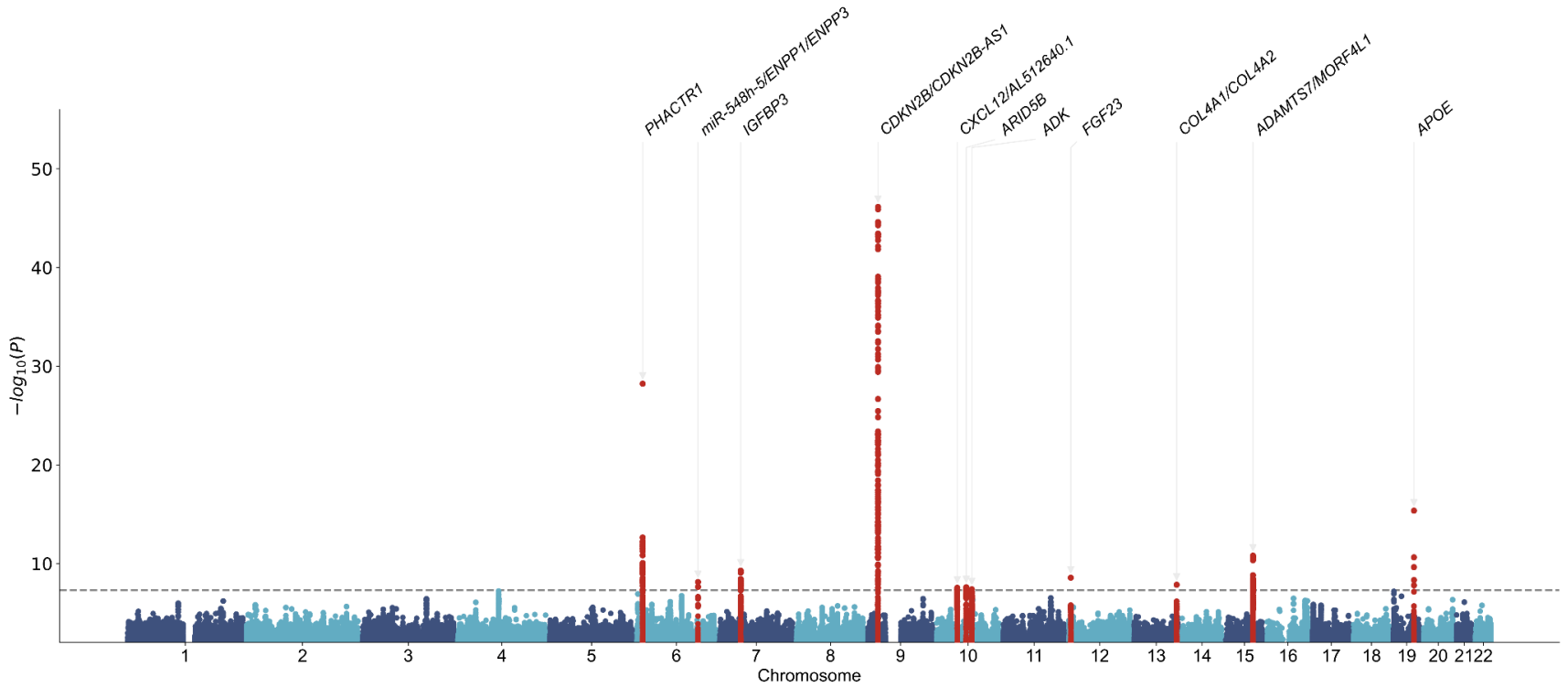
Supplementary Table 20. Overview of the genes identified through various methods

Supplementary Table 21. Study-specific genotyping, quality control, imputation, and analysis

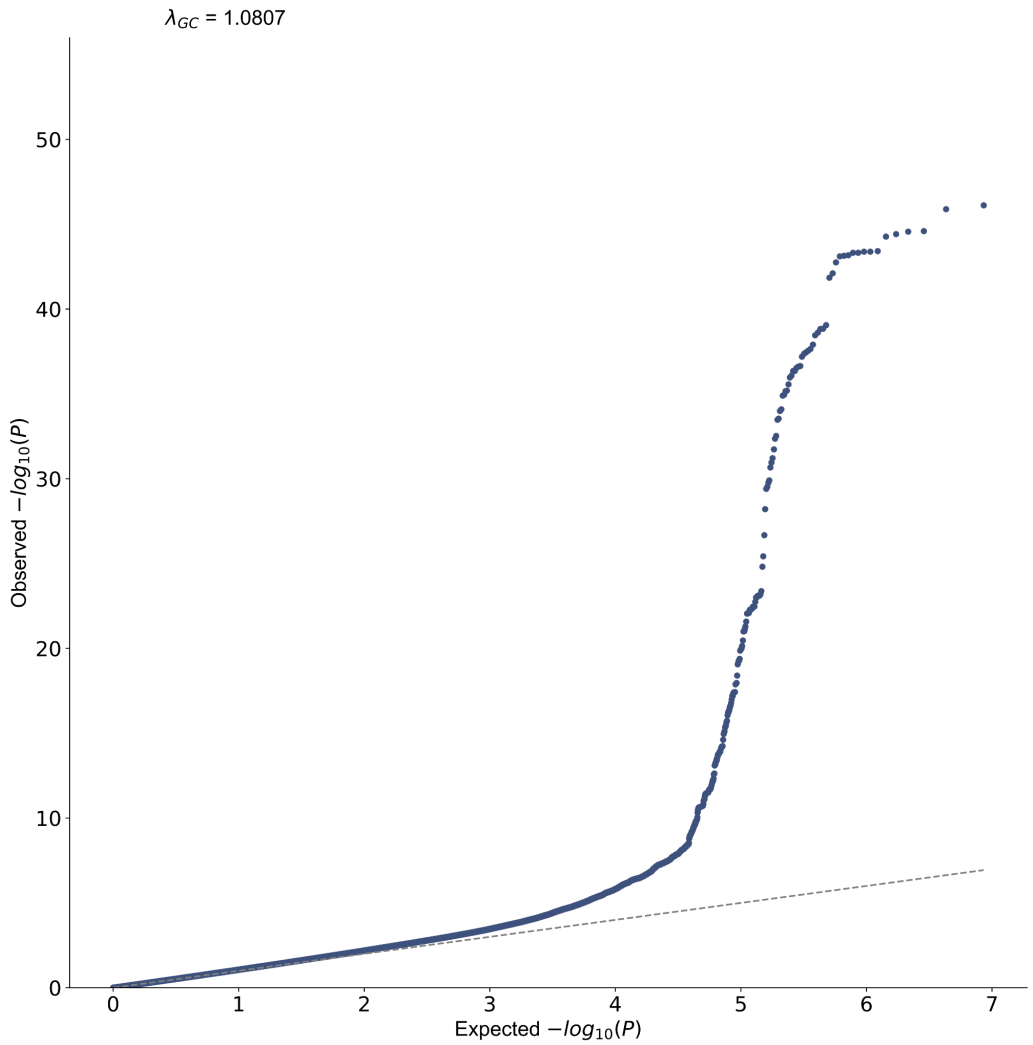
Supplementary Table 22. List of included cohorts and their respective Ethical Review Boards' (ERB) or Institutional Review Boards' (IRB) decisions

Supplementary Table 23. Acknowledgements and grant funding support for each study

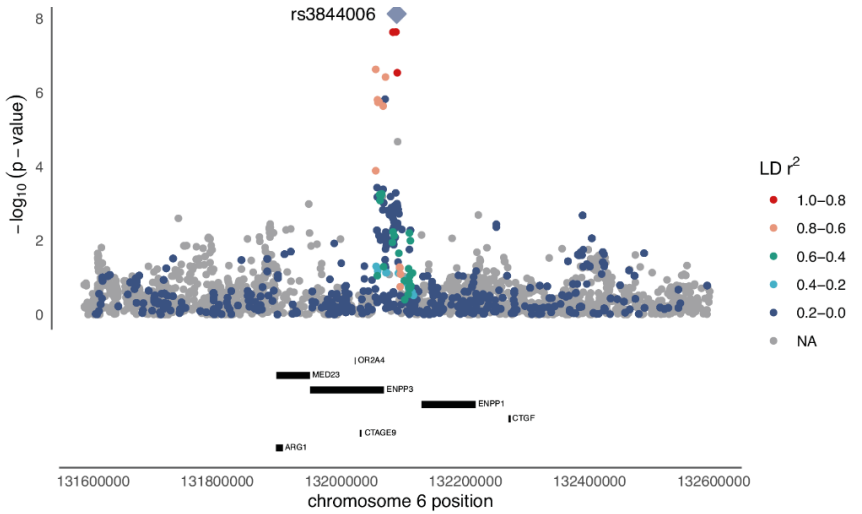
Supplementary figures

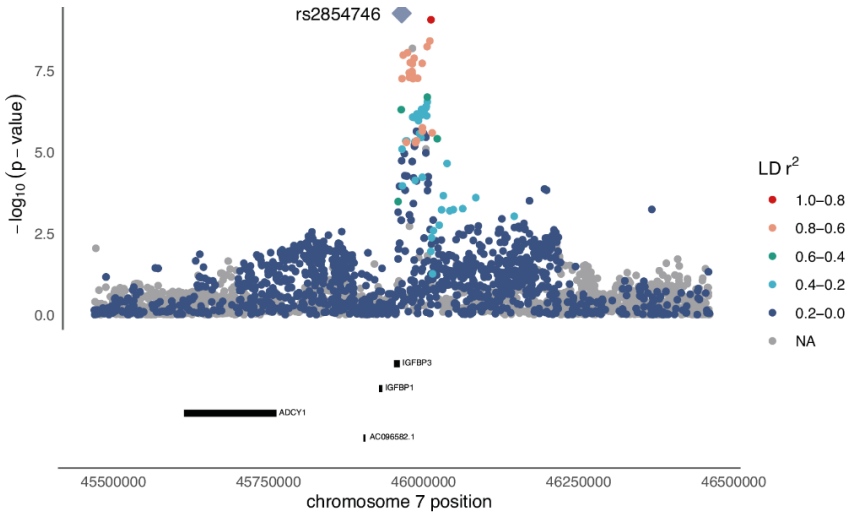


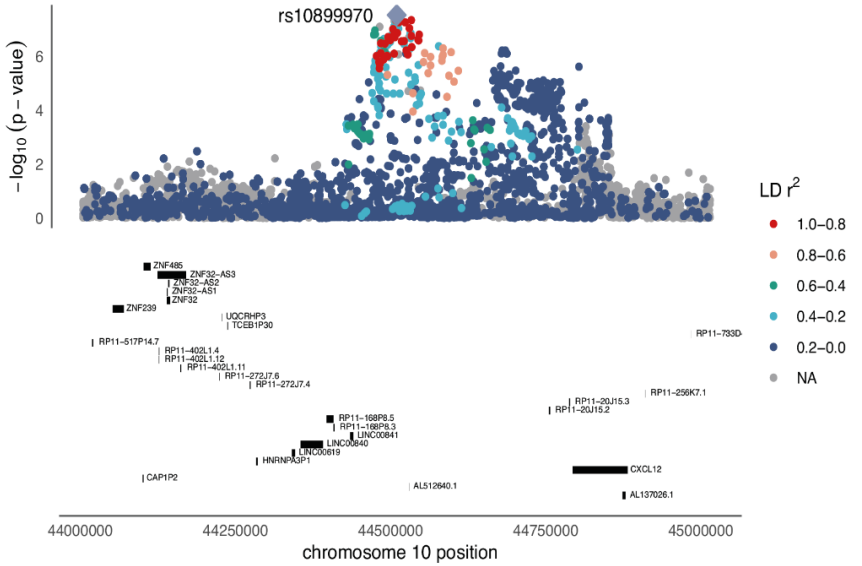
Supplementary Figure 1. GWAS meta-analysis for coronary artery calcification. Manhattan plot displays all associations per variant ordered according to their genomic position on the x-axis and showing the strength of the association with the $-\log_{10}$ transformed p values on the y-axis. The plot shows 11 (including 8 novel and 3 known) genetic loci (all marked red, ±500kb r^2>0.8 from top variant) associated with CAC at a significance level of P $<5\times10^{-8}$ (dotted line), for the combined-ancestry meta-analysis (up to 35,776 individuals from 22 studies). P values (two-sided) were derived from a meta-analysis using a fixed-effects model with an inverse-variance weighted approach. Annotated genes are those mentioned in **Table 1**.

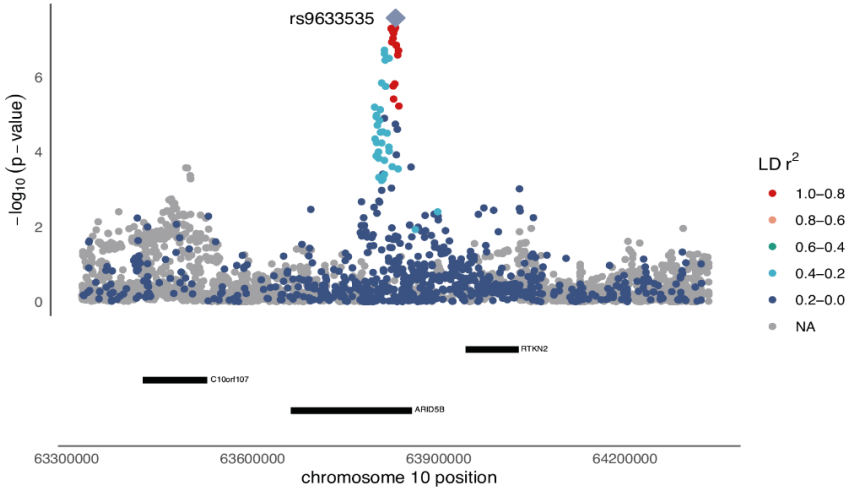


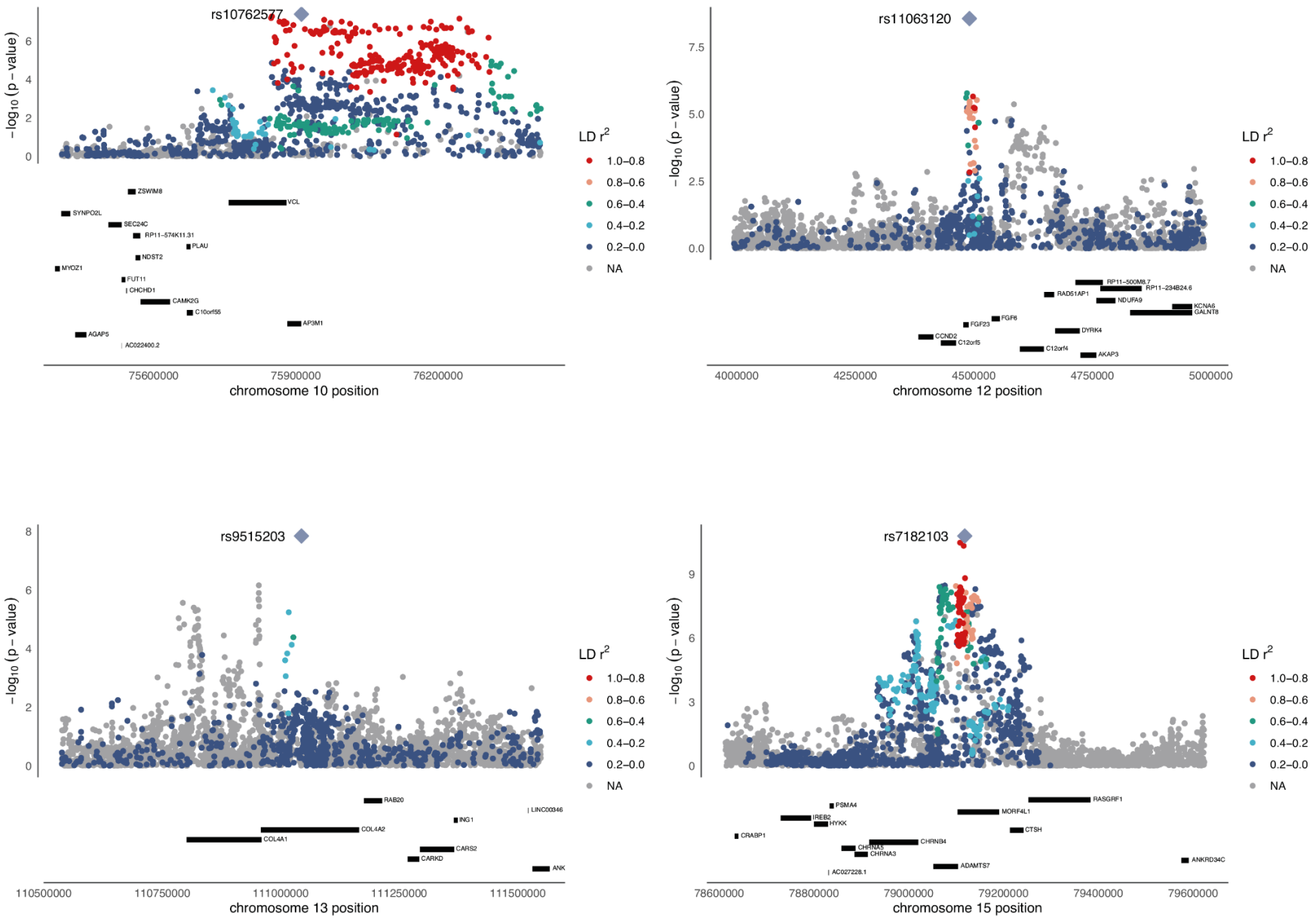
Supplementary Figure 2. Quantile-quantile plot for coronary artery calcification. The genomic inflation factor, λ=1.08, showed adequate correction for population stratification. Meta-analysis p-values determined from weighted z-scores in fixed effects model and central association p-values determined from chi-square test statistic.

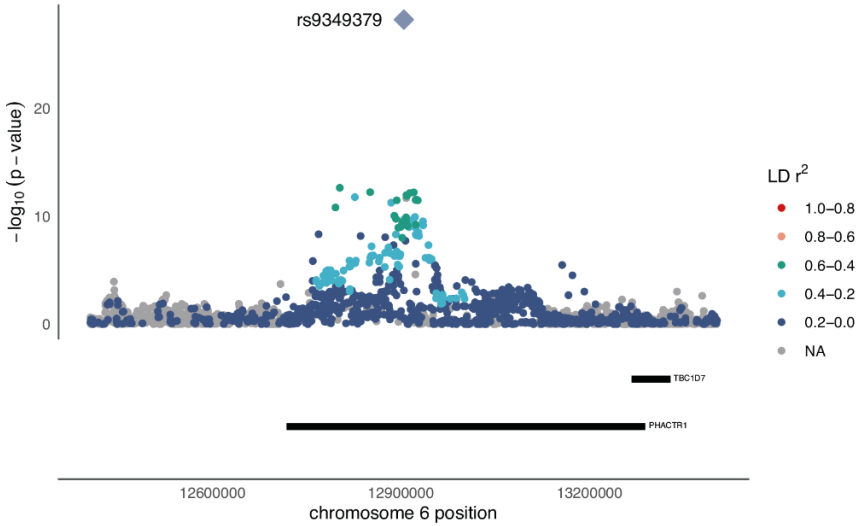

rs3844006
$-\log_{10}$ (p – value)
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA
OR2A4
MED23
ENPP3
ENPP1
CTGF
CTAGE9
ARG1
131600000
131800000
132000000
132200000
132400000
132600000
chromosome 6 position

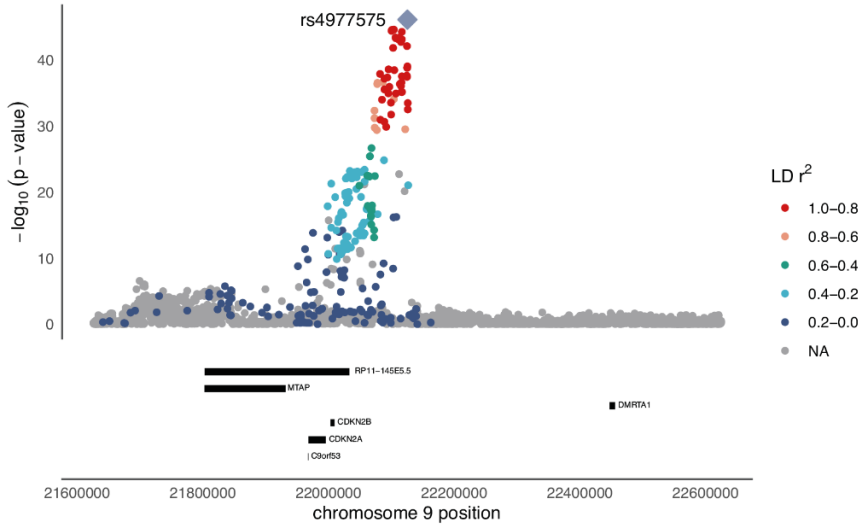

rs2854746
$-\log_{10}$ (p – value)
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA
IGFBP3
IGFBP1
ADCY1
AC096582.1
45500000
45750000
46000000
46250000
46500000
chromosome 7 position

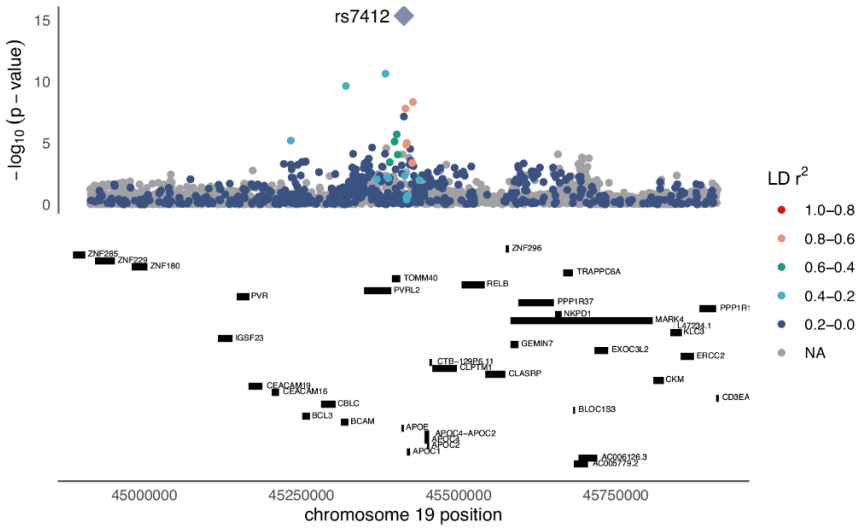

rs10899970
$-\log_{10}$ (p – value)
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA
ZNF485
ZNF32–AS3
ZNF32–AS2
ZNF32–AS1
ZNF32
ZNF239
UQCRHP3
TCEB1P30
RP11–517P14.7
RP11–402L1.4
RP11–402L1.12
RP11–402L1.11
RP11–272J7.6
RP11–272J7.4
RP11–733D
RP11–256K7.1
RP11–20J15.3
RP11–20J15.2
RP11–168P8.5
RP11–168P8.3
LINC00841
LINC00840
LINC00619
HNRNPA3P1
CXCL12
CAP1P2
AL512640.1
AL137026.1
44000000
44250000
44500000
44750000
45000000
chromosome 10 position


rs9633535
$-\log_{10}$ (p – value)
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA
RTKN2
C10orf107
ARID5B
63300000
63600000
63900000
64200000
chromosome 10 position

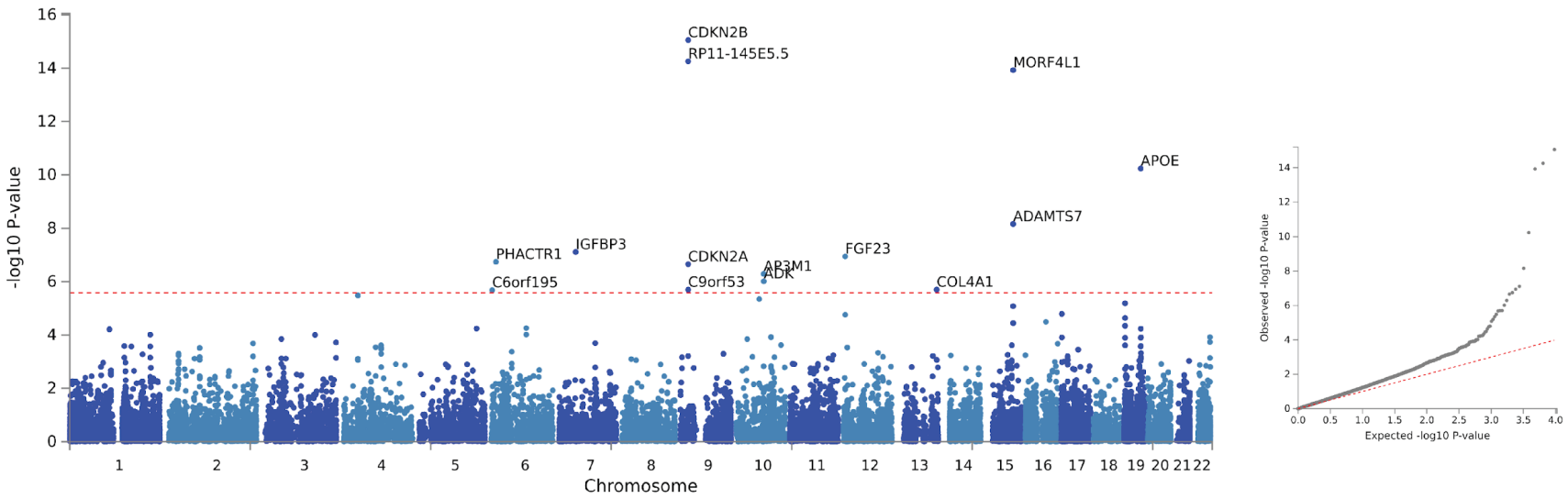

rs10762577
−log10 (p − value)
LD r2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA
ZSWIM8
VCL
SYNPO2L
SEC24C
RP11−574K11.31
PLAU
NDST2
MYOZ1
FUT11
CHCHD1
CAMK2G
C10orf55
AP3M1
AGAP5
AC022400.2
75600000
75900000
76200000
chromosome 10 position
rs11063120
RP11−500M8.7
RP11−234B24.6
RAD51AP1
NDUFA9
KCNA6
GALNT8
FGF6
FGF23
DYRK4
CCND2
C12orf5
C12orf4
AKAP3
4000000
4250000
4500000
4750000
5000000
chromosome 12 position
rs9515203
RAB20
LINC00346
ING1
COL4A2
COL4A1
CARS2
CARKD
ANK
110500000
110750000
111000000
111250000
111500000
chromosome 13 position
rs7182103
RASGRF1
PSMA4
MORF4L1
IREB2
HYKK
CTSH
CRABP1
CHRNB4
CHRNA5
CHRNA3
ANKRD34C
ADAMTS7
AC027228.1
78600000
78800000
79000000
79200000
79400000
79600000
chromosome 15 position


rs9349379
$-\log_{10}$ (p – value)
20
10
0
TBC1D7
PHACTR1
12600000
12900000
13200000
chromosome 6 position
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA

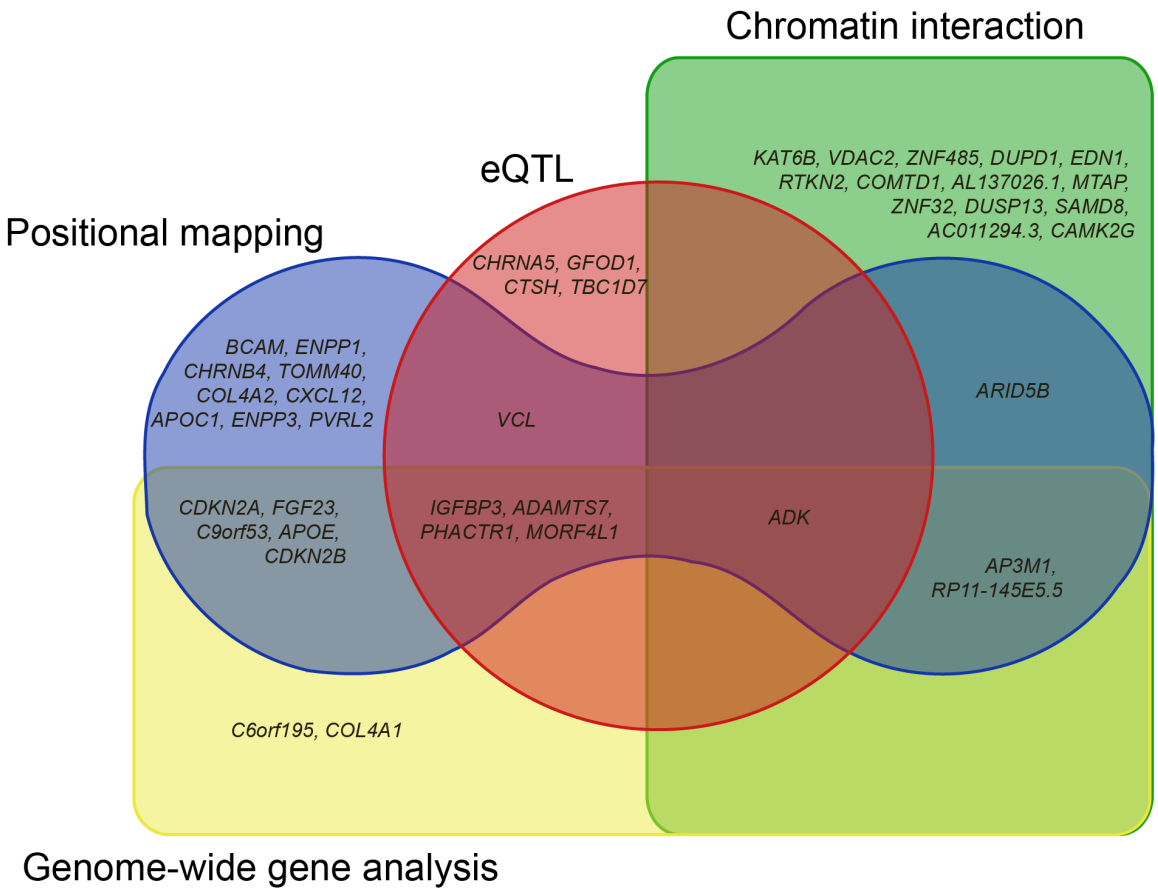

rs4977575
$-\log_{10}$ (p – value)
40
30
20
10
0
RP11–145E5.5
MTAP
DMRTA1
CDKN2B
CDKN2A
C9orf53
21600000
21800000
22000000
22200000
22400000
22600000
chromosome 9 position
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA


rs7412
$-\log_{10}$ (p – value)
15
10
5
0
ZNF296
TRAPPC6A
RELB
PVRL2
PVR
PPP1R37
NKPD1
MARK4
GEMIN7
EXOC3L2
ERCC2
CLASRP
CKM
BLOC1S3
CBLC
BCAM
APOE
APOC4–APOC2
APOC2
APOC1
45000000
45250000
45500000
45750000
chromosome 19 position
LD r^2
1.0–0.8
0.8–0.6
0.6–0.4
0.4–0.2
0.2–0.0
NA

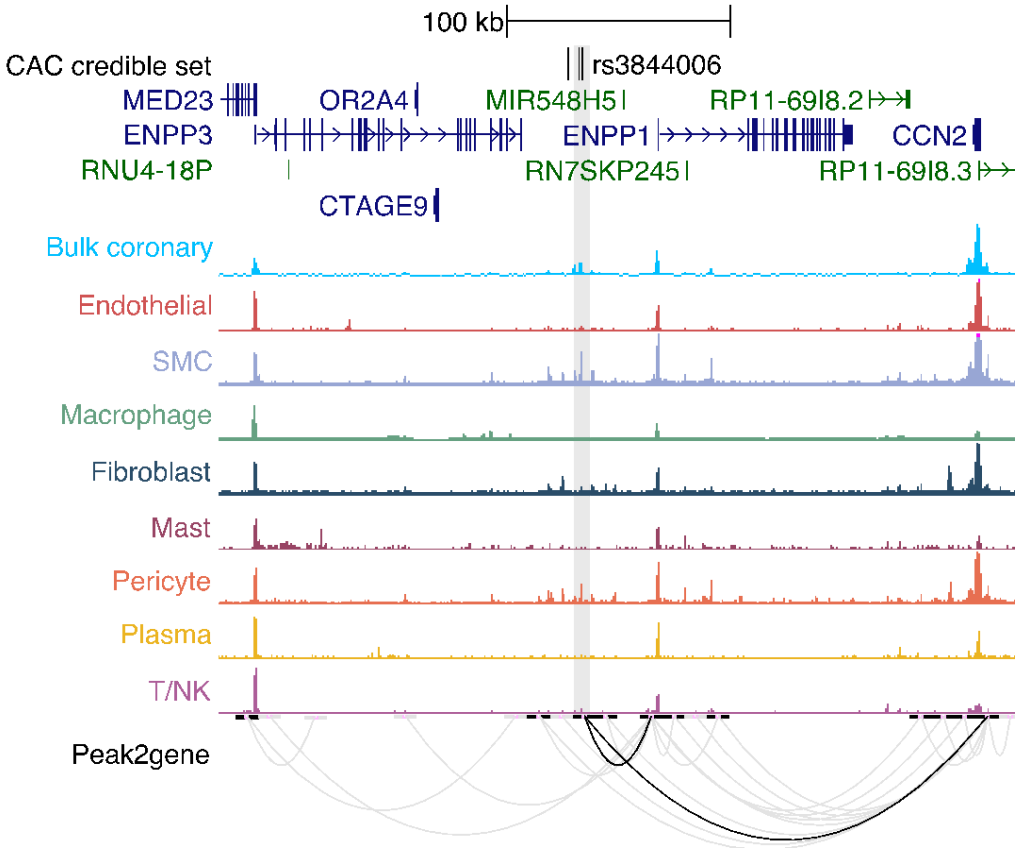
Supplementary Figure 3. Regional association plots of coronary artery calcification for the 11 lead SNPs. Regional association plots showing per variant associations ordered according to their genomic position on the x-axis (top panel) and the protein coding genes (bottom panel, Ensembl GRCh37 Release 105 [Dec 2021]). The y-axis shows the strength of the association as the $-\log_{10}$ transformed meta-analysis p-values. Meta-analysis p-values determined from weighted z-scores in fixed effects model and central association p-values determined from chi-square test statistic. Colors indicate the linkage disequilibrium (LD) correlation coefficient (r^2) of the given variant with the top variant (depicted in purple). Plots are made with RACER (https://github.com/oliviasabik/RACER) which uses LDproxy (https://ldlink.nci.nih.gov/?tab=ldproxy), this only returns variants with non-zero r^2-values values. Thus the LD bin labeled '0.2-0' is not actually inclusive of zero, because LDproxy does not return zero values. NA indicates the r^2 could not be calculated. Also, plotting was done using the trans-ancestral results and EUR population of the LD reference (1000G phase 3), which may influence the r^2 calculations for some variants (the 'ALL'-population is not a valid option in the API for LDproxy used by RACER).



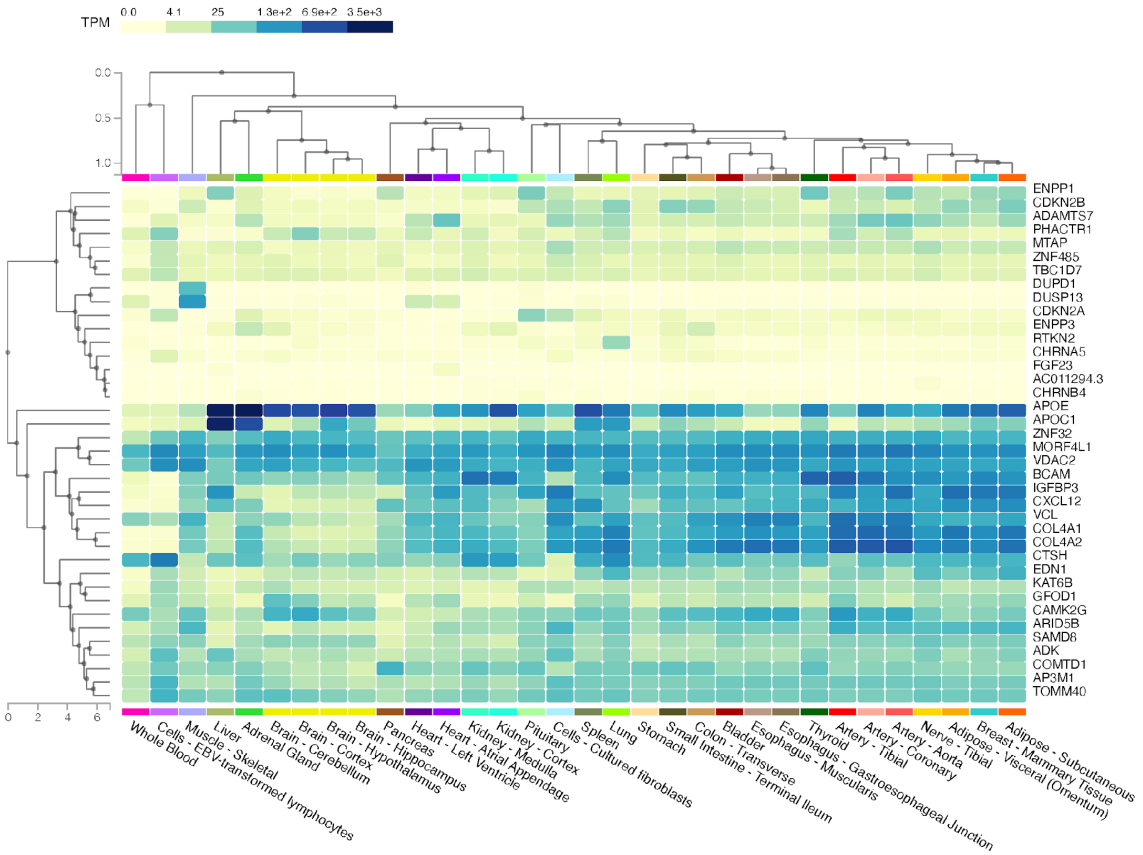
Supplementary Figure 4. Genome-wide gene association analysis (GWGA) for coronary artery calcification. Manhattan plot (left) and quantile-quantile plot (right) of coronary artery calcification quantity based on a GWGA using MAGMA[24]. GWGA provides empirically derived P-values of association with the given trait (CAC) based on all variants located in a gene while considering the underlying linkage disequilibrium structure. MAGMA p-values computed using the T* test statistic as described[24].



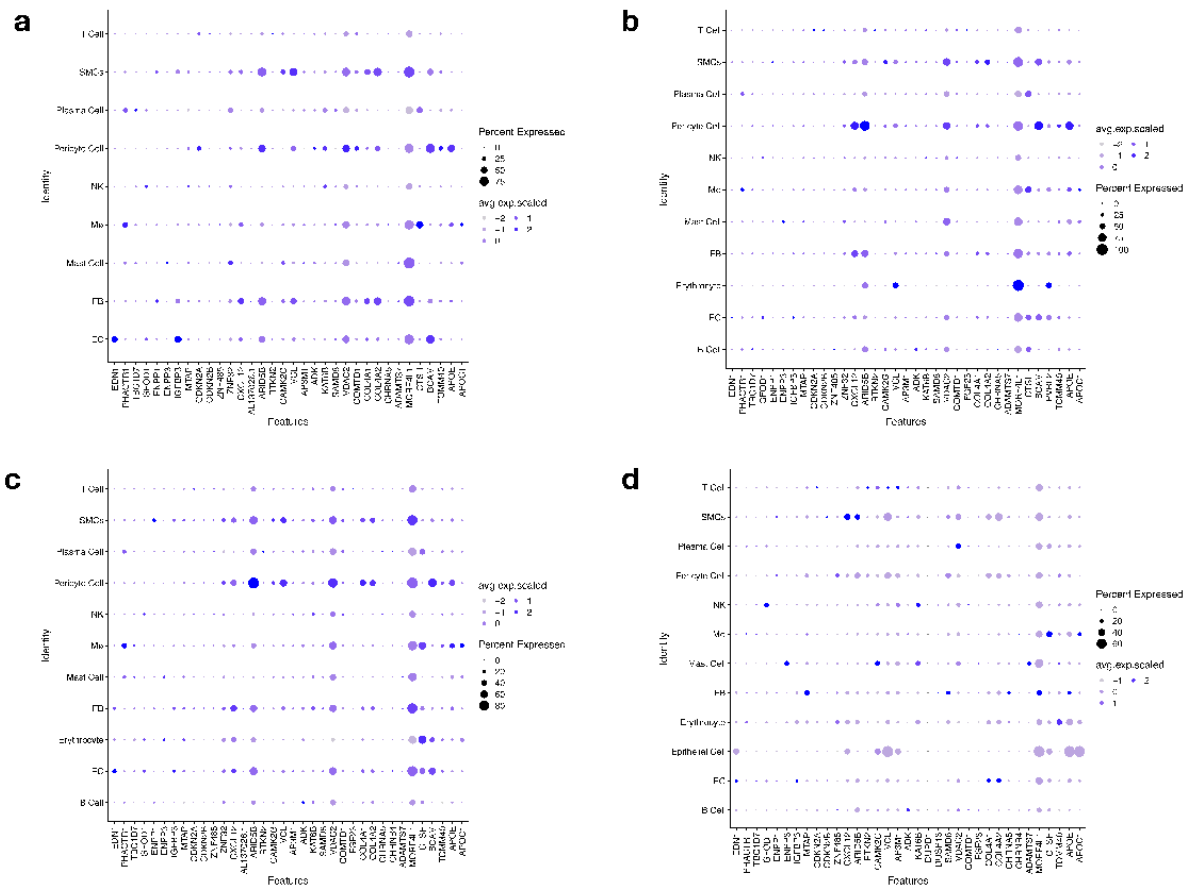
Supplementary Figure 5. Genes implicated in coronary artery calcification. Overview of the prioritized genes based on four FUMA strategies; positional gene mapping, expression quantitative trait loci (eQTL) gene mapping, chromatin interaction[5], and genome-wide gene association analysis (GWGA) using MAGMA[24]. Diagram shows the prioritized genes based on three FUMA strategies; positional gene mapping, expression quantitative trait loci (eQTL) gene mapping, chromatin interaction mapping[5], and genome-wide gene association analysis (GWGA) using MAGMA, as well as adding 3 candidate genes (*ENPP1*, *ENPP3*, and *CXCL12*) by annotating the nearest protein-coding genes as described in the main manuscript and the methods. The locus names labeled within the Venn diagram correspond to those in Tables S5 and S6 and as described in the main manuscript. The circled locus names highlight loci that were implicated by 3 or more FUMA mapping strategies. Created using https://bioinformatics.psb.ugent.be/webtools/Venn/ and Adobe Illustrator 27.5.



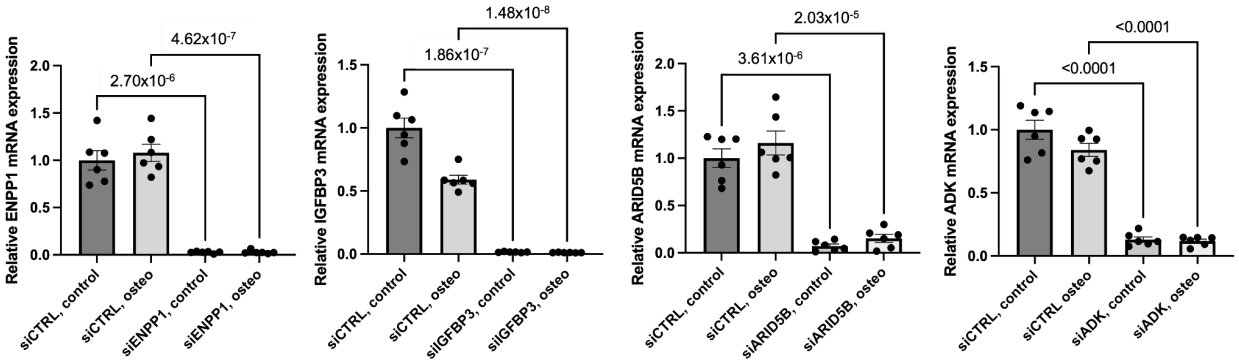
Supplementary Figure 6. UCSC browser tracks for *ENPP1/ENPP3* CAC locus. Browser screenshot shows overlap of CAC credible SNPs (highlighting top credible SNP rs3844006) with accessible chromatin in bulk coronary artery or specific coronary artery cell types from single-nucleus ATAC-seq (snATAC-seq) based profiling. All track heights scaled uniformly for bulk or single-cell datasets. Peak2gene links determined in ArchR from snATAC profiles across all cells, with those overlapping CAC credible SNP rs34844006 highlighted in black. GENCODE protein-coding and non-coding genes annotated in blue and green, respectively.



Supplementary Figure 7. GTEx tissue distribution for CAC candidate genes. Heatmap showing clustering of normalized CAC candidate gene expression across 32 different tissues in GTEx. Reproductive and duplicate tissue types omitted for clarity. TPM: transcripts per million.



Supplementary Figure 8. Single cell expression of GWAS loci associated genes mapped in subclinical coronary arteries and advanced carotid plaques. Panel **A**-**C** shows (subclinical) coronary artery data from Pan *et al.*[25], Wirka *et al.*[26], and Alsaigh *et al.*[27], respectively. Panel **D** shows advanced carotid plaque data from Slenders et al.[28]. On the x-axis are the genes (features) as they are expressed in the given cell type. On the y-axis the cell identity is determined by Seurat based on the Tabula reference[29]. These data are also visible through PlaqView (www.plaqview.com)[30].



Supplementary Figure 9. Efficiency of siRNA-mediated knockdown of *ENPP1*, *IGFBP3*, *ARID5B*, and *ADK.* Relative mRNA expression levels in human coronary VSMCs transfected with siRNA against *ENPP1*, *IGFBP3*, *ARID5B*, or *ADK* and treated in control media and osteogenic media. For all genes, knockdown efficiency was > 85-99%. n=6 independent biological replicates for each experiment. Values represent mean ± standard error of the mean. All statistical comparisons shown were made using a two-tailed student t test.

Supplementary References

1. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–75, S1–3 (2012).
2. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
3. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
4. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
5. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
6. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
7. Fulco, C. P. *et al.* Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669 (2019).
8. Nasser, J. *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).
9. Liu, Y., Sarkar, A., Kheradpour, P., Ernst, J. & Kellis, M. Evidence of reduced recombination rate in human regulatory domains. *Genome Biol.* **18**, 193 (2017).
10. Verhoeven, B. A. N. *et al.* Athero-express: differential atherosclerotic plaque expression of mRNA and protein in relation to cardiovascular events and patient characteristics. Rationale and design. *Eur. J. Epidemiol.* **19**, 1127–1133 (2004).

11. van Lammeren, G. W. *et al.* Atherosclerotic plaque vulnerability as an explanation for the increased risk of stroke in elderly undergoing carotid artery stenting. *Stroke* **42**, 2550–2555 (2011).
12. Verhoeven, B. *et al.* Carotid atherosclerotic plaques in patients with transient ischemic attacks and stroke have unstable characteristics compared with plaques in asymptomatic and amaurosis fugax patients. *J. Vasc. Surg.* **42**, 1075–1081 (2005).
13. Hellings, W. E. *et al.* Intraobserver and interobserver variability and spatial differences in histologic examination of carotid endarterectomy specimens. *J. Vasc. Surg.* **46**, 1147–1154 (2007).
14. van der Laan, S. W. *et al.* Genetic Susceptibility Loci for Cardiovascular Disease and Their Impact on Atherosclerotic Plaques. *Circ Genom Precis Med* **11**, e002115 (2018).
15. Laurie, C. C. *et al.* Quality control and quality assurance in genotypic data for genome-wide association studies. *Genet. Epidemiol.* **34**, 591–602 (2010).
16. Das, S. *et al.* Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
17. Hashimshony, T. *et al.* CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol.* **17**, 77 (2016).
18. Depuydt, M. A. C. *et al.* Microanatomy of the Human Atherosclerotic Plaque by Single-Cell Transcriptomics. *Circ. Res.* **127**, 1437–1455 (2020).
19. Watanabe, K. *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
20. Turner, A. W. *et al.* Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.* (2022)

doi:10.1038/s41588-022-01069-0.

21. Hartiala, J. A. *et al.* Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *Eur. Heart J.* **42**, 919–933 (2021).
22. Malhotra, R. *et al.* HDAC9 is implicated in atherosclerotic aortic calcification and affects vascular smooth muscle cell phenotype. *Nat. Genet.* **51**, 1580–1587 (2019).
23. O'Rourke, C. *et al.* Calcification of Vascular Smooth Muscle Cells and Imaging of Aortic Calcification and Inflammation. *J. Vis. Exp.* (2016) doi:10.3791/54017.
24. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
25. Pan, H. *et al.* Single-Cell Genomics Reveals a Novel Cell State During Smooth Muscle Cell Phenotypic Switching and Potential Therapeutic Targets for Atherosclerosis in Mouse and Human. *Circulation* **142**, 2060–2075 (2020).
26. Wirka, R. C. *et al.* Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289 (2019).
27. Alsaigh, T., Evans, D., Frankel, D. & Torkamani, A. Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. *bioRxiv* 2020.03.03.968123 (2020) doi:10.1101/2020.03.03.968123.
28. Slenders, L. *et al.* Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. *Eur Heart J Open* (2021) doi:10.1093/ehjopen/oeab043.
29. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat.*

Biotechnol. **36**, 411–420 (2018).

30. Ma, W. F. *et al.* Enhanced single-cell RNA-seq workflow reveals coronary artery disease cellular cross-talk and candidate drug targets. *Atherosclerosis* **340**, 12–22 (2022).