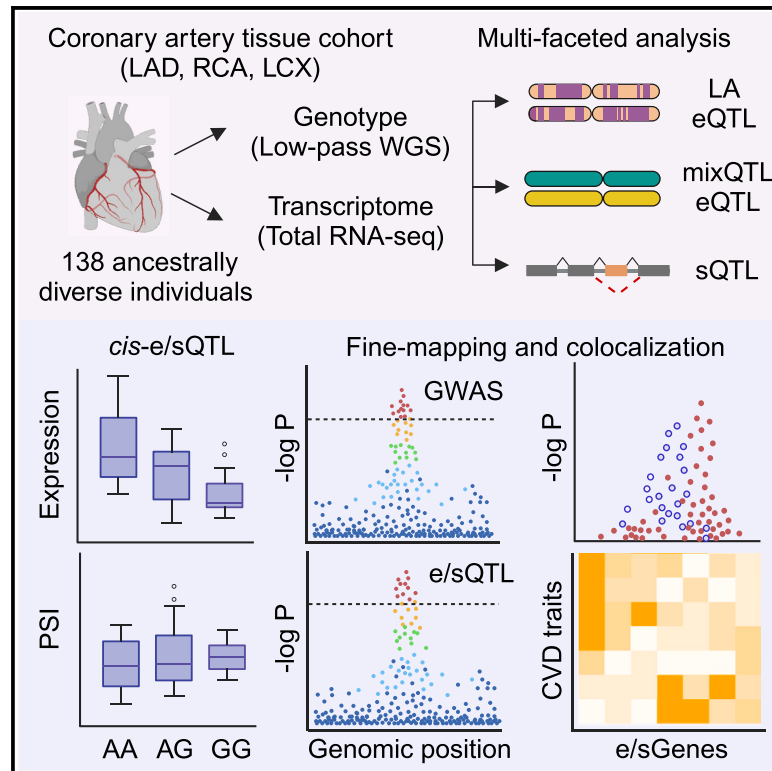


Multi-ancestry genetic analysis of gene regulation in coronary arteries prioritizes disease risk loci

Graphical abstract



Authors

Chani J. Hodonsky, Adam W. Turner, Mohammad Daud Khan, ..., Jason C. Kovacic, Johan L.M. Björkegren, Clint L. Miller

Correspondence

clintm@virginia.edu

In brief

Hodonsky et al. perform gene expression profiling and a multifaceted genetic analysis of gene regulation in coronary artery tissues from an ancestrally diverse patient cohort. This new resource provides further insights into the heritable mechanisms of coronary artery disease risk and could serve as a template for multi-ancestry studies of other complex phenotypes.

Highlights

- Human coronary artery is a relevant tissue for profiling atherosclerosis gene expression
- Multifaceted QTL mapping methodology improves discovery in diverse populations
- Colocalization and epigenomic fine-mapping prioritize candidate risk variants and genes
- Publicly available pipeline is generalizable across phenotypes and study designs



Article

Multi-ancestry genetic analysis of gene regulation in coronary arteries prioritizes disease risk loci

Chani J. Hodonsky,¹ Adam W. Turner,¹ Mohammad Daud Khan,¹ Nelson B. Barrientos,^{1,2} Ruben Methorst,³ Lijiang Ma,⁴ Nicolas G. Lopez,⁵ Jose Verdezoto Mosquera,^{1,6} Gaëlle Auguste,¹ Emily Farber,¹ Wei Feng Ma,^{1,7} Doris Wong,^{1,6} Suna Onengut-Gumuscu,¹ Maryam Kavousi,⁸ Patricia A. Peyser,⁹ Sander W. van der Laan,³ Nicholas J. Leeper,⁵ Jason C. Kovacic,^{10,11,12} Johan L.M. Björkegren,^{4,13} and Clint L. Miller^{1,5,14,15,*}

¹Center for Public Health Genomics, University of Virginia, Charlottesville, VA 22908, USA

²Department of Genetic Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

³Central Diagnostics Laboratory, Division Laboratories, Pharmacy, and Biomedical Genetics, University Medical Center Utrecht, Utrecht University, 3584 CX Utrecht, the Netherlands

⁴Department of Genetics and Genomic Sciences, Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

⁵Division of Vascular Surgery, Department of Surgery, Stanford University, Stanford, CA 94305, USA

⁶Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA 22908, USA

⁷Medical Scientist Training Program, Department of Pathology, University of Virginia, Charlottesville, VA 22908, USA

⁸Department of Epidemiology, Erasmus University Medical Center, 3000 CA Rotterdam, the Netherlands

⁹Department of Epidemiology, University of Michigan, Ann Arbor, MI 48019, USA

¹⁰Cardiovascular Research Institute, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA

¹¹Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010, Australia

¹²St. Vincent's Clinical School, University of New South Wales, Sydney, NSW 2052, Australia

¹³Department of Medicine, Huddinge, Karolinska Institutet, 141 52 Huddinge, Sweden

¹⁴Department of Public Health Sciences, University of Virginia, Charlottesville, VA 22908, USA

¹⁵Lead contact

*Correspondence: clintm@virginia.edu

<https://doi.org/10.1016/j.xgen.2023.100465>

SUMMARY

Genome-wide association studies (GWASs) have identified hundreds of risk loci for coronary artery disease (CAD). However, non-European populations are underrepresented in GWASs, and the causal gene-regulatory mechanisms of these risk loci during atherosclerosis remain unclear. We incorporated local ancestry and haplotypes to identify quantitative trait loci for expression (eQTLs) and splicing (sQTLs) in coronary arteries from 138 ancestrally diverse Americans. Of 2,132 eQTL-associated genes (eGenes), 47% were previously unreported in coronary artery; 19% exhibited cell-type-specific expression. Colocalization revealed subgroups of eGenes unique to CAD and blood pressure GWAS. Fine-mapping highlighted additional eGenes, including *TBX20* and *IL5*. We also identified sQTLs for 1,690 genes, among which *TOR1AIP1* and *ULK3* sQTLs demonstrated the importance of evaluating splicing to accurately identify disease-relevant isoform expression. Our work provides a patient-derived coronary artery eQTL resource and exemplifies the need for diverse study populations and multifaceted approaches to characterize gene regulation in disease processes.

INTRODUCTION

Coronary artery disease (CAD) is the leading cause of death worldwide, and it results from chronic inflammatory processes involving both genetic and environmental risk factors. CAD manifests as the development of atherosclerotic plaques in the coronary arteries of the heart, which can lead to erosion or plaque rupture and ultimately myocardial infarction. Genome-wide association studies (GWASs) have now reported more than 400 independent loci for CAD and related clinical outcomes.^{1–11} As with other common complex traits, the majority of lead CAD GWAS variants reside in non-coding genomic regions, impli-

cating regulatory effects on gene expression.¹² Previous studies have mapped CAD GWAS variants to specific cell types in the vessel wall (e.g., smooth muscle cells [SMCs],¹³ endothelial cells, and immune cells)¹⁴ and refined candidate *cis*-acting regulatory elements responsible for context-specific gene expression patterns.^{14,15} However, cultured vascular cells do not recapitulate the *in vivo* cell phenotype: for instance, high-passage SMCs reprogram toward a fibroblast-like state accompanied by rapid loss of differentiated marker gene expression.

Fine-mapping GWAS loci can help prioritize candidate causal variants within association signals, but identifying the causal variant or target gene within a locus can still be difficult.



Overrepresentation of European- and East Asian-ancestry populations in most GWASs to date has also limited the capacity to identify independent associations within a locus and the generalizability of findings to global populations.^{16–18} Furthermore, genes within most CAD loci have not been associated with traditional risk factors (e.g., lipid levels or cholesterol metabolism), suggesting molecular mechanisms underlying physiological effects on the coronary artery vessel wall itself.

Molecular quantitative trait locus mapping in a disease-relevant tissue or cell line is a powerful approach to prioritize candidate causal genes and underlying mechanisms for complex GWAS loci.¹⁹ Prior studies have identified CAD-relevant expression quantitative trait loci (eQTLs) in bulk arterial tissues²⁰ or vessel wall cell types, including human aortic endothelial cells,²¹ human coronary artery smooth muscle cells (HCASMCs),^{13,22} and monocytes.²³ Similar to CAD GWASs and other eQTL studies, published summary statistics represent exclusively or primarily European-ancestry populations, often lack detailed phenotyping for the patients/participants.

To identify variants associated with coronary artery-specific gene expression and fine-map colocalized CAD GWAS associations, we performed a QTL mapping study in coronary artery tissues from an ancestrally diverse American patient-derived sample. We utilized multiple methods to identify eQTLs and splicing quantitative trait loci (sQTLs) in human coronary artery tissue, followed by bioinformatic characterization of potential eGene-phenotype associations. Our results not only highlight new coronary artery eQTLs at promising GWAS loci such as *TBX20*, but they also replicate and refine eQTLs previously reported in other arterial tissues, including *ARHGAP42*. Our approach represents an improved capacity to characterize gene regulation in coronary artery tissue through all stages of CAD progression. This dataset will therefore be a highly beneficial resource for better characterization of functional variants and molecular mechanisms driving CAD development.

RESULTS

Study overview for transcriptomic profiling of human coronary artery

We conducted transcriptome-wide QTL mapping of autosomal gene expression in human coronary artery tissue samples from explanted transplant tissue as well as samples collected from rejected transplant donors (STAR Methods, Table S1). An overview of our sample characteristics and primary diagnoses are described in Figure 1 and Tables S1–S3. The study comprised 138 individuals from 19 to 72 years old, with 30% being female. While 57% of these individuals were of exclusively European ancestry, 15% were of majority South Asian ancestry, 5% East Asian, 5% Indigenous American, 7% African, and 10% had no majority ancestry, representing the broad genetic diversity of the American population (Figures 1A–1C and S1). Samples were derived from all three major coronary arteries (i.e., left anterior descending coronary artery [LAD], right coronary artery [RCA], and left circumflex artery [LCX], Figure 1D). Majority inferred ancestry groups were represented across diagnoses, but only majority-European and -South Asian genetic ancestries were represented in all primary diagnostic categories (Figure 1E;

Table S1). Up to 5.83 million variants genotyped via low-pass whole-genome sequencing were included depending on method-specific allele frequency and annotation requirements (described in STAR Methods, Table S2).

We performed total RNA sequencing to a median depth of 102.6 million reads per sample (Table S3) to profile both protein-coding and non-coding RNAs (Figure S1). To determine the similarity of our expression profiles to bulk RNA profiles of other tissues in GTEx as well as cultured HCASMCs, we performed multidimensional scaling. Our samples form a cluster located near the left ventricle, muscle, pancreas, fibroblasts, and liver tissues (Figure S1). This distinct but proximal clustering aligns with expectations given differences in sample collection/storage methods and cold ischemia times (time lapsed after cessation of blood flow). Since eQTL studies have primarily been performed in genetically homogeneous populations, information on preferred methods for inclusive study populations is limited. We therefore applied two complementary approaches to attempt to maximize power for identifying associations that may not have globally consistent allele frequencies, i.e., by evaluating haplotype-specific associations (“mixQTL” for the total sample or “mixQTL_{EUR}” for the 100% European-ancestry subset analysis) or ancestry-specific associations (local ancestry adjusted, henceforth referred to as “LA”).^{24,25} For mixQTL analyses, we incorporated three global genetic principal components; for LA, locus-specific genetic ancestry was statistically inferred on a continental scale using a reference panel of genotypes from 1000 Genomes (1000G) participants (see STAR Methods). After filtering for method-specific criteria, up to 20,100 autosomal protein-coding genes and lncRNAs met inclusion criteria for eQTL and sQTL analyses.

Coronary artery eQTL discovery

To identify genetic variants associated with gene expression in our diverse coronary artery tissue cohort, we performed eQTL analyses incorporating haplotype-specific (mixQTL)²⁴ or LA information. Overall, we identified 2,132 and 793 coronary artery eGenes using mixQTL or LA, respectively (Tables 1, S4, and S5; Figure 2A). Between LA and mixQTL analyses, 457 shared eGenes were identified (Figure 2B); 45 lead SNPs were common to both approaches (Table S6). Of note, across all analyses, we report 735 total discovery eGenes (351 mixQTL, 163 LA, 210 mixQTL_{EUR}, 395 mixQTL_{downsample}; 213 protein-coding genes, 514 lncRNAs) with no expression QTLs reported in any arterial tissue in GTEx or Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task (STARNET), including genes with established roles in vascular cell types (e.g., lipase G, endothelial type [*LIPG*] and AKT serine/threonine kinase 3 [*AKT3*]).^{26,27} 40% of discovery eQTLs were >100 kb from the gene transcription start site, in line with long-standing evidence of both short- and long-range *cis*-acting regulatory mechanisms.^{28,29} We report mixQTL results from the entire study sample as our primary findings given the higher statistical power of this method.

Overview of mixQTL results

Of the 2,132 protein-coding or lncRNA mixQTL eGenes, 16% (n = 351) have not been previously reported in published arterial tissue QTL studies (Table S4). In concordance with published studies, most eQTLs were annotated as intergenic or intronic to

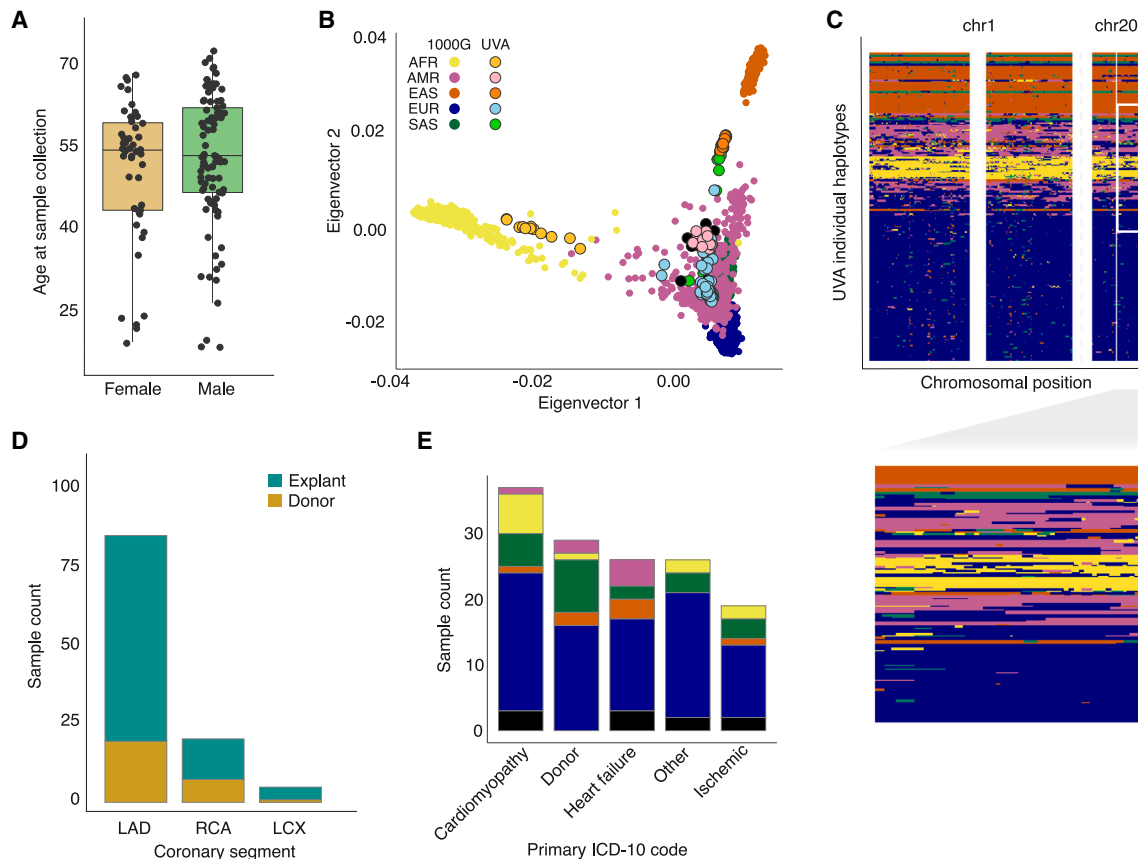


Figure 1. Overview of patient and sample characteristics

(A) Boxplot showing the age range of study participants (y axis) in females and males (x axis; orange and green, respectively). Lines represent median, boxes represent interquartile range (IQR), and upper and lower whiskers represent $1.5 \times$ IQR.
 (B) Genetic ancestry principal components 1 (x axis) and 2 (y axis) mapped onto the 1000 Genomes phase 3 reference population (bold color dots), with color corresponding to Gencove-assigned majority ancestry in our samples (lighter color dots with black outline).
 (C) Local ancestry inference reveals a complex genetic substructure for individuals with ancestral admixture, with each row in the plot representing inferred local ancestry for one haplotype of one study participant and x axis representing position on respective chromosomes. Inset, zoomed in region on chr 20 showing genetic substructure for a subset of the individual haplotypes.
 (D) Number (y axis) of coronary artery segments by type (x axis) used for RNA isolation in samples from explants (turquoise) and rejected donors (gold).
 (E) Number (y axis) of coronary artery segments by primary ICD-10 code diagnosis for explanted hearts or donor hearts. Samples were grouped into main categories listed in [Table S1](#).

their eGenes (2,859 of 3,952 available annotations for 1,122 eGenes, [Figure S2](#)). Only 2% of lead eQTLs ($n = 59$) were protein coding within their respective eGene, implicating the regulation of gene expression through transcriptional, splicing, or epigenetic mechanisms.²⁰ Thirty-nine and 119 eGenes were identified based solely on allele-specific expression and total read count tests, respectively. Overall, 1,779 published arterial eGenes had significant eQTLs in our study sample ([Figure 2C](#); [Table S7](#)). Fewer than 5% of shared associations had the same lead eQTL, but among shared lead eQTLs, we observed 98% directional consistency (64% of all replicated lead eQTLs, [Figures 2D](#) and [S2](#)). Nearly one in five eGenes exhibited cell-type-specific expression in a coronary artery single-cell RNA sequencing reference dataset (cell specificity expression score ≥ 0.7 , [Figure S2](#); [Table S8](#)),³⁰ and functional enrichment analysis revealed several pathways for cell adhesion and inflammation ([Table S9](#)).³¹

Local-ancestry-adjusted and ancestry-specific eQTLs

With regard to LA analyses, 337 eGenes were identified that did not exceed a false discovery rate (FDR) of 5% in the overall mixQTL analysis, demonstrating the merit of incorporating multiple approaches in a diverse study sample with genetic admixture ([Table S5](#)). Among LA-specific eGenes was YY1-associated protein 1 (*YY1AP1*), which has no reported coronary artery eQTLs but has been associated with vascular diseases including Grange syndrome and sudden coronary artery dissection.^{32,33} Seventeen LA lead variants were monomorphic in the 1000G East Asian superpopulation ([Table S10A](#)). Despite high-confidence calling of 1000G continental ancestries in our study sample, the small numbers of shared haplotypes at any given locus likely limited our ability to identify ancestry-specific associations using this method. Using mixQTL, 54 eGenes with lead SNPs monomorphic in one or more 1000G superpopulations were

Table 1. Discovery expression and splicing quantitative trait loci in human coronary artery tissue

Molecular phenotype	Method	Gene type	No. evaluated	No. genes (FDR <0.01)	No. genes (FDR <0.05)	No. genes (FDR <0.1)	No. discovery genes
Gene expression	mixQTL	protein-coding	14,235	1,118	1,457	1,668	127
		lncRNA	5,874	510	682	794	224
	local ancestry	protein-coding	14,274	330	482	602	41
		lncRNA	7,395	234	311	378	122
	mixQTL (Euro)	protein-coding	14,082	671	916	1,087	94
		lncRNA	5,219	285	395	467	116
mixQTL (diverse)	protein-coding	14,084	766	1,075	1,303	127	
	lncRNA	5,222	323	477	569	294	
Splicing	QTLtools	protein-coding	13,103	1,134	1,496	1,735	357
		lncRNA	1,700	152	194	232	93

identified; four overlapped with the LA results, and lead eQTLs were either shared or in high linkage disequilibrium (“LD,” [Figure S2](#); [Tables S10B](#) and [S11](#)).

Among the genes with no eQTLs exceeding genome-wide significance in the mixQTL were several interesting genes with sub-significant associations, including *VPS37B* (vacuolar protein sorting-associated protein 37B). *VPS37B* is involved in endosomal protein binding activity, and the genomic region has been associated with CAD-relevant traits including adiponectin levels, BMI, and cholesterol traits.^{34–38} MixQTL and LA methods resulted in the same lead variant, rs897392, in the intron of neighboring gene *HIP1R*, which exhibits modest to strong LD with reported arterial eQTLs for *VPS37B* ([Table S11](#)) as well as different expression between CAD cases and controls in multiple tissues in the STARNET study population ([Table S12](#)).^{20,39} rs897392 had a mixQTL adjusted p value of 0.14, meaning *VPS37B* would not be considered an eGene using this method alone, despite evidence favoring genetic regulation of this gene in cardiac tissues. In combination, these results indicate the benefit of including genetically diverse individuals as well as multiple approaches to improve effective sample size across the lower end of the global allele frequency range.

Colocalization of eQTLs

We next evaluated the overlap between coronary artery gene expression and genetic associations with CAD and intermediate risk factors including blood pressure, cholesterol, and arterial calcification traits. Across all phenotypes, 108 GWAS association signals colocalized with eQTLs, including 25 discovery eGenes ([Figure 3A](#); [Table S13](#)). Thirty-one eGenes exhibiting cell type specificity colocalized to one or more GWASs, including Rho GTPase-activating protein 42 (*ARHGAP42*) in pericytes and discovery eGenes *LIPG* and adhesion G protein-coupled receptor G6 (*ADGRG6*) in endothelial cells ([Figure 3A](#); [Table S8](#)). We further assessed GWAS associations overlapping our eQTL associations using summarized Mendelian randomization (SMR).⁴⁰ Given LD-dependent restrictions for both colocalization and SMR, as expected, we identified fewer associations but notable overlap between the two methods (25 overlapping signals and 18 unique to SMR, [Table S15](#)).

We observed strong evidence of colocalization and generalization at the *TCF21/TARID* locus. *TCF21*, a known regulator of the SMC phenotype transition to fibromyocytes in plaque,^{41,42} exhibited a significant eQTL for the same variant, rs12190287, in our study as well as in STARNET aorta (AOR) and GTEx coronary artery (COR) samples ([Figures 3B](#) and [S3](#)). The *TCF21* eQTL overlapped with the association for the adjacent lncRNA *TARID* (*TCF21* antisense RNA inducing promoter demethylation). Our lead *TARID* eQTL, rs1535616, was also significant in GTEx COR, though not for other arterial tissues ([Figure S3](#)). Interestingly, GTEx COR exhibited similarly strong associations at both ends of the *TARID* coding region, while AOR showed a much stronger association at the 3' end ([Figure S3](#)). *TCF21* and *TARID* also colocalized with CAD and BP (blood pressure) trait associations but not coronary artery calcification or cholesterol traits ([Figure 3B](#)), suggesting the mechanism for this known CAD locus may be functioning via a causal blood pressure pathway.

TBX20, coding for transcription factor T-box 20, is another gene with established cardiac development and disease associations.⁴³ However, the mechanism of its genetic regulation in CAD risk remains unclear. Our lead eQTL, rs11976145 ([Figure 3C](#)), is located in the fourth intron and colocalizes with a CAD GWAS signal but is independent from the DBP association at the same locus, suggesting a multifaceted approach may be required to ascertain disease-relevant mechanisms underlying transcriptional regulation. While *TBX20* has no reported eQTLs for GTEx arterial tissues, we identify an overlapping association with the eQTL for STARNET (AOR), for which the lead eQTL rs10249005 is in high LD ($r^2 \geq 0.94$) with rs11976145 in all European-ancestry 1000G reference populations ([Table S11](#)). Coronary artery single-cell RNA and assay for transposase-accessible chromatin sequencing (scRNA-seq, snATAC-seq) reference datasets show that *TARID* and *TCF21* are both expressed most highly and most accessible in fibroblasts and SMCs (including SMC-derived fibromyocytes resulting from SMC phenotypic switching during atherosclerosis, [Figures 3D](#), [3E](#), and [S3](#)).^{30,44} *TBX20* is most abundant in cardiac muscle and vascular tissues in GTEx, with coronary artery reference expression predominantly restricted to SMCs and fibroblasts ([Figures 3D](#) and [3E](#)).

Among LA eGenes, colocalization was limited to 27 associations, but four of these were not eGenes in mixQTL, and a

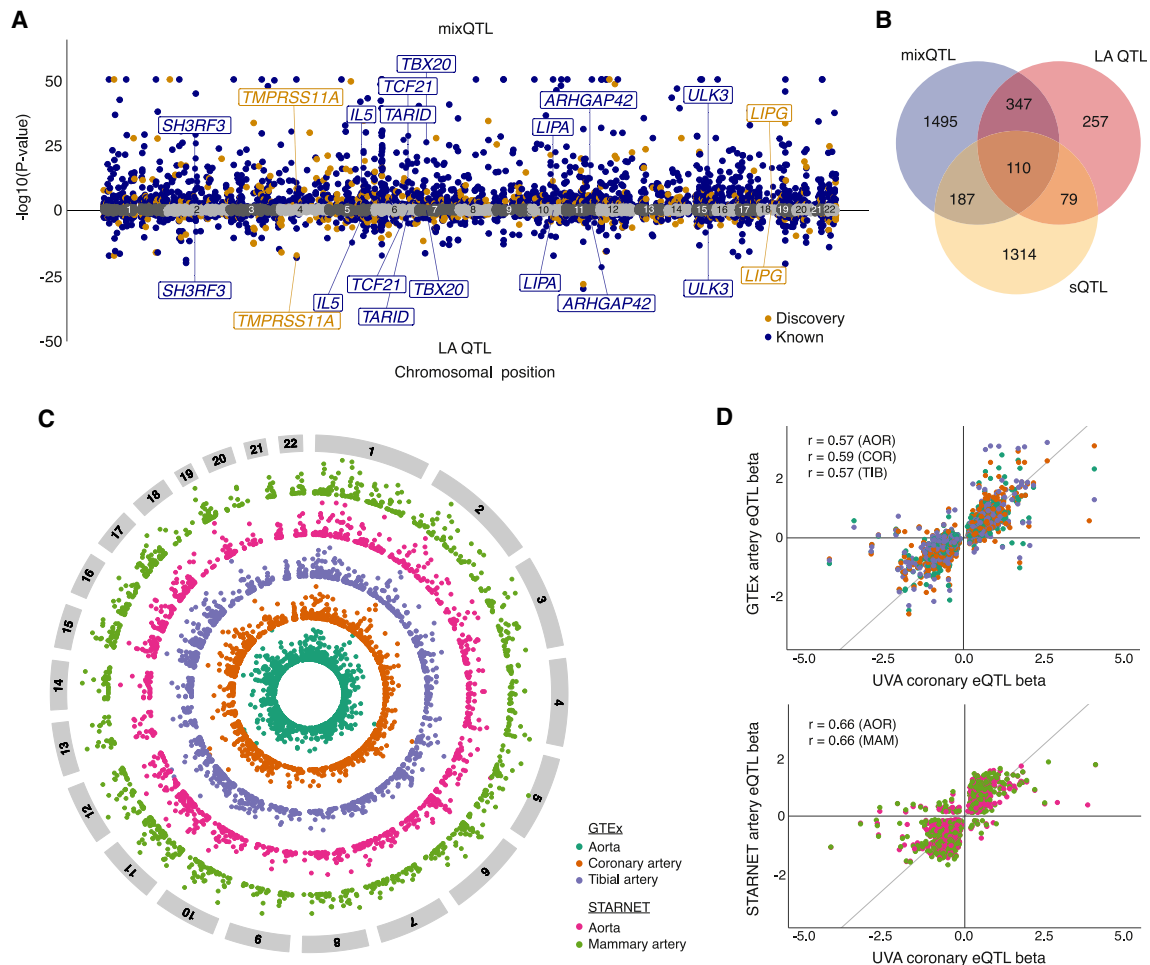


Figure 2. Overview of eQTL analysis and generalization to published arterial eQTLs

(A) Miami plot of lead eQTLs for mixQTL (top) and local ancestry (LA) adjusted (bottom). Navy blue and orange dots represent reported and discovery eGenes. Gray dots represent non-significant genes. A subset of top eGenes are labeled for clarity. (B) Venn diagram showing the overlap of mixQTL and LA-based eGenes and LeafCutter sQTL sGenes. (C) Circos plot portraying generalized (UVA pBH < 0.05) published arterial eGenes from GTEx AOR (blue green), COR (orange), and TIB (purple), and STARNET AOR (pink) and MAM (light green) tissues, with significance increasing toward the outer edge of the circle. (D) Direction of effect for genes in which the UVA lead eQTL was significant (pBH < 0.05) in any of the aforementioned tissues using the same color scheme for GTEx (top) and STARNET (bottom). Pearson's r correlation coefficients shown for overlapping significant UVA coronary eQTL detected in GTEx or STARNET eQTL with tissue indicated in parentheses. AOR: aorta; COR: coronary artery; TIB: tibial artery; MAM: mammary artery.

further three were mixQTL eGenes that did not colocalize to any GWAS trait (Table S15; Figure S4). Of particular interest is *ANAPC13*, a component of an anaphase-associated E3 ubiquitin ligase for which the LA eQTL (led by rs9809619, $p_{LA} = 2.3E-6$) colocalized to the MVP CAD association signal but did not meet the threshold for colocalization ($PPH4 \geq 0.8$) for any trait in micQTL. Rs9809619 is in close proximity to and exhibits near perfect LD globally with the lead mixQTL variant, rs4367113 ($p_{mixQTL} = 1.6E-5$, Table S11), across 1000G populations. LA adjustment resulted in lower p values for *ANAPC13*-associated variants compared to mixQTL (Figure S4), showing the benefit of complementary approaches for a locus with similar associations but different significance between methods.

Fine-mapping significant coronary artery eQTLs

Next, we used a combination of methods to identify both credible sets (CSs) and independent associations within previously unreported coronary artery eGenes. First, we used the Bayesian mixFine function from the mixQTL package to identify independent CSs for significant associations (Table S16). Only 3% of eGenes with converging association signals fine-mapped to a single variant ($n = 44$ of 1,388). As expected, 91% of CSs contained the lead eQTL, which was the sole credible variant for 29 eGenes. Eighty-three eGenes exhibited multiple independent eQTL signals.

Including prior functional annotations in relevant tissues can refine association signals and prioritize variants and candidate *cis*-regulatory mechanisms.^{45,46} Therefore, we employed

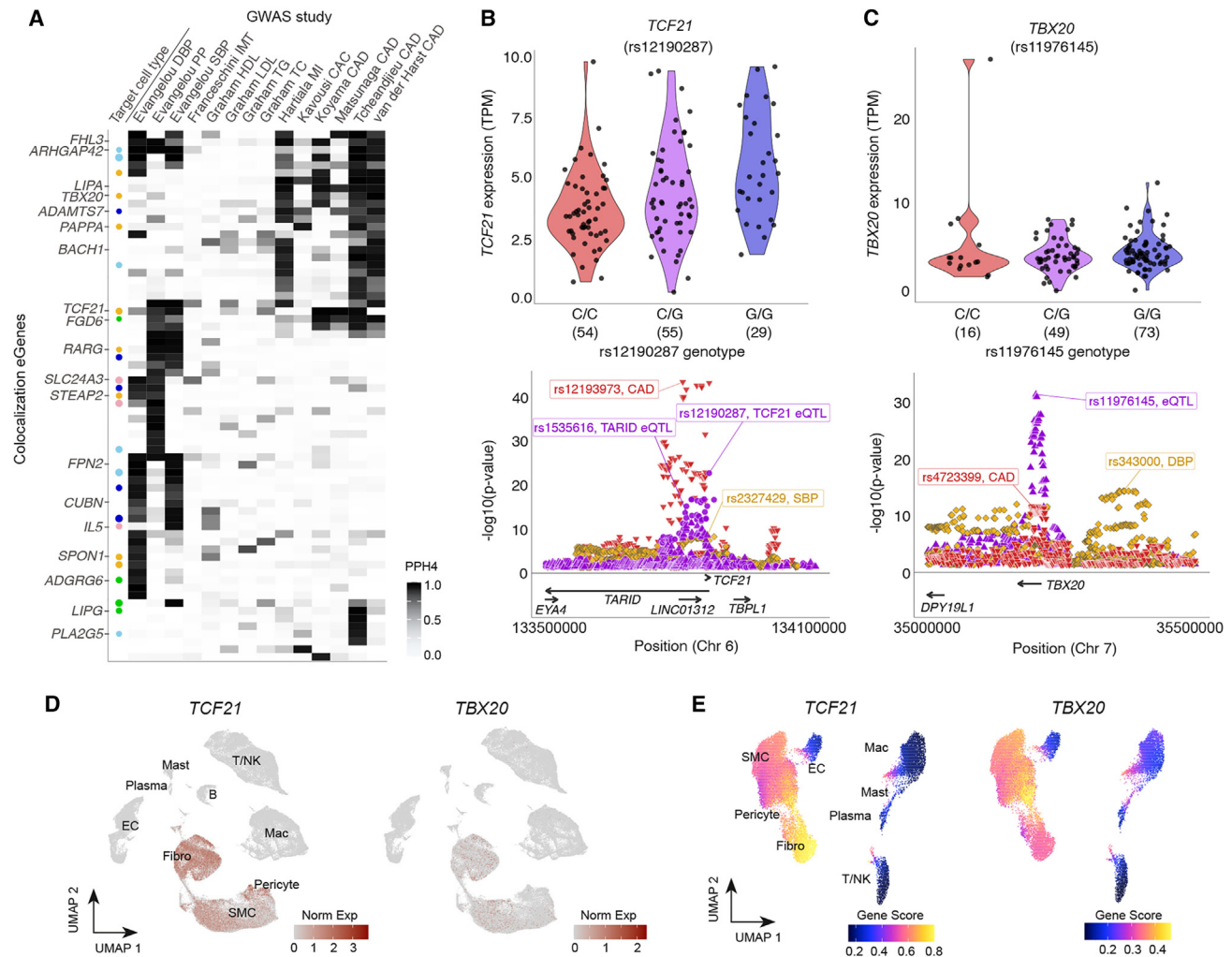


Figure 3. Colocalization reveals trait- and cell-type-specific associations

(A) GWAS colocalization to eGene associations: each column represents the $-\log_{10}(p\text{ value})$ of the study (first author last name) and relevant GWAS trait, with intensity of shading corresponding to a higher posterior probability of a shared association at that locus. Each row represents one protein-coding gene with a PPH4 ≥ 0.8 in at least one GWAS. The leftmost column (“Target cell type”) indicates cell-type specificity for SMC (dark blue), pericyte (light blue), endothelial (green), fibroblast (gold), or blood (pink) cells. Size of the circle represents the CELLEX combined gene score (range: 0.7–1.0). (B and C) Top: normalized expression (y axis) by lead eQTL genotype (x axis) for *TCF21* (B) and *TBX20* (C); bottom: regional association plots showing overlap between our study (violet) and GWAS associations for CAD (red triangle) and blood pressure traits (gold diamonds). (D and E) Human artery atherosclerosis single-cell RNA sequencing (left) and single-nuclear ATAC sequencing (right) showing *TCF21* (D) and *TBX20* (E) gene expression- and chromatin accessibility-based gene scores, respectively, with changes indicated by intensity of red (left) and pink/yellow (right).

FastPainter to fine-map associations with epigenomic annotations (ENCODE coronary artery H3K4me3 and H3K27Ac marks and activity-by-contact scores for human coronary artery SMCs as well as BP and CAD GWAS), which both exhibited strong evidence of colocalization. 1,964 eGenes had a sufficient number of eQTLs to converge for fine-mapping annotation for one or both GWAS traits (Tables S17A and S17B). Across both traits and three epigenetic marks, CSs for 106 eGenes were narrowed to a single likely causal variant. The lead eQTL was included in 11% of CSs. Regardless of epigenetic annotation, most CSs (>80%) were narrowed to five or fewer SNPs (Figure S5). While the majority of CSs contained the same variants

regardless of epigenetic annotation, they often differed by GWAS annotation. Lead eQTLs in the 247 loci for which at least one variant was shared between BP trait and CAD annotations exhibited lower p values and nearer to the eGene transcription start site (TSS) on average (median p values $1\text{E}-08$ vs. $5\text{E}-4$, Figure S5).

We demonstrated the utility of combining multi-omic data with fine-mapped associations in *ARHGAP42*, a Rho-A GTPase-activating protein with functional evidence for disease relevance, and *IL5*, a discovery coronary eGene. Lead *ARHGAP42* eQTL rs2455569 exhibited similar effect direction and magnitude in STARNET aorta tissue (Figure S6), in which decreased *ARHGAP42*

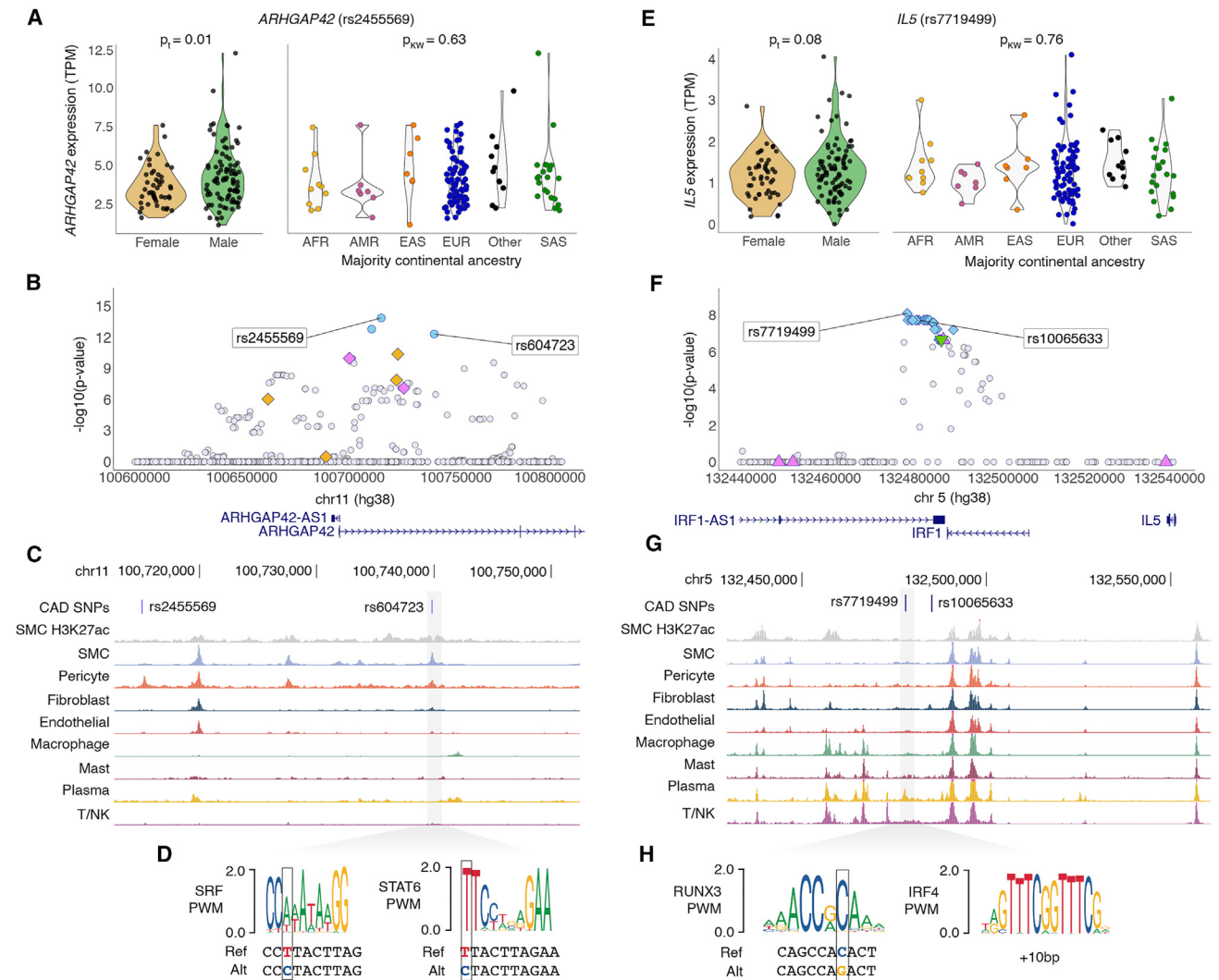


Figure 4. Fine-mapping identifies candidate causal variants for *ARHGAP42* and *IL5* eQTLs at known GWAS loci

(A) *ARHGAP42* association driven by lead eQTL rs2455569, exhibiting differences in expression by genotype but not sex or majority continental ancestry. p values calculated from paired t test (t) or Kruskal-Wallis test (KW).
 (B) Regional association plot depicting variants in PAINTOR credible sets specific to BP (violet) or CAD (golden) GWAS annotations, mixFine only (light blue), and variants not in any credible set (light gray).
 (C) Variants of interest (rs604723 and rs2455569) are indicated by blue lines, and rs604723 is highlighted in gray box and corresponds to UCSC Genome Browser tracks indicating cell-type-specific chromatin accessibility.
 (D) Location of rs604723 in critical nucleotides (in gray outlined boxes) of consensus transcription factor binding sequences for SRF and STAT6 as identified using the JASPAR 2022 database.
 (E) *IL5* association driven by lead eQTL rs7719499, exhibiting differences in expression by genotype but not sex or majority continental ancestry. p values calculated from paired t test (t) or Kruskal-Wallis test (KW).
 (F) Regional association plot depicting variants in PAINTOR credible sets specific to BP (violet) or CAD (golden) GWAS annotations, mixFine only (light blue), mixFine and PAINTOR CS (green), and variants not in any credible set (light gray).
 (G) Variants of interest (rs7719499 and rs10065633) are indicated by blue lines, and rs7719499 is highlighted in gray box and corresponds to UCSC Genome Browser tracks indicating cell-type-specific chromatin accessibility.
 (H) Location of rs7719499 in critical nucleotides (in gray outlined boxes) of consensus transcription factor binding sequences for RUNX3 and nearby IRF4 (+10 bp), as identified using the JASPAR 2022 database.

expression was significantly associated with case status (Table S12). The *ARHGAP42* eQTL overlapped with GTEx AOR and COR associations, and it colocalized with both BP traits and CAD GWAS associations (Table S13; Figure S6). *ARHGAP42*

(also known as GRAF3) regulates vascular tone via expression predominantly in mural cells and fibroblasts (Figure S6),^{30,44,47} and insufficiency causes hypertension.^{48,49} Rs2455569 is located in an intronic pericyte-specific chromatin-accessibility peak, with

fine-mapping highlighting proximal SNPs encompassing a region accessible in multiple vascular cell types (Figures 4A–4C).⁴⁴ Interestingly, rs604723, 25 kb upstream of and in high LD with rs2455569 ($p_{\text{adj}} = 4.6E-13$, Table S11), has been shown to modulate *ARHGAP42* expression in SMCs via SRF (serum response factor) binding.⁵⁰ Rs604723 is also predicted to disrupt a binding motif for STAT6, which has been implicated in the proliferation of vascular SMCs in an injury-response murine model,⁵¹ suggesting multiple functional avenues for genetic regulation of *ARHGAP42* and its downstream effects (Figure 4D).

IL5 encodes for interleukin 5, an inflammatory cytokine with no eQTLs reported in any GTEx tissue but highly significant eQTLs in both STARNET arterial tissues (Figure S7). Lead eQTL rs7719499 lies 75 kb downstream of *IL5* adjacent to a plasma-specific chromatin accessible region for human coronary artery, proximal to *IRF1* (lead eQTL rs72797327, $p_{\text{BH}} = 1$, Figures 4E–4G). Rs7719499 is also predicted to alter a RUNX3 transcription factor binding motif and is proximal to an IRF4 motif, supporting putative *cis*-regulatory mechanisms in plasma cell types (Figure 4H). While *IL5* is expressed at low levels across most tissue types and coronary cell types (Figure S7), it is significantly upregulated in aortic tissues from patients with CAD compared to controls (Figure S7; Table S12).³⁹

Sensitivity analysis: European ancestry only compared to random downsampling

We also assessed whether our inclusive study design affected discovery, colocalization, and fine-mapping by restricting our sample to European-ancestry individuals ($n = 80$), as well as a random subset of 80 members approximating the representation of genetic ancestry of the total sample. In the European-ancestry-only subset, we identified 1,311 eGenes (16% discovery eGenes; 79% eGenes in the combined analysis), compared to 1,469 eGenes in the genetically diverse subset (27% discovery, 71% present in combined analysis, Figure S8; Tables S18A and S18B).

With regard to generalization of published arterial eQTLs, 983 and 1,074 eGenes in the European-only and representative subsets respectively also had eQTLs in a GTEx or STARNET arterial tissue, and directional consistency with GTEx coronary was over 90% (Figure S8; Tables S19A and S19B). Compared to the overall sample, less than half the number of eGenes from either subset colocalized to relevant GWAS traits—fewer than would be expected if colocalization were linearly correlated with sample size (Figure S4; Tables S20A and S20B). The reduction in associations across all analyses in the European-only subset compared to the genetically diverse subset reinforces the benefits of methodological approaches designed to maximize study sample size and diverse genetic ancestry representation.

Coronary artery splicing QTL discovery

Differential isoform expression affects a wide array of complex diseases, and genetic variants affecting splicing events have been shown to be a major and distinct source of regulation underlying disease phenotypes.⁵² However, tissue-specific transcript specificity and isoform switching are not detected through eQTL methods unless total expression is affected. To identify genetic contributions to isoform-specific expression, we therefore evaluated genetic associations with 132,373 splice junctions in

14,815 genes. We identified 3,590 sQTLs ($p_{\text{BH}} < 0.05$) in 1,690 sGenes (Figure 5A; Table S21). Only 296 sGenes (17.5%) were also identified as eGenes using mixQTL or LA approaches (Figure 2B), pointing to the importance of evaluating isoform-related events in addition to total gene expression.

Lead sQTLs for 96 splice junctions (71 total sGenes) were annotated as splice region variants or splice acceptors in SnpEff (Table S22), and nine discovery sGenes (Table S21) had protein-coding variants as lead QTLs. While effect size was not correlated with broadly defined variant category ($p_{\text{ANOVA}} = 0.62$), there was a significant association with p value (Figure S9, $p_{\text{ANOVA}} = 3.65E-11$), suggesting small effects in isoform-specific differences may be functional. While fine-mapping sQTL associations is difficult without isoform-specific expression information, for sGenes with lead sQTLs within 10 kb of the overall TSS, approximately half overlapped enhancer histone modification marks in relevant tissues or cells (Figure 5B).

Generalization and colocalization of lead sQTLs

Tissue-specific differences in isoform proportion cannot necessarily be detected with bulk sequencing, implicating potentially distinct regulatory mechanisms for splicing compared to overall expression. Therefore, we first directly compared the gene sets represented by eQTLs and sQTLs. Only modest overlap was observed (297 eGenes were also sGenes, 23 of which shared the same lead variant, Table S23), suggesting that splicing analyses likely represent unique disease-relevant pathways compared to overall expression. Among 71 sGenes with lead sQTLs reported to have splicing functions in SnpEff, 59 have no eQTLs, with five having no expression variants exceeding even nominal significance (Table S23). This difference is further exemplified by differences in cell type specificity in our scRNA-seq reference dataset: while nearly all cell-type-specific (score > 0.7 as described in STAR Methods) reference genes were eGenes (408 of 411 total), only 37 sGenes (of 83 total) exhibited cell-type specificity. Compared to published arterial sQTL data, 871 (52%) were also sGenes in GTEx coronary artery, while 875 sGenes had no reported sQTLs in GTEx AOR, COR, or tibial artery (TIB) tissues (Figure 5C; Table S21).

Colocalization of sQTL data exhibited a similar pattern to eQTLs, with a small subset of overlapping colocalized signals shared between CAD and BP trait GWAS and a higher number of quantitative trait associations (Figure 5D; Table S24). Despite a similar pattern of colocalization with CAD and BP traits being most well represented, none of the eQTLs for shared associations colocalized with tested GWAS traits, affirming the likely different causal mechanisms underlying genetic effects on gene expression compared to splicing activity.

Splicing contributes to CAD-relevant gene regulation

We highlight two plausible candidates for functional splicing effects in the coronary artery: *ULK3* and *TOR1AIP1* (Figures S9). *ULK3* is a broadly expressed serine/threonine kinase exhibiting autophosphorylation activity.^{53,54} The *ULK3* locus has also been associated with various CAD-relevant traits, including blood pressure, total cholesterol, and estimated glomerular filtration rate.^{18,55–57} Of thirteen junctions tested in our study, common missense variant rs2290572 was the lead sQTL for

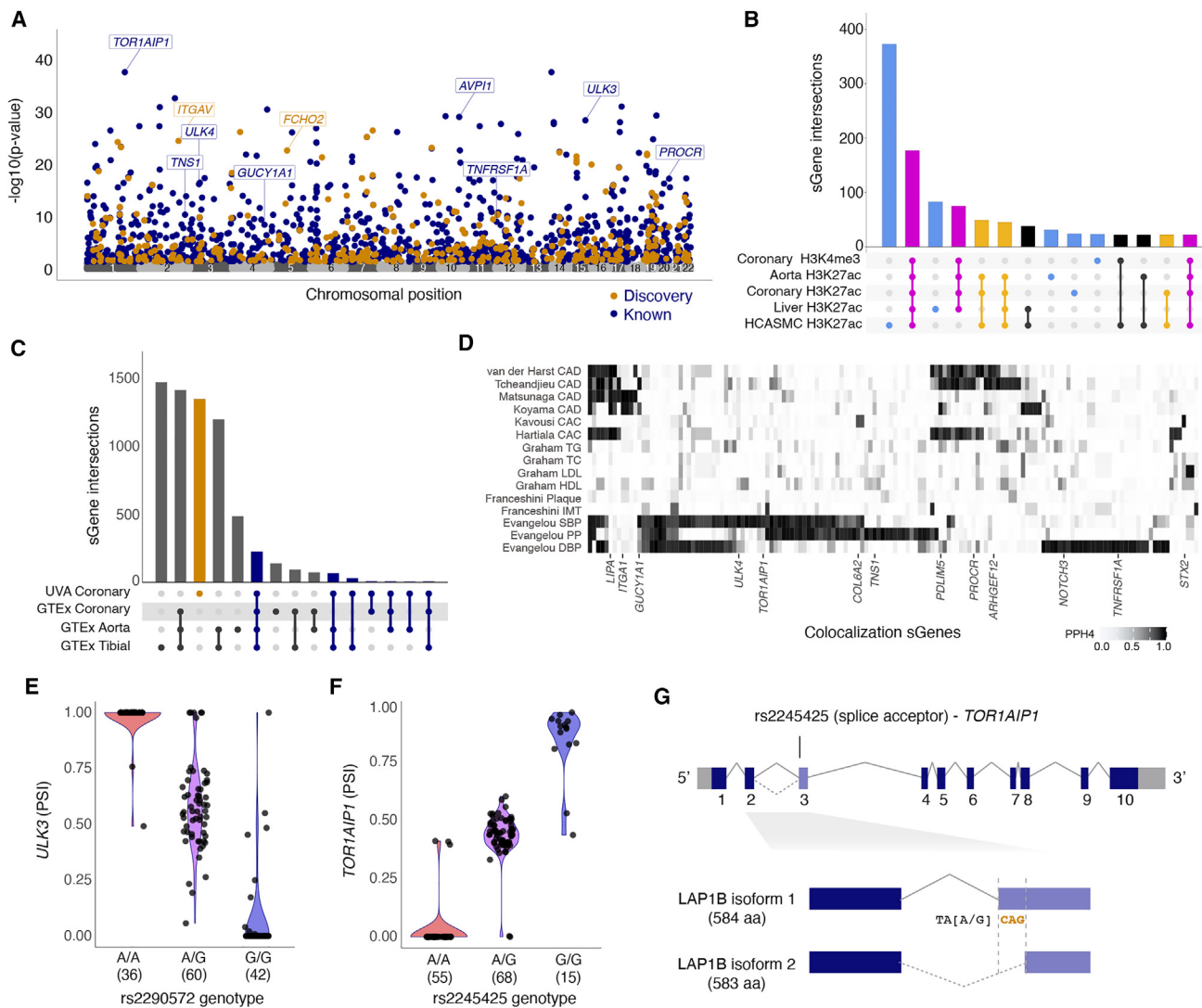


Figure 5. Coronary artery sQTL overview and characterization

(A) Manhattan plot of lead sQTLs. Navy blue and orange dots represent reported and discovery sGenes ($p_{BH} < 0.05$), respectively; gray dots represent non-significant genes.

(B) Upset plot of generalization of GTEX arterial sQTLs. Black bars represent GTEX-specific sGenes; orange bars represent generalized sGenes, and the blue bar represents previously unreported sGenes.

(C) Upset plot of histone modifications in aorta, coronary, and liver from one individual in ENCODE as well as HCASMCs. Light blue bars represent lead sQTLs overlapping modifications specific to one cell or tissue type, gold bars for H3K27 acetylation in coronary and any other tissue or cell type, and magenta bars for both coronary artery H3K4 tri-methylation and H3K27 acetylation.

(D) GWAS colocalization to sQTL associations: each row represents the colocalization PPH4 of the study (first author last name) and relevant GWAS trait, with intensity of shading corresponding to a higher posterior probability of a shared association at that locus. Each column represents one protein-coding gene with a $PPH4 \geq 0.8$ in at least one GWAS.

(E and F) Percent spliced in of *ULK3* and *TOR1AIP1* exons by genotype of their lead sQTLs (rs2290572 and rs2245425, respectively) in our study (left) and GTEX (right).

(G) Schematic of *TOR1AIP1* gene and alternatively spliced isoforms, showing location and effect estimates for top splice acceptor variant (rs2245425) identified in our samples. rs2245425-G creates TAGCAG splice acceptor sequence at 3' end of intron-exon 3 junction. Spliced CAG nucleotides (orange) encoding alanine amino acid distinguish LAP1B isoforms 1 and 2.

both significant associations: chr15:74837435:74837757 and chr15:74837435:74837751 (Figure 5E). Rs2290572 was also significantly associated with chr15:74837435:74837751 in GTEX tissues with directional consistency (Figure S9). The lead sQTL for GTEX arterial tissues, rs12898397, is in high LD with

rs2290572 in reference populations (Table S11) and represents a two-codon difference in an MIT domain in the fourteenth exon for which the T allele is predicted by SpliceAI to cause loss of a splice donor (Δ score = 0.67, <https://spliceailookup.broadinstitute.org>).⁵⁸ Rs12898397 is our lead

eQTL ($p_{BH} = 1.3E-14$), and its contribution to isoform specificity makes it a clear candidate for functional follow-up.

TOR1AIP1 (torsin 1A-interacting protein 1) is a broadly expressed lamin-binding protein that localizes to the inner nuclear membrane. Causal *TOR1AIP1* variants have been found for several autosomal recessive disorders, including limb-girdle muscular dystrophy with cardiac failure.^{59,60} Located in the first intron, splice acceptor variant rs2245425 is the lead sQTL for both significant splice excision events (of six tested in our study; chr1:179884769:179889313 and chr1:179884769:179889310), exhibiting nearly binary effects on the PSI of the third exon in coronary artery (Figures 5F and 5G). Rs2245425 is also the lead sQTL of GTEx arterial tissues (Figure S9); *TOR1AIP1* has no eQTLs in our study or any GTEx artery tissue, pointing to splicing rather than transcription levels as the main effect of genetic regulation for this gene.

DISCUSSION

We report a coronary artery eQTL mapping study accounting for local ancestry and allele-specific expression. Our study sample's representation of both ancestral diversity and phenotypic heterogeneity allowed us to capture eQTLs that likely affect vessel wall integrity and maintenance throughout the life course. The disruption of these transcriptional regulatory networks may be critical to plaque progression in coronary artery atherosclerosis. Our approach maximizes the likelihood of identifying eGenes in a rarely available tissue type, providing a roadmap for future QTL studies of diverse populations.

While tissues from patients with advanced CAD are useful for therapeutic development, profiling gene expression changes in subclinical CAD patients with multiple risk factors may point to avenues to prevent lesion progression. For example, molecular pathways for the strong epidemiologic association between blood pressure and CAD⁶¹ are incompletely described, while our work importantly begins to elucidate these mechanisms by highlighting a subset of eGenes colocalizing to GWAS findings of both traits. Furthermore, eGenes colocalizing to one or more GWAS traits can be prioritized for functional characterization, particularly those exhibiting cell-type specificity, e.g., heparanase 2 (*HPSE2*), which is enriched in SMCs in coronary artery and for which our eQTL colocalizes with SBP and DBP.³⁰ *HPSE2* would not be a strong candidate gene in a QTL study focused on disease outcomes rather than intermediate phenotypes, but recent characterization of this extracellular matrix protein in endothelial maintenance suggests a contribution to vascular function during inflammation.^{62,63}

Statistical fine-mapping analyses may prioritize causal genes and mechanisms of CAD loci, and we highlight example eGenes *TBX20* and *TCF21*. Here we identify the first genetic evidence of *TBX20* regulation in the human coronary artery, filling in a key knowledge gap in characterizing this gene in atherosclerosis. *TBX20* is required for normal cardiovascular development and was recently identified as regulating *PROK2*, a critical component of angiogenesis.⁶⁴ *TBX20* has been implicated in causal pathways for congenital heart defects and continuous traits involving the great vessels.⁶⁵⁻⁶⁷ However, the role of *TBX20* in atherosclerosis has been minimally investigated, despite expression in arterial tis-

ues and a well-replicated CAD GWAS signal.^{5,10,11} Conflicting evidence about cell-type-specific expression of *TBX20* related to vascular function and neointimal hyperplasia⁶⁸ demands functional characterization of regulatory elements in this locus that may modulate vessel wall pathways both during development and in a diseased state. Our *TCF21/TARID* findings also build on existing work by our lab and others characterizing an established CAD locus.^{13,69-71} This relationship is supported by research demonstrating regulation of *TCF21* expression by long non-coding RNA *TARID* in the context of CAD via promoter demethylation.^{72,73} The mural cell-enriched expression and chromatin accessibility of both genes, as well as overlapping association signals, compel a deeper look into the potential co-regulation of *TCF21* and *TARID* during SMC phenotypic transition.

It is now well appreciated that a large fraction of candidate regulatory variants are predicted to function via transcription factor binding-independent mechanisms, requiring comprehensive fine-mapping of candidate loci to prioritize likely causal variants.²⁰ Our results expand the current focus on disease outcomes such as CAD by incorporating intermediate phenotype summary statistics into fine-mapping analyses. Using this approach, we narrowed colocalized GWAS signals to CSs of candidate variants, highlighting *ARHGAP42* and *IL5*. The *ARHGAP42* locus is associated with blood pressure and cIMT: carotid intima-media thickness,⁷⁴ and gene expression changes affect smooth muscle cell contractility.⁷⁵ Multi-omic fine-mapping revealed two candidate causal SNPs, of which the rs604723 risk allele has been shown to generate a cryptic SRF binding site to increase expression.⁵⁰ We also highlight a putative STAT6 binding site created by the risk allele, suggesting a potential IL-4/IL-13-mediated activation of SMCs, both of which are normally lowly expressed in the coronary artery.

Another candidate gene identified through our eQTL discovery and fine-mapping analyses, *IL5*, resides in a gene-dense SBP GWAS locus.⁷⁶ While *IL5* is reported to function in Th2 cells or eosinophils associated with atherosclerosis progression,⁷⁷⁻⁷⁹ its precise role and regulation in the coronary artery remains unknown. Notably, we did not identify eQTLs for neighboring inflammatory genes *IL4*, *IL13*, and *CSF2*,⁸⁰ despite proposed overlapping functions and regulatory mechanisms.^{81,82} Aorta-specific upregulation of *IL5* in STARNET CAD cases supports a potential disease-specific effect for *IL5* in multiple arterial tissues.³⁹ Our lead eQTL, rs7719499, overlaps a plasma cell-specific chromatin-accessibility peak in the coronary artery. Given the known influence of inflammatory cytokines on endothelial and SMC activation, future studies are warranted to investigate the immune cell-derived IL5-mediated vascular wall injury in cell and animal models.

Finally, we demonstrated distinct genetic contributions for expression and splicing activity in the coronary artery. The relationship between isoform specificity and overall expression is complex.^{83,84} We observed a modest sharing of eGenes and sGenes (14%) and <10% overlap for GWAS colocalization, supporting orthogonal effects of genetic variation on total transcription compared to isoform regulation. Development in the areas of single-cell long-read sequencing will provide a more complete understanding of isoform-specific regulatory mechanisms in coronary artery.

We explored two sGenes with potential roles in coronary artery disease—*ULK3* and *TOR1AIP1*. *TOR1AIP1* mutations have been reported for monogenic dystrophic developmental disorders via dysregulation of necessary protein complex formation at the inner nuclear membrane.^{59,85} The *TOR1AIP1* sQTL regulates the addition of an alanine residue to the third exon (position 185). While this variant has not been associated with disease, its high allele frequency suggests differential isoform expression may have an important functional role, particularly given the lack of *TOR1AIP1* eQTLs. Regarding *ULK3*, though mainly described in the context of cancers,⁸⁶ this gene may mediate vascular disease through autophagy dysregulation and interactions in the Shh signaling pathway.^{87,88} Although the *ULK3* locus has also been reported in BP and cholesterol GWAS, the published associations are located downstream of the *ULK3* coding region and are statistically independent of our sQTL and eQTL.^{9,55,76} Publicly available data do not implicate a particular cell type of interest for functional studies, pointing to the future benefits of long-read sequencing to characterize isoform-specific and cell-type-specific mechanisms relevant to disease pathways.^{89,90} Our findings for both transcriptional and isoform-specific regulation by variants with no GWAS signal point to the importance of considering multiple ‘omics datasets when evaluating candidate genes for complex diseases and traits.

Overall, our findings emphasize the importance of context in interpreting genetic associations with disease. Characterizing loci within a comprehensive genomic and physiologic setting helps prioritize top candidate genes uniquely relevant to atherosclerosis disease processes. While coronary artery tissue remains the most relevant single tissue type for prioritizing CAD candidate genes, the gene expression program changes during atherosclerosis likely involve multi-tissue and multi-cellular gene-regulatory networks. Multi-tissue network analyses may further resolve underlying paracrine signaling pathways and regulatory mechanisms for eGenes without an obvious role in predominant intimal or medial cell types.⁹¹ For instance, inflammatory processes driven by cytokine signaling may be difficult to detect in target cell types, but significantly different expression of *IL5* specifically in the aorta in STARNET cases compared to controls provides both validation of our approach and options for identifying other genes functioning in the same molecular pathway.³⁹

A major strength of our work is demonstrating the feasibility and promise of incorporating local genetic ancestry and allele-specific expression into eQTL analyses to discover disease-relevant genes/pathways. Our inclusive study design increased statistical power in both our diverse downsampled subset and overall study population compared to a genetically homogeneous European-ancestry-only subset. This is significant given the predominantly European genetic architecture of GTEx and published GWASs: while these resources have been crucial for genomics discovery to date, work highlighting the limitations of genetically restricted samples⁹² and technologies developed based on those samples⁹³ points to the necessity of new, more expansive approaches.^{94–97} This also aligns with current appeals in basic science and public health to promote equitable research benefiting all populations, rather than studies that may

extend the health disparity gap.^{17,98,99} Furthermore, incorporation of single-cell chromatin accessibility datasets in coronary artery complemented our eQTL-based gene prioritization approach to nominate *cis*-regulatory mechanisms underlying complex diseases associations.⁴⁴ Finally, our findings benefited from comparing epigenomic and genetic annotations, providing a more tenable suite of candidate variants for future functional work. Our multiple-phenotype fine-mapping approach will be particularly relevant as the field moves toward functional characterization of disease-associated lncRNAs and splice isoforms, for which traditional metrics such as evolutionary conservation cannot be consistently applied.

In summary, we present a genetically diverse evaluation of coronary artery gene expression across the phenotypic spectrum of atherosclerosis. Our inclusive study design with respect to ancestry and robust pipeline facilitated the discovery of atherosclerosis-associated genes with plausible functional roles in the vascular wall. Molecular characterization of these genes in environments representing subclinical atherosclerosis will improve the identification of therapeutic targets for CAD patients.

Limitations of the study

It is worth noting limitations both common to eQTL studies and unique to our approach. First, restricting to 5% minor-allele frequency variants limited detection of ancestry-specific associations; previous work has shown that effectively capturing lower frequency variants increases discovery both across and within ancestral populations.¹⁰⁰ A second limitation relates to interpretability of our colocalization and fine-mapping results based on genetically homogeneous public datasets.¹⁷ Individuals with genetic ancestry from Africa and the Americas in particular continue to be severely underrepresented in both GWASs and ‘omics reference datasets. Homogeneous study populations not only prevent identification of ancestry- or haplotype-specific associations, but they also limit generalizability of global associations when fine-mapping is restricted to genetic architecture from a single ancestry group.^{92,101} With the generation of new eQTL datasets from admixed populations, establishing best practices such as minimizing LD mismatch and using local ancestry estimates is needed to improve data standards, integration, and replication efforts. Additionally, we acknowledge that both sample quality and disease status may affect interindividual cell-type proportions and therefore eQTL detection.¹⁰² Adjusting for estimated cell type proportions using single-cell reference-based deconvolution may improve discovery across tissues¹⁰³ and complement cell-type-specific QTL studies.¹⁰⁴

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
 - Lead contact
 - Materials availability
 - Data and code availability

- **EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS**
 - Ethics statement
 - Sample acquisition
- **METHOD DETAILS**
 - DNA genotyping
 - Bulk RNA sequencing and processing
 - RNA-seq read mapping and quality control
 - Detection of RNA-seq sample swaps
 - RNA sequencing data phasing for allele-specific expression
 - Multidimensional scaling
 - Quantitative trait loci mapping
 - Characterization and fine-mapping of QTLs

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xgen.2023.100465>.

ACKNOWLEDGMENTS

This work was supported by grants from the following: the National Institutes of Health (grant numbers R01HL148239 and R01HL164577 to C.L.M.; T32HL007284 to C.J.H.; and R01HL125863 to J.L.M.B.; R01HL130423, R01HL135093, and R01HL148167 to J.C.K.), the American Heart Association (grant number 20POST35120545 to A.W.T.; AHA909150 to J.V.M.; and A14SFRN20840000 to J.L.M.B.), the Swedish Research Council and Heart Lung Foundation (grant numbers 2018-02529 and 20170265 to J.L.M.B.), the Fondation Leducq (grant number “PlaQOmics” 18CVD02 to C.L.M. and J.L.M.B.), and the Single-Cell Data Insights award from the Chan Zuckerberg Initiative, LLC, and Silicon Valley Community Foundation (to C.L.M.). This work was also supported by fellowship grants from the Bench to Bassinet Pediatric Cardiac Genomics Consortium (PCGC) and Cardiovascular Development Data Resource Center (CDDRC) (to C.J.H.), as well as the UVA MSTP training grant (NIH T32GM007267, to W.F.M.). S.W.v.d.L. is supported by EU H2020 TO_AITON (grant number: 848146). We are thankful for the support of the Netherlands CardioVascular Research Initiative of the Netherlands Heart Foundation (CVON 2011/B019 and CVON 2017-20: Generating the best evidence-based pharmaceutical targets for atherosclerosis [GENIUS I&II]) and the ERA-CVD program “druggable-MI-targets” (grant number: 01KL1802). The research for this contribution was made possible in part by the AI for Health working group of the EWUU alliance and the CZI Foundation. The authors would like to thank Yipei Song and Wesley Craig for assistance with scripting and troubleshooting; Catherine Robertson for assistance with fine-mapping study design; Katia Sol-Church and Yongde Bao for assistance with library preparation and sequencing; Peter Chiu, Paul Chang, A.J. Pedroza, Tiffany Koyano, Euan Ashley, Tom Quertermous, and all of the transplant recipients, heart donors, family members, study coordinators, and transplant procurement team at Stanford for coronary artery tissue procurement.

AUTHOR CONTRIBUTIONS

We used CRediT taxonomy to determine the authors' contributions. C.J.H., A.W.T., J.L.M.B., and C.L.M. conceptualized the study. C.J.H., A.W.T., M.D.K., and N.G.L. curated the data. C.J.H., A.W.T., M.D.K., N.B.B., R.M., L.M., and J.V.M., performed formal data analysis. C.J.H., A.W.T., N.J.L., J.C.K., J.L.M.B., and C.L.M. acquired funding. C.J.H., A.W.T., and E.F. performed the experiments. M.D.K., N.B.B., R.M., E.F., D.W., and S.O.-G. contributed to methods development. G.A. and S.O.-G. performed project administration. A.W.T., L.M., N.G.L., J.V.M., W.F.M., M.K., P.A.P., S.W.v.d.L., N.J.L., J.C.K., J.L.M.B., and C.L.M. contributed biospecimens, datasets, scripts, or other resources. C.J.H. and W.F.M. contributed software tools. C.L.M. supervised the project. L.M., J.V.M., and J.L.M.B. contributed

to validation of the results. C.J.H., A.W.T., M.D.K., N.B.B., J.V.M., and C.L.M. contributed to data visualization. C.J.H., A.W.T., and C.L.M. prepared the manuscript draft. All authors reviewed and edited the manuscript.

DECLARATION OF INTERESTS

J.L.M.B. is a shareholder in Clinical Gene Network AB and has an invested interest in STARNET. J.C.K. is the recipient of an Agilent Thought Leader Award, which includes funding for research that is unrelated to the current manuscript. S.W.v.d.L. has received Roche funding for unrelated work. C.L.M. has received AstraZeneca funding for unrelated work.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: April 17, 2023

Revised: September 7, 2023

Accepted: November 19, 2023

Published: December 15, 2023

REFERENCES

1. Khera, A.V., and Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.* **18**, 331–344.
2. Watkins, H., and Farrall, M. (2006). Genetic susceptibility to coronary artery disease: from promise to progress. *Nat. Rev. Genet.* **7**, 163–173.
3. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F.R., Barbalic, M., Gieger, C., et al. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**, 333–338.
4. Webb, T.R., Erdmann, J., Stirrups, K.E., Stitzel, N.O., Masca, N.G.D., Jansen, H., Kanoni, S., Nelson, C.P., Ferrario, P.G., König, I.R., et al. (2017). Systematic evaluation of pleiotropy identifies 6 further loci associated with coronary artery disease. *J. Am. Coll. Cardiol.* **69**, 823–836.
5. van der Harst, P., and Verweij, N. (2018). Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ. Res.* **122**, 433–443.
6. Nikpay, M., Goel, A., Won, H.-H., Hall, L.M., Willenborg, C., Kanoni, S., Saleheen, D., Kyriakou, T., Nelson, C.P., Hopewell, J.C., et al. (2015). A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130.
7. Nelson, C.P., Goel, A., Butterworth, A.S., Kanoni, S., Webb, T.R., Marouli, E., Zeng, L., Ntalla, I., Lai, F.Y., Hopewell, J.C., et al. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391.
8. Koyama, S., Ito, K., Terao, C., Akiyama, M., Horikoshi, M., Momozawa, Y., Matsunaga, H., Ieki, H., Ozaki, K., Onouchi, Y., et al. (2020). Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat. Genet.* **52**, 1169–1177.
9. Hartiala, J.A., Han, Y., Jia, Q., Hliser, J.R., Huang, P., Gukasyan, J., Schwartzman, W.S., Cai, Z., Biswas, S., Trégouët, D.A., et al. (2021). Genome-wide analysis identifies novel susceptibility loci for myocardial infarction. *Eur. Heart J.* **42**, 919–933.
10. Matsunaga, H., Ito, K., Akiyama, M., Takahashi, A., Koyama, S., Nomura, S., Ieki, H., Ozaki, K., Onouchi, Y., Sakaue, S., et al. (2020). 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.

11. Tcheandjieu, C., Zhu, X., Hilliard, A.T., Clarke, S.L., Napolioni, V., Ma, S., Lee, K.M., Fang, H., Chen, F., Lu, Y., et al. (2022). Large-scale genome-wide association study of coronary artery disease in genetically diverse populations. *Nat. Med.* **28**, 1679–1692.
12. Kichaev, G., Bhatia, G., Loh, P.-R., Gazal, S., Burch, K., Freund, M.K., Schoech, A., Pasaniuc, B., and Price, A.L. (2019). Leveraging polygenic functional enrichment to improve GWAS power. *Am. J. Hum. Genet.* **104**, 65–75.
13. Liu, B., Pjanic, M., Wang, T., Nguyen, T., Gloudemans, M., Rao, A., Castano, V.G., Numberg, S., Rader, D.J., Elwyn, S., et al. (2018). Genetic regulatory mechanisms of smooth muscle cells map to coronary artery disease risk loci. *Am. J. Hum. Genet.* **103**, 377–388.
14. Örd, T., Öunap, K., Stolze, L.K., Aherrahrou, R., Nurminen, V., Toropainen, A., Selvarajan, I., Lönnberg, T., Aavik, E., Ylä-Herttuala, S., et al. (2021). Single-Cell Epigenomics and Functional Fine-Mapping of Atherosclerosis GWAS Loci. *Circ. Res.* **129**, 240–258.
15. ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
16. van de Geijn, B., Finucane, H., Gazal, S., Hormozdiari, F., Amariuta, T., Liu, X., Gusev, A., Loh, P.-R., Reshef, Y., Kichaev, G., et al. (2020). Annotations capturing cell type-specific TF binding explain a large fraction of disease heritability. *Hum. Mol. Genet.* **29**, 1057–1067.
17. Bien, S.A., Wojcik, G.L., Hodonsky, C.J., Gignoux, C.R., Cheng, I., Matise, T.C., Peters, U., Kenny, E.E., and North, K.E. (2019). The Future of Genomic Studies Must Be Globally Representative: Perspectives from PAGE. *Annu. Rev. Genom. Hum. Genet.* **20**, 181–200.
18. Graham, S.E., Clarke, S.L., Wu, K.-H.H., Kanoni, S., Zajac, G.J.M., Ramdas, S., Surakka, I., Ntalla, I., Vedantam, S., Winkler, T.W., et al. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675–679.
19. Wong, D., Turner, A.W., and Miller, C.L. (2019). Genetic insights into smooth muscle cell contributions to coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **39**, 1006–1017.
20. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330.
21. Stolze, L.K., Conklin, A.C., Whalen, M.B., López Rodríguez, M., Öunap, K., Selvarajan, I., Toropainen, A., Örd, T., Li, J., Eshghi, A., et al. (2020). Systems Genetics in Human Endothelial Cells Identifies Non-coding Variants Modifying Enhancers, Expression, and Complex Disease Traits. *Am. J. Hum. Genet.* **106**, 748–763.
22. Slenders, L., Landsmeer, L.P.L., Cui, K., Depuydt, M.A.C., Verwer, M., Mekke, J., Timmerman, N., van den Dungen, N.A.M., Kuiper, J., de Winther, M.P.J., et al. (2022). Intersecting single-cell transcriptomics and genome-wide association studies identifies crucial cell populations and candidate genes for atherosclerosis. *Eur. Heart J. Open* **2**, oeab043.
23. Rotival, M., Zeller, T., Wild, P.S., Maouche, S., Szymczak, S., Schillert, A., Castagné, R., Deiseroth, A., Proust, C., Brocheton, J., et al. (2011). Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. *PLoS Genet.* **7**, e1002367.
24. Liang, Y., Aguet, F., Barbeira, A.N., Ardlie, K., and Im, H.K. (2021). A scalable unified framework of total and allele-specific counts for cis-QTL, fine-mapping, and prediction. *Nat. Commun.* **12**, 1424.
25. Gay, N.R., Gloudemans, M., Antonio, M.L., Abell, N.S., Balliu, B., Park, Y., Martin, A.R., Musharoff, S., Rao, A.S., Aguet, F., et al. (2020). Impact of admixture and ancestry on eQTL analysis and GWAS colocalization in GTEx. *Genome Biol.* **21**, 233.
26. Robert, J., Osto, E., and von Eckardstein, A. (2021). The endothelium is both a target and a barrier of hdl's protective functions. *Cells* **10**.
27. Yu, H., Littlewood, T., and Bennett, M. (2015). Akt isoforms in vascular disease. *Vasc. Pharmacol.* **71**, 57–64.
28. Lappalainen, T., Sammeth, M., Friedländer, M.R., 't Hoen, P.A.C., Monlong, J., Rivas, M.A., González-Porta, M., Kurbatova, N., Griebel, T., Ferreira, P.G., et al. (2013). Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **507**, 506–511.
29. GTEx Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660.
30. Mosquera, J.V., Wong, D., Auguste, G., Turner, A.W., Hodonsky, C.J., Lino Cardenas, C.L., Theofilatos, K., Bos, M., Kavousi, M., Peyser, P., et al. (2022). Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. Preprint at bioRxiv.
31. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J. (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198.
32. Carss, K.J., Baranowska, A.A., Armisen, J., Webb, T.R., Hamby, S.E., Premawardhana, D., Al-Hussaini, A., Wood, A., Wang, Q., Deevi, S.V.V., et al. (2020). Spontaneous coronary artery dissection: insights on rare genetic variation from genome sequencing. *Circ. Genom. Precis. Med.* **13**, e003030.
33. Guo, D.-C., Duan, X.-Y., Regalado, E.S., Mellor-Crummey, L., Kwartler, C.S., Kim, D., Lieberman, K., de Vries, B.B.A., Pfundt, R., Schinzel, A., et al. (2017). Loss-of-Function Mutations in YY1AP1 Lead to Grange Syndrome and a Fibromuscular Dysplasia-Like Vascular Disease. *Am. J. Hum. Genet.* **100**, 21–30.
34. Lee, S.-B., Choi, J.-E., Park, B., Cha, M.-Y., Hong, K.-W., and Jung, D.-H. (2022). Dyslipidaemia-Genotype Interactions with Nutrient Intake and Cerebro-Cardiovascular Disease. *Biomedicines* **10**, 1615.
35. Richardson, T.G., Leyden, G.M., Wang, Q., Bell, J.A., Elsworth, B., Davey Smith, G., and Holmes, M.V. (2022). Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol.* **20**, e3001547.
36. Christakoudi, S., Evangelou, E., Riboli, E., and Tsilidis, K.K. (2021). GWAS of allometric body-shape indices in UK Biobank identifies loci suggesting associations with morphogenesis, organogenesis, adrenal cell renewal and cancer. *Sci. Rep.* **11**, 10688.
37. Wu, Y., Gao, H., Li, H., Tabara, Y., Nakatochi, M., Chiu, Y.-F., Park, E.J., Wen, W., Adair, L.S., Borja, J.B., et al. (2014). A meta-analysis of genome-wide association studies for adiponectin levels in East Asians identifies a novel locus near WDR11-FGFR2. *Hum. Mol. Genet.* **23**, 1108–1119.
38. Zhu, Z., Guo, Y., Shi, H., Liu, C.-L., Panganiban, R.A., Chung, W., O'Connor, L.J., Himes, B.E., Gazal, S., Hasegawa, K., et al. (2020). Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J. Allergy Clin. Immunol.* **145**, 537–549.
39. Koplev, S., Seldin, M., Sukhvasi, K., Ermel, R., Pang, S., Zeng, L., Bankier, S., Di Narzo, A., Cheng, H., Meda, V., et al. (2022). A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat. Cardiovasc. Res.* **1**, 85–100.
40. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487.
41. Wirka, R.C., Wagh, D., Paik, D.T., Pjanic, M., Nguyen, T., Miller, C.L., Kundu, R., Nagao, M., Collier, J., Koyano, T.K., et al. (2019). Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat. Med.* **25**, 1280–1289.
42. Nagao, M., Lyu, Q., Zhao, Q., Wirka, R.C., Bagga, J., Nguyen, T., Cheng, P., Kim, J.B., Pjanic, M., Miano, J.M., and Quertermous, T. (2020). Coronary Disease-Associated Gene TCF21 Inhibits Smooth Muscle Cell

Differentiation by Blocking the Myocardin-Serum Response Factor Pathway. *Circ. Res.* 126, 517–529.

43. Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. *Science* 313, 1922–1927.
44. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E., et al. (2022). Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.* 54, 804–816.
45. Çalıřkan, M., Manduchi, E., Rao, H.S., Segert, J.A., Beltrame, M.H., Trizino, M., Park, Y., Baker, S.W., Chesi, A., Johnson, M.E., et al. (2019). Genetic and epigenetic fine mapping of complex trait associated loci in the human liver. *Am. J. Hum. Genet.* 105, 89–107.
46. Trynka, G., Sandor, C., Han, B., Xu, H., Stranger, B.E., Liu, X.S., and Raychaudhuri, S. (2013). Chromatin marks identify critical cell types for fine mapping complex trait variants. *Nat. Genet.* 45, 124–130.
47. Bai, X., Mangum, K., Kakoki, M., Smithies, O., Mack, C.P., and Taylor, J.M. (2020). GRAF3 serves as a blood volume-sensitive rheostat to control smooth muscle contractility and blood pressure. *Small GTPases* 11, 194–203.
48. Fjorder, A.S., Rasmussen, M.B., Mehrjouy, M.M., Nazaryan-Petersen, L., Hansen, C., Bak, M., Grarup, N., Nørremølle, A., Larsen, L.A., Vestergaard, H., et al. (2019). Haploinsufficiency of ARHGAP42 is associated with hypertension. *Eur. J. Hum. Genet.* 27, 1296–1303.
49. Li, Q., Dibus, M., Casey, A., Yee, C.S.K., Vargas, S.O., Luo, S., Rosen, S.M., Madden, J.A., Genetti, C.A., Brabek, J., et al. (2021). A homozygous stop-gain variant in ARHGAP42 is associated with childhood interstitial lung disease, systemic hypertension, and immunological findings. *PLoS Genet.* 17, e1009639.
50. Bai, X., Mangum, K.D., Dee, R.A., Stouffer, G.A., Lee, C.R., Oni-Orisan, A., Patterson, C., Schisler, J.C., Viera, A.J., Taylor, J.M., and Mack, C.P. (2017). Blood pressure-associated polymorphism controls ARHGAP42 expression via serum response factor DNA binding. *J. Clin. Invest.* 127, 670–680.
51. Baetta, R., Soma, M., De-Fraja, C., Comparato, C., Teruzzi, C., Magrassi, L., and Cattaneo, E. (2000). Upregulation and activation of Stat6 precede vascular smooth muscle cell proliferation in carotid artery injury model. *Arterioscler. Thromb. Vasc. Biol.* 20, 931–939.
52. Xiong, H.Y., Alipanahi, B., Lee, L.J., Bretschneider, H., Merico, D., Yuen, R.K.C., Hua, Y., Gueroussov, S., Najafabadi, H.S., Hughes, T.R., et al. (2015). RNA splicing. The human splicing code reveals new insights into the genetic determinants of disease. *Science* 347, 1254806.
53. Maloverjan, A., Piirsoo, M., Michelson, P., Kogerman, P., and Osterlund, T. (2010). Identification of a novel serine/threonine kinase ULK3 as a positive regulator of Hedgehog pathway. *Exp. Cell Res.* 316, 627–637.
54. Caballe, A., Wenzel, D.M., Agromayor, M., Alam, S.L., Skalicky, J.J., Kloc, M., Carlton, J.G., Labrador, L., Sundquist, W.I., and Martin-Serrano, J. (2015). ULK3 regulates cytokinetic abscission by phosphorylating ESCRT-III proteins. *Elife* 4, e06547.
55. Levy, D., Ehret, G.B., Rice, K., Verwoert, G.C., Launer, L.J., Dehghan, A., Glazer, N.L., Morrison, A.C., Johnson, A.D., Aspelund, T., et al. (2009). Genome-wide association study of blood pressure and hypertension. *Nat. Genet.* 41, 677–687.
56. Sinnott-Armstrong, N., Tanigawa, Y., Amar, D., Mars, N., Benner, C., Aguirre, M., Venkataraman, G.R., Wainberg, M., Ollila, H.M., Kiiskinen, T., et al. (2021). Genetics of 35 blood and urine biomarkers in the UK Biobank. *Nat. Genet.* 53, 185–194.
57. Hoffmann, T.J., Ehret, G.B., Nandakumar, P., Ranatunga, D., Schaefer, C., Kwok, P.-Y., Iribarren, C., Chakravarti, A., and Risch, N. (2017). Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. *Nat. Genet.* 49, 54–64.
58. Jaganathan, K., Kyriazopoulou Panagiotopoulou, S., McRae, J.F., Darbandi, S.F., Knowles, D., Li, Y.I., Kosmicki, J.A., Arbelaez, J., Cui, W., Schwartz, G.B., et al. (2019). Predicting Splicing from Primary Sequence with Deep Learning. *Cell* 176, 535–548.e24.
59. Ghaoui, R., Benavides, T., Lek, M., Waddell, L.B., Kaur, S., North, K.N., MacArthur, D.G., Clarke, N.F., and Cooper, S.T. (2016). TOR1AIP1 as a cause of cardiac failure and recessive limb-girdle muscular dystrophy. *Neuromuscul. Disord.* 26, 500–503.
60. Shin, J.-Y., Méndez-López, I., Wang, Y., Hays, A.P., Tanji, K., Lefkowitz, J.H., Schulze, P.C., Worman, H.J., and Dauer, W.T. (2013). Lamina-associated polypeptide-1 interacts with the muscular dystrophy protein emerin and is essential for skeletal muscle maintenance. *Dev. Cell* 26, 591–603.
61. Lewington, S., Clarke, R., Qizilbash, N., Peto, R., and Collins, R.; Prospective Studies Collaboration (2002). Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 360, 1903–1913.
62. Pinhal, M.A.S., Melo, C.M., and Nader, H.B. (2020). The Good and Bad Sides of Heparanase-1 and Heparanase-2. *Adv. Exp. Med. Biol.* 1221, 821–845.
63. Kiyani, Y., Tkachuk, S., Kurselis, K., Shushakova, N., Stahl, K., Dawodu, D., Kiyani, R., Chichkov, B., and Haller, H. (2019). Heparanase-2 protects from LPS-mediated endothelial injury by inhibiting TLR4 signalling. *Sci. Rep.* 9, 13591.
64. Meng, S., Gu, Q., Yang, X., Lv, J., Owusu, I., Matrone, G., Chen, K., Cooke, J.P., and Fang, L. (2018). TBX20 Regulates Angiogenesis Through the Prokineticin 2-Prokineticin Receptor 1 Pathway. *Circulation* 138, 913–928.
65. Tcheandjieu, C., Xiao, K., Tejada, H., Lynch, J.A., Ruotsalainen, S., Bellomo, T., Palnati, M., Judy, R., Klarin, D., Kember, R.L., et al. (2022). High heritability of ascending aortic diameter and trans-ancestry prediction of thoracic aortic disease. *Nat. Genet.* 54, 772–782.
66. Škorić-Milosavljević, D., Tados, R., Bosada, F.M., Tessadori, F., van Weerd, J.H., Woudstra, O.I., Tjong, F.V.Y., Lahrouchi, N., Bajolle, F., Cordell, H.J., et al. (2022). Common genetic variants contribute to risk of transposition of the great arteries. *Circ. Res.* 130, 166–180.
67. Luyckx, I., Kumar, A.A., Reyniers, E., Dekeyser, E., Vanderstraeten, K., Vandeweyer, G., Wünnemann, F., Preuss, C., Mazzella, J.-M., Goudot, G., et al. (2019). Copy number variation analysis in bicuspid aortic valve-related aortopathy identifies TBX20 as a contributing gene. *Eur. J. Hum. Genet.* 27, 1033–1043.
68. Ji, Y., Ma, Y., Shen, J., Ni, H., Lu, Y., Zhang, Y., Ma, H., Liu, C., Zhao, Y., Ding, S., et al. (2021). TBX20 Contributes to Balancing the Differentiation of Perivascular Adipose-Derived Stem Cells to Vascular Lineages and Neointimal Hyperplasia. *Front. Cell Dev. Biol.* 9, 662704.
69. Miller, C.L., Haas, U., Diaz, R., Leeper, N.J., Kundu, R.K., Patlolla, B., Assimes, T.L., Kaiser, F.J., Perisic, L., Hedin, U., et al. (2014). Coronary heart disease-associated variation in TCF21 disrupts a miR-224 binding site and miRNA-mediated regulation. *PLoS Genet.* 10, e1004263.
70. Miller, C.L., Pjanic, M., Wang, T., Nguyen, T., Cohain, A., Lee, J.D., Perisic, L., Hedin, U., Kundu, R.K., Majmudar, D., et al. (2016). Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* 7, 12092.
71. Miller, C.L., Anderson, D.R., Kundu, R.K., Raiesdana, A., Nürnberg, S.T., Diaz, R., Cheng, K., Leeper, N.J., Chen, C.-H., Chang, I.-S., et al. (2013). Disease-related growth factor and embryonic signaling pathways modulate an enhancer of TCF21 expression at the 6q23.2 coronary heart disease locus. *PLoS Genet.* 9, e1003652.
72. Arab, K., Park, Y.J., Lindroth, A.M., Schäfer, A., Oakes, C., Weichenhan, D., Lukanova, A., Lundin, E., Risch, A., Meister, M., et al. (2014). Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. *Mol. Cell* 55, 604–614.
73. Cheng, Z., Zhang, Y., Zhuo, Y., Fan, J., Xu, Y., Li, M., Chen, H., and Zhou, L. (2022). LncRNA TARID induces cell proliferation through cell cycle

- pathway associated with coronary artery disease. *Mol. Biol. Rep.* **49**, 4573–4581.
74. Yeung, M.W., Wang, S., van de Vegte, Y.J., Borisov, O., van Setten, J., Snieder, H., Verweij, N., Said, M.A., and van der Harst, P. (2022). Twenty-Five Novel Loci for Carotid Intima-Media Thickness: A Genome-Wide Association Study in >45 000 Individuals and Meta-Analysis of >100 000 Individuals. *Arterioscler. Thromb. Vasc. Biol.* **42**, 484–501.
 75. Bai, X., Lenhart, K.C., Bird, K.E., Suen, A.A., Rojas, M., Kakoki, M., Li, F., Smithies, O., Mack, C.P., and Taylor, J.M. (2013). The smooth muscle-selective RhoGAP GRAF3 is a critical regulator of vascular tone and hypertension. *Nat. Commun.* **4**, 2910.
 76. Evangelou, E., Warren, H.R., Mosen-Ansorena, D., Mifsud, B., Pazoki, R., Gao, H., Ntritsos, G., Dimou, N., Cabrera, C.P., Karaman, I., et al. (2018). Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat. Genet.* **50**, 1412–1425.
 77. Marx, C., Novotny, J., Salbeck, D., Zellner, K.R., Nicolai, L., Pekayvaz, K., Kilani, B., Stockhausen, S., Bürgener, N., Kupka, D., et al. (2019). Eosinophil-platelet interactions promote atherosclerosis and stabilize thrombosis with eosinophil extracellular traps. *Blood* **134**, 1859–1872.
 78. Fernández-Gallego, N., Castillo-González, R., Méndez-Barbero, N., López-Sanz, C., Obeso, D., Villaseñor, A., Escribese, M.M., López-Melgar, B., Salamanca, J., Benedicto-Buendía, A., et al. (2022). The impact of type 2 immunity and allergic diseases in atherosclerosis. *Allergy* **77**, 3249–3266.
 79. Knutsson, A., Björkbacka, H., Dunér, P., Engström, G., Binder, C.J., Nilsson, A.H., and Nilsson, J. (2019). Associations of Interleukin-5 With Plaque Development and Cardiovascular Events. *JACC. Basic Transl. Sci.* **4**, 891–902.
 80. Wenderfer, S.E., Slack, J.P., McCluskey, T.S., and Monaco, J.J. (2000). Identification of 40 genes on a 1-Mb contig around the IL-4 cytokine family gene cluster on mouse chromosome 11. *Genomics* **63**, 354–373.
 81. Chang, L., Yang, H.-W., Lin, T.-Y., and Yang, K.D. (2021). Perspective of immunopathogenesis and immunotherapies for kawasaki disease. *Front. Pediatr.* **9**, 697632.
 82. Laporte, J.C., Moore, P.E., Baraldo, S., Jouvin, M.H., Church, T.L., Schwartzman, I.N., Panettieri, R.A., Kinet, J.P., and Shore, S.A. (2001). Direct effects of interleukin-13 on signaling pathways for physiological responses in cultured human airway smooth muscle cells. *Am. J. Respir. Crit. Care Med.* **164**, 141–148.
 83. Garrido-Martín, D., Borsari, B., Calvo, M., Reverter, F., and Guigó, R. (2021). Identification and analysis of splicing quantitative trait loci across multiple tissues in the human genome. *Nat. Commun.* **12**, 727.
 84. Li, Y.I., van de Geijn, B., Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., and Pritchard, J.K. (2016). RNA splicing is a primary link between genetic variation and disease. *Science* **352**, 600–604.
 85. Rebelo, S., da Cruz E Silva, E.F., and da Cruz E Silva, O.A.B. (2015). Genetic mutations strengthen functional association of LAP1 with DYT1 dystonia and muscular dystrophy. *Mutat. Res. Rev. Mutat. Res.* **766**, 42–47.
 86. Liu, B., Gao, W., Sun, W., Li, L., Wang, C., Yang, X., Liu, J., and Guo, Y. (2020). Promoting roles of long non-coding RNA FAM83H-AS1 in bladder cancer growth, metastasis, and angiogenesis through the c-Myc-mediated ULK3 upregulation. *Cell Cycle* **19**, 3546–3562.
 87. Han, Y., Wang, B., Cho, Y.S., Zhu, J., Wu, J., Chen, Y., and Jiang, J. (2019). Phosphorylation of cI/gli by fused family kinases promotes hedgehog signaling. *Dev. Cell* **50**, 610–626.e4.
 88. Goruppi, S., Procopio, M.-G., Jo, S., Clocchiatti, A., Neel, V., and Dotto, G.P. (2017). The ULK3 Kinase Is Critical for Convergent Control of Cancer-Associated Fibroblast Activation by CSL and GLI. *Cell Rep.* **20**, 2468–2479.
 89. Hardwick, S.A., Hu, W., Joglekar, A., Fan, L., Collier, P.G., Foord, C., Ballacco, J., Lanjewar, S., Sampson, M.M., Koopmans, F., et al. (2022). Single-nuclei isoform RNA sequencing unlocks barcoded exon connectivity in frozen brain tissue. *Nat. Biotechnol.* **40**, 1082–1092.
 90. Gupta, I., Collier, P.G., Haase, B., Mahfouz, A., Joglekar, A., Floyd, T., Koopmans, F., Barres, B., Smit, A.B., Sloan, S.A., et al. (2018). Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. *Nat. Biotechnol.* **36**, 1197–1202.
 91. Lempiäinen, H., Brænne, I., Michael, T., Tragante, V., Vilne, B., Webb, T.R., Kyriakou, T., Eichner, J., Zeng, L., Willenborg, C., et al. (2018). Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Sci. Rep.* **8**, 3434.
 92. Wojcik, G.L., Graff, M., Nishimura, K.K., Tao, R., Haessler, J., Gignoux, C.R., Highland, H.M., Patel, Y.M., Sorokin, E.P., Avery, C.L., et al. (2019). Genetic analyses of diverse populations improves discovery for complex traits. *Nature* **570**, 514–518.
 93. Joshi, A., and Mayr, M. (2018). Aptamers They Trust: The Caveats of the SOMAScan Biomarker Discovery Platform from SomaLogic. *Circulation* **138**, 2482–2485.
 94. Katz, D.H., Tahir, U.A., Bick, A.G., Pampana, A., Ngo, D., Benson, M.D., Yu, Z., Robbins, J.M., Chen, Z.-Z., Cruz, D.E., et al. (2022). Whole genome sequence analysis of the plasma proteome in black adults provides novel insights into cardiovascular disease. *Circulation* **145**, 357–370.
 95. Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D.J., Martin, I., Kullo, I.J., Kenny, E.E., Pasaniuc, B., Witte, J.S., et al. (2023). Polygenic Risk Methods in Diverse Populations (PRIMED) Consortium Methods Working Group. Principles and methods for transferring polygenic risk scores across global populations. *Nat. Rev. Genet.* **25**, 8–25.
 96. Liao, W.-W., Asri, M., Ebler, J., Doerr, D., Haukness, M., Hickey, G., Lu, S., Lucas, J.K., Monlong, J., Abel, H.J., et al. (2023). A draft human pangenome reference. *Nature* **617**, 312–324.
 97. Zhou, W., Kanai, M., Wu, K.-H.H., Rasheed, H., Tsuo, K., Hirbo, J.B., Wang, Y., Bhattacharya, A., Zhao, H., Namba, S., et al. (2022). Global Biobank Meta-analysis Initiative: Powering genetic discovery across human disease. *Cell Genom.* **2**, 100192.
 98. Polygenic Risk Score Task Force of the International Common Disease Alliance (2021). Responsible use of polygenic risk scores in the clinic: potential benefits, risks and gaps. *Nat. Med.* **27**, 1876–1884.
 99. Browning, S.R., Waples, R.K., and Browning, B.L. (2023). Fast, accurate local ancestry inference with FLARE. *Am. J. Hum. Genet.* **110**, 326–335.
 100. Mohammadi, P., Castel, S.E., Brown, A.A., and Lappalainen, T. (2017). Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Res.* **27**, 1872–1884.
 101. Long, E., García-Closas, M., Chanock, S.J., Camargo, M.C., Banovich, N.E., and Choi, J. (2022). The case for increasing diversity in tissue-based functional genomics datasets to understand human disease susceptibility. *Nat. Commun.* **13**, 2907.
 102. Newman, A.M., Steen, C.B., Liu, C.L., Gentles, A.J., Chaudhuri, A.A., Scherer, F., Khodadoust, M.S., Esfahani, M.S., Luca, B.A., Steiner, D., et al. (2019). Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat. Biotechnol.* **37**, 773–782.
 103. Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Wucher, V., Kasela, S., Castel, S.E., Hamel, A.R., Viñuela, A., Roberts, A.L., et al. (2019). Cell type specific genetic regulation of gene expression across human tissues. Preprint at bioRxiv.
 104. Maria, M., Pouyanfar, N., Örd, T., and Kaikkonen, M.U. (2022). The Power of Single-Cell RNA Sequencing in eQTL Discovery. *Genes* **13**, 502.
 105. Kavousi, M., Bos, M.M., Barnes, H.J., Lino Cardenas, C.L., Wong, D., Lu, H., Hodonsky, C.J., Landsmeer, L.P.L., Turner, A.W., Kho, M., et al. (2023). Multi-ancestry genome-wide analysis identifies effector genes and druggable pathways for coronary artery calcification. *Nat. Genet.* **55**, 1651–1664.

106. Franceschini, N., Giambartolomei, C., de Vries, P.S., Finan, C., Bis, J.C., Huntley, R.P., Lovering, R.C., Tajuddin, S.M., Winkler, T.W., Graff, M., et al. (2018). GWAS and colocalization analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat. Commun.* **9**, 5141.
107. Pan, H., Xue, C., Auerbach, B.J., Fan, J., Bashore, A.C., Cui, J., Yang, D.Y., Trignano, S.B., Liu, W., Shi, J., et al. (2020). 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.
108. Alsaigh, T., Evans, D., Frankel, D., and Torkamani, A. (2020). Decoding the transcriptome of atherosclerotic plaque at single-cell resolution. Preprint at [bioRxiv](https://arxiv.org/abs/2008.00000).
109. Hu, Z., Liu, W., Hua, X., Chen, X., Chang, Y., Hu, Y., Xu, Z., and Song, J. (2021). Single-Cell Transcriptomic Atlas of Different Human Cardiac Arteries Identifies Cell Types Associated With Vascular Physiology. *Arterioscler. Thromb. Vasc. Biol.* **41**, 1408–1427.
110. Zheng, X., Levine, D., Shen, J., Gogarten, S.M., Laurie, C., and Weir, B.S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics* **28**, 3326–3328.
111. Browning, B.L., Zhou, Y., and Browning, S.R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *Am. J. Hum. Genet.* **103**, 338–348.
112. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
113. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760.
114. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842.
115. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al. (2011). The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158.
116. Castel, S.E., Mohammadi, P., Chung, W.K., Shen, Y., and Lappalainen, T. (2016). Rare variant phasing and haplotypic expression from RNA sequencing with phASER. *Nat. Commun.* **7**, 12817.
117. van de Geijn, B., McVicker, G., Gilad, Y., and Pritchard, J.K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nat. Methods* **12**, 1061–1063.
118. Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., Daly, M.J., Bustamante, C.D., and Kenny, E.E. (2017). Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am. J. Hum. Genet.* **100**, 635–649.
119. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermizakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. *Nat. Commun.* **8**, 15452.
120. Wallace, C. (2021). A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet.* **17**, e1009440.
121. Kichaev, G., Roytman, M., Johnson, R., Eskin, E., Lindström, S., Kraft, P., and Pasaniuc, B. (2017). Improved methods for multi-trait fine mapping of pleiotropic risk loci. *Bioinformatics* **33**, 248–255.
122. Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* **30**, 2811–2812.
123. Yin, L., Zhang, H., Tang, Z., Xu, J., Yin, D., Zhang, Z., Yuan, X., Zhu, M., Zhao, S., Li, X., and Liu, X. (2021). rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Dev. Reprod. Biol.* **19**, 619–628.
124. Gu, Z., Eils, R., and Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849.
125. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat. Protoc.* **7**, 500–507.
126. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. *Nat. Genet.* **50**, 151–158.
127. Maples, B.K., Gravel, S., Kenny, E.E., and Bustamante, C.D. (2013). RFMix: a discriminative modeling approach for rapid and robust local-ancestry inference. *Am. J. Hum. Genet.* **93**, 278–288.
128. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7.
129. Delaneau, O., Howie, B., Cox, A.J., Zagury, J.-F., and Marchini, J. (2013). Haplotype estimation using sequencing reads. *Am. J. Hum. Genet.* **93**, 687–696.
130. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. *Nature* **526**, 68–74.
131. Frankish, A., Diekhans, M., Ferreira, A.-M., Johnson, R., Jungreis, I., Loveland, J., Mudge, J.M., Sisu, C., Wright, J., Armstrong, J., et al. (2019). GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773.
132. Auwera, G.A.V. der, and O'Connor, B.D. (2020). *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, 1st ed. (O'Reilly Media).
133. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinform.* **12**, 323.
134. Jun, G., Flickinger, M., Hetrick, K.N., Romm, J.M., Doheny, K.F., Abecasis, G.R., Boehnke, M., and Kang, H.M. (2012). Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am. J. Hum. Genet.* **91**, 839–848.
135. Lee, S., Lee, S., Ouellette, S., Park, W.-Y., Lee, E.A., and Park, P.J. (2017). RNSCheckMate: software for validating sample identity in next-generation sequencing studies within and across data types. *Nucleic Acids Res.* **45**, e103.
136. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. *Nat. Genet.* **48**, 206–213.
137. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354.
138. Zhong, Y., Perera, M.A., and Gamazon, E.R. (2019). On using local ancestry to characterize the genetic architecture of human traits: genetic regulation of gene expression in multiethnic or admixed populations. *Am. J. Hum. Genet.* **104**, 1097–1115.
139. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.
140. Cotto, K.C., Feng, Y.-Y., Ramu, A., Skidmore, Z.L., Kunisaki, J., Conrad, D.F., Lin, Y., Chapman, W., Uppaulri, R., and Govindan, R. (2023). RegTools: Integrated analysis of genomic and transcriptomic data for discovery of splicing variants in cancer. Preprint at [bioRxiv](https://arxiv.org/abs/2308.00000).
141. GTEx Consortium; Laboratory, Data Analysis & Coordinating Center LDACC—Analysis Working Group; Statistical Methods groups—Analysis Working Group; Enhancing GTEx eGTEx groups; NIH Common Fund; NIH/NCI; NIH/NHGRI; NIH/NIMH; NIH/NIDA; Biospecimen Collection Source Site—NDRI (2017). Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213.
142. Franzén, O., Ermel, R., Cohain, A., Akers, N.K., Di Narzo, A., Talukdar, H.A., Foroughi-Asl, H., Giambartolomei, C., Fullard, J.F., Sukhvasi, K.,

- et al. (2016). Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases. *Science* 353, 827–830.
143. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
144. Shabalin, A.A. (2012). Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28, 1353–1358.
145. Zhu, Y., Orre, L.M., Zhou Tran, Y., Mermelekas, G., Johansson, H.J., Malyutina, A., Anders, S., and Lehtiö, J. (2020). Deqms: A method for accurate variance estimation in differential protein expression analysis. *Mol. Cell. Proteomics* 19, 1047–1057.
146. Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., and Ruden, D.M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92.
147. Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 38, e164.
148. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A simple new approach to variable selection in regression, with application to genetic fine mapping. *J. R. Stat. Soc. Series B Stat. Methodol.* 82, 1273–1300.
149. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature* 593, 238–243.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genotypes from low-pass whole genome sequencing	This paper	dbGaP: phs002855.v1.p1
Bulk RNA sequencing (raw files and TPMs)	This paper	GEO: GSE225650
eQTL and sQTL summary statistics	This paper	Zenodo: https://doi.org/10.5281/zenodo.7581778
snATAC sequencing (raw and processed data)	This paper	GEO: GSE175621 and GEO: GSE188422
ATAC sequencing and H3K27 Hi-ChIP for human coronary artery smooth muscle cells (HCASMCs)	This paper	GEO: GSE113348 and GEO: GSE101498
GWAS summary statistics for coronary artery disease	Tcheandjieu et al. ¹¹	GWAS catalog: GCST90132305
GWAS summary statistics for myocardial infarction	Hartiala et al. ⁹	GWAS catalog: GCST011365
GWAS summary statistics for coronary artery disease	Matsunaga et al. ¹⁰	GWAS catalog: GCST010480
GWAS summary statistics for coronary artery disease	Koyama et al. ⁸	Biobank Japan: https://biobankjp.org
GWAS summary statistics for coronary artery disease	van der Harst et al. ⁵	GWAS catalog: GCST005194
GWAS summary statistics for coronary artery calcification	Kavousi et al. ¹⁰⁵	GWAS catalog: GCST90278455 and GCST90278456
GWAS summary statistics for lipid traits (HDL, LDL, TC, logTG)	Graham et al. ¹⁸	GWAS catalog: GCST90239649, GCST90239655, GCST90239661, and GCST90239673
GWAS summary statistics for blood pressure traits (DBP, SBP, PP)	Evangelou et al. ⁷⁶	GWAS catalog: GCST006624, GCST006629, and GCST006630
GWAS summary statistics for carotid plaque and IMT	Franceschini et al. ¹⁰⁶	GWAS catalog: GCST001231
ENCODE CTCF, H3K27Ac, and H3K4me3 annotations for coronary artery	ENCODE ¹⁵	ENCODE Pproject: https://www.encodeproject.org/
GTEX summary statistics for eQTLs in AOR, COR, and TIB tissues	GTEX ²⁰	GTEX Portal: https://www.gtportal.org/home/
STARNET eQTL summary statistics, gene expression, and clinical trait enrichment for AOR & MAM tissues	Koplev et al. ³⁹	dbGaP: phs001203.v2.p1 http://starnet.mssm.edu/
scRNAseq data for coronary and carotid tissues (integrated single-cell reference)	Verdezoto et al. ³⁰ ; Wirka et al. ⁴¹ ; Pan et al. ¹⁰⁷ ; Alsaigh et al. ¹⁰⁸ ; Hu et al. ¹⁰⁹	GEO: GSE131778, GSE155512, and GSE159677. Zenodo: https://doi.org/10.5281/zenodo.6032099
Software and algorithms		
Custom R and shell scripts (data processing; plot generation; coloc, Paintor, SnpEff, SMR, RfMix implementation)	This paper	GitHub: https://github.com/MillerLab-CPHG/CAD_QTL and zenodo: https://doi.org/10.5281/zenodo.10095581
R 4.0.3	R Core Team	https://www.r-project.org/
R package SNPRelate v1.24.0	Zheng et al. ¹¹⁰	https://github.com/zhengxwen/SNPRelate
Beagle v5.2	Browning et al. ¹¹¹	https://bioinformatics.home.com/tools/imputation/descriptions/BEAGLE.html
STAR	Dobin et al. ¹¹²	https://github.com/alexdobin/STAR/releases
samtools v1.10	Li & Durbin ¹¹³	http://samtools.sourceforge.net/
bedtools v2.29.2	Quinlan et al. ¹¹⁴	https://github.com/arq5x/bedtools2
VCftools v0.1.16	Danecek et al. ¹¹⁵	https://vcftools.github.io
BCftools v1.9	https://doi.org/10.1093/gigascience/giab008	http://www.htslib.org/
Tabix v0.6	https://doi.org/10.1093/gigascience/giab008	https://www.htslib.org/doc/tabix.html
Python library phaser (requires python 3.7)	Castel et al. ¹¹⁶	https://github.com/secastel/phaser

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
WASP v0.3.4	van de Geijn et al. ¹¹⁷	https://github.com/bmvdgeijn/WASP
SnEff (requires java 1.11.0)	https://pcingola.github.io/SnpEff/adds/SnpEff_paper.pdf	https://pcingola.github.io/SnpEff/
Local ancestry inference R pipeline	Martin et al. ¹¹⁸	https://github.com/armartin/ancestry_pipeline
QTLtools v1.3.1	Delaneau et al. ¹¹⁹	https://qtltools.github.io/qtltools/
R package mixqtl v0.2	Liang et al. ²⁴	https://github.com/hakyimlab/mixqtl
R package coloc v5.1.1	Wallace et al. ¹²⁰	https://github.com/chr1swallace/coloc
SMR v1.03	Zhu et al. ⁴⁰	https://yanglab.westlake.edu.cn/software/smr/
paintor v3.0	Kichaev et al. ¹²¹	https://github.com/gkichaev/PAINTOR_V3.0
R package tidyverse v1.3.1		https://cran.r-project.org/
R package dplyr v1.0.7		https://cran.r-project.org/
R package data.table v1.14.2		https://cran.r-project.org/
R package ggplot2 v3.4.0		https://cran.r-project.org/
R package susie v0.11.92		https://cran.r-project.org/
R package BiocManager v1.30.6		https://cran.r-project.org/
R package devtools v2.4.3		https://cran.r-project.org/
R package reshape2 v1.4.4	https://www.jstatsoft.org/article/view/v021i12	https://cran.r-project.org/
R package ggrepel v0.9.2		https://cran.r-project.org/
R package VennDiagram v1.7.3		https://cran.r-project.org/
R package UpSetR v1.4.0		https://cran.r-project.org/
R package circlize v0.4.15	Gu et al. ¹²²	https://cran.r-project.org/
R package GetoptLong v1.0.5		https://cran.r-project.org/
R package RColorBrewer v1.1-3		https://cran.r-project.org/
R package CMplot v4.2.0	Yin et al. ¹²³	https://cran.r-project.org/
R package extrafont v0.19		https://cran.r-project.org/
R package ComplexHeatmap v2.14.0	Gu et al. ¹²⁴	https://github.com/jokergoo/ComplexHeatmap
R package peer v1.0	Stegle et al. ¹²⁵	https://github.com/PMBio/peer
LeafCutter v1.0	Yang et al. ¹²⁶	https://github.com/davidaknowles/leafcutter
RFMix v2.0	Maples et al. ¹²⁷	https://github.com/slowkoni/rfmix
Plink v1.9	Chang et al. ¹²⁸	https://www.cog-genomics.org/plink/
Custom QTL scripts	This paper	GitHub: https://github.com/MillerLab-CPHG/CAD_QTL and zenodo: https://doi.org/10.5281/zenodo.10095581

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Clint L. Miller (clintm@virginia.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Data

All raw and processed bulk RNA-sequencing data are available in the Gene Expression Omnibus database (GEO: GSE225650). Low-pass whole-genome sequencing-based genotyping data are available on dbGaP: phs002855.v1.p1. The full summary statistics for the mixQTL eQTL analyses (full sample as well as subsamples for sensitivity analyses), as well as the local ancestry eQTL and the sQTL analyses are available on zenodo: <https://doi.org/10.5281/zenodo.7992146>

The single-cell RNA-seq datasets from coronary and carotid artery were re-analyzed and integrated from the original datasets available through GEO: GSE131778, GSE155512, GSE159677, and zenodo: <https://doi.org/10.5281/zenodo.6032099>.¹⁰⁹ The raw and processed single-nucleus ATAC-seq datasets are available through GEO: GSE175621 and GEO: GSE188422. The reprocessed and analyzed human scRNA-seq datasets are also available on PlaqView (<https://plaqview.com>). GTEx gene expression and eQTL data were obtained from the v8 portal website (<https://gtexportal.org>). STARNET gene expression, eQTL, and clinical trait enrichment data were obtained from dbGaP: phs001203.v2.p1 and are also available at <http://starnet.mssm.edu>. The HCASMC ATAC-seq and H3K27ac HiChIP data used to calculate ABC scores are available through GEO: GSE113348 and GEO: GSE101498).

Code

All custom scripts used to generate the results are available on GitHub: https://github.com/MillerLab-CPHG/CAD_QTL and zenodo: <https://doi.org/10.5281/zenodo.10095581>. Specific parameters and versions for published software tools are also included in the [key resources table](#) and [method details](#).

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Ethics statement

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

Sample acquisition

Freshly explanted hearts from heart transplant recipients were obtained at Stanford University under approved Institutional Review Board protocols and written informed consent. Hearts were arrested in cardioplegic solution and rapidly transported from the operating room to the lab on ice. The proximal 5–6cm of three major coronary vessels (left anterior descending, left circumflex, and right coronary artery) were dissected from the epicardium on ice, trimmed of surrounding adipose and adventitia, rinsed in cold phosphate buffered saline, and rapidly snap frozen in liquid nitrogen. Similarly, aortic root and left ventricular free-wall tissues were also processed and stored at -80°C until processing. Throughout the manuscript, these tissues will be referred to as "Explants." Normal coronary artery, aorta, and left ventricle tissues were obtained by Stanford University (from Donor Network West and California Transplant Donor Network) from donor hearts rejected for transplantation, procured for research studies and were treated following the same protocol as the explanted hearts. The collected tissues will be referred to hereby as "Donors." Tissues were de-identified and clinical information (e.g., ICD-10 codes) was used to classify normal, ischemic and non-ischemic hearts. Frozen tissues were transferred to the University of Virginia through a material transfer agreement and IRB-approved protocols.

METHOD DETAILS

DNA genotyping

Genomic DNA isolation and sequencing

Approximately 20–25mg of frozen left ventricle or coronary artery tissue was used to isolate genomic DNA for each donor sample following the manufacturer's protocol (Qiagen DNeasy Blood and Tissue Kit, cat# 69504). Genomic DNA samples for all donors in the study were diluted using TE buffer to [5–15 ng/ μL] in skirted 96-well PCR plates. Plates were sealed and shipped to Gencove (New York, USA) for 0.4X low-pass genomic DNA sequencing.

Genomic DNA sequence processing

Phasing and imputation. Unphased low-pass whole genome sequencing files were provided by Gencove in build b37. VCFs provided by Gencove included just over 38 million variants imputed to 1000G using a proprietary pipeline—imputation quality scores were not provided. Because phasing was necessary for downstream analyses, samples were phased and subsequently re-imputed to 1000G phase 3 b37 reference panel using Beagle with `impute = true` and `gp = true` options.^{111,129} No additional variants had been imputed at the conclusion of this process.

Liftover of genomic coordinates. Phased autosomal VCFs were lifted over from b37 to hg19 to hg38 using Picard ("Picard Toolkit", 2019, Broad Institute. GitHub Repository: <http://broadinstitute.github.io/picard/>). After excluding approximately 10,000 variants that could not be mapped, approximately 38 million total variants were available for consideration in analyses.

Principal component estimation

We calculated ancestral principal components using the R package SNPrelate.¹¹⁰ Briefly, SNPrelate uses LD pruning to restrict genotype data to ~500,000 independent biallelic SNPs with a MAF >1% across the autosomal chromosomes. We used all participants from 1000 Genomes Phase 3 as a reference panel given the diversity of our sample.¹³⁰ Eigenvectors (EVs) 1 through 3 demonstrated clustering and correlation with Gencove-assigned majority continental ancestry, but subsequent EVs were driven by one or a small number of individuals (Figure S1). Analyses adjusting for global ancestry were therefore restricted to the first three EVs.

Local ancestry estimation

Self-reported race/ethnicity and Gencove-reported regional ancestry estimates suggested that adjustment for local ancestry (LA) might improve discovery compared to global EVs. We adapted the local-ancestry pipeline developed by Alicia Martin

(https://github.com/armartin/ancestry_pipeline), including incorporation of several Python scripts.¹¹⁸ We used RFmix2 (<https://github.com/slowkoni/rfmix>) to calculate ancestry from one of five continental reference populations¹²⁷; ancestry was recorded as “missing” for a region if >90% probability of concordance with a reference was not attained. We randomly selected genotypes representing continental ancestries from each superpopulation for 1,200 total 1000 Genomes participants: 400 AFR (African), and 200 each of AMR (Indigenous to the Americas), EAS (East Asian), EUR (European), and SAS (South Asian) superpopulations.¹³⁰ Because the most recent dataset with a 1000G genetic map is b37, we performed LA estimation and all downstream analyses using hg19 positions and gencode ‘v37lift37’ annotations.¹³¹

Bulk RNA sequencing and processing

Coronary artery tissue processing and RNA isolation

Coronary artery samples were selected for RNA sequencing based tissue availability (>50mg) and disease status (prioritizing capturing a range of phenotypes). Total RNA was extracted from using the QIAGEN miRNeasy Mini RNA Extraction kit (catalog #217004). Approximately 20 mg of frozen tissue pulverized using a pre-chilled mortar and pestle under liquid nitrogen was added to 1.5 mL RINO tubes (Next Advance, SKU TUBE1R5-S) which were stored on dry ice. As expected, ease of sample processing under liquid nitrogen varied depending on sample calcification. Tissue powder was then further homogenized in Qiazol lysis buffer using stainless steel beads in a Bullet Blender (Next Advance) homogenizer, followed by column-based purification according to the manufacturer’s instructions. RNA concentration was determined using Qubit 3.0; RNA quality was determined using Agilent 4200 TapeStation. Three to five samples were processed per day by one of three individuals. Samples with RNA Integrity Number (RIN) > 4.5 and Illumina DV₂₀₀ values > 75 were included for library construction.

RNA library sequencing

Total RNA libraries were constructed using the Illumina TruSeq Stranded Total RNA Gold kit (catalog #20020599) and barcodes were added to RNA libraries using the IDT for Illumina-TruSeq RNA Unique Dual Indexes (96 indices, 96 samples) Kit (IDT, Illumina, catalog #20022371). This library preparation captured coding RNAs and some noncoding RNAs, while depleting ribosomal RNAs. After re-evaluating library quality using TapeStation, individually barcoded libraries were sent to Novogene for next generation sequencing. After passing additional QC, libraries were multiplexed and subjected to paired-end 150bp read sequencing on an Illumina NovaSeq S4 Flowcell to a median depth of 100 million total reads (>30G) per library.

RNA-seq mapping

Our RNA sequencing QC pipeline and scripts pertaining to all analyses described can be publicly accessed at: (https://github.com/MillerLab-CPHG/CAD_QTL)

RNA-seq read mapping and quality control

The raw passed filter sequencing reads obtained from Novogene were demultiplexed using the bcl2fastq script. Read quality was assessed using FastQC (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, version 0.11.9) and the adapter sequences were trimmed using Trim Galore version 0.6.5 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads with Phred scores <20 were also removed, after which all samples passing the quality control analysis were considered for inclusion. Trimmed reads were mapped to the hg38 human reference genome using STAR v2.7.3a according to the GATK Best Practices for RNA-seq.¹³² To increase mapping efficiency and sensitivity, new splice junctions discovered in a first alignment pass with high stringency were used as annotation in a second pass to permit lower stringency alignment and therefore increase sensitivity. PCR duplicates were marked using Picard; WASP was used to filter reads prone to mapping bias. Total read counts and RPKM were calculated with RNA-SeQC v1.1.8 using default parameters and additional flags “-n 1000 -noDoC -strictMode” and GENCODE v32 reference annotation. Transcript and isoform expression levels were estimated using RSEM v1.3.3.¹³³

Detection of RNA-seq sample swaps

Using known genotypes for an individual from our cohort, we used verifyBamID¹³⁴ to check for contaminated reads or sample swaps, we identified four swapped samples. To crosscheck these swaps, we ran NGSCheckMate, which uses a depth-dependent correlation model of allele fractions of known SNPs to identify samples from the same individual.¹³⁵ Consistent with VerifyBamID, we found four swapped samples, two of which we rematched and two of which were duplicates and therefore excluded from downstream analysis.

RNA sequencing data phasing for allele-specific expression

In order to capture allele-specific expression within RNAseq data, we phased all reads using RASQUAL.¹³⁶ Haplotype phasing of RNA sequencing reads was performed using phASER v1.1.1.¹¹⁶ Known sites for allelic mapping bias and *HLA* genes were excluded because of their high mapping error rates and introduction of bias using the ENCODE Unified GRCh38 Blacklist (as of 5-5-2020).¹³⁷ The phASER pipeline was performed using the guided tutorial with additional flags; “-paired_end 1 -mapq 255 -baseq 10”. Generation of the haplotype expression quantifications was performed with the companion tool called “phASER Gene AE” using the standard pipeline and the GENCODE v36 GRCh38 gene coordinates for haplotypic expression calculation.¹³¹

Multidimensional scaling

We performed nonparametric clustering in transcriptomic space to determine whether the transcriptomes of our coronary artery samples were unique from GTEx tissues as well as previously published HCASMC transcriptomes. We performed multidimensional scaling on log-transformed RPKM values (excluding genes in our study which did not have ≥ 0.1 RPKM in ≥ 10 individuals) of all GTEx tissues, our samples, and HCASMC cells using Kruskal's non-parametric method. The 54 unique GTEx tissue sites were grouped into 27 broad tissue categories as previously described.¹³

Quantitative trait loci mapping

eQTL analysis

PEER factor calculation and optimization. Probabilistic Estimation of Expression Residuals (PEER) was used to account for unmeasured confounders related to RNA sequencing (e.g., batch effects).¹²⁵ No covariates were included. Based on recommendations regarding sample size and statistical models, we expected to incorporate up to five PEER factors in all models. In order to determine how the number of PEER factors may affect our results, we selected a random subset of 400 genes on chr17 (~30% of annotated lncRNAs or protein-coding genes on that chromosome). We then performed linear regression using QTLtools (described below but with an allele frequency cutoff of 1% rather than 5%) with no additional covariates and compared the number of eGenes with an FDR < 0.05 adjusting for 1, 2, 3, 5, 8, 10, 15, 20, and 25 PEER factors (Figure S1). The number of eGenes was not meaningfully affected by the number of PEER factors, hence we continued with five as expected.

Regression analyses. To maximize true positive associations and minimize Type I error, we excluded variants that did not meet the following criteria: biallelic single-nucleotide variants with a minor allele frequency $> 5\%$ and HWE p value $> 1E-6$ (<http://samtools.github.io/bcftools/bcftools.html>). Approximately 6,100,000 variants were available for inclusion in each analysis. Bed files were generated using the criteria implemented by GTEx: all lncRNA and protein-coding genes exceeding 0.1 TPM for at least 20% of samples were included.²⁹ Genes were annotated with name, genomic coordinates, and strand using Gencode v32 (https://www.encodegenes.org/human/release_32.html) for hg38 analyses and gencode v37liftb37 for hg19 analyses.¹³¹ Due to strong selection and high likelihood of population stratification contributing to false-positive identification of genes in the MHC region which is under high selective pressure, HLA genes and other MHC components were not considered candidates for fine-mapping in downstream analyses.

Statistical significance reporting. For all regression analyses, the following standards will be used for reporting p values throughout the manuscript. For single variant association tests, p_{nom} refers to the p value reported in the "pval_meta" column for mixQTL or the 12th output column for QTLtools run under the nominal pass model. For variants that were not lead QTLs, p_{adj} refers to the Bonferroni-corrected p_{nom} , where the correction is adjusting for the number of SNPs tested for each gene or splice junction respectively. For lead QTLs, p_{BH} refers to the value obtained by correcting for the total number of genes or splice junctions tested within each method using the Benjamini-Hochberg FDR correction applied to the p_{adj} for mixQTL analyses and the 19th output column for QTLtools analyses.

Local ancestry adjustment. Local-ancestry-adjusted cis-eQTL analyses were performed using a new pipeline which incorporates QTLtools (a computationally efficient implementation of MatrixQTL which allows adjustment for SNP-level covariates).^{119,138} Models adjusted for age, sex, local ancestry, and five PEER factors. Because local ancestry designations occur at the SNP level and cannot be included as traditional covariates in genome-wide eQTL regression analyses, proportions of estimated continental ancestry for each individual were calculated for each gene, and gene-specific covariate and bed files were used in QTLtools. Local ancestry interpolation scripts as well as scripts to run each gene individually and select lead SNPs with adjusted p values can be found on the lab GitHub repository referenced above.

Permutation pass mode was run in QTLtools using up to 100,000 iterations per gene based on the method described by Gay et al.,²⁵ to generate a lead variant for each gene tested. "Nominal-pass" results were run for significant eGenes only to obtain a non-permuted p value for each variant within 500kb up- or downstream of the transcription start site (TSS). Isoforms were not evaluated separately, therefore the TSS used for each gene was the most upstream TSS for transcripts with multiple annotated isoforms.

Combined global-ancestry-adjusted and allele-specific expression analysis. We implemented the R package 'mixQTL' (<https://github.com/hakyimlab/mixqtl>) to identify eQTLs incorporating allele-specific expression in our data.²⁴ All mixQTL analyses were performed using genome build hg38; RNA transcripts were mapped to Gencode v32 (https://www.encodegenes.org/human/release_32.html). Briefly, mixQTL tests for an association between a genotype and total read count; an association between specific alleles and corresponding haplotype expression (allele-specific expression); and the meta-analysis of both scores when inclusion criteria are met for both methods. MixQTL inputs include phased genotypes, total read counts, allele-specific read counts, and covariate information. Because mixQTL has to be run separately for each gene and does not have a way to account for missing data, all data frames (covariate, haplotype, and expression) would need to be re-generated for each gene in the course of running the R script to account for "missing" values for any covariate, haplotype, or expression value for one or more individuals. We did not have missing data for any included covariates or genetic variants. Most genes with any missing expression values were lowly expressed, therefore we do not expect inflation that would not be accounted for by adjusting the resulting p values and set all missing values in the phased RNAseq data to zero. We adjusted the example script on the mixQTL Github (<https://github.com/lianggy/mixqtl>) to limit our window to variants within 500kb of the transcription start sites of each gene for a maximum window size of 1Mb, and analyzed associations all genes with ≥ 0.1 mean TPM. MixQTL models included three ancestry EVs, age, sex, and five PEER factors as covariates.

Sensitivity analyses in European-ancestry-specific and ancestrally proportional sample subsets. Most statistical fine-mapping methods and publicly available genomics references continue to over-represent ancestrally homogeneous European-ancestry populations. To evaluate whether discovery or replication could be improved in a genetically homogeneous study sample despite a meaningful decrease in sample size, we performed mixQTL and downstream analyses in subsets of our sample restricted either to individuals with 100% European ancestry ($n = 80$) or an 80-person downsample of the total study population randomly selected within each majority-assigned-ancestry group. Regression, generalization, and colocalization were performed within both subsets as described above for the combined sample.

Splicing analysis

Generation and processing of BAM and junction files for splice QTL analysis. We aligned to the human genome the FASTQ files from RNAseq data of 138 ancestrally diverse heart transplant patients using the aligner tool STAR (version 2.7.2b) in 2-pass mode.¹¹² An STAR index was generated using Gencode v37 annotations and the UCSC hg38 reference genome. BAM files were generated using the `-twopassmode basic` and `-outSAMtype BAM SortedByCoordinate` options with a minimum overhang of 8 bp for spliced alignments. Additionally, the WASP correction option, `-waspOutputMode SAMtag`, was used to mitigate allelic mapping bias.¹¹⁷ The generated BAM files were further indexed using the option `-index` from the package Samtools.¹³⁹ Next, junction files were generated using the package RegTools.¹⁴⁰ The option “-junctions extract” was used with the following parameters: (1) minimum intron size of 50 bp, (2) maximum intron length (100 kbp), (3) minimum overlap between junctions of 8 bp, and (4) leaving the strand specificity as ‘unstranded’ to analyze both positive and negative strands. We removed unknown or unwanted chromosomes from junction files.

LeafCutter. QTL input files were generated using the modified LeafCutter-GTEX pipeline.^{126,141} Junction files were used as input to generate a count matrix of intron excision ratios. Clusters of alternatively spliced introns were identified using split reads that mapped with a minimum of 8 bp into each exon. Singleton introns that did not cluster with other introns were discarded. LeafCutter iteratively analyzed each intron cluster and removed introns in two ways: (1) introns with fewer than 50 reads across all samples, or (2) intron reads present in less than 0.1% of the total number of reads in the entire cluster. LeafCutter re-clustered introns and only included those with a maximum length of 100,000 bp. The count matrix was further processed to (1) generate a BED file and its index in a format compatible for QTL mappers, such as QTLtools; and (2) to calculate splice principal components. The LeafCutter protocol specifically outputs percent spliced in (PSI) for each splice junction, which adjusts for splicing events with overlapping start or stop positions to account for gene-level variation.

sQTL association. Genetic variants affecting the quantity of proximal splice junctions (*cis*-sQTLs) were estimated using QTLtools as described above, with age, sex, and four ancestry principal components included in the model. Nominal-pass was used to obtain non-permuted p values for each variant within 250kb up- or downstream of the LeafCutter-generated BED file of splice junction expression. 100,000 permutations were run with a permutation pass to generate an adjusted p value for each lead sQTL.

Characterization and fine-mapping of QTLs

Characterization of coronary artery QTLs

STARNET gene and protein expression analysis. In order to compare magnitude and direction of effect at eGenes of interest, we generated expression plots by genotype for normalized expression and applied the FDR of 0.05 in the published work as a threshold for significance. Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task study (STARNET) subject recruitment and tissue collection were performed as described previously.¹⁴² Briefly, patients with coronary artery disease (CAD) who were eligible for open-thorax surgery at the Department of Cardiac Surgery, Tartu University Hospital in Estonia were enrolled after informed consent. Venous blood was drawn, and DNA was isolated for genotyping using the Illumina Infinium assay. Tissue biopsies from aorta and mammary artery were obtained to study tissue-specific gene expression and the disease. RNA was extracted from tissues and sequenced for whole transcriptome as described.¹⁴² In addition, 250 subjects who were eligible for open-heart surgery for reasons other than coronary artery bypass graft surgery were consented and recruited as controls.³⁹ For case control study cohort, whole transcriptome sequencing was performed on selected CAD cases and non-CAD controls with matched age, gender and BMI. Transcriptome data was processed and differential gene expression between cases and controls was estimated using DESeq2.¹⁴³ Association of gene expression with clinical phenotypes of subjects was calculated by Pearson’s correlation coefficients. Matrix eQTL was used for the inference of *cis*-regulated eQTLs analysis on genotype and expression data from individuals with CAD cases.¹⁴⁴ Association of gene expression with genotypes of SNPs were plotted using ggplot2 in R. For protein study, EDTA plasma was collected after blood drawn. In total, 304 cases and 217 controls were selected with matched age, sex, and BMI. Full-scale screening of Olink’s library of 11 panels targeting 1,012 proteins was performed by multiplex Proximity Extension Assay (Olink Proteomics, Uppsala, Sweden). R package DEqMS was used for quality control, data normalization, and statistical analysis for differential protein expression.¹⁴⁵

Generalization of previously published eQTLs in our study. We compared LA and mixQTL eGenes to published eQTL datasets in relevant tissues. Specifically, we assessed whether eGenes in our study also had significant eQTLs—and whether our lead variants were correlated using European-ancestry LD—in GTEXv8 coronary, thoracic aorta, and tibial artery tissues²⁹; and STARNET internal mammary artery and aortic root tissues.³⁹ Due to differences in annotation, overlap between each dataset and genes meeting our inclusion criteria was identified using the maximum set of gene names or primary ENSG identifiers (excluding post-decimal identifying numbers).

Annotation of eQTLs and sQTLs. To identify potential causal mechanisms of QTLs, we annotated lead eQTLs and sQTLs to their respective eGenes and sGenes using the “GRCh38.p13” database in SnpEff to correspond to the Gencode v32 annotation we aligned our RNAseq data to.¹⁴⁶ SnpEff reports variant annotations by gene (i.e., specific variant-gene annotations are not provided in the main output). For this reason, we utilized the verbose (-v) option and the log file to identify gene-variant-specific annotations to ensure the eQTL-eGene association was maintained. Corresponding scripts are available on our Github. We further evaluated gene-based and region-based annotations using the `annotate_variation.pl` script with Annovar, also for build hg38.¹⁴⁷

Co-localization of eQTLs and sQTLs with GWAS variants. We tested for colocalization of eQTLs with significant ($p < 5 \times 10^{-8}$) published genetic associations in GWAS of coronary artery disease and myocardial infarction.^{5,8–11} Due to the phenotypic heterogeneity within our study population, we additionally evaluated quantitative traits that are traditional risk factors for CAD including blood pressure traits (systolic [SBP], diastolic [DBP], and pulse [PP] pressures),⁷⁶ cholesterol traits (total cholesterol [TC], high-density lipoprotein [HDL], low-density lipoprotein [LDL], and triglycerides (reported as a log concentration, logTG)),¹⁸ carotid artery calcification (CAC),¹⁰⁵ and intima media thickness (IMT).¹⁰⁶ The GWAS catalog (<https://www.ebi.ac.uk/gwas>) was most recently accessed on November 1, 2022. We considered PPH3 >0.8 to implicate two independent causal SNPs associated with eQTL and GWAS association signals, and PPH4 >0.8 to support evidence of a shared causal variant.^{120,148}

Summary-data-based Mendelian randomization (SMR). SMR evaluates evidence for statistical pleiotropy by comparing summary statistics from GWAS to eQTL results from expression data in relevant tissues; i.e., to test whether the association between a genetic variant and CAD-related phenotypes is mediated by gene expression in our study sample.⁴⁰ Due to the complexity and low informativeness likely in LD generated from our genetically diverse but modestly sized sample, we generated bed files in Plink using 1000G EUR population (excluding Finnish samples due to genetic distinction of that population, which was not represented in our samples) as our LD ref.^{128,130} We then performed SMR (<https://yanglab.westlake.edu.cn/software/smr>) using mixQTL results and GWAS summary statistics from CAD and BP traits included in colocalization analyses.^{5,9,11,76,106} We were not able to evaluate Japanese CAD GWAS^{8,10} due to the large number of variants with allele frequency differences >30% between the GWAS study populations and ours. Options included a minor allele frequency threshold of 0.01, a 1Mb window surrounding the most significant eQTL per locus, and a maximum mixQTL p value of $5E-06$ required for variant inclusion.

Fine-mapping of eGene association signals

mixFine. To identify independent associations and prioritize credible sets for discovery coronary artery eGenes, we used the command line implementation of mixFine (<https://github.com/hakyimlab/mixqtl>).²⁴ Mixfine runs SuSiE internally to accommodate the possibility of multiple causal variants in a dataset where many of the input variables (SNPs) are likely highly correlated.¹⁴⁸ Input files utilized by mixFine are the same as those for mixQTL and were prepared as described above.

Fine-mapping with GWAS and variant annotations. To combine the effects of GWAS association signals, eQTLs identified in our coronary artery tissue samples, and tissue-relevant variant annotations, we fine-mapped mixQTL eGene associations using fast Paintor (v3.0, https://github.com/gkichaev/PAINTOR_V3.0).¹²¹ Due to FastPaintor input requirements and the ancestral representation in published GWAS of relevant traits, we restricted published datasets included in these analyses to European-only ancestry. MixQTL results were combined with overlapping SNPs from GWAS summary statistics for two groups of traits—CAD/MI and blood pressure traits.^{5,9,10} For each eGene (+/–500kb from TSS), VCFs and LD matrices were generated with Plink, using hg38 1000G EUR samples excluding FIN samples (which are minimally represented in our study or published GWAS).^{115,128,130} We evaluated four annotations based gene repression, activation, transcription, and chromatin-contact-based enhancer activity: CTCF only, H3K4me3 only, and H3K4me3/H3K27Ac combined. To do this, we included ENCODE binary SNP annotations for chromatin accessibility (CTCF, H3K27ac, and H3K4me3 in coronary artery tissue from one 53yo female).¹⁵ To apply enriched enhancer regulatory activity in coronary artery tissue, we also used the activity-by-contact model (ABC; <https://github.com/broadinstitute/ABC-Enhancer-Gene-Prediction>). ABC scores were obtained from previous H3K27ac HiChIP and ATAC data in HCASMC and coronary artery from our previous work are publicly available (see “Data and Code Availability”).^{13,149}

Supplemental information

**Multi-ancestry genetic analysis of gene
regulation in coronary arteries
prioritizes disease risk loci**

Chani J. Hodonsky, Adam W. Turner, Mohammad Daud Khan, Nelson B. Barrientos, Ruben Methorst, Lijiang Ma, Nicolas G. Lopez, Jose Verdezoto Mosquera, Gaëlle Auguste, Emily Farber, Wei Feng Ma, Doris Wong, Suna Onengut-Gumuscu, Maryam Kavousi, Patricia A. Peyser, Sander W. van der Laan, Nicholas J. Leeper, Jason C. Kovacic, Johan L.M. Björkegren, and Clint L. Miller

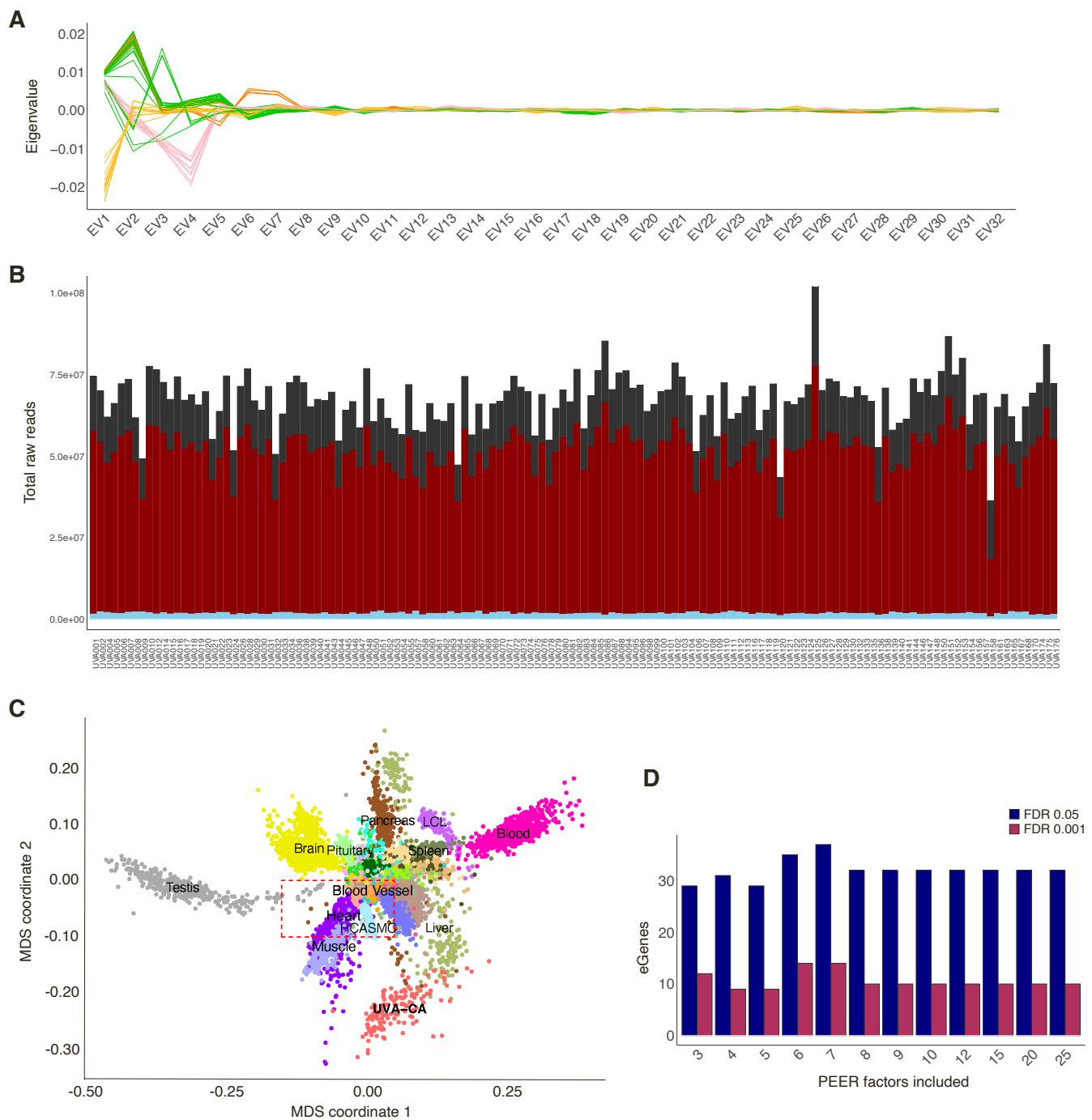


Figure S1. UVA sample genome-based ancestry and transcriptome-based expression values, Related to Figure 1. (A) Line plot displays the values (y-axis) of 32 Eigenvectors (x-axis) calculated using the 1000G phase 3 reference panel, with each individual in our study population represented by a solid line. Colors correspond to Gencove-assigned majority continental ancestry: pink = Amerindigenous; blue = European; green = South Asian; orange = East Asian; yellow = African. (B) Cumulative raw read counts (y-axis) for lncRNA (light blue), protein-coding (red), and other (gray) genes annotated to GENCODE v32 for individuals considered for eQTL analyses (x-axis). (C) Multidimensional scaling (MDS) plot shows UVA transcriptomes clustered separately from other GTEx tissue transcriptomes. Red dashed box includes overlapping tissue types of heart, muscle, and blood vessel and nearby HCASMC. (D) Bar plot represents the number (y-axis) of significant eQTLs identified among common variants within 500kb of 400 randomly selected chr17 genes adjusted for an increasing number of PEER factors included in the crude model (x-axis). FDRs of 5% and 0.1% are shown in navy blue and maroon, respectively.

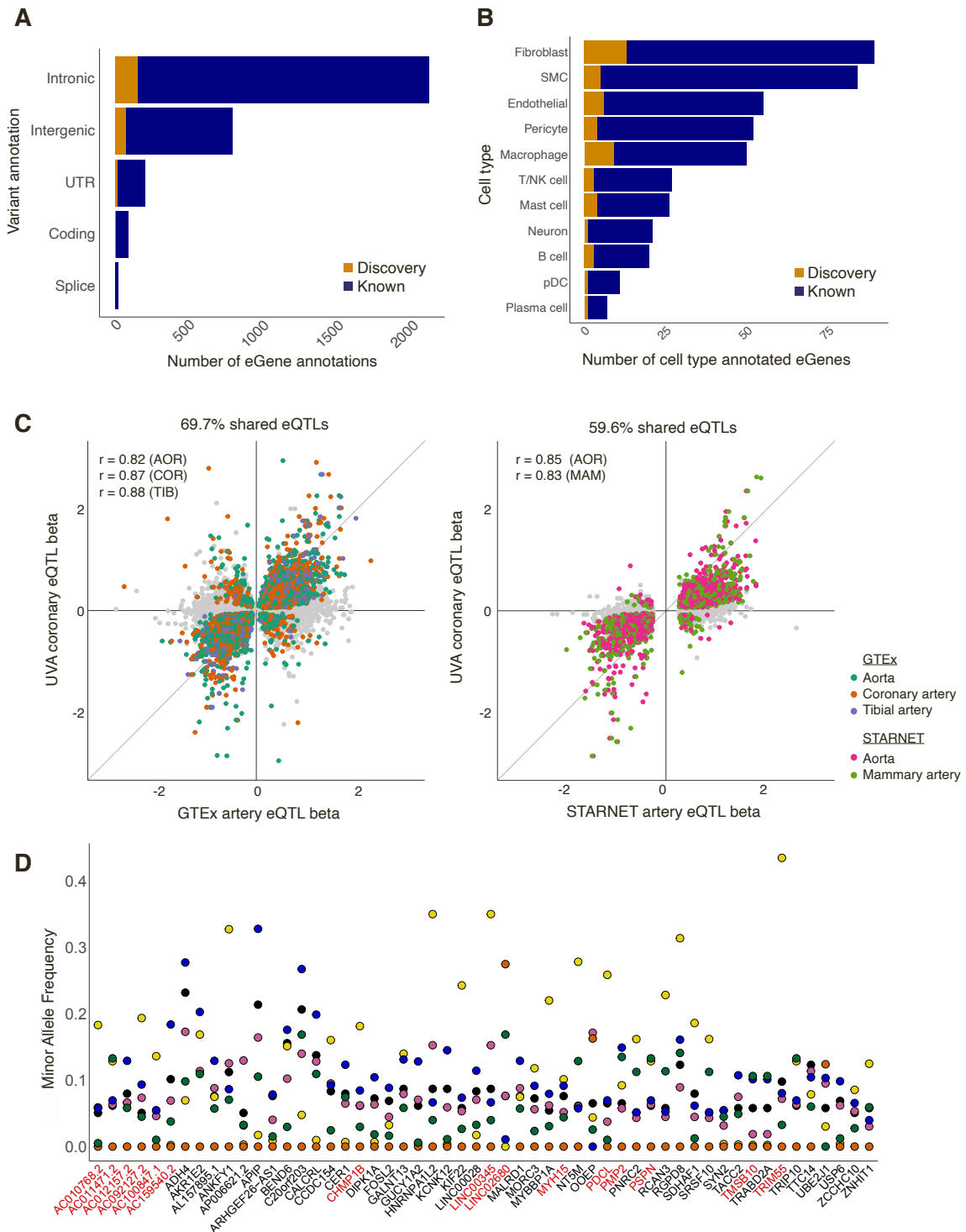


Figure S2. Annotation of mixQTL eGenes and shared effects in GTEx and STARNET artery tissues, Related to Figure 2. (A) SNPEff eQTL-eGene annotations for lead eQTLs. Navy blue and orange represent numbers of known and discovery eGenes, respectively. (B) Number of eGenes exhibiting cell-type-specific expression in an arterial single-cell RNA sequencing reference (CELLEX combined gene score >0.7).¹ Navy blue and orange represent numbers of reported and discovery eGenes, respectively. (C) Direction of effect for genes in which the UVA lead eQTL was significant ($p_{BH} < 0.05$) in the respective tissue. Pearson correlation coefficients (r) shown for overlapping significant UVA coronary eQTL detected in GTEx or STARNET eQTL with tissue indicated in parentheses. GTEx AOR: aorta (turquoise); COR: coronary artery (orange); TIB: tibial artery (purple); and STARNET AOR (pink); MAM: mammary artery (green). (D) Ancestry-specific eQTL allele frequency distribution: each point represents the effect allele frequency (y-axis) of a lead eQTL (x-axis) in the respective 1000 Genomes Phase 3 continental genetic ancestry superpopulations. Ancestries shown are AFR (yellow), AMR (pink), EAS (orange), EUR (blue), and SAS (green).

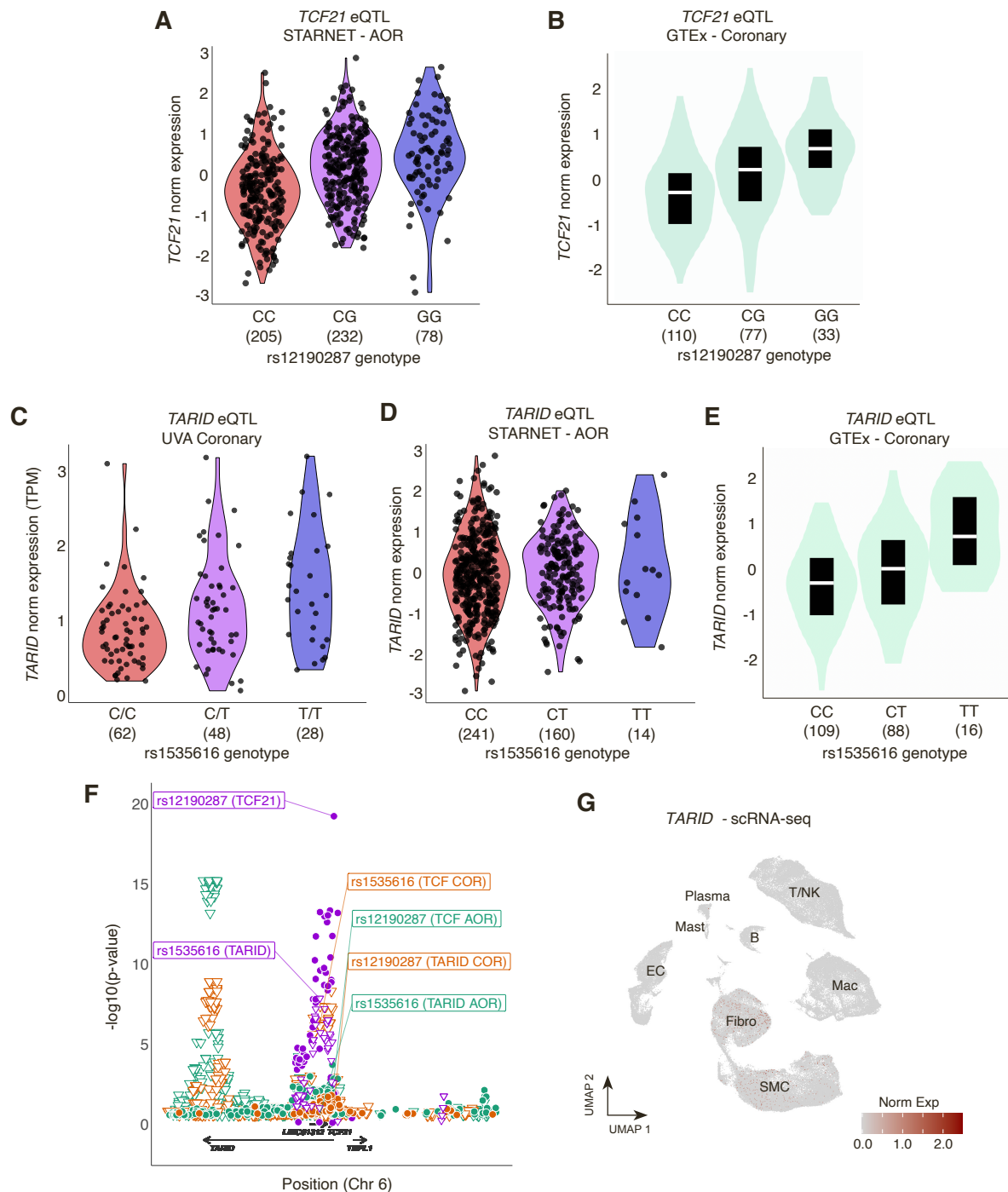
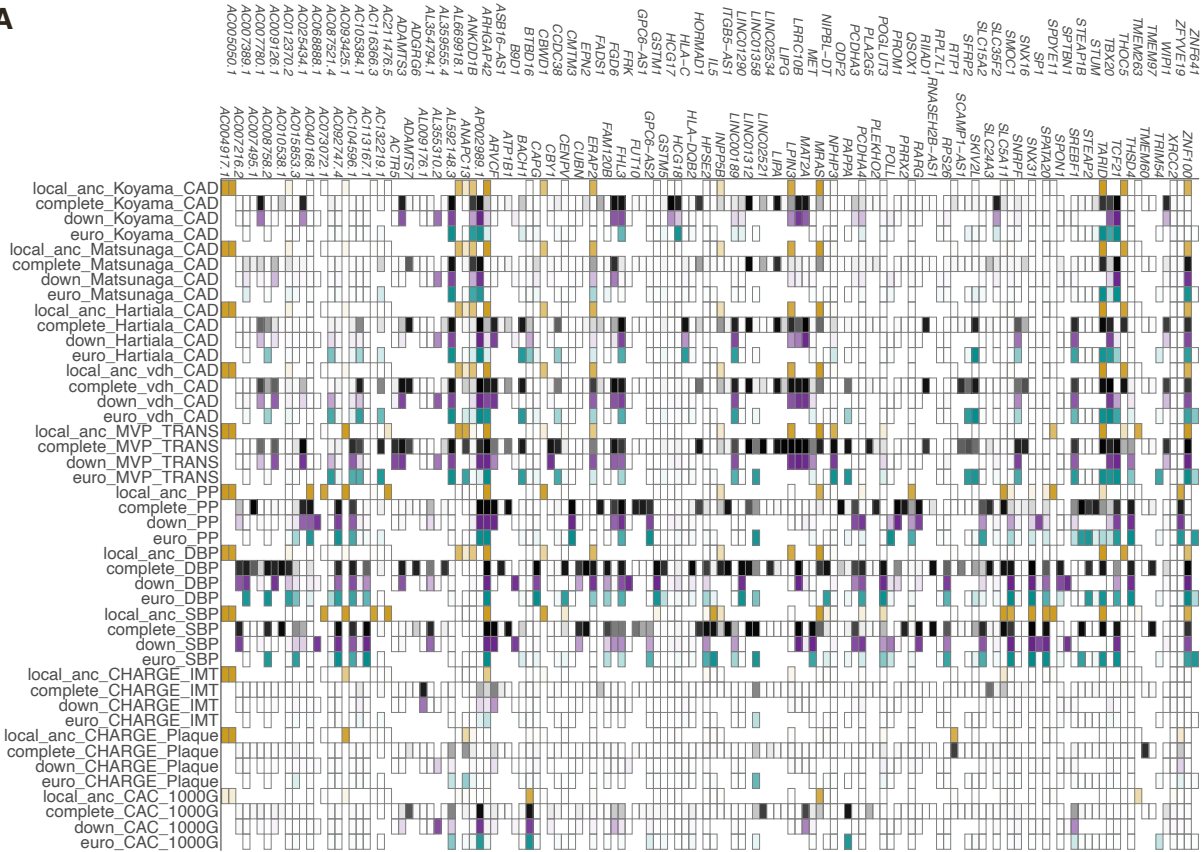
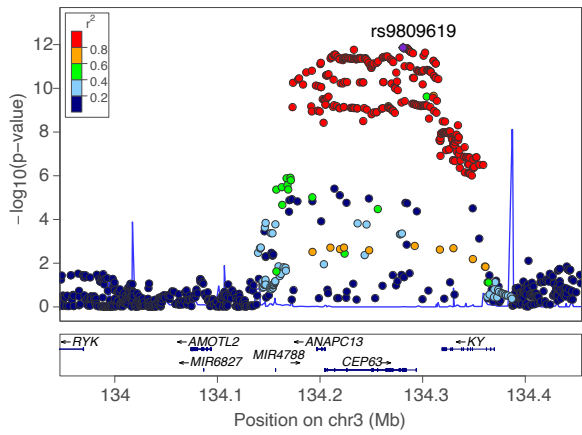


Figure S3. GWAS-eQTL colocalization eGenes in Europeans and supporting evidence for *TCF21*/*TARID* association, Related to Figure 3. (A,B) Violin plots for normalized expression of *TCF21* (y-axis) in STARNET aortic root (AOR) tissue (A) or GTEx coronary artery tissue (B) shown by genotype (x-axis) for lead UVA eQTL rs12190287. (C-E) Violin plots for expression of *TARID* lncRNA (y-axis) in UVA coronary artery (C), STARNET AOR (D), or GTEx coronary artery tissue (E) shown by genotype (x-axis) for lead UVA eQTL rs1535616. (F) Associations of variants with gene expression for UVA coronary (purple), GTEx AOR (green), and GTEx COR (orange) for *TCF21* (circle) and *TARID* (triangle). Position and direction of gene coding regions shown below x-axis. (G) UMAP plot showing relative expression of *TARID* in single-cell RNA sequencing data from human artery reference dataset.¹ Each point represents a single cell; intensity of red color corresponds to higher relative expression of *TARID*. Broad cell type clusters are labeled. SMC: smooth muscle cells; EC: Endothelial cells, Fibro: Fibroblasts; Mac: Macro-phage; B: B-cells; T/NK: T-cells and Natural Killer cells.

A



B



C

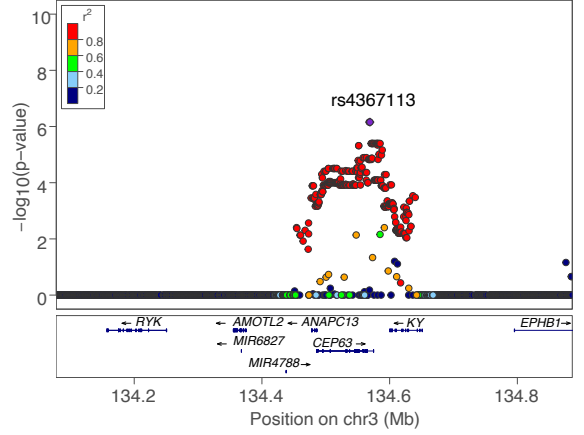


Figure S4. Comparison of colocalized associations between mixQTL, local ancestry-adjusted, and population-subsample coronary artery eQTLs, Related to STAR Methods. (A) Colocalization with CAD and BP GWAS traits for all eQTL methods. Heatmap indicates colocalization posterior probabilities (PPH4) for mixQTL total study population (dark gray), European-ancestry-only subset (turquoise), diverse down-sample subset (purple), or local ancestry-adjusted (gold). (b,c) Locus-Zoom plots for regional associations with gene expression for *ANAPC13* using mixQTL (B) or local-ancestry adjustment (C). SNP r^2 values are using the 1000G EUR reference.

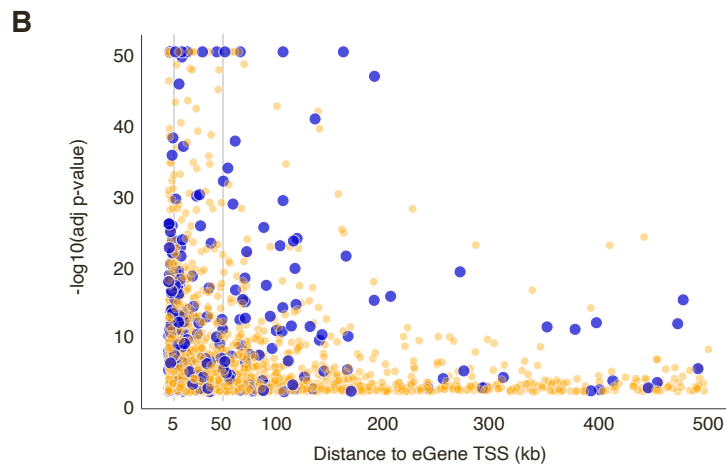
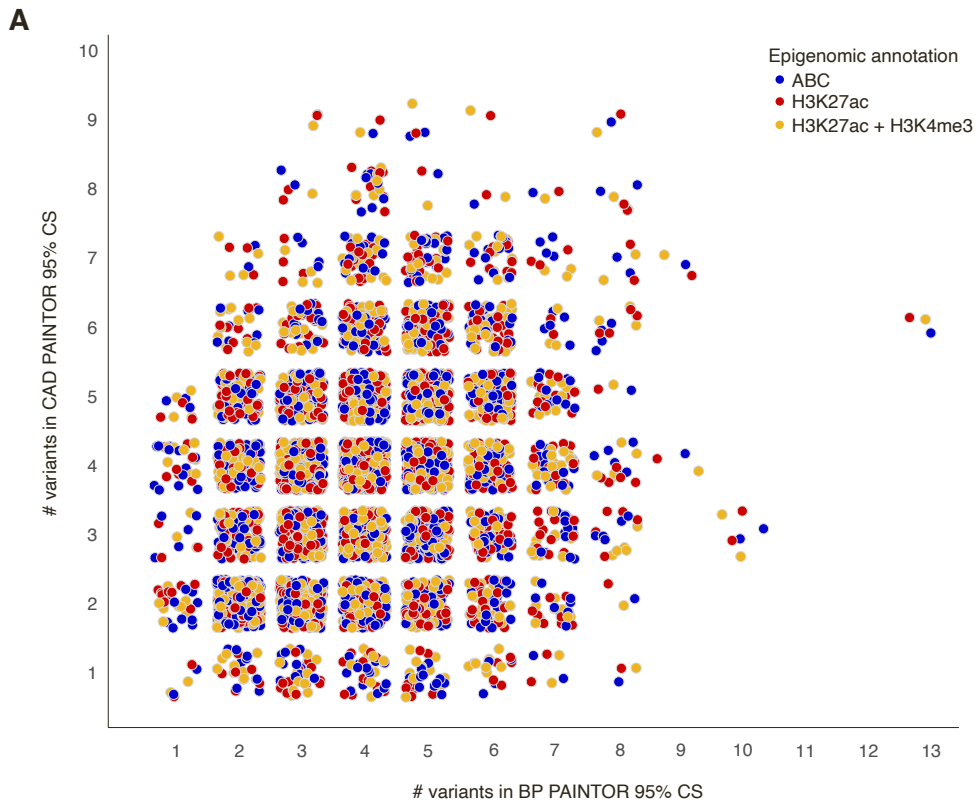


Figure S5. Fine-mapped credible set size and distance to TSS for CAD and BP trait GWAS, Related to STAR Methods. (A) Each point represents the number of variants in PAINTOR 95% credible set (CS) incorporating blood pressure trait GWAS data (x-axis) or CAD GWAS data (y-axis) for one UVA coronary artery eGene. All eGenes were fine-mapped using both GWAS datasets as well as annotation to ENCODE coronary artery activity-by-contact (ABC) chromatin contacts (blue), H3K27 acetylation (red), or H3K27 acetylation and H3K4 tri-methylation (yellow). For eGenes with no difference in credible set size only ABC annotation is shown. (B) Scatterplot shows significance of lead eQTLs (adjusted $-\log_{10}(p\text{-value})$, y-axis) versus distance from the corresponding eGene transcription start site (in kilobases, x-axis). Large blue points represent lead eQTLs for which at least one variant overlapped in the paintor CAD and BP GWAS-annotated credible sets; small orange points represent lead eQTLs for which no variants overlapped between the two credible sets for that eGene.

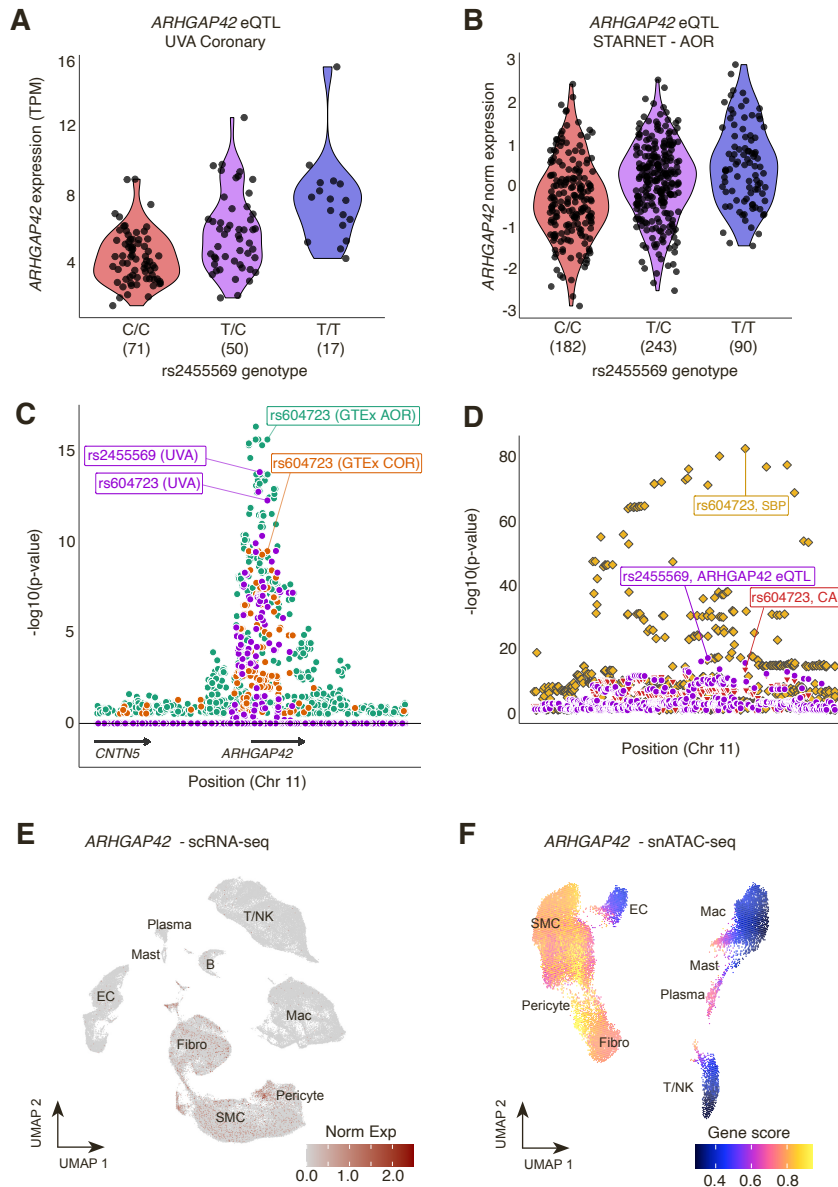


Figure S6. *ARHGAP42* eQTL and GWAS associations and cell type expression in human artery, Related to Figure 4. (A,B) Violin plots for normalized expression of *ARHGAP42* (y-axis) in UVA coronary artery tissue (A) and STARNET aortic root tissue (AOR) (B) shown by genotype (x-axis) for lead UVA eQTL rs2455569. (C) Associations of individual variants with gene expression are plotted for UVA coronary (purple), GTEx AOR (green), and GTEx COR (orange) for *ARHGAP42* (circle) and *CNTN5* (triangle). (D) Association of individual variants are plotted for coronary artery gene expression (purple) or GWAS traits (red [CAD] and gold [SBP]). Position of variants on chromosome 11 is shown on the x-axis, with position and direction of gene coding regions shown underneath the x-axis. Significance of each variant shown as the $-\log_{10}(p\text{-value})$ of the association with expression of the respective gene on the y-axis. UVA lead variant rs2455569 is labeled. (E) UMAP plot showing relative expression of *ARHGAP42* in single-cell RNA sequencing data from human artery atherosclerosis reference dataset.¹ Each point represents a single cell; intensity of red color corresponds to higher relative expression of *ARHGAP42*. General cell type clusters are labeled accordingly. (F) UMAP plot showing imputed gene score activity at *ARHGAP42* calculated from a single-nucleus chromatin accessibility sequencing dataset in human coronary artery.² SMC: smooth muscle cells; EC: Endothelial cells, Fibro: Fibroblasts; Mac: Macrophage; B: B-cells; T/NK: T-cells and Natural Killer cells.

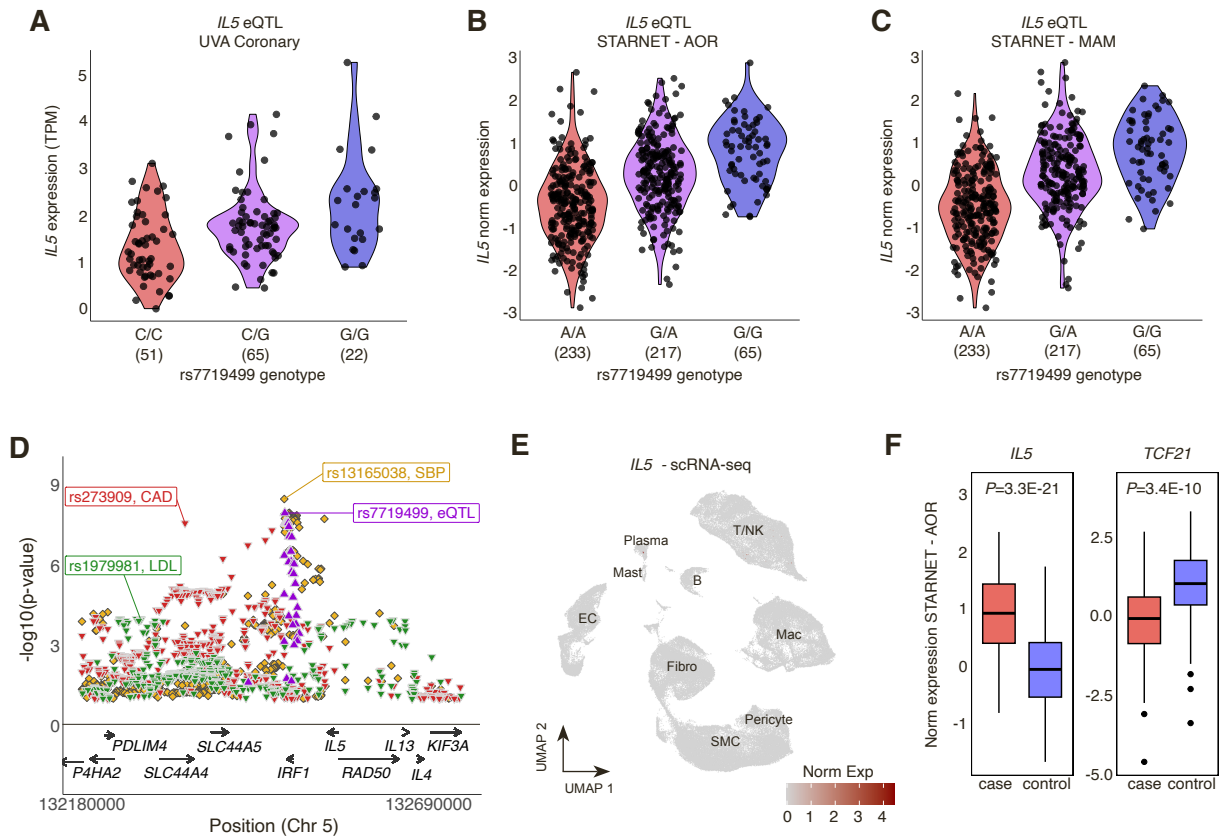


Figure S7. *IL5* eQTL associations, cell type and differential expression in human artery, Related to Figure 4. (A-C) Violin plots for normalized expression of *IL5* (y-axis) in UVA coronary artery tissue (a), STARNET aortic root tissue (AOR) (B), and STARNET mammary artery tissue (C) shown by genotype (x-axis) for lead UVA eQTL rs7719499. (D) Associations with individual variants are plotted for coronary artery gene expression (purple) or GWAS traits (red [CAD], gold [SBP], and green [LDL]). Position of variants on chromosome 5 is shown on the x-axis, with position and direction of gene coding regions shown underneath the x-axis. Significance of each variant shown as the $-\log_{10}(\text{p-value})$ of the association with expression of the respective gene on the y-axis. UVA lead variant rs7719499 is labeled. (E) UMAP plot showing relative expression of *IL5* in single-cell RNA sequencing data from human artery atherosclerosis reference dataset.¹ Each point represents a single cell; intensity of red color corresponds to higher relative expression of *IL5*. General cell type clusters are labeled accordingly. SMC: smooth muscle cells; EC: Endothelial cells, Fibro: Fibroblasts; Mac: Macrophage; B: B-cells; T/NK: T-cells and Natural Killer cells. (F) Normalized expression of *IL5* and *TCF21* in STARNET AOR tissues stratified by cases (individuals with coronary artery disease) or controls (individuals without coronary artery disease). Boxes represent upper and lower quartiles, line represents the median, whiskers as $1.5 \times \text{IQR}$, and outliers as individual points. P-values were determined from DEseq2 and adjusted for metabolic phenotypes, drug treatments and principal components of ancestry.³

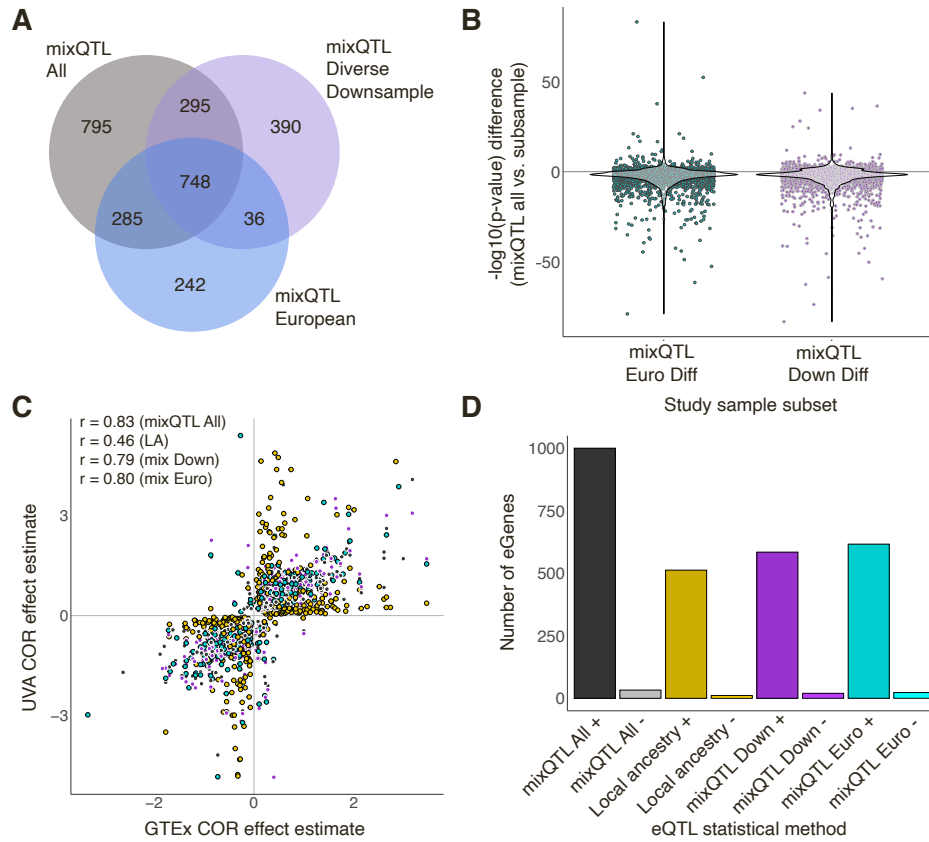


Figure S8. Sensitivity analyses of European-ancestry-only and proportionally representative down-sampled subsets compared to mixQTL and local-ancestry-adjusted results, Related to STAR Methods.

Methods. In all panels, colors represent mixQTL total study population (dark gray), the subset including 80 individuals of exclusively European ancestry (turquoise), a diverse subset including 80 individuals sampled to represent the broad genetic ancestries present in the total sample (purple), or local-ancestry-adjusted (gold) (A) Venn Diagram of shared and unique eGenes between mixQTL performed in the total study population and downsampled subsets. (B) Violin plot of differences in lead eQTL significance by regression method compared to total study sample. Each point represents the difference in $-\log_{10}(\text{p-value})$ for the lead eQTL of the respective method and the mixQTL lead eQTL for the same eGene. Positive values reflect a lower p-value in the subsample; negative values reflect a lower p-value in the combined overall sample. (C) Direction of effect comparison for GTEx coronary artery genes for which the UVA lead eQTL was significant ($p_{\text{BH}} < 0.05$). Pearson correlation coefficients (r) shown for overlapping GTEx COR-UVA coronary eQTLs with method indicated in parentheses. (D) Directional consistency (number of eGenes, y-axis) of lead eQTLs by regression method (x-axis). Darker and lighter bars represent consistent (+) and opposite (-) directions of effect by method, respectively.

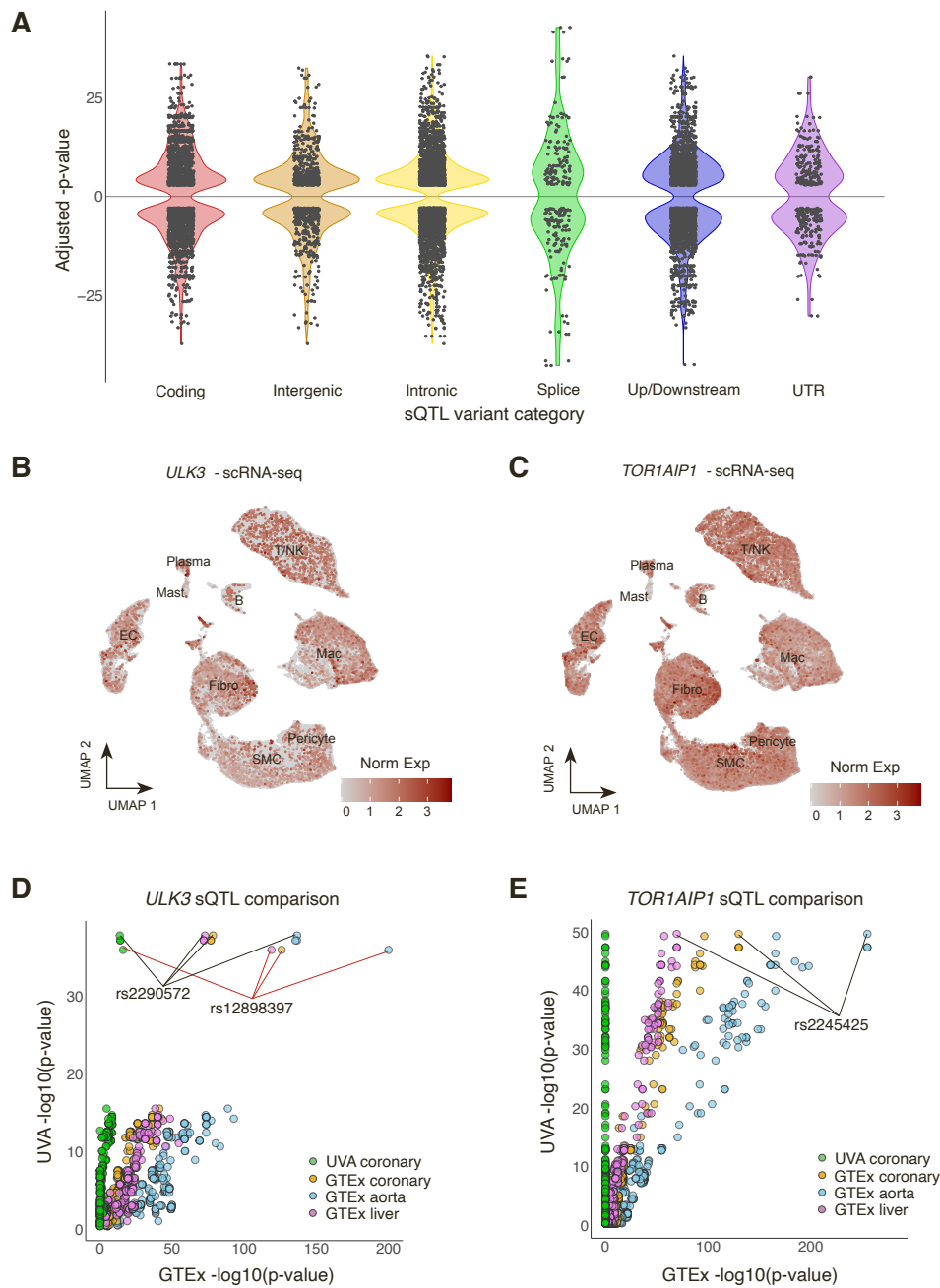


Figure S9. sQTL annotations highlight candidate sGenes *ULK3* and *TOR1AIP1*, Related to Figure 5. (A) Violin plot shows the adjusted significance incorporating direction of effect (y-axis) for lead sQTLs by SnpEff category of variant annotation (x-axis, p). (B,C) UMAP plots showing relative expression of *ULK3* (B) and *TOR1AIP1* (C) in single-cell RNA sequencing data from a human artery atherosclerosis reference dataset.¹ Each point represents a single cell; intensity of red color corresponds to higher relative expression of *IL5*. General cell type clusters are labeled accordingly. SMC: smooth muscle cells; EC: Endothelial cells, Fibro: Fibroblasts; Mac: Macrophage; B: B-cells; T/NK: T-cells and Natural Killer cells. (D) Scatterplot shows variants with UVA coronary sQTL significance (-log₁₀(p-value) of association with chr15:74837435: 74837757 at *ULK3*, y-axis) compared to significance of the same variant in UVA eQTL for the same gene (green) and GTEx aorta (blue), coronary (orange), and liver (violet) tissues. All p-values are adjusted for the number of variants tested for that gene in the corresponding tissue. Lead UVA and GTEx QTLs (rs2290572 and rs12898397, respectively) are labeled. (E) Scatterplot shows variants with UVA coronary sQTL significance (-log₁₀(p-value) of association with chr1:179884769:179889313 at *TOR1AIP1*, y-axis) compared to significance of the same variant in UVA eQTL for the same gene (green) and GTEx aorta (blue), coronary (orange), and liver (violet) tissues. All p-values are adjusted for the number of variants tested for that gene in the corresponding tissue. Lead UVA sQTL rs2245425 is labeled.

References for Supplemental Figures

1. Mosquera, J.V., Wong, D., Auguste, G., Turner, A.W., Hodonsky, C.J., Lino Cardenas, C.L., Theofilatos, K., Bos, M., Kavousi, M., Peyser, P., et al. (2022). Integrative single-cell meta-analysis reveals disease-relevant vascular cell states and markers in human atherosclerosis. *BioRxiv*. 10.1101/2022.10.24.513520.
2. Turner, A.W., Hu, S.S., Mosquera, J.V., Ma, W.F., Hodonsky, C.J., Wong, D., Auguste, G., Song, Y., Sol-Church, K., Farber, E., et al. (2022). Single-nucleus chromatin accessibility profiling highlights regulatory mechanisms of coronary artery disease risk. *Nat. Genet.* 54, 804–816. 10.1038/s41588-022-01069-0
3. Koplev, S., Seldin, M., Sukhavasi, K., Ermel, R., Pang, S., Zeng, L., Bankier, S., Di Narzo, A., Cheng, H., Meda, V., et al. (2022). A mechanistic framework for cardiometabolic and coronary artery diseases. *Nat. Cardiovasc. Res.* 1, 85–100. 10.1038/s44161-021-00009-1.