Cell Genomics, Volume *4*

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

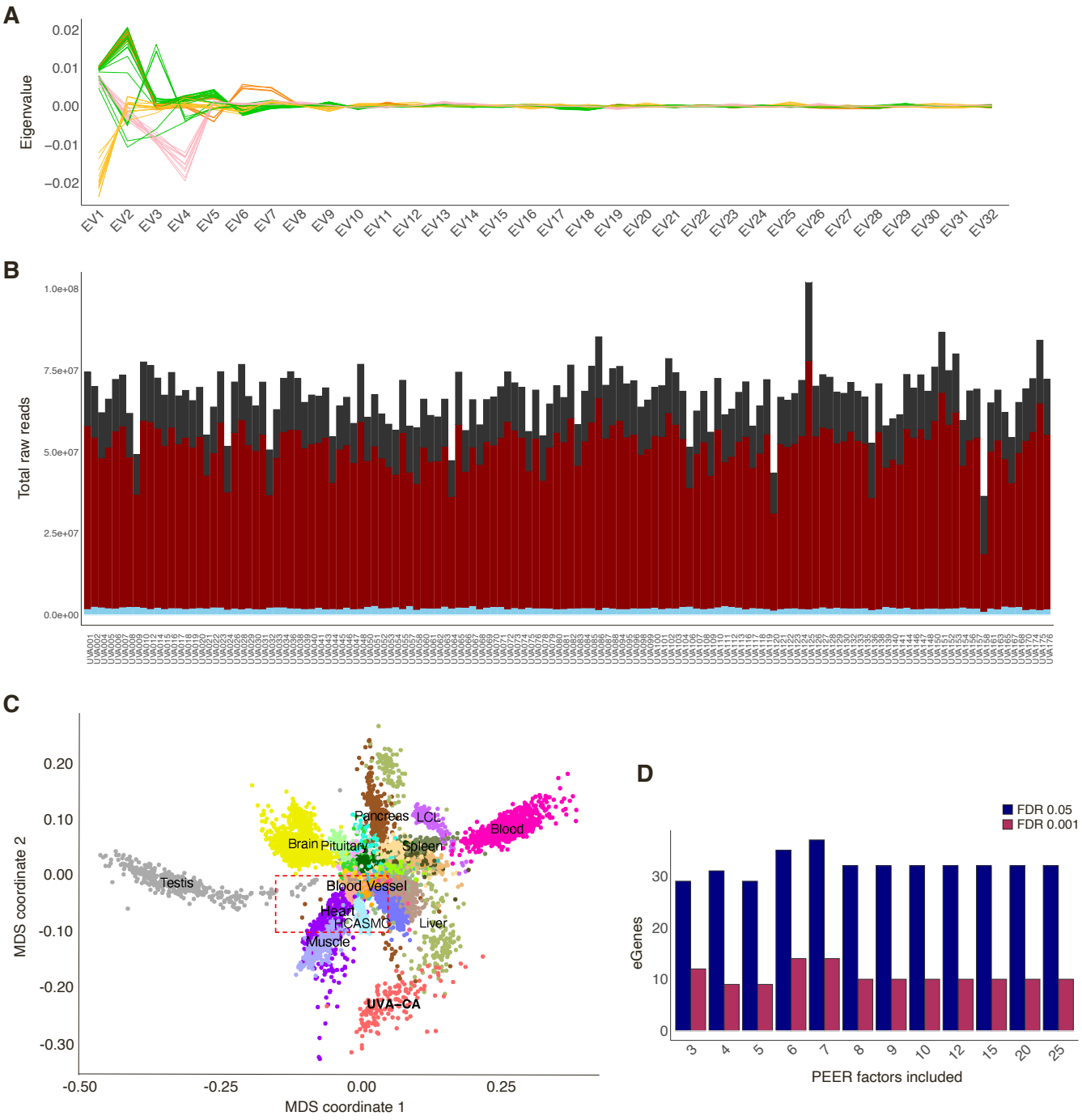

Figure S1. UVA sample genome-based ancestry and transcriptome-based expression values, Related to Figure 1. (A) Line plot displays the values (y-axis) of 32 Eigenvectors (x-axis) calculated using the 1000G phase 3 reference panel, with each individual in our study population represented by a solid line. 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 seprately 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.

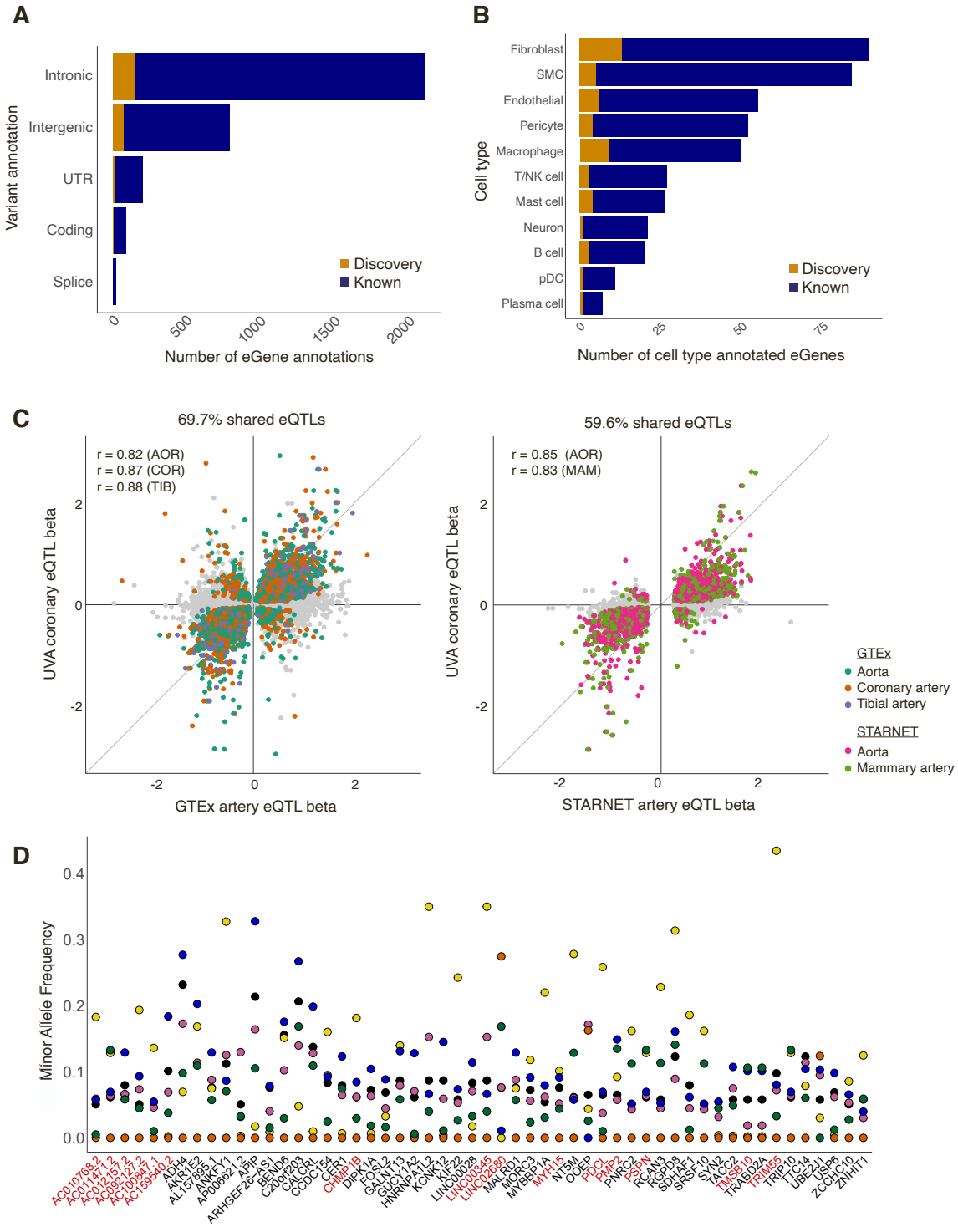

Figure S2. Annotation of mixQTL eGenes and shared effects in GTEx and STARNET artery tissues, Related to Figure 2. (A) SNPEff eQTL-eGene annotations for lead eQTLs. Navy blue and orange represent numbers of known and discovery eGenes, respectively. (B) Number of eGenes exhibiting cell-type-specific expression 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 (pBH<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 parenthe-ses. 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 an-cestry superpopulations. Ancestries shown are AFR (yellow), AMR (pink), EAS (orange), EUR (blue), and SAS (green).

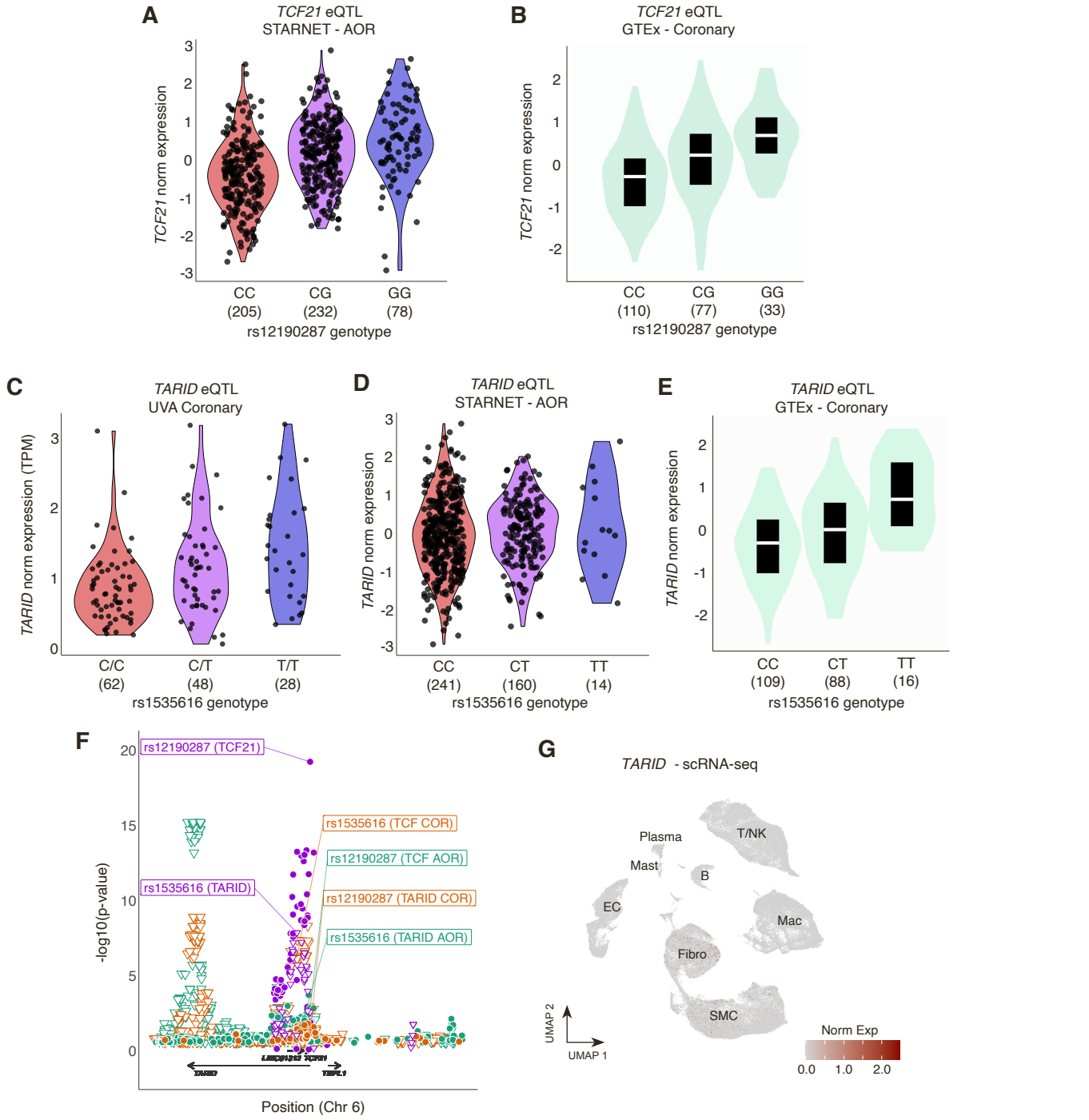

Figure S3. GWAS-eQTL colocalization eGenes in Europeans and supporting evidence for *TCF21/TARID* association, Related to Figure 3. (A,B) Violin plots for normalized expression of *TCF21* (y-axis) in STARNET aortic root (AOR) tissue (A) or GTEx coronary artery tissue (B) shown by genotype (x-axis) for 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.

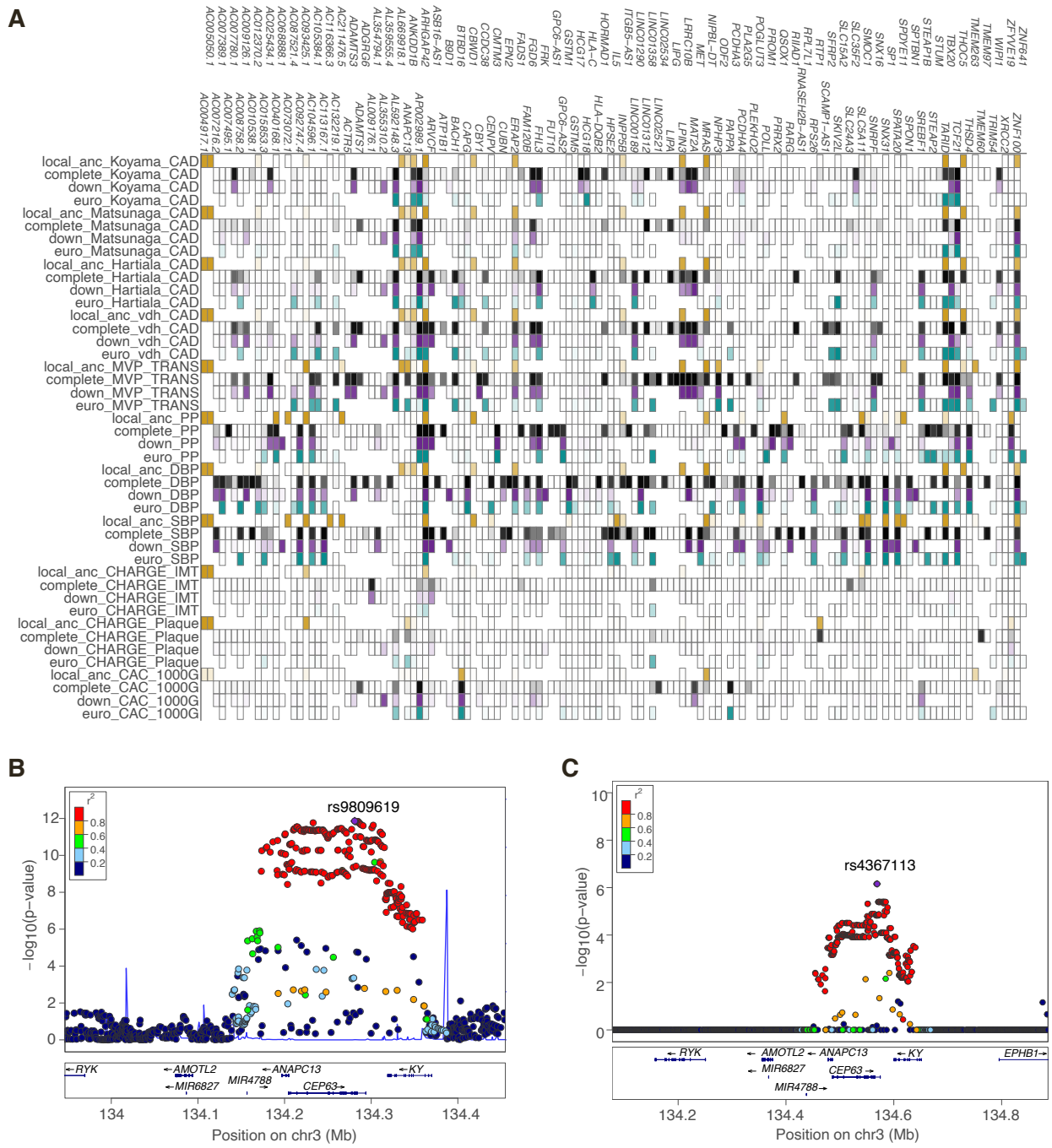

Figure S4. Comparison of colocalized associations between mixQTL, local ancestry-adjusted, and population-subsample coronary artery eQTLs, Related to STAR Methods. (A) Colocalization with CAD and BP GWAS traits for all eQTL methods. Heatmap indicates colocalization posterior probabilities (PPH4) for 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.

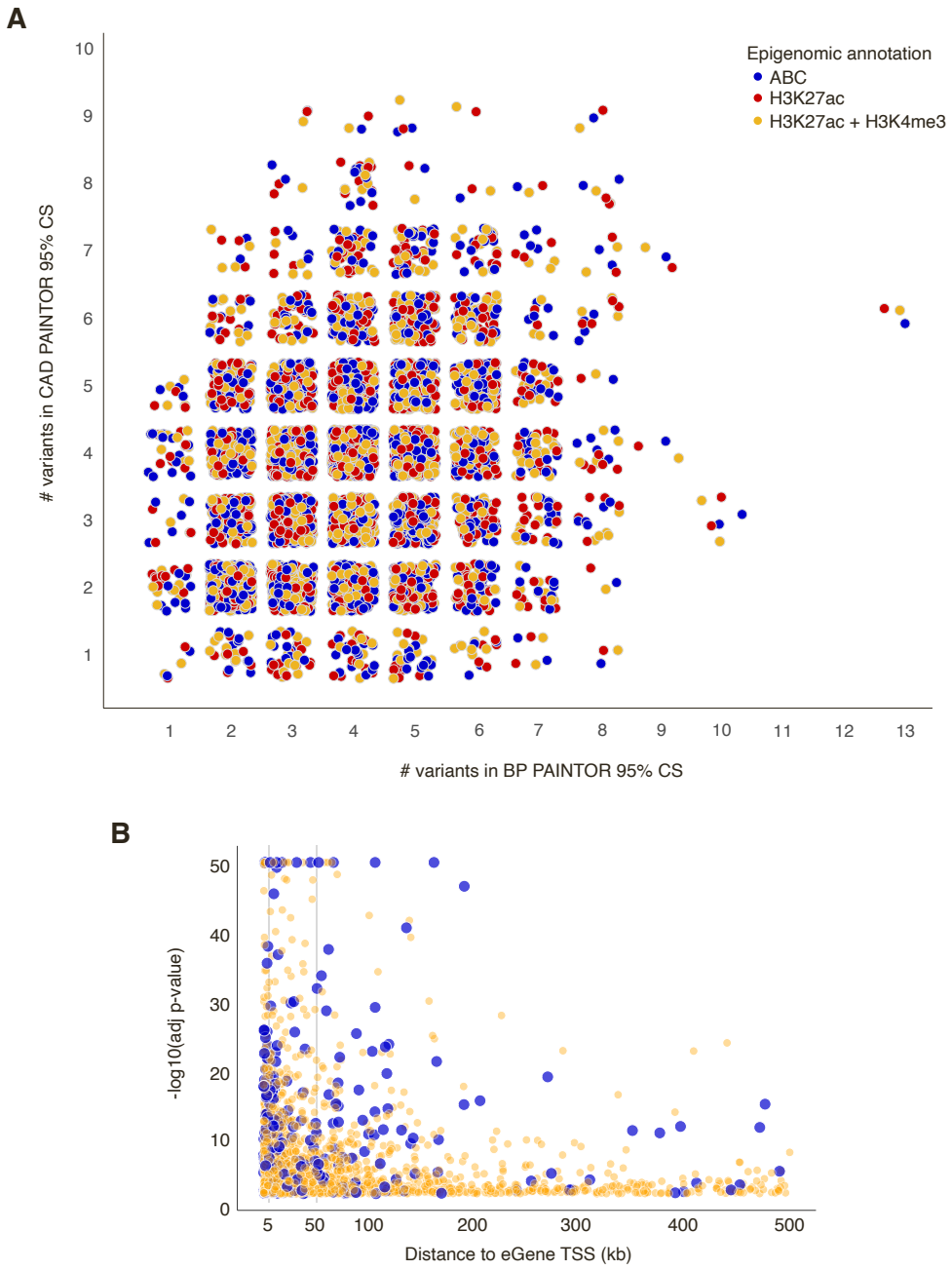

Figure S5. Fine-mapped credible set size and distance to TSS for CAD and BP trait GWAS, Related to STAR Methods. (A) Each point represents the number of variants in PAINTOR 95% credible set (CS) incorporating blood pressure trait GWAS data (x-axis) or CAD GWAS data (y-axis) for one UVA coronary artery 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 -log10(p-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.

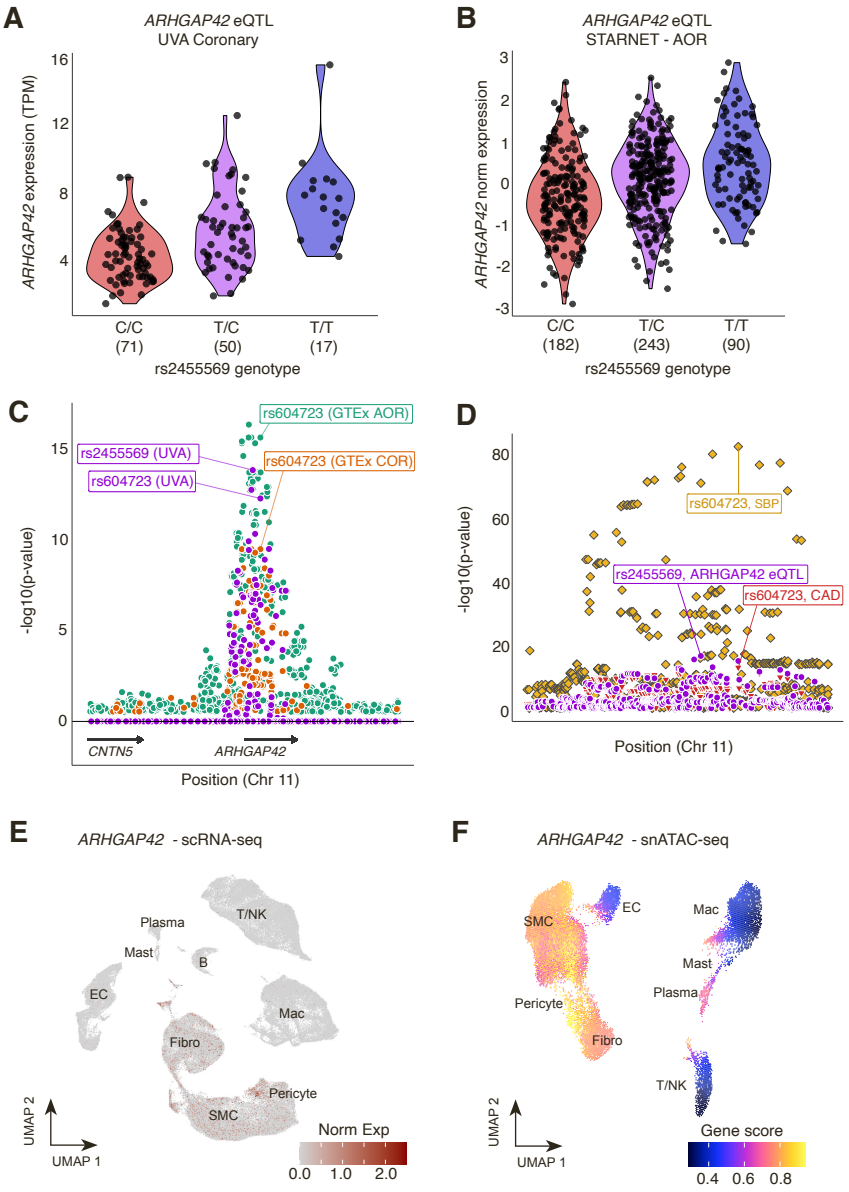

Figure S6. *ARHGAP42* eQTL and GWAS associations and cell type expression in human artery, Related to Figure 4. (A,B) Violin plots for normalized expression of *ARHGAP42* (y-axis) in UVA coronary artery tissue (A) and STARNET aortic root tissue (AOR) (B) shown by genotype (x-axis) for lead UVA eQTL rs2455569. (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 -log10(p-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.

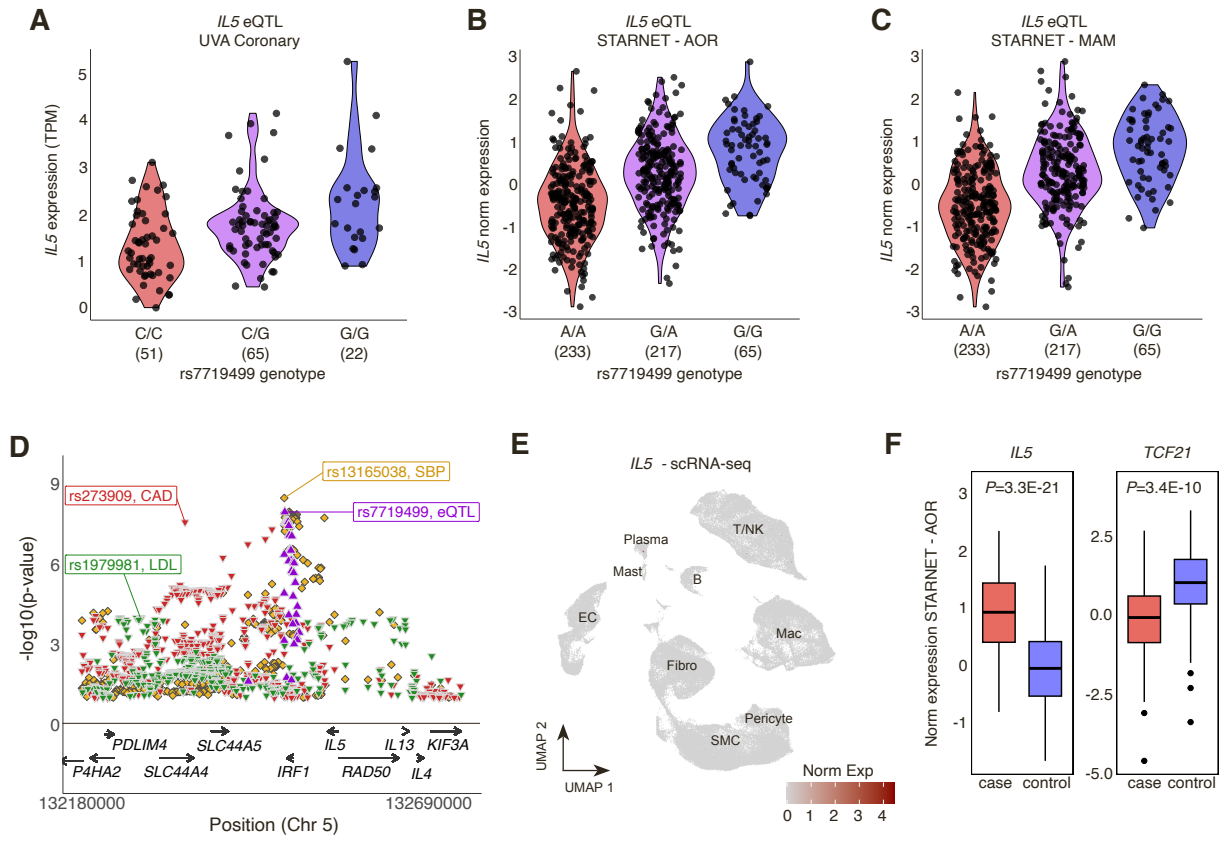

Figure S7. *IL5* eQTL associations, cell type and differential expression in human artery, Related to Figure 4. (A-C) Violin plots for normalized expression of *IL5* (y-axis) in UVA coronary artery tissue (a), STARNET aortic root tissue (AOR) (B), and STARNET mammary artery tissue (C) shown by genotype (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 -log10(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 * 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]

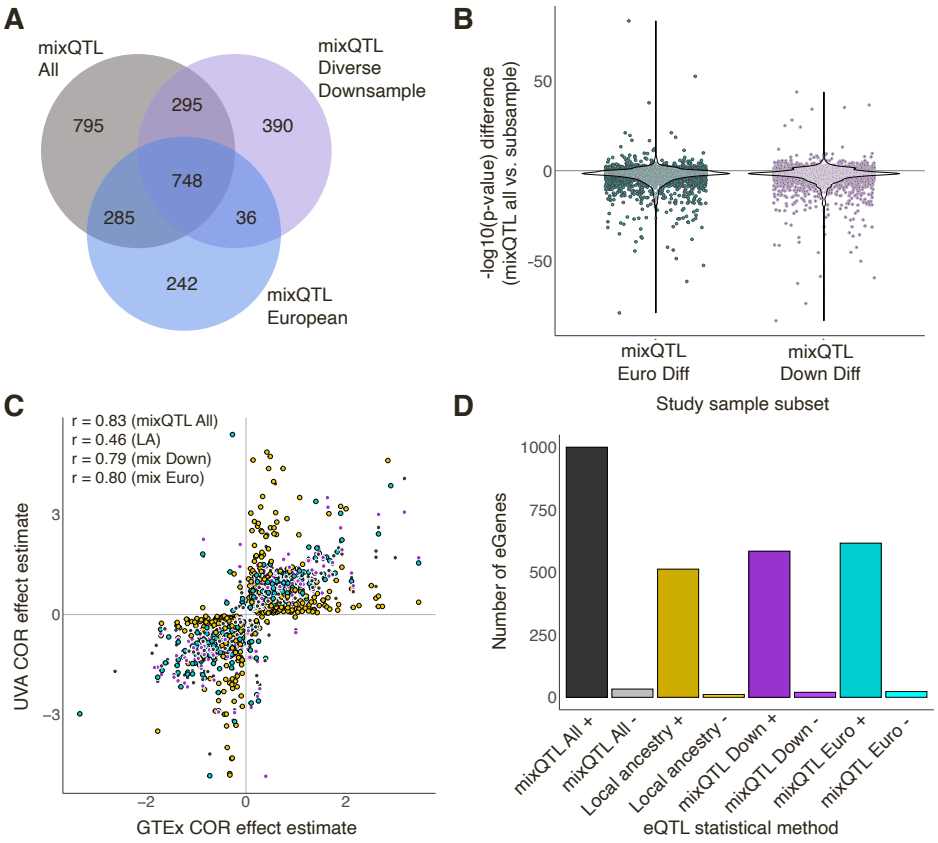

Figure S8. Sensitivity analyses of European-ancestry-only and proportionally representative down-sampled subsets compared to mixQTL and local-ancestry-adjusted results, Related to STAR Methods. In all panels, colors represent mixQTL total study population (dark gray), the subset including 80 individuals 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 -log10(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 sub-sample; 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 (pBH<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.

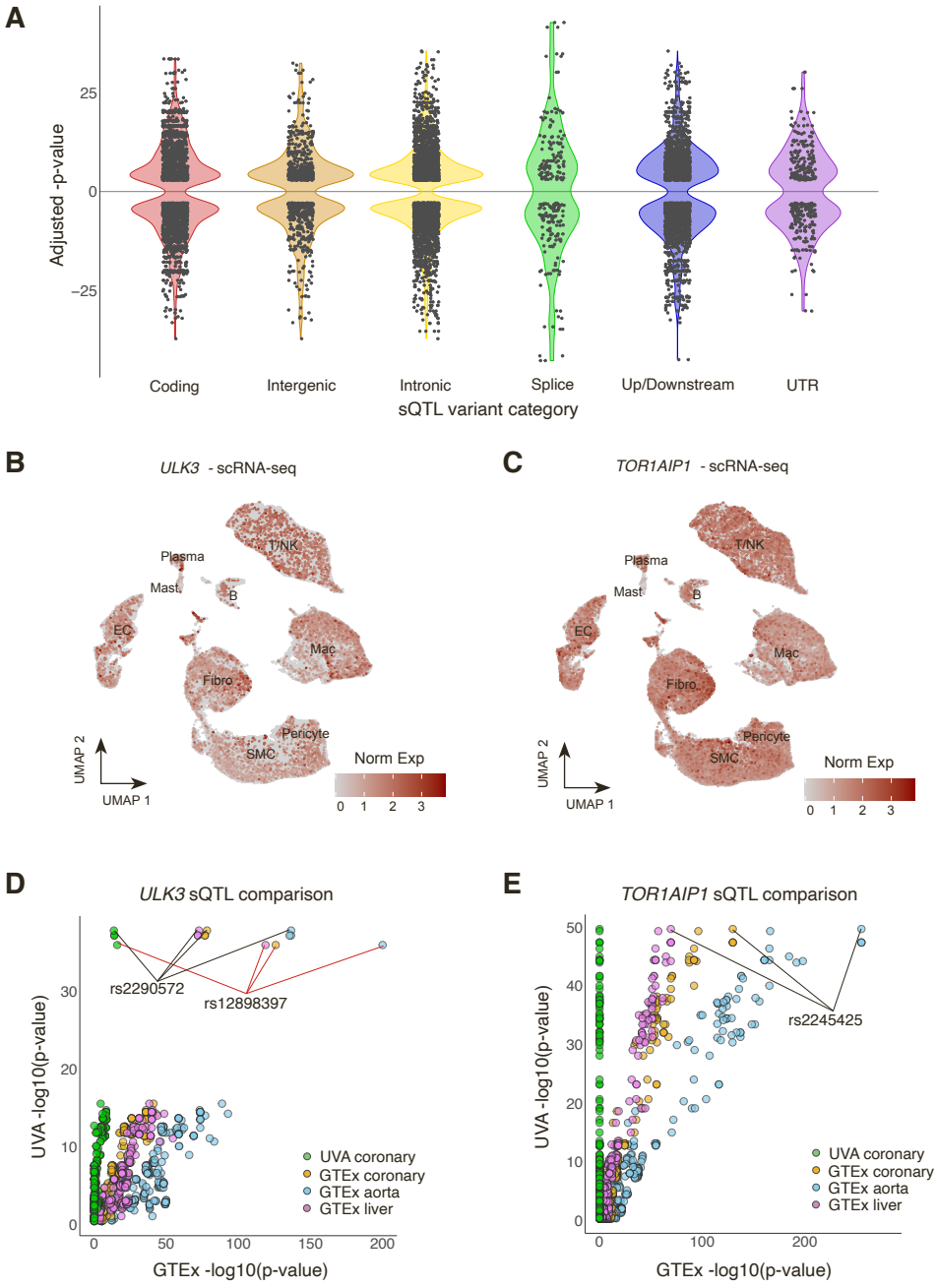

Figure S9. sQTL annotations highlight candidate sGenes *ULK3* and *TOR1AIP1,* Related to Figure 5. (A) Violin plot shows the adjusted significance incorporating direction of effect (y-axis) for lead sQTLs by SnpEff category of variant annotation (x-axis, p). (B,C) UMAP plots showing relative expression 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.