The American Journal of Human Genetics, Volume *106*

Supplemental Data

Systems Genetics in Human Endothelial Cells Identifies Non-coding Variants Modifying Enhancers, Expression, and Complex Disease Traits

Lindsey K. Stolze, Austin C. Conklin, Michael B. Whalen, Maykel López Rodríguez, Kadri Õunap, Ilakya Selvarajan, Anu Toropainen, Tiit Örd, Jin Li, Anna Eshghi, Alice E. Solomon, Yun Fang, Minna U. Kaikkonen, and Casey E. Romanoski

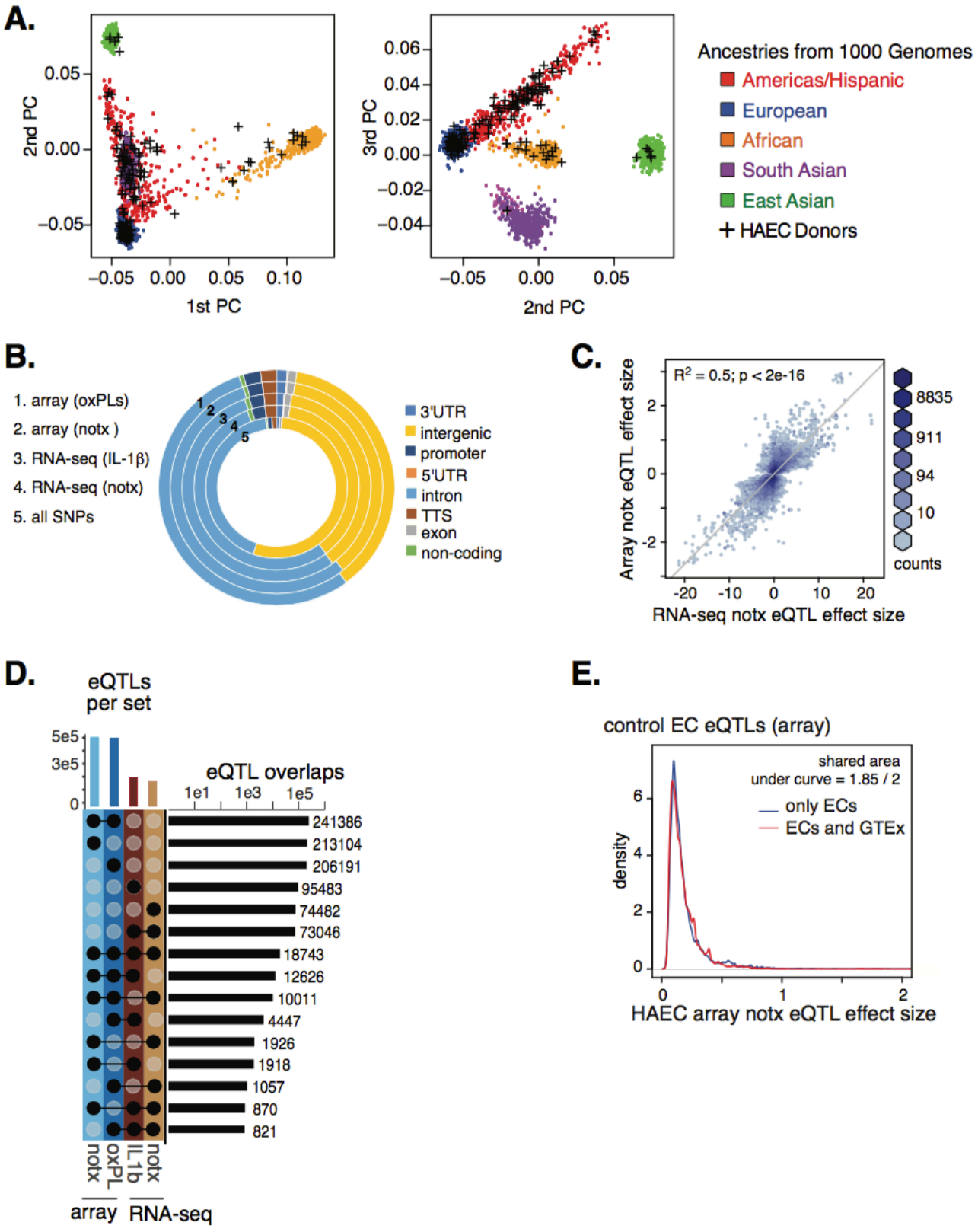

Figure S1. Characteristics of EC donors and EC eQTLs. A. Principal Component Analysis of genomic distances among individuals in 1000 genomes and HAEC donors in this study are depicted for

Principal Components (PCs) 1-3. Genotypes on chromosome 1 were used for this analysis. **B.** Genomic annotations of eQTL SNPs in HAEC datasets and all SNPs tested. Datasets indicated by number with key. **C.** Hexbin plot showing relationship between effect sizes of eQTLs discovered in array notx eQTLs (y-axis) and corresponding effect size in RNA-seq notx eQTL data (x-axis). Statistics are from linear regression. **D.** Upset plot showing overlap of eQTLs detected between datasets at 5% FDR. Along the right are the total eQTLs discovered and the number of unique transcripts associated in parentheses. **E.** Density of the HAEC eQTL effect sizes from notx array data (x-axis) for eQTLs unique to ECs (blue) and eQTLs discovered in GTEx and ECs (red).

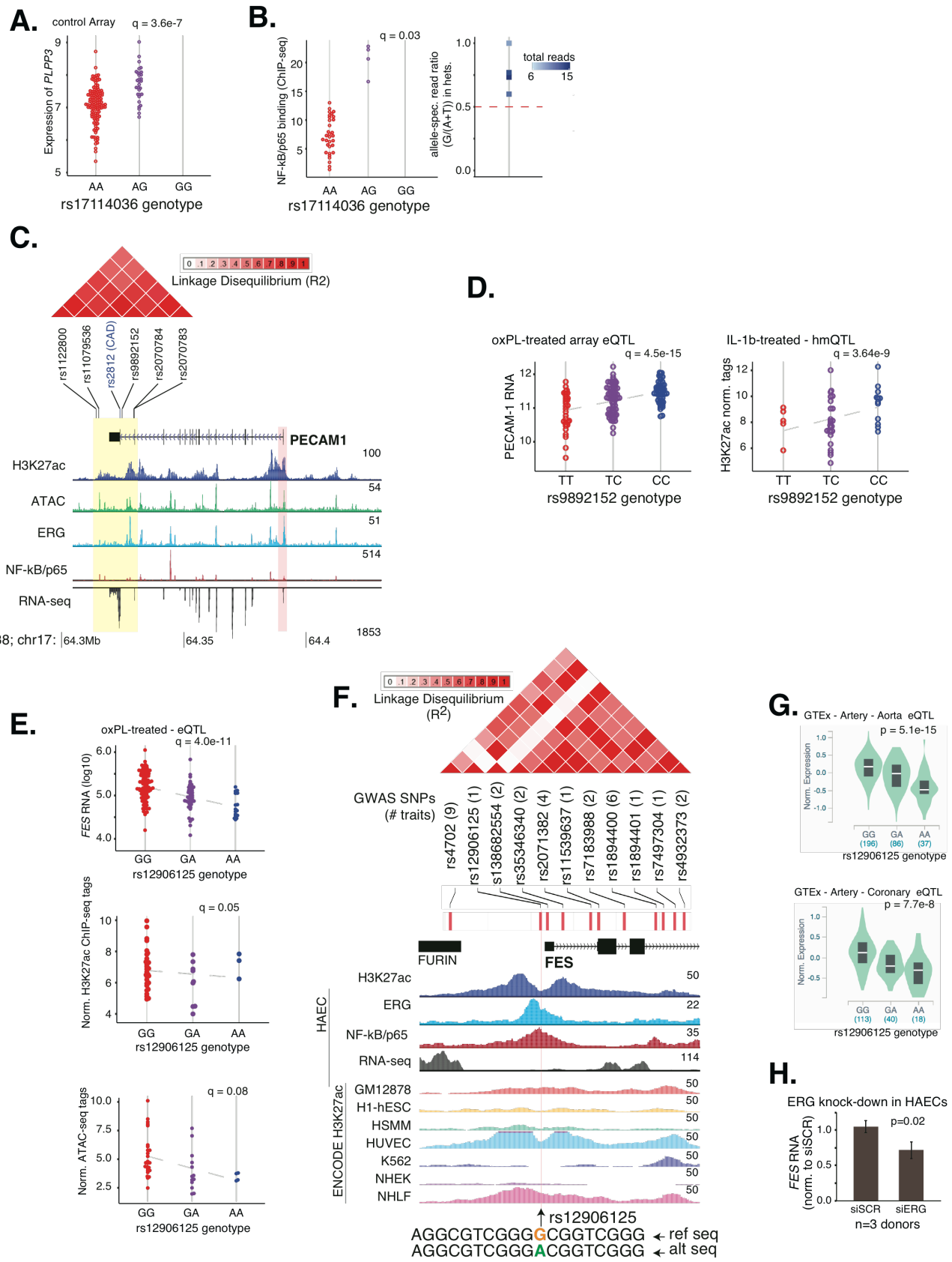

Figure S2. Effects of regulatory SNPs and target genes across multiple loci in ECs. A. eQTL for *PLPP3* at rs17114036 in the array notx dataset. **B.** NF-kB/p65 binding QTL for the enhancer region

over rs17114036 with allele-specific ratios in heterozygotes to right and results on *PLPP3* expression from CRISPRi targeting rs17114036. **C.** The *PECAM-1* locus with UCSC browser-style tracks from HAECs below. Above shows the LD structure for genome-wide significant CAD-associated GWAS SNPs. Yellow bar highlights region with association and enhancer marks having molQTLs; pink highlights *PECAM-1* promoter. **D.** The *PECAM-1* eQTL for rs9892152 in the oxPL-treated dataset, with corresponding hmQTL in IL-1b HAECs to right. **E.** The *FES* eQTL for rs12906125 in the oxPL dataset is shown above, with molQTLs at this SNP for H2K27ac and ATAC-seq below. **F.** A browser-style view zoomed to the FES promoter shows epigenetic data below from HAECs and ENCODE. Above is LD structure for GWAS SNPs. Numbers of associated GWAS traits shown in parentheses. Local sequence with alleles of promoter SNP rs12906125 is shown beneath. **G.** The *FES* eQTL at rs12906125 is shown in Aorta and Coronary Arteries from GTEx. **H.** *FES* RNA expression in ERG siRNA knock-down siRNA in three HAEC donors is shown. P-value is from 2-tailed unpaired t-test.

LEGENDS FOR TABLES S1-S7

Table S1. Sequence tag characteristics are summarized per HAEC donor. The number of unique mapped reads after removal of duplicates and mapping bias correction for each HAEC donor (rows), with summary data (at bottom) for each assay (columns). These values were used as co-variates in association mapping.

Table S2: Shared effect sizes between HAEC and GTEx eQTLs. Area under the curve for density of shared effect sizes between GTEx tissues (rows) and HAEC eQTL datasets (columns).

Table S3: Motifs enriched for mutation in allele-specific molecular HAEC traits. Motifs tested for Motif Mutation Analysis in rows, with corresponding summary statistics in columns for the molecular QTLs of ATAC-seq (notx and IL-1b), ERG binding, and NF-kB/p65 binding. Difference in medians was calculated by subtracting median distributions of i) sequencing counts on alleles where the indicated motif was in-tact versus, ii) counts from the allele that mutated the motif. P-values are from unpaired, 2-tailed t-test.

Table S4. 18 variants are associated with CAD, EC gene expression, and EC epigenetics. Each row represents a SNP for which there may be multiple gene expression traits associated.

Region gene(s) with association	CAD GWAS SNPs that share an eQTL and molQTL with Ecs	Posterior probability that EC eQTL (array notx) and CAD GWAS locus share the same causal variant
PPAP2B	rs17114036	0.998
FES	rs1894400, rs35346340, rs7497304	0.997
AXL	rs61459202	0.985
MLH3	rs175071, rs175065, rs4903284	0.945
NEK9	rs175071, rs175065, rs4903284	0.991
EIF2B2	rs175071, rs175065, rs4903284	0.835
BCKDHA	rs2241709, rs11670757, rs2241708	0.823
PECAM1	rs1122800, rs11079536, rs2812, rs2070784, rs2070783, rs9892152	0.974
LDLR	negative control: Doesn't have an eQTL in ECs. Does have CAD GWAS hits (rs6511720 and rs2738447).	0.072

Table S5. Posterior Probability from R package coloc, that the gene locus and CAD GWAS share underlying causal SNP(s)

Table S6. GWAS traits with enrichments in EC QTLs. Enrichment between co-mapped EC e/molQTLs and sets of GWAS SNPs. P-values were derived by Fisher's Exact Test.

Table S7. Top genes whose expression in ECs maps to an eQTL and molQTL, and is associated to multiple GWAS traits.

SUPPLEMENTAL METHODS

Cell Culture and Collection: Human aortic endothelial cells (HAECs) were isolated from de-identified deceased heart donor aortic trimmings at the University of California Los Angeles Hospital. The cells were isolated according to institutional guidelines as described previously[1]. These cells were cultured at passage 6 or less. The cells were cultured in M-199 (ThermoFisher Scientific, Waltham, MA, MT-10–060-CV) supplemented with 1.2% sodium pyruvate (ThermoFisher Scientific, Catalog# 11360070), 1% 100X Pen Strep Glutamine (ThermoFisher Scientific Cat# 10378016), 20% fetal bovine serum (FBS, GE Healthcare, Hyclone, Pittsburgh, PA), 1.6% Endothelial Cell Growth Serum (Corning, Corning, NY, Product #356006), 1.6% heparin, and 10 µL/50 mL Amphotericin B (ThermoFisher Scientific #15290018). Donor cells from up to 53 individuals were expanded at 5% CO2, at 37degC. Approximately 5 million cells were used per ChIP-seq assays, 500 thousand cells for RNA isolation, and 50 thousand cells for ATAC-seq. Cells were treated prior to harvest for 4 hours in media containing 1% FBS and either no additional protein, or with 10 ng/mL human recombinant IL-1B protein (R&D Systems Cat# 201-LB-005/CF).

RNA-seq, ChIP-seq, and ATAC-seq: Total RNA was extracted using the Zymo Quick-RNA MicroPrep (Zymo Research Cat# R1051) 1 ug of total RNA was submitted to polyA selection ahead of library construction using previously described methods. Cells used for ERG and H3K27ac ChIP-seq were cross-linked with 1% formaldehyde for 10 min at room temperature and then quenched with 2.65M Glycine (Fisher BioReagents Cat# BP381). ChIP-seq samples for p65/NF-kb were additionally fixed with Disuccinimidyl glutarate (DSG ; ProteoChem Cat# c1104) for 30 min. Chromatin was sheered using the Bioruptor Pico (Diagenode) with 8- 30s on/60s off cycles. IPs were performed using antibodies conjugated to Protein A/G Dynabeads (Invitrogen Cat# 10002D; Invitrogen Cat# 10004D) with the following antibodies: EPR3863 (abcam #ab110639) for ERG, NFkB p65 (c-20)x (Santa Cruz Antibodies sc-372x, Lot #E0916) for p65, and Histone H3K27ac (Active Motif #39135) for H3K27ac. All other details of ChIP were described previously[2]. ATAC-seq was performed according to the originally published protocol[3] with the exception of size selection from 175-225 bp prior to sequencing to enrich for regulatory elements. Sequence libraries were prepared as previously described[2] and sent to The University of Chicago's Genetics core for sequencing on an Illumina HiSeq 4000 for single-end 50bp reads. Sequencing depth is summarized in **Table S1**.

Microarray data was provided from previous publications[4; 5] utilizing HAECs a subset of which was used for the previously mentioned RNA-seq, ATAC-seq, and ChIP-seq. Microarrays were used to examine gene expression in 157 and 156 donors in untreated and Oxidized 1-palmitoyl-2-arachidonoyl-sn-glycero-3-phosphocholine (oxPAPC) treated conditions respectively. Data is in GEO Accessions GSE30169 and GSE139377.

Mapping and Processing: The sequencing data mapped utilizing the mapping bias correction and duplicate read removal from software package WASP[6], and mapping software Bowtie2[7] with default mapping parameters. This process involved mapping the sequencing data to the reference genome first, then remapping the sequencing data to a version of the genome with the alternate alleles at sites of common genetic variation. If the read did not align to both versions of the genome, the read was removed from further analysis. The correction resulted in a mapped bam file for use in allele specific analysis. The resulting sequencing data have an average of 11,390,742 unique reads per sample. (Table S1). HOMERs preferred file type (tag directories) were made using the command makeTagDirectory from software suite HOMER[8]. These tag directories were used in expression and peak file creation (see **RNA-seq QTL analysis** and **ChIP-seq and ATAC-seq QTL analysis**).

Genotyping: Genomic DNA was isolated from HAECs with the DNeasy extraction kit including optional DNase treatment (QIAGEN) and quantified with NanoDrop 2000 (Thermo Fisher Scientific, Waltham, MA). All samples were randomly arrayed into three 96-well microtiter plates at 50 ng/ul. Per Affymetrix Genome wide Human SNP Array 6.0 assay protocol, 2 x 250 ng of gDNA were digested by restriction enzymes NspI and StyI separately and products were ligated to respective adaptors (Affymetrix Human SNP 6.0 assay). PCR was used for amplifying ligation products and checked for size and quality by QIAxcel (QIAGEN). Labeled PCR products were hybridized to the Human SNP 6.0 array. Array hybridization, washing and scanning were performed according to the Affymetrix recommendations. Scanned images were subjected to visual inspection and a chip quality report was generated by the Affymetrix GeneChip Operating System (command console) and the Genotyping console (Affymetrix). The image data was processed as described previously[5] with the Affymetrix Genotyping Console or Birdsuite algorithm24 for determining the specific hybridizing signal for each SNP call and copy-number detection.

Imputation: HAEC genotypes were used in conjunction with reference genomic sequence from 1000 Genomes Project Data to impute missing genotypes. Genotype and haplotype data were downloaded from 1000 Genomes Phase 3 data (1000G)[9] (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/; accessed 04/2017). Tri-allelic and very rare variants were removed from the reference panel using plink1.9 '--biallelic-only strict and --maf 0.0000000001' options[10]. SHAPEIT [11] was used to align reference alleles between HAEC and 1000G files and to pre-phase HAEC alleles prior to imputation. We included all populations in the 1000G reference panel and used the 'k_hap 600 -Ne 20000' options for imputation using IMPUTE2[12; 13] as to improve representation of shared haplotypes across recently admixed populations. Genotypes were called for imputed SNPs with allelic R2 values greater than 0.9.

SNPs used in association analysis were included if they met the following criteria: Autosomal SNPs; Minor allele frequency (MAF) greater than 5%; and no missing genotypes across donors. SNP filtering was performed with vcftools[14] and PLINK[10]. Sex was determined from heterozygous genotype calls on the X chromosome using PLINK.

Quality Control of Sequencing Samples: Samples were removed from further analysis if: the sample contained less than 3 million unique reads, the sample had an average of 6 or more duplicate tags per site, the sample was an extreme outlier in PCA, or if the sample could not replicate the genotype identity on file for that individual. Genotype identity was assigned by calling genotypes from sequencing reads using samtools[15] function “mpileup” followed by the bcftools function “call” and comparing the results to the SNP Chip data for each individual.

VCF file Preparation: To reduce multiple testing as well as avoid false positives, VCF files were restricted to only include genetic variants with a minor allele frequency of at least 0.05 using vcftools (option --maf 0.05)[14]. This was performed for each set individually to account for the differences in donor presence due to the quality control measures from above. Therefore, each data set was analyzed using SNPs with sufficient variance within the individuals contained in the data set.

Covariate Discovery: Since the HAECs were de-identified, biological sex and ethnicity were determined from the genotyping data. Ethnicity was determined by comparing genotype data to 1000 Genomes data by PCA analysis. Biological sex was determined by heterozygosity on the X chromosome using PLINK[10].

RNA-seq eQTL analysis: Expression QTL analysis was done using linear regression via the R package MatrixeQTL[16] (options “pvOutputThreshold.cis = 1”, “cisDist = 1e6”, “verbose = TRUE”, “min.pv.by.genesnp = FALSE”, “noFDRsaveMemory = FALSE”) . To obtain the rpkm normalized expression matrix used in testing, the command analyzeRepeats from software package HOMER[8] was used (options “-rpkm”, “-tbp 1”, “-count exons”, “-condenseGenes”, and “-strand +”). MatrixeQTL was run using the rpkm normalized expression matrix, a VCF file using the number code for genotypes (Homozygous reference = 0, Heterozygous = 1, Homozygous alternate = 2), and a covariate file containing biological sex, unique total tag counts, the top four PCs from a PCA performed on the genotypes to account for ancestry, and fifteen factors discovered by PEER[17]. The number of hidden factors discovered by PEER was determined by the version 7 GTEx protocols (15 factors for sample sizes <150). To be consistent with GTEx, SNPs were tested against a gene’s expression if they were within 1Mb of the start or end of the gene.

Microarray gene expression: Cytoplasmic RNA was extracted with the RNeasy kit including optional DNase treatment (QIAGEN). RNA concentrations were measured with the NanoDrop 2000 (Thermo Fisher Scientific) and quality checked with the Agilent 2100 Bioanalyzer (Agilent) so that RNA Integrity numbers were greater than 8.5. RNA was prepared for hybridization to Affymetrix HT-HU133A microarrays using the standard protocol from the manufacturer. Intensity values were normalized using the robust multi-array average (RMA)[18; 19] normalization method in R 2.5.0 with the justRMA() function of the affy package of Bioconductor[20]. We utilized an alternative CDF file that excluded misaligned probes that were artifacts of the previous transcriptome build that was used for creating the publicly available Affymetrix CDF file. To create an updated CDF, we used the custom CDF created by Zhang J et. al.[21] that was created for the Affymetrix U133A array. Because of the differences in the underlying physical location of probes between the U133A and HT-U133A arrays, we created a probe-to-probe map between arrays and updated the alternate CDF file to the HT-U133A format.

Microarray QTL analysis: Normalized microarray values were obtained as described above[4], and run through PEER[22] using the following known covariates: culture and treatment batch, sex, and the first four principal components from principal component analysis on genotypes. Expression QTL analysis was performed using the R package MatrixeQTL (options pvOutputThreshold.cis = 1, cisDist = 1e6, verbose = TRUE, min.pv.by.genesnp = FALSE, noFDRsaveMemory = FALSE). MatrixeQTL was run using the normalized expression matrix, a VCF file using the number code for genotypes (Homozygous reference = 0, Heterozygous = 1, Homozygous alternate = 2), and a covariate file containing biological sex, unique total tag counts, the top four PCs from a PCA performed on the genotypes to account for ancestry, and thirty factors discovered by PEER. The number of hidden factors discovered by PEER was determined by the version 7 GTEx protocols (30 factors for sample sizes >150). SNPs were tested against a gene's expression if they were within 1Mb of the start or end of the gene.

To determine the similarity between the eQTL results from the RNA-seq data sets and the Microarray datasets, each eQTL dataset was compared to all other datasets individually. The effect sizes of significant SNPs in one set were graphically compared to the effect sizes of the same SNPs in the second data set regardless of significance in this second set. The R package "hexbin" was used to plot this comparison.

Multiple correction for eQTL Analysis: Results were then restricted to a gene-level Benjamini-Hochberg (correction for all tests done for a single transcript) adjusted p-value of less than 0.05. For the majority of analyses performed in this paper, we used the gene-level correction. However, for comparison to GTEx, additional correction was done. The lowest adjusted pvalue at each gene was selected to 'represent' the gene in a second Benjamini-Hochberg correction. The resulting adjusted

pvalue was restricted to 0.05, and the genes remaining were considered the eGenes for this analysis. SNP-Gene associations were only kept if they 1) were associations with an eGene by the secondary correction, and 2) if their gene-level adjusted pvalue was less than or equal to the most significant gene-level pvalue of the least significant eGene.

Lead SNP determination: To report the number of lead SNPs present in each eQTL dataset, a combination of custom code in R, vcftools, and PLINK was used to LD prune the results. The following was done on each eQTL dataset separately. Custom R code was used to create SNP lists containing all SNPs associated with a single gene. The program vcftools [14] was used to filter the vcf file to the SNPs in the SNP list for each gene (option --snps). The program PLINK was then used to LD prune on an R2 of 0.8 for each gene's associations (option "--indep-pairwise 2000000 1 0.6"), keeping the SNP with the highest significance in each LD block.

ChIP-seq and ATAC-seq QTL analysis: Molecular QTL analysis was done using an allele specific method via software package RASQUAL[23]. To obtain allele specific data, the function createASVCF.sh from RASQUAL was used. This process utilized the mapping bias corrected bam files for each data set and counted the number of tags which contained each allele at a given genetic variant on an individual basis. This information was added into the VCF file for each set to be used by RASQUAL for allele specific QTL testing. Non-allele specific tag counts for each region in the genome were obtained and RPKM normalized using commands findpeaks and annotatePeaks from software package HOMER[24] to create an individual by peak matrix. The options used for findpeaks for the transcription factors was "-style factor" and "-o auto". The options used in findpeaks for H3K27ac were "-style histone" and "-o auto". The options used in findpeaks for ATAC-seq were "-style histone", "-L 8", "-F 8", "-size 75", "-minDist 75", minTagThreshold 6", and "-o auto". All annotatePeaks commands were run with the option "-rpkm". The QTL analysis portion of RASQUAL[23] was run using the allele-specific VCF files, RPKM normalized tag matrix, and the covariates: sex, unique total tag counts, and the four top principal components from a PCA run on the genotypes to adjust for ancestry. SNPs were tested against a peak's intensity if they were found within the boundaries of the peak. The results were filtered using a per site false discovery rate of 0.05.

Enrichment Analysis: To determine enrichment of a QTL dataset within either a GWAS or eQTL dataset, we took the unique significant SNPs of a test QTL dataset (FDR <0.05) and pulled the corresponding p-values (GWAS data) or FDRs (QTL analyses) from the target dataset. These values were binned by cumulative cut off points (1e-8, 1e-7, 1e-6, 1e-5, 1e-4, 1e-3, 1e-2, 0.1, 1). The number of test dataset SNPs in each cumulative bin were then compared to the number of SNPs that would be

found in that bin by random chance. This was determined by randomly taking a subsample of the same size as the significant test dataset from the full list of SNPs tested in the QTL analysis of the test dataset. To get the 'enrichment score' used in the paper, the number of SNPs found experimentally was divided by the average number of SNPs found by random sampling.

Comparison to GWAS Catalog Data: To find diseases and traits that are enriched for EC QTLs, the summary statistics for all lead SNPs of all the diseases and traits available on the GWAS catalog[25] (as of August 2019). Four datasets were created from this study for comparison to the GWAS Catalog data: Untreated RNA-seq eQTLs with any molQTL, IL1B treated RNA-seq eQTLs with any molQTL, Untreated Microarray eQTLs with any molQTL, and oxPL treated Microarray eQTLs with any molQTL. Enrichment in a trait was tested via a Fishers Exact test performed in R. This was done by creating a two by two contingency table (in R: matrix(c(**x,y,z,a**),2,2)) where **x** is the number of overlaps between the eQTL dataset and the GWAS trait of interest, **y** is the number of overlaps between the SNPs tested in the eQTL dataset (that were not significant eQTLs) and the GWAS trait of interest, **z** is the number of eQTL SNPs that do not overlap with the GWAS trait of interest, and **a** is the number of SNPs tested in the eQTL dataset (that were not significant eQTLs) that do not overlap with the GWAS trait of interest. This two by two matrix was input into the function fisher.test() in R. Enrichment was concluded if there was significant overlap (pval<0.05) by this test.

Enrichment of specific genes in multiple traits was tested by assessing the frequency at which a gene's expression was seen associated across multiple diseases/traits by a HAEC eQTL.

Candidate SNP-gene pair selection: The SNP-gene pairs that eventually became our candidate SNPs for functionality with CAD were selected via the fine-mapping with molQTLs. The eQTLs that also were significant molQTLs were compared against the genome-wide significant CAD associated SNPs[26] resulting in the 18 SNPs seen in **Table S4**.

Annotation and SNP localization Analysis: To determine the annotated localization of eQTLs in the genome in comparison to all tested SNPs, we used HOMER command "annotatePeaks.pl" to assign an annotation to location of the SNP of "intergenic", "coding", "non-coding", "intronic", "exonic", "Transcription Start Site" (TSS), or "Transcription Termination Site" (TTS). The proportions of each of the annotations found in each of the eQTL datasets and the total tested SNPs were compared via circular bar graph in Microsoft Excel.

To determine the location of eQTLs with respect to TSSs, we used HOMER command "annotatePeaks.pl" using options "-hist 1000" and "-size 1000000". The data sets were only compared

to gene TSSs of genes that had at least one associated significant eQTL within that dataset. The total tested SNPs were compared to all genes.

Motif Mutation Analysis: Motif mutations were detected when the local sequence altered by alleles of a SNP such that one allele dropped the match to the motifs position weight matrix (PWM) below the motif detection threshold that is defined in the HOMER motif database[8]. The analysis pipeline has been described previously[2].

GTEx analysis: To compare GTEx eQTLs to the EC eQTLs, EC RefSeq IDs were converted to Ensembl IDs using Biomart. Next, variant IDs tested in both GTEx and EC eQTL studies were overlapped on chromosome, position, and nucleotide. Any variants tested in only one study were discarded, and then EC variant IDs were converted to match those in GTEx. Finally, intersection of datasets was performed using unique eQTL/eGene pairs as identifiers.

The distribution comparisons were compared using a shared area under curve metric. To calculate this, we performed kernel density estimation in R using the density function for both of the datasets to be compared. Following this, the area between curves for the two density estimates was calculated. The distance metric is equal to 2 - area between curves.

Graphical Packages: R was used for the visualization of results using packages: “graphics”, “ggbeeswarm”, “beeswarm”, “hexbin”, and “UpSetR”. Additional images were used from UCSC Genome Browser[27] and the 3D Genome Browser[28].

Dual Luciferase Reporter Assay: For the dual luciferase reporter assay, 198 bp fragments of the enhancer regions, ordered from Agilent Technologies, were cloned into ClaI and SalI sites of Addgene plasmid #99297[29]. The integrity of plasmids was verified by Sanger sequencing. The control vector or the luciferase constructs were co-transfected with the pGL4.75 vector (Promega) that encodes the luciferase gene hRluc (*Renilla reniformis*) in TeloHAEC cells using Lipofectamine Stem Transfection Reagent (ThermoFisher Scientific) according to manufacturer instructions. A total of 500 ng of DNA per well in 24 well plates was transfected, keeping a 10:1 molar ratio between the constructs or the vector with respect the pGL4.75 plasmid. The molar ratio between the target constructs and the control was 1:1. Luciferase activity was measured 48 h post transfection with a Dual-luciferase Reporter Assay (Promega) in a CLARIOstar (BMG Labtech) plate reader coupled with a dual injector system. The firefly luciferase signal was normalized to *Renilla* signal and the data is presented proportional to the control vector. Three independent experiments with four technical replicates were performed. Intra-haplotype or haplotype-control statistical analyses were performed with two-tailed *t*-test.

CRISPR Interference (CRISPRi):

CRISPRi experiments were conducted in HAECs to determine the enhancer activity of *cis*-regulatory elements of interest. A fusion protein of catalytically dead Cas9 (dCas9) fused to KRAB repressor protein (addGene cat#46911) was expressed in HAEC using transfection of *in vitro* transcripts. In a 24-well plate, 25 ng of dCas9-KRAB *in vitro* transcripts and 3 pmol of sgRNA were diluted in 25 µl of opti-MEM. Next, 0.75 µl of Lipofectamine Messenger MAX (Invitrogen) was diluted into 25 µl of opti-MEM. Each dilution pool was combined, incubated for 10 minutes at RT, and added to cells. After 8 hours at 37degC, cells were lysed for RNA collection, cDNA synthesis, and analysis via qPCR. Non-targeting control guide RNA was purchased from IDT. Guide RNA sequence targeted to a previously-identified endothelial enhancer in PLPP3 rs17114036 locus was used as a positive control (5'-GTTGATATCACTAAGTTTTCAGG- 3', 5' -CAAGAGCTGAAGTCAGGCAGTGG- 3'). Guide RNA sequences for targeted loci of interest are listed here: (PPAP2B-rs17114036-sgRNA-1: 5'-GTTGATATCACTAAGTTTTCAGG-3';
PPAP2B-rs17114036-sgRNA-2: 5'-CAAGAGCTGAAGTCAGGCAGTGG-3';
FGD6_intronic_enhancer_sg1: TATTCTGAGCCCCTTTACCA;
FGD6_intronic_enhancer_sg2: TGGAATCTGCAGTCCTATAA;
VEGFC_upstream_enhancer_sg1: TGCGAGATGCACACATTCCC;
VEGFC_upstream_enhancer_sg2: ACCCTAAACACCCATAATGA;
KIF26B_upstream_enhancer_sg1: GGAGTGATAACTCCTATTGT;
KIF26B_upstream_enhancer_sg2: TTCTTATAACGGGAAAGTGT). qPCR primers: (FGD6_F2 5'-CTGTTCGAGAGATTGGGCAGT-3',
FGD6_R2 5'-TCATTGCTCTGATTGCCTTCAT-3',
VEGFC_F2 5'-GAGGAGCAGTTACGGTCTGTG-3',
VEGFC_R2 5'-TCCTTTCCTTAGCTGACACTTGT-3',
KIF26B_F2 5'-TTCTCGGCTGTGATTCACGAC-3',
KIF26B_R2 5'-AGGTGAGTGGCGCAAATGT-3', PPAP2B_F2 5'-AAGTCCAGGAAGCCAGGAAGT-3',
PPAP2B_R2 5'-GACAGTCCCGTGTAGAAGGC-3').

SUPPLEMENTAL REFERENCES

1. Navab, M., Hough, G.P., Stevenson, L.W., Drinkwater, D.C., Laks, H., and Fogelman, A.M. (1988). Monocyte migration into the subendothelial space of a coculture of adult human aortic endothelial and smooth muscle cells. J Clin Invest 82, 1853-1863.
2. Hogan, N.T., Whalen, M.B., Stolze, L.K., Hadeli, N.K., Lam, M.T., Springstead, J.R., Glass, C.K., and Romanoski, C.E. (2017). Transcriptional networks specifying homeostatic and inflammatory programs of gene expression in human aortic endothelial cells. Elife 6.
3. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods 10, 1213-1218.
4. Romanoski, C.E., Che, N., Yin, F., Mai, N., Pouldar, D., Civelek, M., Pan, C., Lee, S., Vakili, L., Yang, W.P., et al. (2011). Network for activation of human endothelial cells by oxidized phospholipids: a critical role of heme oxygenase 1. Circ Res 109, e27-41.
5. Romanoski, C.E., Lee, S., Kim, M.J., Ingram-Drake, L., Plaisier, C.L., Yordanova, R., Tilford, C., Guan, B., He, A., Gargalovic, P.S., et al. (2010). Systems genetics analysis of gene-by-environment interactions in human cells. Am J Hum Genet 86, 399-410.
6. 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.
7. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357-359.
8. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576-589.
9. Genomes Project, C., 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., et al. (2015). A global reference for human genetic variation. Nature 526, 68-74.
10. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-575.
11. Delaneau, O., Marchini, J., and Zagury, J.F. (2011). A linear complexity phasing method for thousands of genomes. Nat Methods 9, 179-181.
12. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat Genet 44, 955-959.
13. Howie, B., Marchini, J., and Stephens, M. (2011). Genotype imputation with thousands of genomes. G3 (Bethesda) 1, 457-470.
14. 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.
15. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078-2079.
16. Shabalin, A.A. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 28, 1353-1358.

17. 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. Nature protocols 7, 500-507.
18. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31, e15.
19. Bolstad, B.M., Irizarry, R.A., Astrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 19, 185-193.
20. Gautier, L., Cope, L., Bolstad, B.M., and Irizarry, R.A. (2004). affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20, 307-315.
21. Zhang, J., Finney, R.P., Clifford, R.J., Derr, L.K., and Buetow, K.H. (2005). Detecting false expression signals in high-density oligonucleotide arrays by an in silico approach. Genomics 85, 297-308.
22. 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.
23. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat Genet 48, 206-213.
24. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38, 576-589.
25. Buniello, A., MacArthur, J.A.L., Cerezo, M., Harris, L.W., Hayhurst, J., Malangone, C., McMahon, A., Morales, J., Mountjoy, E., Sollis, E., et al. (2019). The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res 47, D1005-D1012.
26. 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.
27. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Res 12, 996-1006.
28. Wang, Y., Song, F., Zhang, B., Zhang, L., Xu, J., Kuang, D., Li, D., Choudhary, M.N.K., Li, Y., Hu, M., et al. (2018). The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 19, 151.
29. Muerdter, F., Boryn, L.M., Woodfin, A.R., Neumayr, C., Rath, M., Zabidi, M.A., Pagani, M., Haberle, V., Kazmar, T., Catarino, R.R., et al. (2018). Resolving systematic errors in widely used enhancer activity assays in human cells. Nat Methods 15, 141-149.