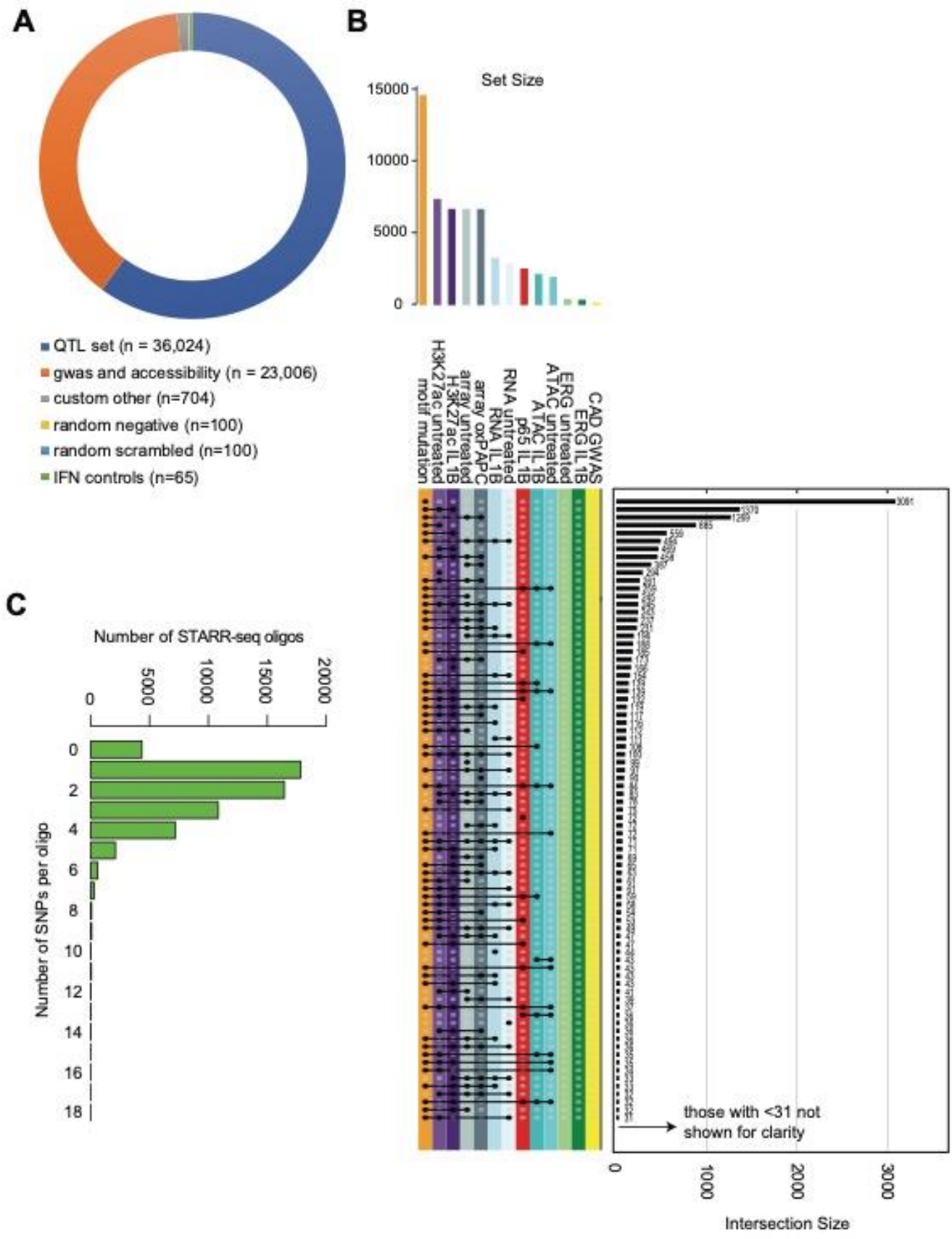
Supplementary Figures S1-S12 for:

Functional non-coding SNPs in human endothelial cells fine-map vascular trait associations

Anu Toropainen[1]*, Lindsey K. Stolze[2,3]*, Tiit Örd[1]*, Michael Whalen[2], Paula Martí Torrell[1], Verena M. Link[4], Minna U Kaikkonen[1#], & Casey Romanoski[2,3#]

1) A. I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio 70211, Finland. 2) The Department of Cellular and Molecular Medicine, The University of Arizona, Tucson, AZ 85721, USA. 3) The Genetics Interdisciplinary Graduate Program, The University of Arizona, Tucson, AZ, 85721, USA 4) Metaorganism Immunity Section, Laboratory of Host Immunity and Microbiome, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD 20892, USA.
* co-first/equal contribution # co-last/corresponding: cromanoski@arizona.edu and minna.kaikkonen@uef.fi

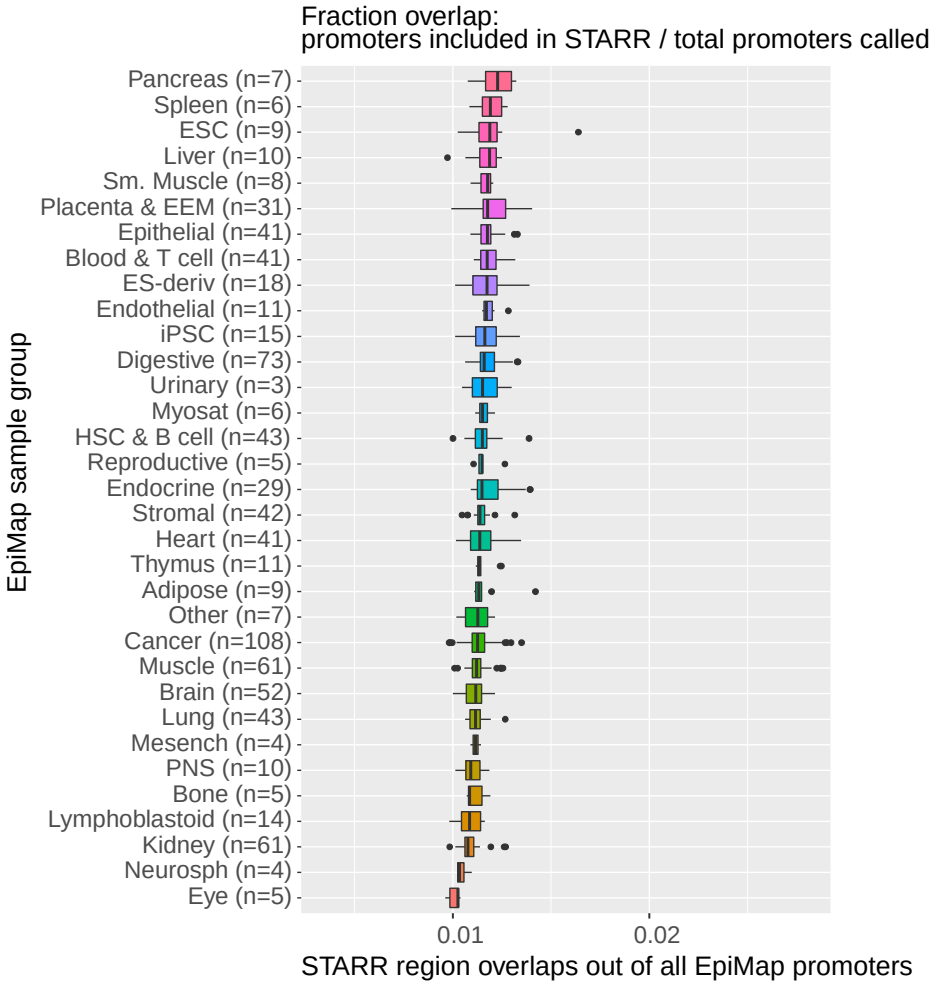


Supplemental Figure S1. Contents of STARR-seq library

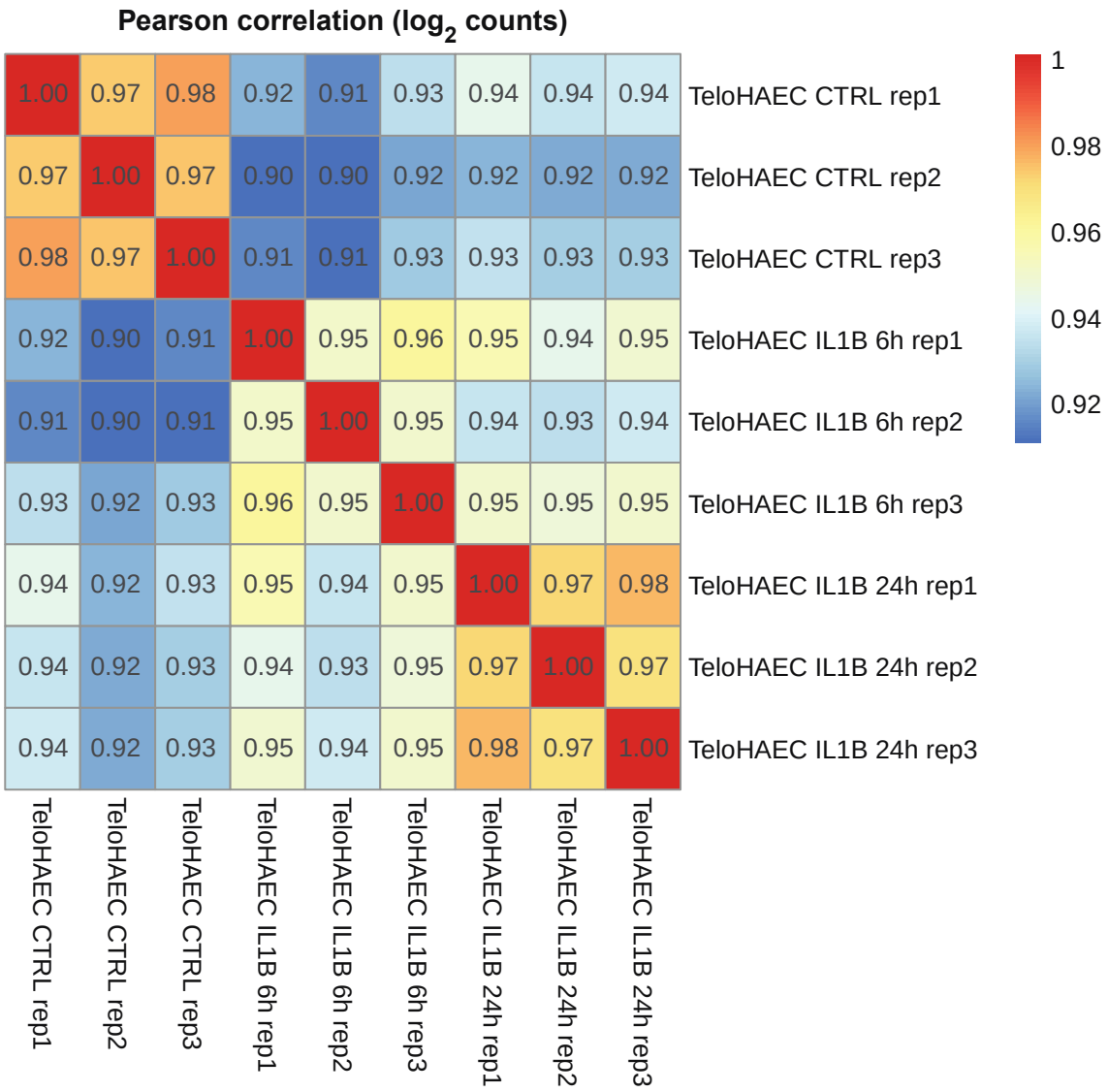
A) Annotations of the regions included in the STARR-seq library

B) Upset plot of the QTL_set regions showing the QTL, CAD GWAS, and motif mutation identities of the SNPs included

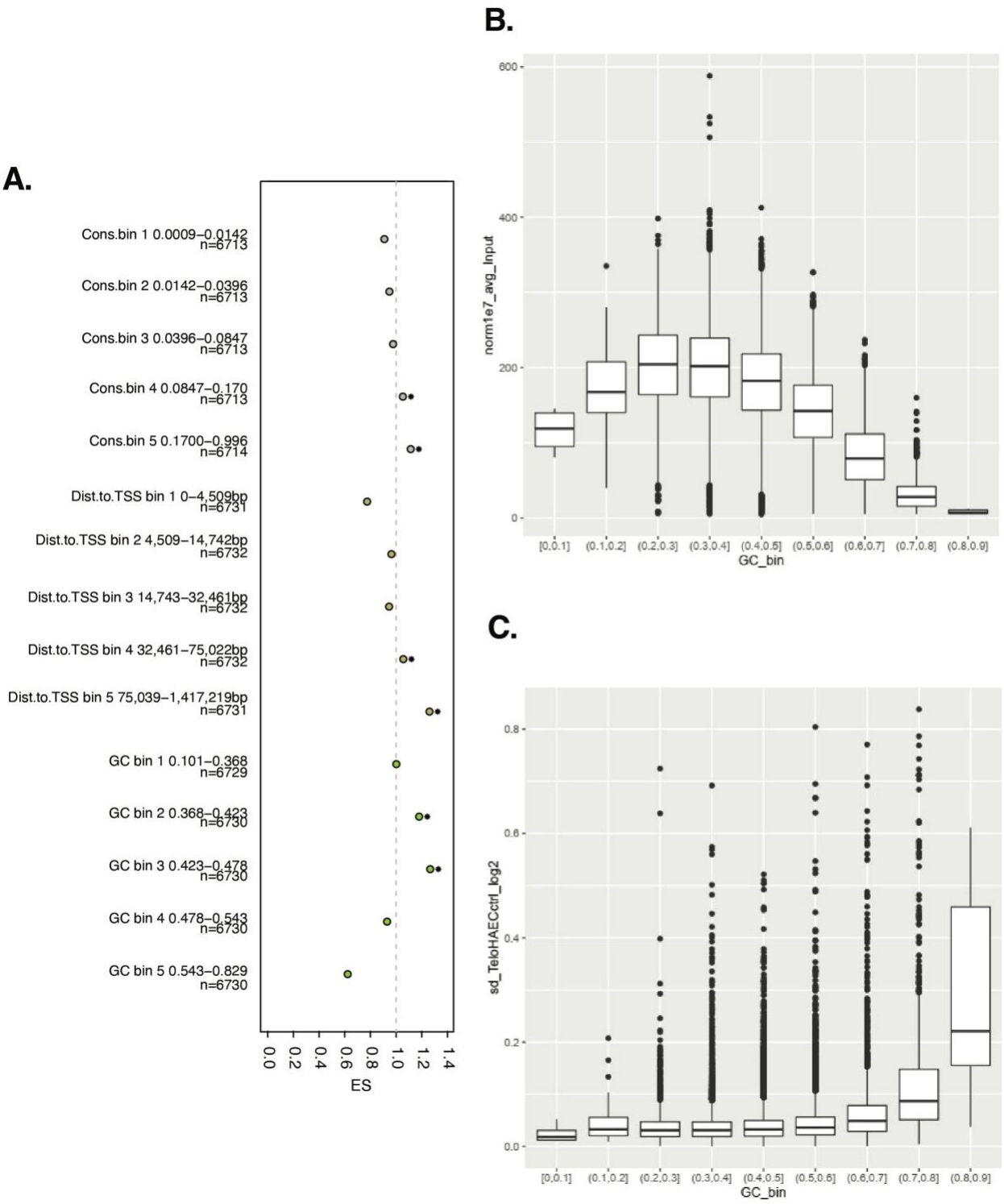
C) Histogram of the number of SNPs within each region in the STARR-seq library



Supplemental Figure S2. STARR library regions that overlap an active promoter, relative to the total number of active promoters called in the sample.
Promoter coordinates and sample grouping were obtained from EpiMap.



Supplemental Figure S3. Pearson correlation of log2-transformed counts from individual STARR-seq experiments.



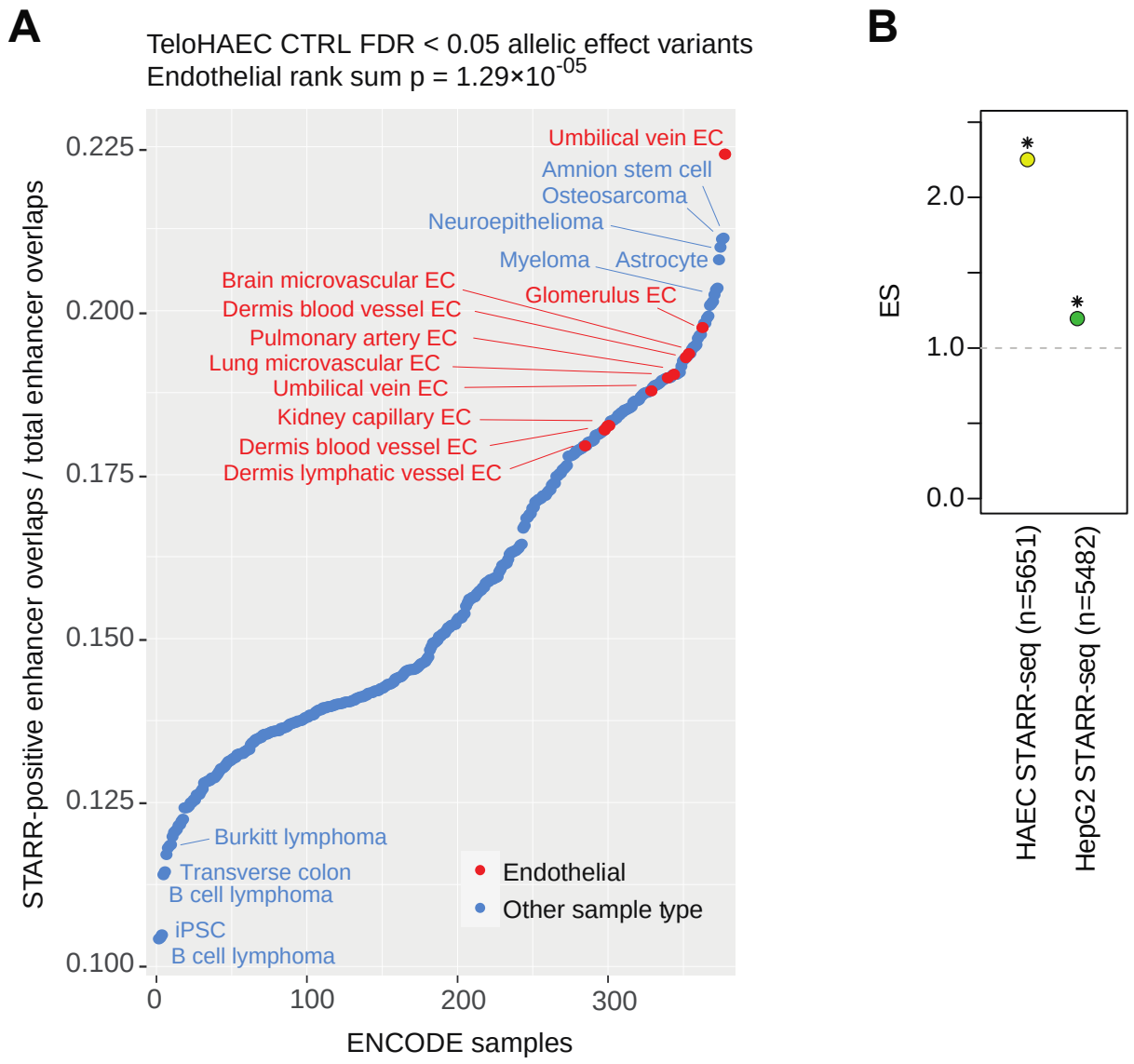
Supplemental Figure S4. Genomic characteristic trends with validation and representation in the STARR-seq library

A) Enrichments of general genomic characteristics ranked by the characteristic and binned into equal number of SNPs. Characteristics included are average GC content of the DNA segments in the STARR-seq library that included the SNP, Distance to nearest TSS in the genomic context of the SNP, and average conservation of the regions surrounding the SNP that were input into the STARR-seq library.

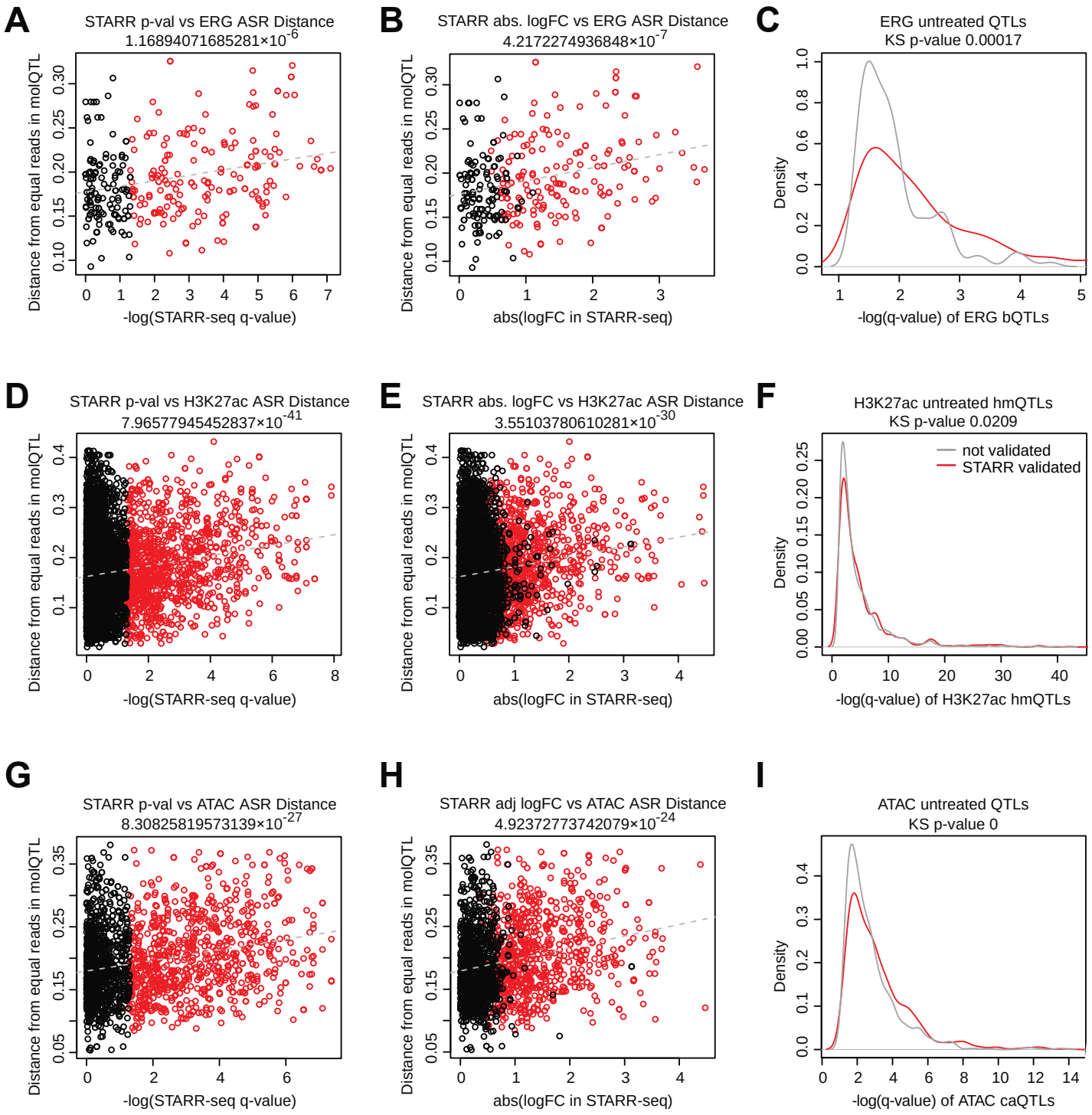
B) Oligo proportions in the plasmid DNA library as a function of GC%.

C) Standard deviation of STARR-seq results between replicate experiments as a function of GC%. Reporter activity from library transfected into teloHAECs (untreated) is presented as RNA normalized to DNA and log-transformed.

Supplemental Figure S5. Allele-specific enhancer function is associated with SVM-determined motif mutations for particular transcription factors. The fraction of deltaSVM-predicted TF binding disruptions (x-axis) that exhibited significant allelic effects in STARR-seq in teloHAEC. The horizontal line at 17% signifies the average proportion of validated (allelic effect *p* adj < 0.05) STARR-seq variants across the entire library.



Supplemental Figure S6. Allele-specific enhancer function is associated with cell type-specific chromatin state.
A) For each epigenome, the fraction of enhancer-overlapping STARR-seq variants that revealed a significant allelic effect in teloHAEC STARR-seq. The enhancers of each ENCODE sample were intersected with STARR-seq library variants to obtain the epigenome-specific set of enhancer overlapping variants, which was then checked for significant (*p* adj < 0.05) allelic effect variants in STARR-seq of control-treatment teloHAEC cells. The ranking of EC samples, compared to all others, was tested using the Wilcoxon rank sum test.
B) Enrichment of SNPs in HAEC accessible chromatin from STARR-seq performed in teloHAECs and STARR-seq performed in HepG2 cells.



Supplemental Figure S7. Linear associations and KS validation of enrichments observed for molQTLs in STARR-seq.

A) Scatter plot of the STARR-seq 0h –log(adjusted p-value) [x-axis] and the absolute value of the distance from equal reads at both alleles for ERG bQTL analysis [y-axis]

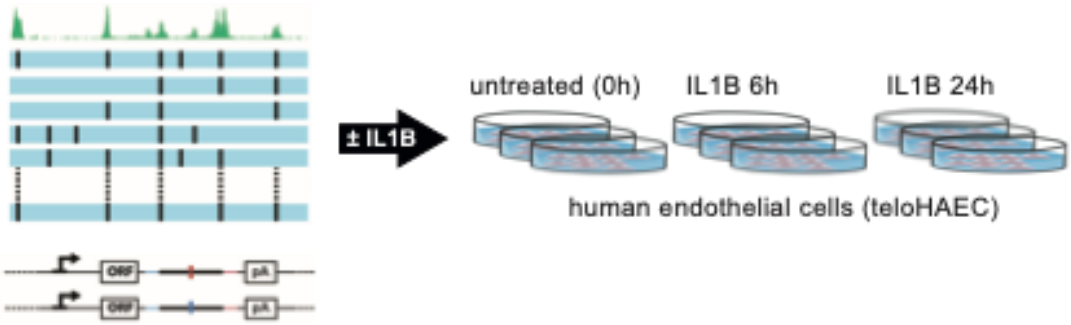
B) Scatter plot of the absolute value log fold change from STARR-seq 0h [x-axis] compared to the absolute value of the distance from equal reads at both alleles for ERG bQTL analysis [y-axis]

C) Density of SNPs that validated (red) and did not validate (grey) in the –log(q-value) from ERG bQTL analysis

D) Scatter plot of the STARR-seq 0h –log(adjusted p-value) [x-axis] and the absolute value of the distance from equal reads at both alleles for H3K27ac hmQTL analysis [y-axis]

E) Scatter plot of the absolute value log fold change from STARR-seq 0h [x-axis] compared to the absolute value of the distance from equal reads at both alleles for H3K27ac hmQTL analysis [y-axis]
F) Density of SNPs that validated (red) and did not validate (grey) in the –log(q-value) from H3K27ac hmQTL analysis
G) Scatter plot of the STARR-seq 0h –log(adjusted p-value) [x-axis] and the absolute value of the distance from equal reads at both alleles for ATAC caQTL analysis [y-axis]
H) Scatter plot of the absolute value log fold change from STARR-seq 0h [x-axis] compared to the absolute value of the distance from equal reads at both alleles for ATAC caQTL analysis [y-axis]
I) Density of SNPs that validated (red) and did not validate (grey) in the –log(q-value) from ATAC caQTL analysis

A



B

C

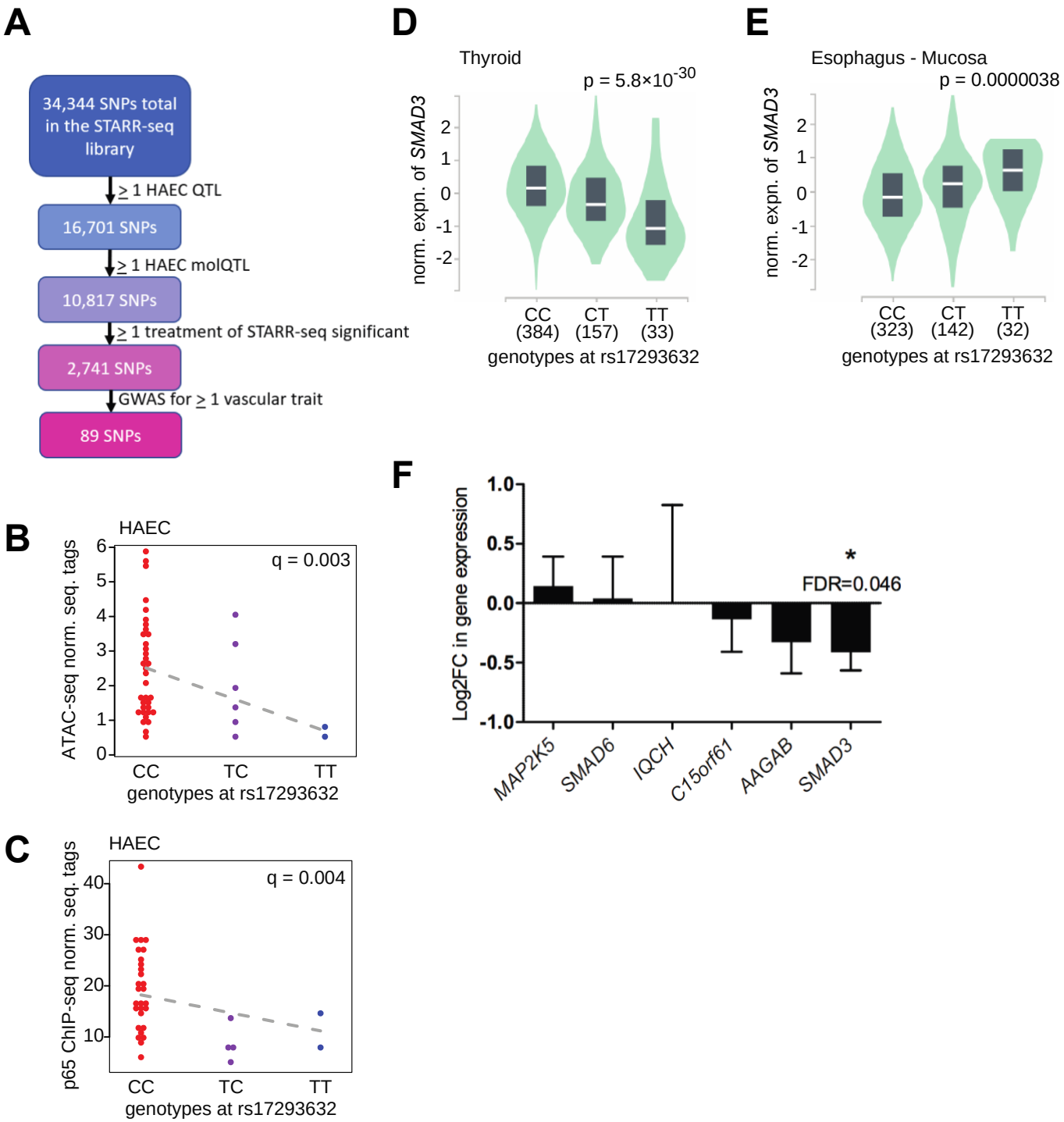
Upregulated

Rank	Motif	Best match	p-value	% of Targets	% Background
1		NFKB-p65	1×10^{-281}	20.11	2.36
2		FRA1 (bZIP)	1×10^{-246}	41.86	14.28
3		PCBP3 (KH)	1×10^{-77}	56.80	38.50
4		ISRE (IRF)	1×10^{-72}	8.03	1.61
5		JUN-CRE (bZIP)	1×10^{-42}	15.87	7.64

Downregulated

Rank	Motif	Best match	p-value	% of Targets	% Background
1		ERG (ETS)	1×10^{-50}	46.51	19.29
2		TP53	1×10^{-47}	11.46	0.96
3		ATF4 (bZIP)	1×10^{-26}	6.31	0.53
4		NEUROD1	1×10^{-22}	4.82	0.33
5		GATA15	1×10^{-20}	3.16	0.10

Supplemental Figure S8. Analysis of the effect of inflammatory stimulus on STARR-seq enhancer activity in cultured human endothelial cells.
A) Schematic of the experimental design
B) Venn diagram of the oligos exhibiting differential enhancer activity in response to inflammatory treatment analyzed using DEseq2.
C) *De novo* motif analysis of the oligos differentially expressed between control and interleukin-stimulated conditions.



Supplemental Figure S9. Selection of candidates and the Coronary Artery Disease associated *SMAD3* locus.

A) Pipeline for selection of all candidate SNPs

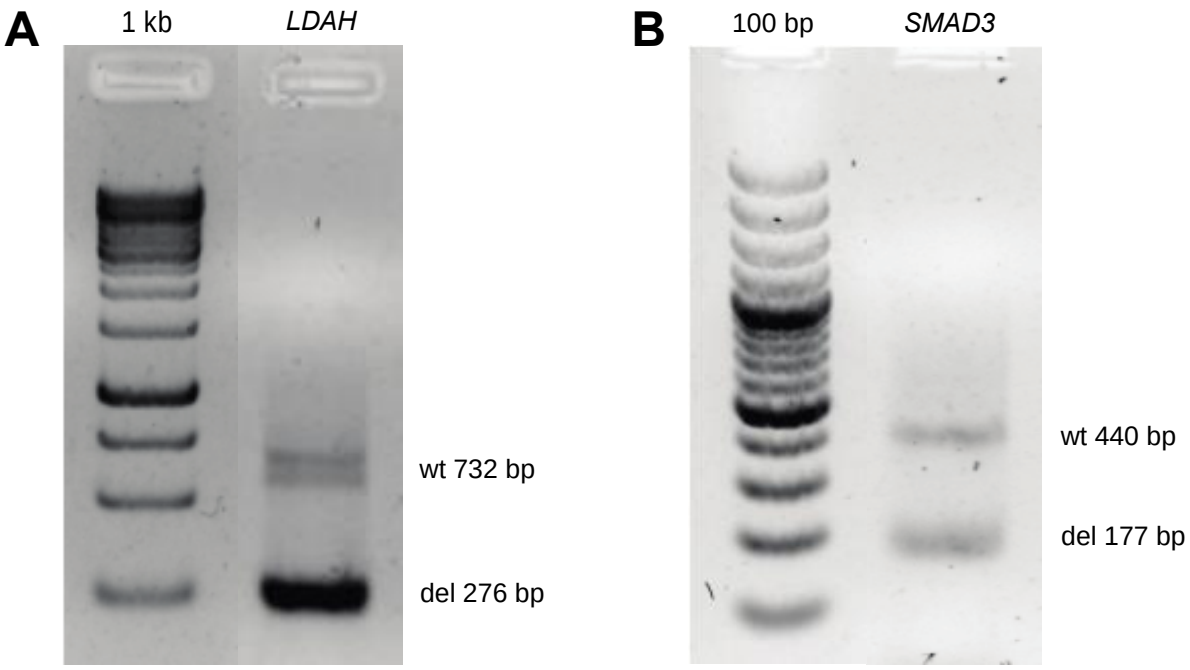
B) ATAC-seq tags in HAECs by genotypes at rs17293632 (RPKM normalized)

C) p65 tags in HAECs by genotypes at rs17293632 (RPKM normalized)

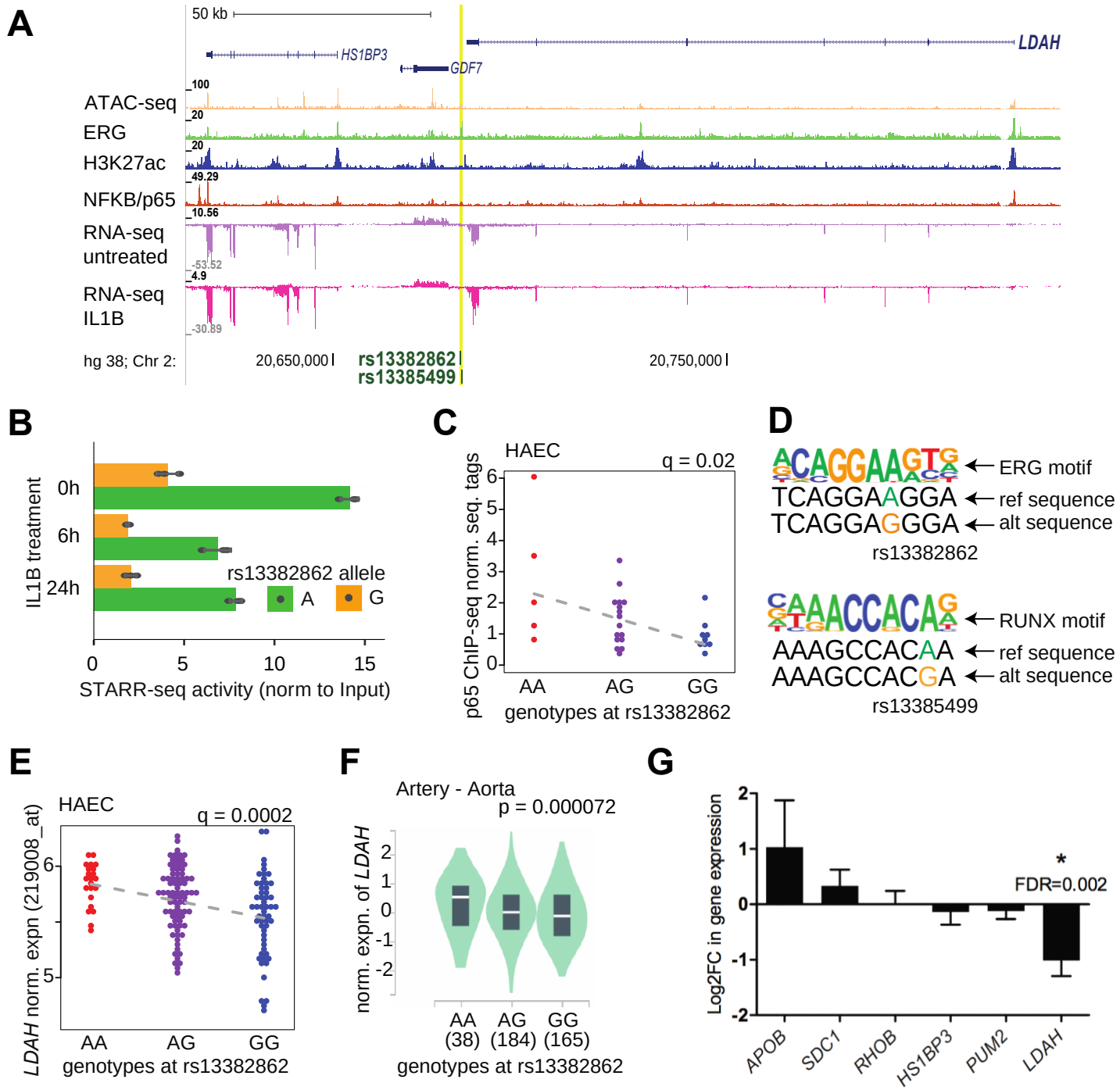
D) *SMAD3* gene expression by genotypes at rs17293632 in Thyroid from GTEx

E) *SMAD3* gene expression by genotypes at rs17293632 in Esophagus - mucosa from GTEx

F) Effect of rs17293632-carrying enhancer deletion on the expression of genes within 1 Mb. Only *SMAD3* demonstrates significant repression.

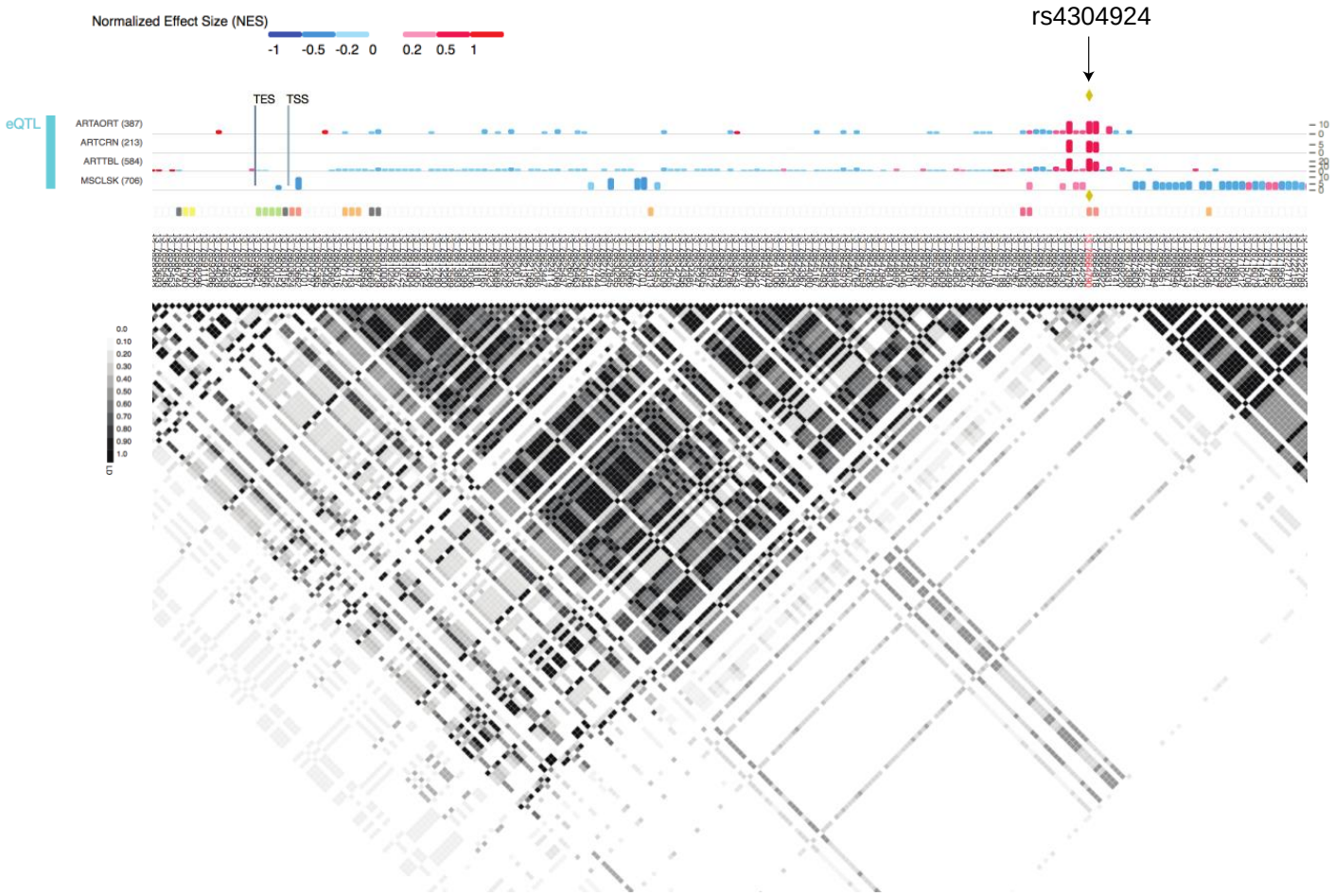


Supplemental Figure S10. CRISPR-mediated deletion of the validated endothelial enhancers.
Analysis of deletion efficiency of the variant-containing enhancer regions predicted to regulate *LDAH* and *SMAD3* in teloHAEC genomic DNA.
A) PCR primers targeting the enhancer proposed to regulate *LDAH* amplify a 732 base pair-long region, if deletion did not occur and the genomic region represents the wild type (wt). Upon successful deletion the band length is 276 bp. Here, the intensity difference between deleted and wild type bands demonstrates a strong deletion.
B) Wild type genomic region for the *SMAD3* enhancer is represented by a 440 bp band, whilst a successfully deleted enhancer produces a 177 bp band. Here, band intensities suggest a ~50% deletion efficiency.



Supplemental Figure S11. Abdominal aortic aneurysm associated locus at gene *LDAH.*
A) Genomic region containing the putative enhancer where rs13382862 and rs13385499 (yellow) reside and the surrounding genes. Tracks below display H3K27ac, ATAC-seq, ERG binding, p65 binding, and RNA-seq expression from HAECs.
B) STARR-seq allele specific expression for rs13382862 across the different treatment points.
C) p65 (NF-kB) tags in HAECs compared between genotypes at rs13382862 (RPKM normalized).
D) PWM for the ERG motif and the motif sequences created by the two alleles at rs13382862
E) *LDAH* gene expression in HAECs by microarray (Probe set ID: 219008_at) compared between genotypes at rs13382862
F) *LDAH* gene expression by genotypes at rs13382862 within artery-aorta from GTEx
G) Effect of rs13385499 and rs13382862 carrying enhancer deletion on the expression of genes within 1 Mb. Only *LDAH* demonstrates significant repression. The *GDF7* gene was not expressed at a quantifiable level.

GTEx eQTLs in Arteries: *POU4F1* ~ rs4304924



Supplemental Figure S12. Similarity of eQTL signal strength for *POU4F1* trancript levels at the *POU4F1* locus in GTEx tissues.
eQTL -$\log_{10}$ p-values are shown on the y-axes for SNPs across the POU4F1 locus (x-axis) for the following tissues in GTEx: Artery Aorta (ARTAORT); Artery Coronary (ARTCRN); Artery Tibial (ARTTBL); and Muscle Skeletal (MSCLSK). Normalized eQTL effect sizes are indicated by color. The location of rs4304924 is highlighted, with LD structure (R^2) shown by heatmap below.