

High-resolution Regulatory Maps Connect Vascular Risk Variants to Disease Related Pathways

Running title: Åkerborg, Spalinska & Pradhananga et al.; Assigning target genes to cardiovascular risk SNPs

Örjan Åkerborg, PhD^{1*}; Rapolas Spalinskas, PhD^{1*}; Sailendra Pradhananga, MS^{1*};

Anandashankar Anil, MS¹; Pontus Höjer, MS¹; Flore-Anne Poujade, PhD²;

Lasse Folkersen, PhD³; Per Eriksson, Prof.², Pelin Sahlén, PhD¹

¹Science for Life Laboratory, School of Engineering Sciences in Chemistry, Biotechnology & Health, Division of Gene Technology, KTH Royal Institute of Technology, Solna; ²Cardiovascular Medicine Unit, Center for Molecular Medicine, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; ³Department of Bioinformatics, Technical University of Denmark, Copenhagen, Denmark
*Share first authorship

Correspondence:

Pelin Sahlén, PhD

Assistant Professor

KTH School of Engineering Sciences in

Chemistry, Biotechnology and Health

Science for Life Laboratory

Tomtebodavägen 23A

171 65 Solna, Sweden

E-mail: pelin.akan@scilifelab.se

Journal Subject Terms: Gene Expression and Regulation; Coronary Artery Disease; Cardiovascular Disease; Epigenetics; Genetic, Association Studies

Abstract:

Background: Genetic variant landscape of coronary artery disease (CAD) is dominated by non-coding variants among which many occur within putative enhancers regulating the expression levels of relevant genes. It is crucial to assign the genetic variants to their correct genes both to gain insights into perturbed functions and better assess the risk of disease.

Methods: In this study, we generated high-resolution genomic interaction maps (~750 bases) in aortic endothelial, smooth muscle cells and THP-1 macrophages stimulated with lipopolysaccharide using Hi-C coupled with sequence capture targeting 25,429 features including variants associated with CAD. We also sequenced their transcriptomes and mapped putative enhancers using chromatin immunoprecipitation with an antibody against H3K27Ac.

Results: The regions interacting with promoters showed strong enrichment for enhancer elements and validated several previously known interactions and enhancers. We detected interactions for 727 risk variants obtained by genome-wide association studies (GWAS) and identified novel as well as established genes and functions associated with cardiovascular diseases. We were able to assign potential target genes for additional 398 GWAS variants using haplotype information, thereby identifying additional relevant genes and functions. Importantly, we discover that a subset of risk variants interact with multiple promoters and their expression levels were strongly correlated.

Conclusions: In summary, we present a catalogue of candidate genes regulated by coronary artery disease related variants and believe that it will be an invaluable resource to further the investigation of cardiovascular pathologies and disease.

Key words: gene regulation; genome-wide analysis; cardiovascular genomics; linkage disequilibrium

Introduction

Coronary artery disease (CAD) labels medical problems of the circulatory system (heart, blood vessels and arteries) often due to build-up of fatty cell debris (plaques) deposited inside the blood vessels. It is the leading cause of disability and death globally ¹. Atherosclerosis, the main underlying mechanism leading to the acute events of CAD, is characterized by a lipid driven chronic inflammation of the arterial intima, a process that includes all major cells in the vascular wall, i.e. endothelial, smooth muscle and inflammatory cells. The acute complication of atherosclerosis such as myocardial infarction and stroke is due to rupture of the fibrous cap with subsequent thrombus formation that totally or partially occludes the vessel and thereby stops the nutrient-rich blood flow. Traditional risk assessment methods based on age, gender, smoking, diabetes, hypertension and dyslipidemia tracks the disease incidence well but underestimates its occurrence since almost half the population classified as low or intermediate risk end up developing cardiovascular disease ²⁻⁴ as these methods fail to inform on the underlying pathological processes that may have been going on for years ⁵. In addition, ethnic differences in cholesterol and blood lipid levels complicate the assessment of the individual risk of disease ⁶. Heritability for CAD is estimated between 40 to 60% based on twin and family studies, therefore genetic risk contributors at play can be utilized in its early diagnosis and treatment ⁷.

Genome-wide association studies (GWAS) have emerged as an important tool in the search for disease causing genomic variants ⁸. Coronary artery disease-specific and other atherosclerosis related indications have been addressed by large GWAS meta-analyses enabled by consortia such as the CARDIoGRAMplusC4D ⁹ and the MEGASTROKE consortium ¹⁰. At its current state, just over 300 independent variants explain 21% of coronary artery disease heritability ¹¹. According to GWAS, a locus on chromosome 9p21 has the strongest association

signal^{12,13}. Although it is established that the risk allele is associated with formation and progression of plaques but not with their rupture^{14,15}, the mechanistic understanding of the conferred risk by these loci remains elusive¹⁶⁻¹⁸. Pathways such as cholesterol and triglyceride metabolism, blood pressure, inflammation, vascular proliferation and remodeling, nitric oxide signaling, vascular tone, extracellular matrix integrity and axon guidance and signaling are also enriched for target genes of GWAS variants¹⁷⁻²⁰.

GWAS studies do not in themselves provide functional insight for the large subset of hits that are non-coding^{21,22}: only one-third of the time a variant affects the expression level of its nearest gene, highlighting the limitations of the nearest-gene assignment approach^{23,24}. The target gene mappings can be refined using various layers of genome annotation information as well as gene expression profiles. To alleviate the problem of complex linkage structures between variants, vast amounts of public datasets of epigenetic marks and transcription factor binding profiles employed to help prioritize the causal/functional variant²⁵⁻²⁷. Expression quantitative trait loci (eQTL) analyses based on gene expression and genotype datasets are also used to locate potentially functional variants that are in linkage disequilibrium (LD) with top association variants²⁸⁻³⁰.

Pathway or gene-set based approaches using canonical pathways and gene ontology (GO) terms goes beyond single variant-based analyses and investigate the combined effect of multiple disease/trait variants on biological functions in terms of the perturbations on pathways or cellular processes³¹⁻³⁵. Such pathway based analyses revealed the functional GWAS variants in cases such as Crohn's disease³⁶, multiple sclerosis³⁷, schizophrenia³⁸ and breast cancer³⁹. Functional gene sets built using co-expression and protein-protein interaction datasets are also used successfully to interpret the GWAS variants⁴⁰⁻⁴².

Many promoters require regulatory elements called enhancers, to drive and regulate gene expression. Enhancers can be located at long distances from their cognate promoters and brought into contact via chromatin looping. Many enhancers carry tissue specific epigenetic marks such as H3K4me1 or H3K27Ac, facilitating their discovery, however not providing information on gene(s) they act upon. Studies of the chromatin interaction landscape were revolutionized by the invention of chromosome conformation capture coupled with next-generation sequencing (Hi-C) enabling the study of genome structure and folding. Combining Hi-C with sequence capture (HiCap), the improvement in resolution required to study of individual promoter-enhancer interactions can be obtained. A recent study also employed high-resolution chromatin conformation capture to obtain promoter-anchored regulatory landscape of induced pluripotent stem cells (iPSCs) and iPSC-derived cardiomyocytes (CM), providing a valuable resource for the cardiovascular biology ⁴³.

In this study, we employed HiCap on three cell types relevant for vascular diseases, particularly atherosclerosis and aortic diseases, to discover novel biological processes and pathways related to onset and pathology of the disease. We utilized chromatin contacts of promoters to GWAS variants or those that are in linkage disequilibrium (LD) to assign potential target genes.

Methods

The authors declare that all supporting process data are available within its online supplementary files. The raw data files (fastq and bam files) that support the findings of this study are available from the corresponding author upon reasonable request. This study does not involve animal studies. The study was approved by the Human Research Ethics Committee at Karolinska

Institutet (application number 2006/784-31/1 and 2012/1633-31/4), Stockholm, Sweden; written informed consent was obtained from all the patients according to the Declaration of Helsinki, and methods were carried out in accordance with relevant guidelines. Please see supplementary material for detailed method descriptions.

Results

Using high-resolution chromatin interactions, we mapped genomic interaction of promoters and variants associated with traits and conditions related to cardiovascular diseases in particular coronary artery and aortic diseases. Three cell types for the investigation are human aortic endothelial cells (AEC), human aortic smooth muscle cells (ASMC) and macrophage-THP-1 cells (mTHP-1-LPS) stimulated with lipopolysaccharide (LPS) for two hours. For AEC, we obtained two technical replicates from two individuals; the technical replicates were pooled, and the two individuals were held separate and constituted biological replicates. For ASMC and mTHP-1-LPS cells, two technical replicates were obtained. We employed HiCap using a probe set targeting 21,479 promoters and 3,950 variants (suppl. table 1). There were 199, 300 and 289 million read pairs uniquely mapped to probes in AEC, ASMC and mTHP-1-LPS experiments, respectively (suppl. table 2). We made a distinction between interactions between promoters (promoter-promoter or Prom-Prom) and those between promoters and elsewhere in the genome (promoter-distal or Prom-Dist). For the sake of clarity, we also defined interactions between disease/trait associated variants and promoters (GWAS-Promoter or GWAS-Prom) (figure 1a).

We called interactions using HiCapTools requiring each interaction present in both replicates, and p-value cutoffs deployed yielded interaction sets of sizes 69,753 (AEC Prom-Dist), 38,759 (ASMC Prom-Dist), 19,920 (mTHP-1-LPS Prom-Distal), and 5,671 (AEC Prom-

Prom) and 4,293 (ASMC Prom-Prom) and 1,698 (mTHP-1-LPS Prom-Prom), respectively (suppl. table 3A, 3B and 3C, suppl. methods). Importantly, we were able to detect many long-range (>500Kb) interactions across the three cell-types (suppl. figure 1). Equally important, the distal elements (DEs) as well as the interacting promoters were short; average length being 749 and 776 bases, respectively. In total, the interaction datasets covered around 3.4% of the genome. Most promoters (65%) were found to interact with less than five distal regions, whereas the interactome of extreme hub-promoters contain several hundred DEs (suppl. fig. 1b-d). We identified several interconnected units of promoters and enhancers, figure 1b displays the largest connected subsection of chromosome 9 (i.e. giant component). Interestingly, two CVD associated GWAS SNPs in chr9p21 region (rs1333042 and rs944797) were part of this network.

We profiled expression levels of genes in AEC, ASMC and mTHP-1-LPS cell lines using RNA sequencing (Suppl. Table 4). Utilizing principal component analysis, we show that the promoter-interaction profiles of individual cell types are specific and can separate individual cell types independent of gene expression information (figure 1c, suppl. figure 3).

Promoter-interacting distal elements were enriched for functional elements

Promoter-interacting DEs were previously shown to be highly enriched for enhancer marking features. To confirm that is also the case for this study, we overlapped DEs with H3K27ac enriched regions obtained through ChIP-seq in the same cells, as well as relevant DNaseI and transcription factor binding datasets from the ChipAtlas (list of public datasets can be found in suppl. table 5 and www.chip-atlas.org). Our interactor sets were indeed enriched relative to size- and genomic context controlled random sets and the enrichment was stronger for the better matching cell types (figure 1d, suppl. figure 4a-d, supplementary material section 7).

Interestingly, enrichment levels for the GWAS-Prom set was much higher in AECs and ASMCs

(figure 1d, suppl. figure 4a-d). Further, promoters interacting with DEs carrying H3K27ac marks were expressed at higher levels as expected (suppl. fig 4e).

Figure 1e and 1f shows two examples of promoter interactions (MTAP-rs944797 and BMP6-rs9328448) where the interactor overlaps with both CVD GWAS variants and cell-specific H3K27Ac enhancer marks.

Promoter interacting GWAS-variants were often contained within regulatory elements

We next turned our attention to variants associated with vascular disease phenotypes according to GWA studies and asked if any variants or in LD with those are contained within DEs found in this study. First, we took all single nucleotide variants associated with “cardiovascular disease” resulting in 3,814 SNPs (minimum p value for association is 10^{-6}), which we called CVD_GWAS (cardiovascular disease GWAS) (suppl. table 6A). As previously shown, a substantial portion of GWAS variants are themselves located within potential regulatory regions. We therefore targeted a subset (723, 19%) of these variants using probes to increase the probability of obtaining a signal without the need for deep sequencing (suppl. table 6B). Of those 723 targeted GWAS variants, 295 (41%) interacted with at least one promoter in at least one cell type, constituting the GWAS-Prom dataset. We assigned 423 target genes to 295 variants (54% interacting with only one promoter). To rule out the possibility that GWAS-Prom hits occur merely as result of the probing of variants we investigated the corresponding Prom-Dist set, i.e. we studied the same promoters and their interactions with non-probed distal regions. There should be vastly more Prom-Dist hits close to the variant site, than to a site on the same distance from the promoter but at the other side of it (hence keeping the distance from the probed feature the same). This is indeed the case as shown in figure 2a (supplementary material section 8).

A large fraction of GWAS-Prom interactions spanned distances above 500 kb (Figure S1). Consequently, many of the GWAS variants (68%) are found to interact with non-closest genes thus ‘jumped over’ by the interaction loop formed (figure 2b). On the other hand, only 43% of the Prom-Dist interactions involved non-closest gene.

Whole groups of genes frequently interact with the same GWAS variant. Promoters of genes USP24, PLPP3, PRKAA2 and FYB2 thus share a putative enhancer containing variant rs12239436. The rs12239436-PLPP3 interaction is particularly interesting due to its large 1.3Mb distance as well as the fact that PLPP3 was already identified as a CAD disease risk gene (figure 2c).

The GWAS variant rs9349379, associated with five vascular diseases, was recently shown to regulate expression of the Endothelin-1 gene. We see this interaction in one of the two AEC investigated individuals. Interestingly, the individual with the stronger interaction is heterozygous for rs9349379 (A/G), whereas the other individual is homozygous reference (A/A). In ASMC, where there is no interaction, rs9349379 variant is homozygous alternative (G/G) (figure 2d).

Discovery of target genes of GWAS variants using shared haplotype and interaction information

To detect further interactions of CVD_GWAS variants with promoters, we looked at the fraction of DEs containing such variants. Of the 3,814 associated variants in the CVD_GWAS, there were 216 (5.7%) variants within unique DEs. One complication of GWA studies is that the association signal many GWAS variants possess are due to their sharing of haplotype with the functional variants. If the functional variant can indeed modulate a distal promoter via looping, it should also be possible to locate it in our DE datasets. We therefore looked at the fraction of DEs

that contain variants that are in LD with those in CVD_GWAS. Due to the sheer number of SNPs in LD, we devised a double randomization scheme to assign statistical significance to the observed overlap between LD SNPs and DE datasets using both size- and context-matched random interaction datasets and random SNP datasets matched with respect to allele frequencies and surrounding LD structure of the real set (Supplementary Material section 9). We obtained LD and allele frequencies from 1000 Genomes (Phase 3 v5) using European population and set LD threshold of 0.8. Figure 2e shows that the DE dataset is enriched for SNPs in LD with CVD_GWAS that are within 80 kb, 20 kb and 30 kb in AEC, ASMC and mTHP-1-LPS cells respectively, beyond which no enrichment can be seen (suppl. figure 5a and b). There were 559 SNPs in LD with variants in CVD_GWAS located within DE fragments (Prom-Dist_LD). DEs carrying either the GWAS variant themselves or those in LD showed higher enrichments for open chromatin, transcription factor binding sites and enhancer marks compared to the entire DE set, supporting their potential for expression modulation (suppl. figure 6).

The eQTL technique was deployed to examine the expression modulation capacity of variants contained in DEs. The GWAS-Prom set was extended with Prom-Dist hits and likewise selected hits in linkage disequilibrium with those (Prom-Dist_LD). The comparison was performed relative to the afore mentioned size- and distance controlled random set and yielded the Q-Q plot presented in figure 3b (merged set) and suppl. figure 7a-c (GWAS-Prom, Prom-Dist and Prom-Dist_LD separate). The deviation from the diagonal is striking and concerns not only extreme cases but large fractions of the entire sets.

Only 30% of the DEs containing these variants interacted with the closest gene and the average interaction distance is 301 kb (suppl. table 6). In terms of trait categorization, variants for 167 of the 247 traits (0.67) related to CVD in EBI GWAS catalogue was found, figure 3c lists

overrepresented traits in our dataset. We also compared enrichment of promoter-interacting GWAS variants for enhancer marks with respect to trait they belong. We chose an CVD-related trait (“coronary artery disease”) and compared to five variant size-matched non-CVD traits to demonstrate the specificity of such enrichments for the disease, and found that the enrichment for enhancer marks was significantly higher in CVD-related trait (supplementary material section 10).

Gene enrichment analysis of target genes of GWAS variants for discovery of disease associated cellular processes

We next asked if genes interacting with fragments carrying disease associated variants are enriched for particular functions or pathways. To discover cell context dependent signal, we performed a gene set enrichment analysis using genes interacting with GWAS variants themselves and those in LD for each cell type. We only included LD SNPs up to 80 kb, 20 kb and 30 kb to the proxy SNP in AEC, ASMC and mTHP-1-LPS cells. In order to assess the success of discovering novel processes or functions, we input the DEs containing these variants to GREAT software package to retrieve the gene sets independent of interaction information to perform the same enrichment analysis for comparison (supplementary material section 11). We performed enrichment analysis separately for each cell and also combined to assess the specific contribution of each cell type. Comparison of the enriched terms by interacting or closest gene information (GREAT package) using a GO term semantic similarity measure revealed little overlap in between (suppl. Methods section 11, suppl. table 8). Figure 3d shows enriched biological processes when target genes from all cell types are merged. We located novel genes associated with known CVD complications including “response to lipopolysaccharide” (GO:0032496, in AEC and mTHP1-LPS), “phosphatidylinositol-3-kinase signaling”

(GO:0014065, in AEC) and “SMAD protein signal transduction” (GO:0060395, in AEC and ASMC). Moreover, we discovered genes and functions not previously associated with CVD onset and or progress such as “cilium assembly” (GO:0060271, in AEC, 14 genes).

Ectopic deposition of calcium in arterial vessel walls leading to vascular calcification is a main feature of atherosclerosis and similar to the ossification process. Concordantly, terms such as “endochondral ossification”, “regulation of chondrocyte differentiation”, “regulation of osteoblast differentiation”, “positive regulation of ossification” were among the enriched functions. Eleven genes (BMP6, SMAD3, JAG1, PDLIM7, SLC8A1, DLX5, TEK, HOXA2, EFEMP1, RARB and IL6) responsible for the above enrichments and only four (BMP6, TEK, SMAD3 and JAG1) interacted with lead SNPs, while the rest interacted with variants in LD with lead SNPs.

Fourteen target genes were involved in cilium assembly, including IFT74, a component of endothelial intraflagellar transport, which interacts with a CVD associated variant in the chr9p21 region (rs944797). It has been shown that endothelial cells can sense and respond to shear stress levels using their cilia and endothelial cilia were shown to deflect in response to blood flow rates. The deflection angle is regulated by calcium levels. Moreover, endothelial cilia inhibit onset of atherosclerosis in mouse models.

We identified several genes involved in leukocyte adhesion and vascular inflammation, key processes of atherosclerotic development. Examples of target genes include Cadherin 13 (CDH13 interacting with rs8055236) which has previously been shown to protect against atherosclerosis in experimental models, AMP-activated protein kinase (PRKAA2 interacting with rs12239436) whose activity inhibits cell migration via phosphorylation of Pdlim5 and

BACH1 (interacting with rs2832227), a transcriptional regulator which has been shown to be involved in atherosclerosis development in apolipoprotein E deficient mice.

Other examples of plausible candidate genes for inflammatory cardiovascular disease include CD86 (interacting with rs13083990), a receptor involved in the costimulatory signal essential for T-lymphocyte proliferation and interleukin-2 production and AKIRIN2 (interacting with rs6900057) a gene that has been shown to stimulate a pro-inflammatory gene in macrophages during innate immune responses.

Expression levels of interacting promoters sharing enhancers are correlated

Sometimes variants are contained within enhancers controlling multiple genes, suggesting such gene sets to be group- and pairwise co-expressed. Using data from the ASAP-Heart study we were able to test 75 AEC gene pairs sharing enhancers and could conclude that 34 (45.3%) are co-expressed at p-value level 10^{-3} ; 18 of 75 (24.0%) also at p-value level 10^{-10} and there is a strong enrichment over random background of gene-pairs where only 23% are co-expressed ($P=9.6e-06$ by chi-square test). Excluding a large cluster of genes all interacted upon by the same variant rs13083990, these percentages rises to 64.1% ($p < 10^{-3}$) and 38.4% ($p < 10^{-10}$), respectively.

In figure 4a this is exemplified with co-expression plots for the MTAP and IFT74 genes, both interacted upon by the chr9p21 locus variant rs944797 in AEC. The genomic distance between the two gene promoters is in excess of 5 million bases. Even more extreme is the co-expression of genes SMARCD1 and BMP2K both located on chromosome 4 more than 15 Mb apart in AEC (figure 4b). Both genes also interact a GWAS variant associated with diastolic blood pressure (rs16998073, p value = 10^{-21}), which itself interacts two other genes (LINC01094 and PAQR3) and 27 DEs (11 overlapping with H3K27ac marks) (figure 4c). Most

of these interactions were specific to AEC (figure 4d). LINC01094 is a non-coding RNA and its expression levels are correlated with serum albumin (p value = 10^{-9}) and co-expressed with BMP2K (p value = $1.26 * 10^{-93}$). Serum albumin levels are positively correlated with blood pressure. Moreover, PAQR3 levels modulates leptin signaling in mouse models and leptin is found to mediate the increase in blood pressure associated with obesity.

Discussion

Our aim in this study was to evaluate the contribution of high-resolution promoter-anchored regulatory interaction maps to locate the target genes of noncoding GWAS variants associated with vascular diseases. Although many GWAS variants are merely tags for the functional SNP within the same haplotype, some may still be or be very close to the “functional” variant as suggested by their enrichment for enhancer marks and re-sequencing studies. Here we show that targeting GWAS variants in capture Hi-C experiments can be a useful strategy in conquest for target gene associations due to lesser need for sequencing. We also uncover several enhancers regulating multiple genes and a strong correlation signal between such sharing enhancers, implying the underlying complexity of regulatory networks. A recent study showed the implicit wiring of enhancer redundancy in regulatory networks, where the system can tolerate loss of enhancers by connecting promoters to multiple enhancers⁴⁵. However, the case when multiple genes connected to the same enhancer could negatively affect the resilience of the network in the case of enhancer malfunction, potentially disturbing the co-regulation of multiple genes.

We tackle the difficulty of locating the “functional” variant using LD information. When a GWAS SNP is associated with a trait, essentially any other variant on the same haplotype could be responsible for the association. However, due to sheer number of variants in LD, it is

not straightforward to locate the “functional” one. The resolution in this study was around 750 bases, which allowed us to locate DEs containing variants in LD with CVD GWAS variants. Extending the genomic window around the DEs to 250 kb, we found that, it is possible to discriminate between functional and tagged variants using a double randomization procedure. DEs containing variants in LD with CVD GWAS variants showed better enrichments for histone enhancer marks and TF binding sites. We found that HiCap-identified loci are strongly enriched for genes identified by eQTL-based investigations of GWAS hits. While the two methods are conceptually different, this overlap supports the idea that a diverse set of methods for functional genetics is advantageous when identifying causal genes from GWAS disease loci.

We confirm that it is only one-third of the time the enhancer is connected to the promoters of its nearest gene. We take on the challenge of assigning the correct genes to GWAS variants using promoter- and variant-anchored regulatory maps produced in two cell types. Indeed, we discover multiple biological processes and cellular structures that are associated with vascular disease pathology not by genomic but by functional studies. We were able to suggest variants that could be responsible for the perturbations of such processes or structures. Here it is important to note that by only mining the variants associated with vascular disease trait, we will be able to discover the genes that are perturbed in a given pathway, process or structure. Discovery of the full network of genes within such processes or structure is beyond the scope of this study.

In summary, we provide high-resolution promoter-anchored regulatory networks of three cell types and list novel genes, processes and cellular structures relevant for vascular disease pathologies, in particular coronary artery and aortic diseases. We hope that the data and the

methodologies in this study will aid us in our mission to further understand the contribution of non-coding genomic variation to complex disease biology.

Sources of Funding: This work has been supported by the Swedish Research Council [Grant Agreement No. 78081].

Disclosures: None

References:

1. Roth GA, et al. Global, Regional, and National Burden of Cardiovascular Diseases for 10 Causes, 1990 to 2015. *J Am Coll Cardiol*. 2017;70:1-25.
2. European Association for Cardiovascular P, et al. ESC/EAS Guidelines for the management of dyslipidaemias: the Task Force for the management of dyslipidaemias of the European Society of Cardiology (ESC) and the European Atherosclerosis Society (EAS). *Eur Heart J*. 2011;32:1769-818.
3. Perk J, et al. European guidelines on cardiovascular disease prevention in clinical practice (version 2012) : the fifth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of nine societies and by invited experts). *Int J Behav Med*. 2012;19:403-88.
4. Wang TJ. Assessing the role of circulating, genetic, and imaging biomarkers in cardiovascular risk prediction. *Circulation*. 2011;123:551-65.
5. Duprez DA, et al. Identifying early cardiovascular disease to target candidates for treatment. *J Clin Hypertens (Greenwich)*. 2008;10:226-31.
6. Chaturvedi N. Ethnic differences in cardiovascular disease. *Heart*. 2003;89:681-6.
7. Zdravkovic S, et al. Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J Intern Med*. 2002;252:247-54.
8. Erdmann J, et al. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res*. 2018;114:1241-1257.
9. Consortium CAD, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25-33.

10. Malik R, et al. Multiancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet.* 2018;50:524-537.
11. Nelson CP, et al. Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet.* 2017;49:1385-1391.
12. Samani NJ, et al. Genomewide association analysis of coronary artery disease. *N Engl J Med.* 2007;357:443-53.
13. Helgadottir A, et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science.* 2007;316:1491-3.
14. Reilly MP, et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet.* 2011;377:383-92.
15. Horne BD, et al. Association of variation in the chromosome 9p21 locus with myocardial infarction versus chronic coronary artery disease. *Circ Cardiovasc Genet.* 2008;1:85-92.
16. Almontashiri NA, et al. 9p21.3 Coronary Artery Disease Risk Variants Disrupt TEAD^{can} Transcription Factor-Dependent Transforming Growth Factor beta Regulation of p16 Expression in Human Aortic Smooth Muscle Cells. *Circulation.* 2015;132:1969-78.
17. Ghosh S, et al. Systems Genetics Analysis of Genome-Wide Association Study Reveals Novel Associations Between Key Biological Processes and Coronary Artery Disease. *Arterioscler Thromb Vasc Biol.* 2015;35:1712-22.
18. Kessler T, et al. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Mol Med.* 2016;8:688-701.
19. Khera AV, et al. Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat Rev Genet.* 2017;18:331-344.
20. Prashar Y, et al. Emerging role of various signaling pathways in the pathogenesis and therapeutics of atherosclerosis. *Reviews in Vascular Medicine.* 2017;10-11:1-12.
21. Balding DJ. A tutorial on statistical methods for population association studies. *Nat Rev Genet.* 2006;7:781-91.
22. Visscher PM, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet.* 2017;101:5-22.
23. Dostie J, et al. Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 2006;16:1299-309.

24. Gusev A, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet.* 2016;48:245-52.
25. Cantor RM, et al. Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am J Hum Genet.* 2010;86:6-22.
26. Chung D, et al. GPA: a statistical approach to prioritizing GWAS results by integrating pleiotropy and annotation. *PLoS Genet.* 2014;10:e1004787.
27. Morris JA. Using epigenomic data to inform genome-wide association studies of bone mineral density. *Ann Transl Med.* 2016;4:487.
28. Hauberg ME, et al. Large-Scale Identification of Common Trait and Disease Variants Affecting Gene Expression. *Am J Hum Genet.* 2017;100:885-894.
29. Ma C, et al. The integrated landscape of causal genes and pathways in schizophrenia. *Transl Psychiatry.* 2018;8:67.
30. Zhu Z, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016;48:481-7.
31. Jia P, et al. Network-assisted analysis to prioritize GWAS results: principles, methods and perspectives. *Hum Genet.* 2014;133:125-38.
32. Leiserson MD, et al. Network analysis of GWAS data. *Curr Opin Genet Dev.* 2013;23:602-10.
33. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102:15545-50.
34. Torkamani A, et al. Pathway analysis of seven common diseases assessed by genome-wide association. *Genomics.* 2008;92:265-72.
35. Yang J, et al. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat Genet.* 2011;43:519-25.
36. de Lange KM, et al. Understanding inflammatory bowel disease via immunogenetics. *J Autoimmun.* 2015;64:91-100.
37. Baranzini SE, et al. Pathway and network-based analysis of genome-wide association studies in multiple sclerosis. *Hum Mol Genet.* 2009;18:2078-90.
38. Askland K, et al. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. *Hum Genet.* 2009;125:63-79.

39. Wang W, et al. Pathway-based discovery of genetic interactions in breast cancer. *PLoS Genet.* 2017;13:e1006973.
40. Calabrese GM, et al. Integrating GWAS and Co-expression Network Data Identifies Bone Mineral Density Genes SPTBN1 and MARK3 and an Osteoblast Functional Module. *Cell Syst.* 2017;4:46-59 e4.
41. Gustafsson M, et al. A validated gene regulatory network and GWAS identifies early regulators of T cell-associated diseases. *Sci Transl Med.* 2015;7:313ra178.
42. Huan T, et al. Integrative network analysis reveals molecular mechanisms of blood pressure regulation. *Mol Syst Biol.* 2015;11:799.
43. Montefiori LE, et al. A promoter interaction map for cardiovascular disease genetics. *Elife.* 2018;7.
44. Hahne F, et al. Visualizing Genomic Data Using Gviz and Bioconductor. *Methods Mol Biol.* 2016;1418:335-51.
45. Osterwalder M, et al. Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature.* 2018;554:239-243.

Circulation: Genomic and Precision Medicine

Figure Legends:

Figure 1 (A) Three types of HiCap-established interactions; Prom-Dist, Prom-Prom and GWAS-Prom, respectively; for the Gene A promoter. In Prom-Dist is analyzed probed promoters' interaction with distal element (DEs); the latter separated by restriction sites. In Prom-Prom and GWAS-Prom both ends of the interaction are probed. (B) The largest connected subgraph in chr9 in AEC cells. There are 22 promoters, three GWAS SNPs and 271 DE (only those overlapping with H3K27ac marks are included). The blue track represents gene expression levels, gray boxes represents transcripts, innermost layers represent H3K27Ac marks (two

replicates). Purple arcs represent GWAS-Prom or Prom-Prom, orange arcs represent Prom-Dist interactions respectively. (C) Principal component analysis (PCA) of Prom-Dist interaction of genes not expressed in none of the cell types (D) Overlap enrichment relative to a segment-length and distance-from-promoter controlled random set. The AEC Prom-Dist and GWAS-Prom datasets were overlapped with general Cardiovascular, HUVEC and HAEC transcription factor marker data from ChipAtlas. (E) 357kb region containing MTAP-rs944797 interaction and (F) 393 kb region containing BMP6-rs9328448 interaction visualized using Gviz package ⁴⁴. Overlap with H3K27Ac marks are shown in the lower panes as well as the signal from the input chromatin. Prom-distal (including Prom-GWAS) and prom-prom interactions are colored as green and purple respectively.



Figure 2: (A) The AEC Prom-Dist set was searched for interactions between the promoter and distal elements close to a variant interacting with the same promoter (green curve). The comparison was made relative to a site at the same distance from the promoter it but located at the other side of it (light blue curve). The latter is nearly a horizontal line as expected whereas the blue curve strongly deviates from that at distances not too far from the variant site. The bin size used to count interactions is 5 kb. (B) Variants in the AEC, ASMC and mTHP-1-LPS merged GWAS-Prom set and their interaction preferences with genes at distance 0 (no gene-jumping), 1 (nearest gene is jumped-over) etc. Distal regions in the corresponding Prom-Dist set are shown for reference. The GWAS-Prom interaction set is further split in equally sized halves depending the variants distance to its nearest gene (GWAS-Prom without- and with a nearby gene, respectively). (C) The CAD-related variant rs12239436 (red box) interacts with 62 kb distant gene USP24 (grey box) in AEC (dotted line), ASMC (dashed) and mTHP-1-LPS (solid).

Strengths of interactions are represented with the p-value recorded and indicated by arrow thickness. Our result set further include very strong interactions with the 1,3Mb distant previously CAD associated gene PLPP3 (also known as PPAP2B). Less significant interactions with nearby FYB2 (ASMC) and PRKAA2 (mTHP-1-LPS) are potentially bystanders. (D) Earlier reported interaction between variant rs934937 and the EDN1 gene is, to a varying degree, present in both AEC patients. Not so in less relevant tissues ASMC and mTHP-1-LPS. (E) Comparison of overlaps between Prom-Dist dataset (AEC) vs 100 matched random datasets and CVD_GWAS and matched SNP datasets (see methods) shows that there is an enrichment for variants in LD with CVD_GWAS found in Prom-Dist dataset when the genomic distance between SNP and its LD proxy is ≤ 80 kb. No such enrichment was seen for random Prom-Dist datasets vs real or random SNP sets.

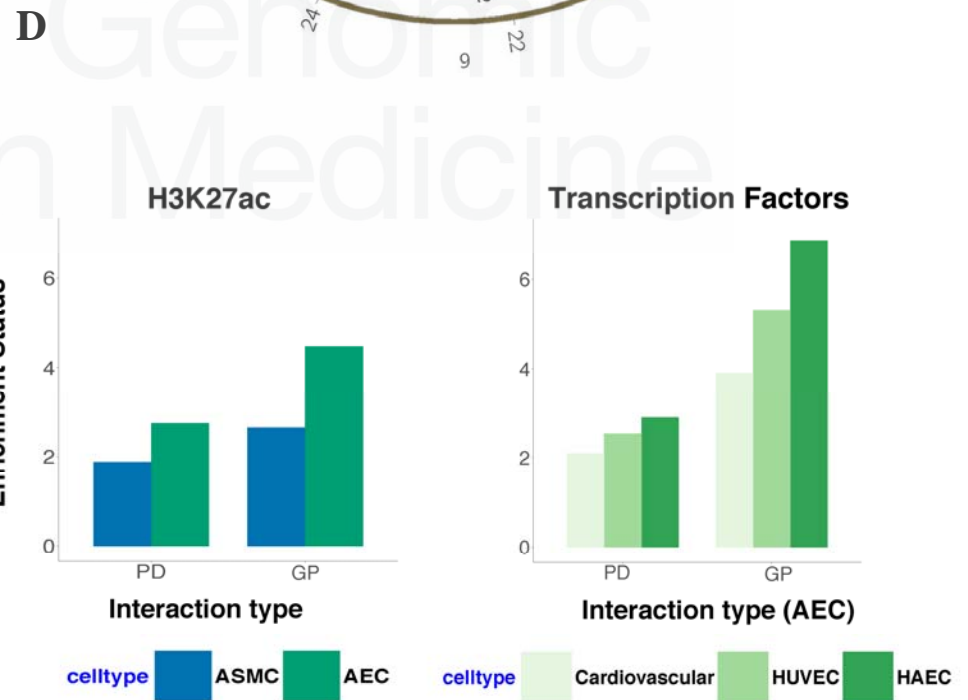
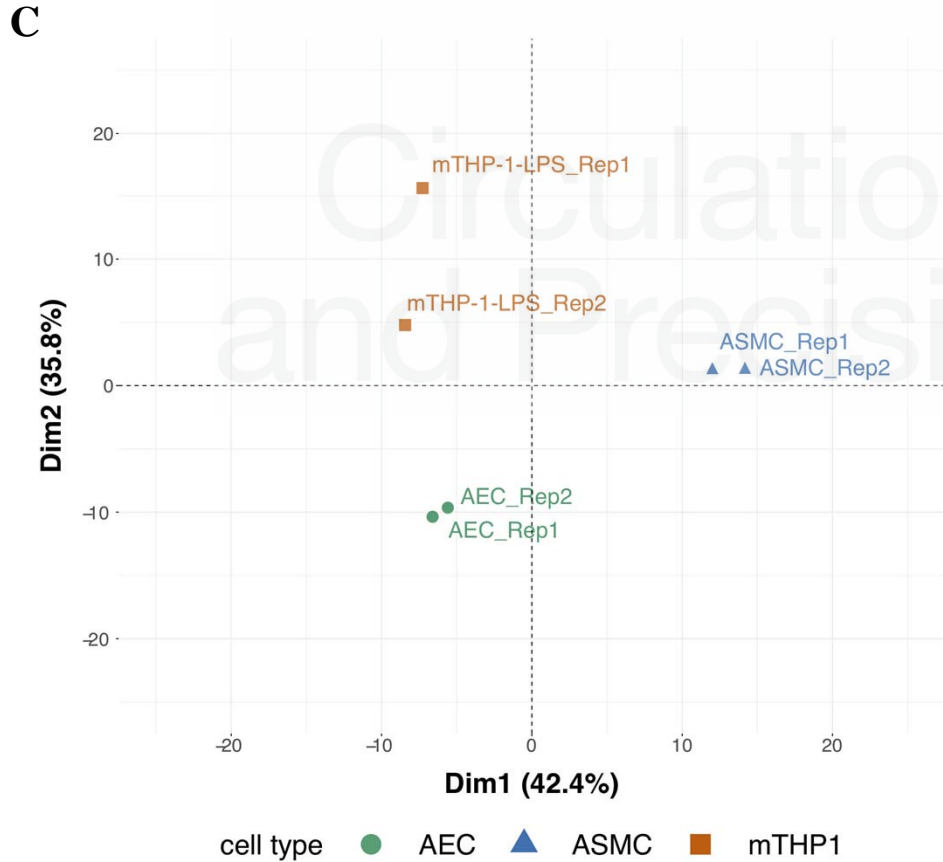
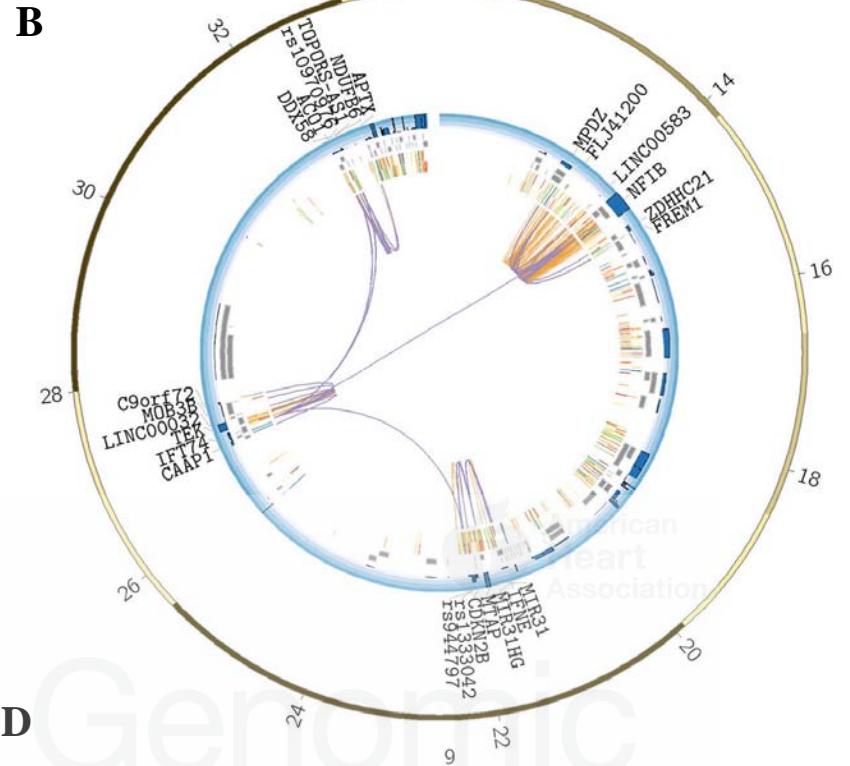
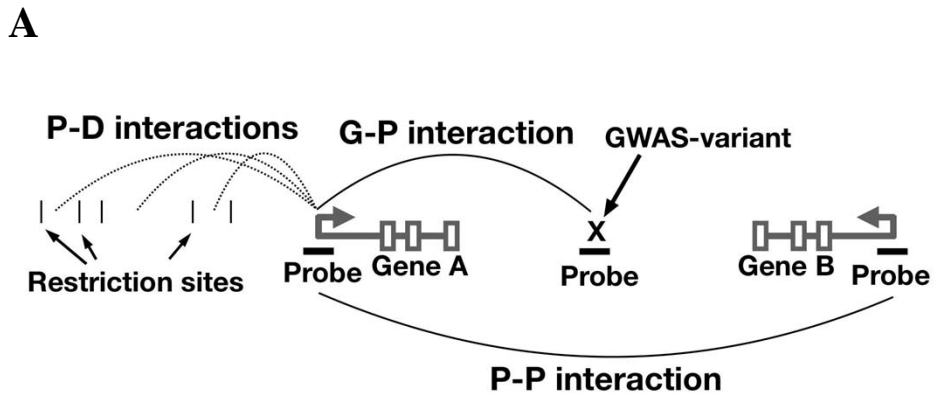


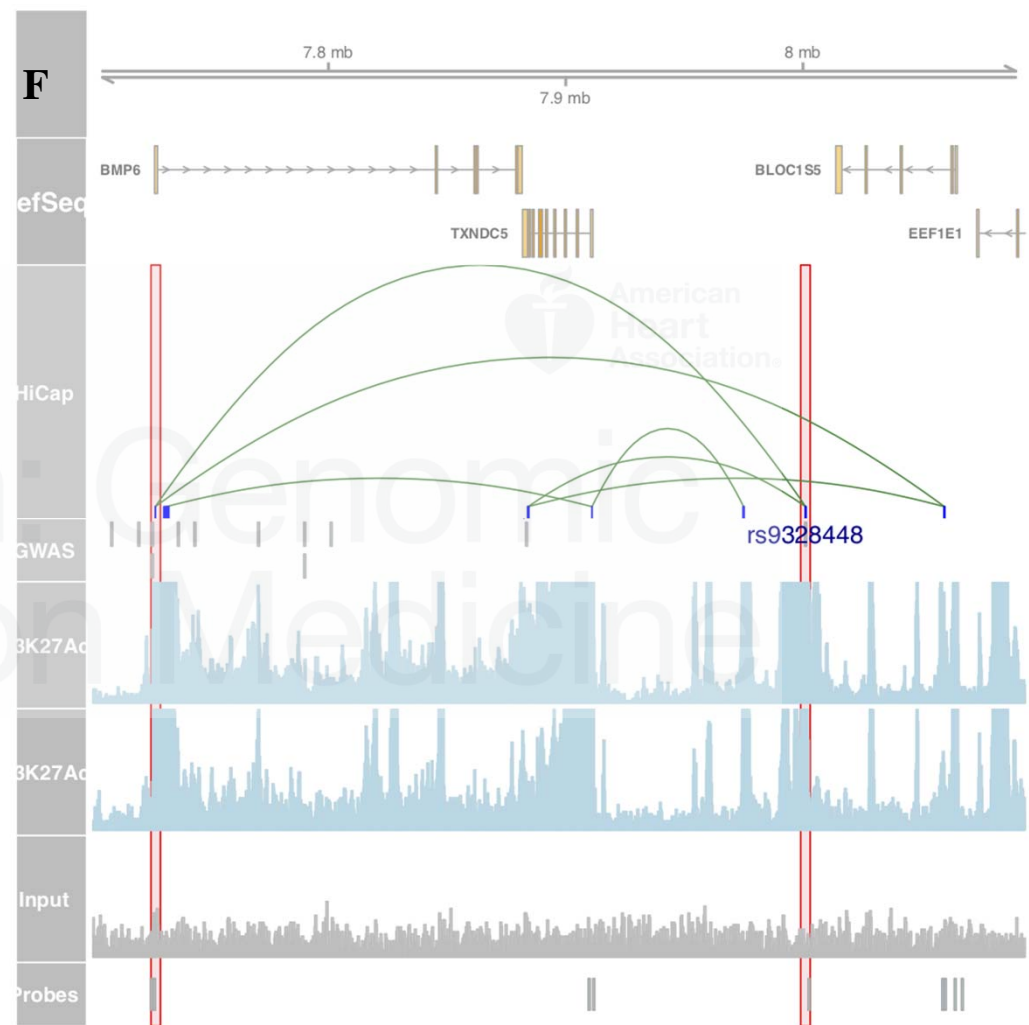
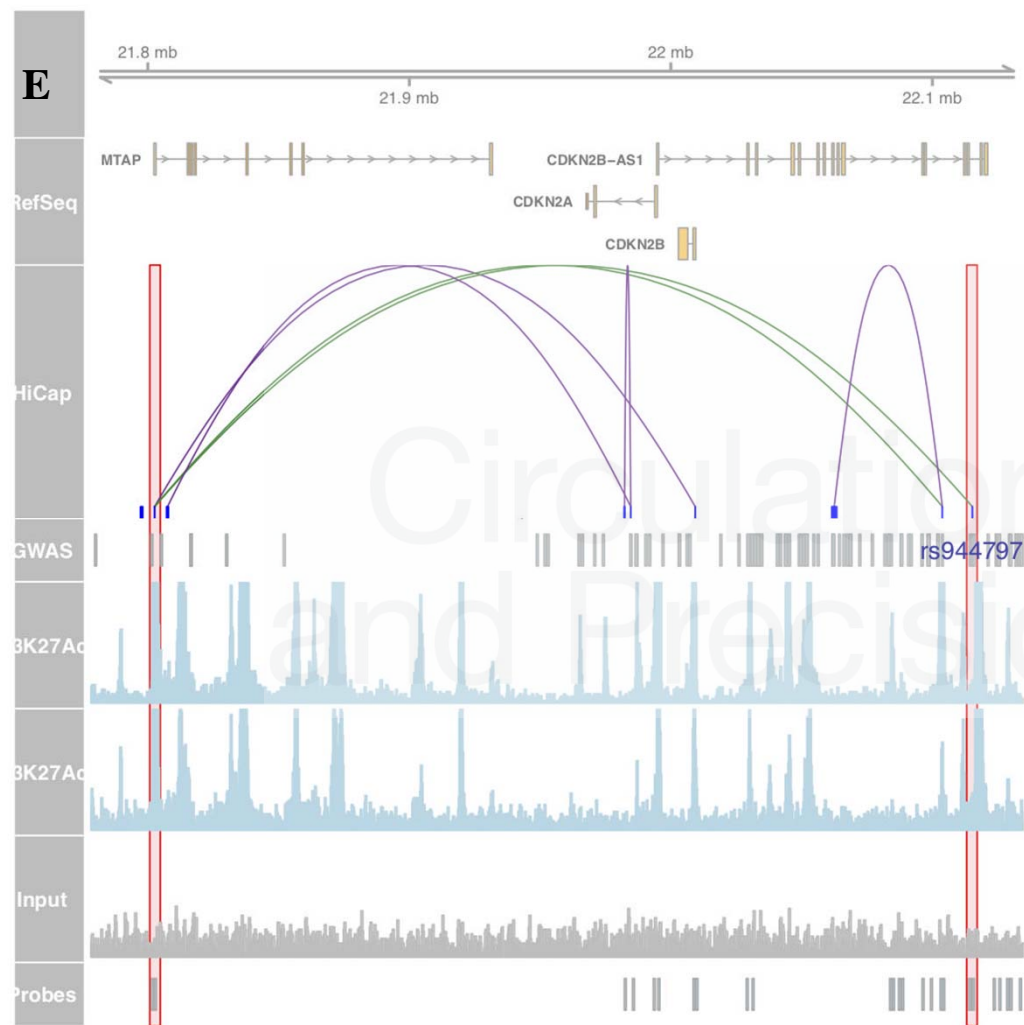
Figure 3. (A) eQTLs contained in the merged CVD GWAS-Prom, Prom-Dist and Prom-Dist_LD datasets plotted versus a size and distance-corrected random set. Deviation from diagonal is present among approximately 30% of the data. (B) GWAS traits that are overrepresented in the interaction datasets. Only traits containing at least 14 variants were taken forward. Fold enrichment is calculated by dividing the actual number of trait variants in the interaction dataset to that of expected (fraction of trait variants in the full trait set). The bar labels denote the fraction of variants found in the interaction datasets. (C) GO Term enrichment analysis of genes interacting with variants or those in LD with CVD_GWAS set using TopGO package. GO terms enriched using only nearest genes to the variants are not reported. Terms with less than five genes and enrichment score less than 0.05 were not included.

Figure 4. Expression correlation between (A) genes MTAP and IFT74 (p-value 4.2×10^{-16}) and (B) genes SMARCAD1 and BMP2K (p-value 2.1×10^{-45}) using aortic intima media expression from 131 individuals. The values on axes are RPKM values and each dot corresponds to each individual where both expression information are obtained from. C,D) Circos plot representation of interactions between rs16998073 and rest of the genome in c) AEC and d) ASMC for comparison. The plot spans chr4:73000000-103000000. The blue track represents gene expression levels, gray boxes represents transcripts, innermost layers represent H3K27Ac marks in AEC cells (two replicates). Purple arcs represent GWAS-Prom or Prom-Prom and orange arcs represent Prom-Dist interactions.



Circulation: Genomic
and Precision Medicine

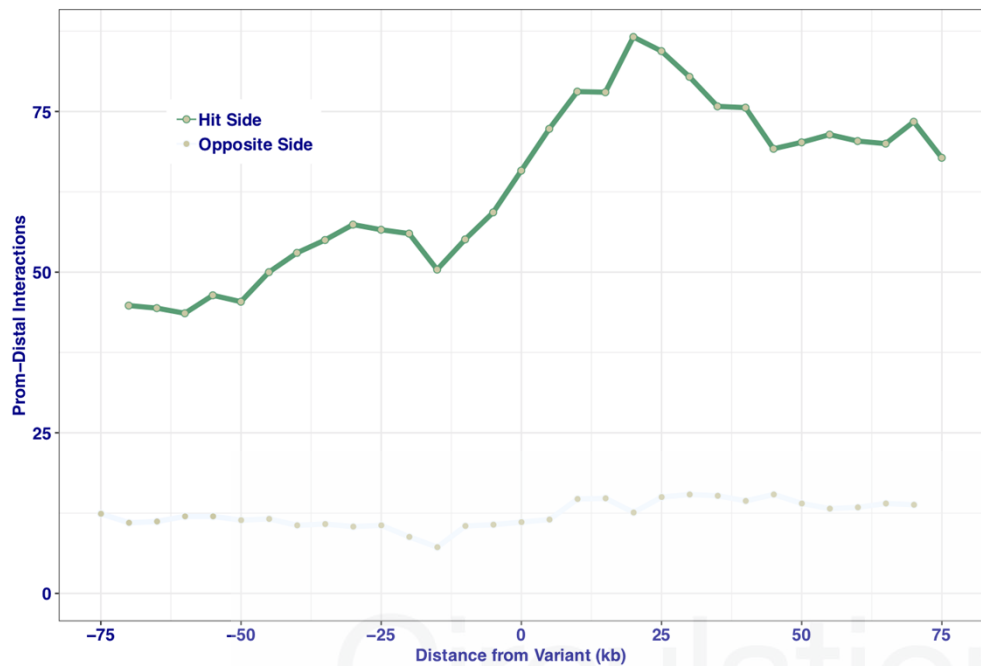




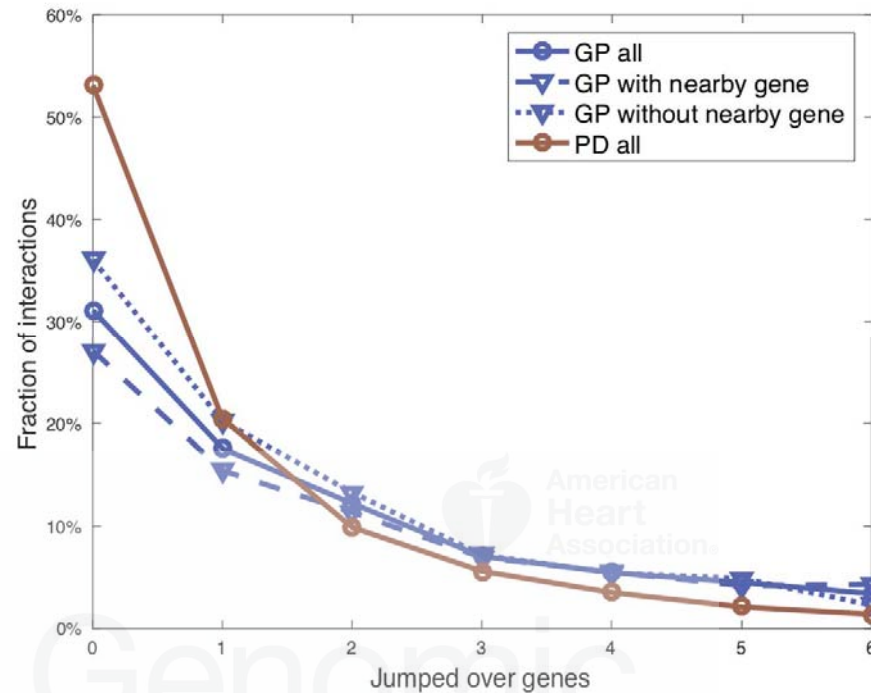
Circulation Genomic and Precision Medicine

American Heart Association

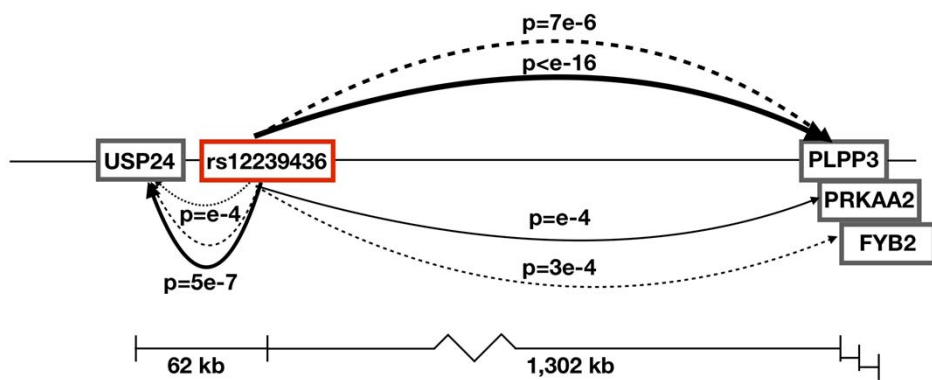
A



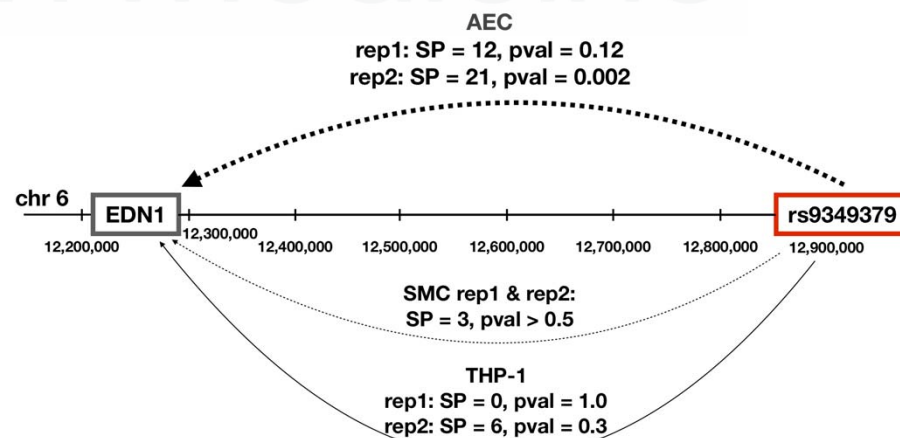
B



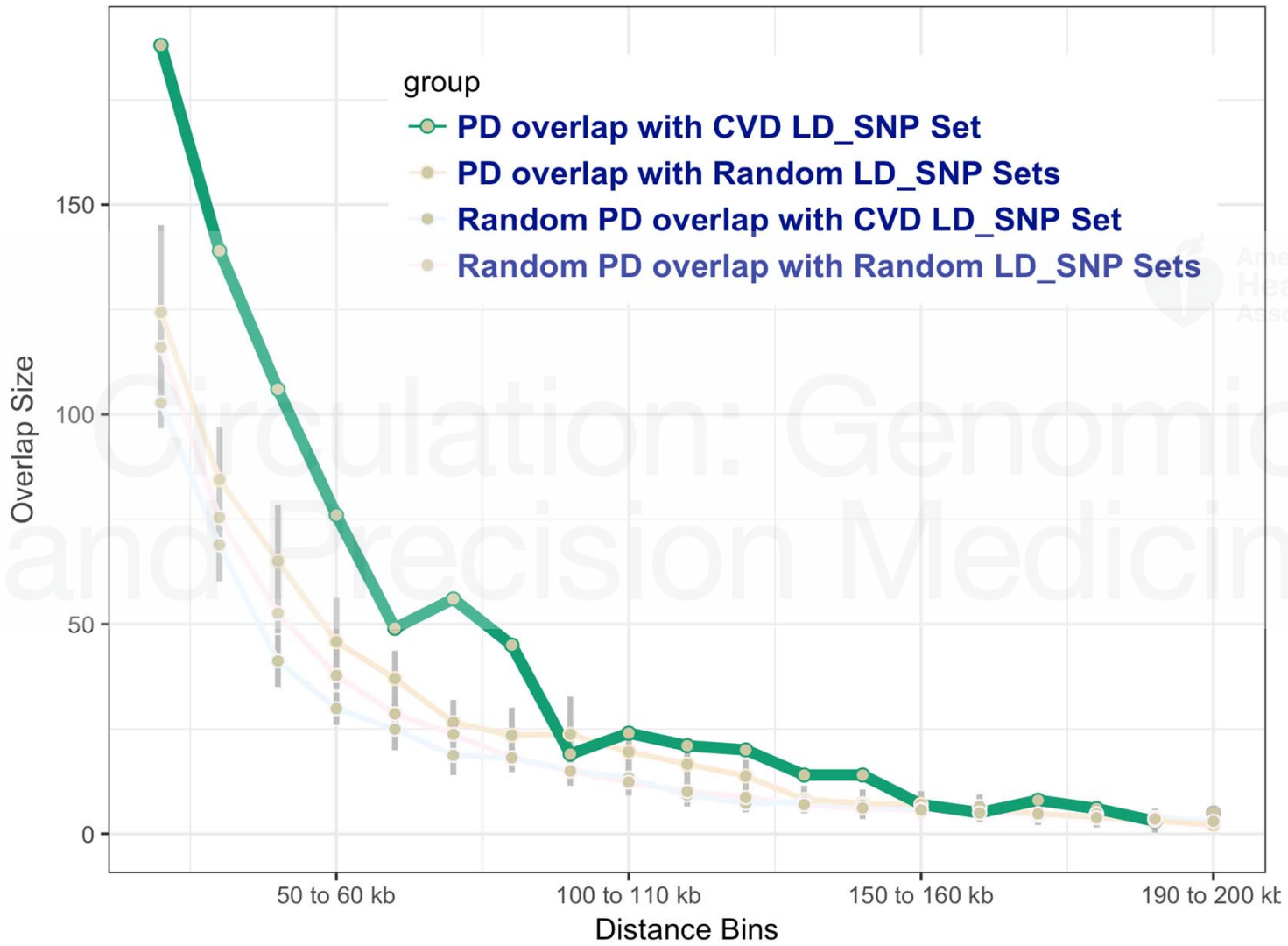
C



D



E



American Heart Association

Circulation: Genomic and Precision Medicine

