5-2020

Using S-Predixcan To Identify Associations Of Genetically Determined Gene Expression Levels With Carotid Intima Media Thickness

Andy Bryan A Castaneda
UTHealth School of Public Health

USING S-PREDIXCAN TO IDENTIFY ASSOCIATIONS OF GENETICALLY DETERMINED GENE EXPRESSION LEVELS WITH CAROTID INTIMA MEDIA THICKNESS

by

ANDY BRYAN CASTANEDA, BA

APPROVED:

HAN CHEN, PHD

PAUL DE VRIES, PHD

USING S-PREDIXCAN TO IDENTIFY ASSOCIATIONS OF GENETICALLY DETERMINED GENE EXPRESSION LEVELS WITH CAROTID INTIMA MEDIA THICKNESS

by

ANDY CASTANEDA
BA, Williams College, 2018

Presented to the Faculty of The University of Texas

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF PUBLIC HEALTH

THE UNIVERSITY OF TEXAS
SCHOOL OF PUBLIC HEALTH
Houston, Texas
May 2020

ACKNOWLEDGEMENTS

I would like to thank Dr. de Vries for his continual support, mentorship, and education throughout the whole process. I want to thank Lauren Petty, Dr. Eric Gamazon, and Dr. Jennifer Below for their collaboration and help with applying S-PrediXcan. A special thanks to the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium Coronary Heart Disease and Subclinical Atherosclerosis Working Group for providing us with the summary statistics of both the GWAS and TWAS that were used in our analysis. Lastly, thank you to Dr. Chen for helping me figure out how to begin this whole process.

USING S-PREDIXCAN TO IDENTIFY ASSOCIATIONS OF GENETICALLY DETERMINED GENE EXPRESSION LEVELS WITH CAROTID INTIMA MEDIA THICKNESS

Andy Bryan Castañeda, MPH, BA
The University of Texas
School of Public Health, 2020

Thesis Chair: Paul de Vries, PhD, MSc

Abstract: The aim of the study was to identify associations between predicted gene expression levels using S-PrediXcan and carotid intima media thickness (cIMT) in order to elucidate the etiology of atherosclerosis. Further, we wanted to compare our results to those from a transcriptome wide association study (TWAS) that used actual gene expression levels in whole blood. For our analysis we used summary statistics from a GWAS of cIMT from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium, tissue-specific gene expression profiles from the GTEx project, and S-PrediXcan prediction models. The TWAS of cIMT that we compared our results to used gene expression levels in peripheral whole blood of 5,647 individuals from the CHARGE consortium. After quality control measures and limiting our results to a single tissue-gene pair per locus in arterial tissues, we identified 4 novel loci that had not previously been associated with cIMT including locus near *RP11-227D13.1*, *CCDC34*, *CD52*, and *RAB6A*. When restricting analysis to whole blood, correlation between predicted gene expression levels using S-

PrediXcan and actual gene expression using TWAS had low correlation (Pearson's r 0.0241).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF APPENDICES

BACKGROUND

Literature Review

In 2010, Cardiovascular Disease (CVD) was the cause of about 31% of all global deaths or over 17.3 million deaths[1]. Generally, CVD refers to a broad category of disease that can lead to narrow or blocked blood vessels[1]. Currently, there is great interest in identifying individuals that are at increased risk for CVD, and more specifically, the development of atherosclerosis that results in coronary heart disease (CHD). Atherosclerosis refers to a general disease process whereby arteries become thick, stiff, or narrow and is usually attributed to buildup of plaque from fatty deposits, calcification, or cholesterol along arterial walls[1, 2]. Atherosclerosis can develop to the point where arteries are blocked and arterial blood flow is compromised increasing one's risk for a myocardial infarction or stroke[2, 3].

Measurement of carotid intima-thickness (cIMT) is a noninvasive method healthcare providers could use to identify the presence of atherosclerotic plaque and quantify one's risk for CVD related events[4]. cIMT measures the thickness of the intima and media walls of the carotid artery which supply the brain with blood. cIMT is thought to provide an earlier sign of atherosclerotic burden compared to measuring coronary artery calcification because pathologic intimal thickening precedes the development of atherosclerotic plaque[1, 5]. cIMT can be used as an early marker for atherosclerosis and has proven useful in helping predict future CVD events in those without traditional risk factors such as diabetes and smoking[4, 6]. Large genome-wide association studies (GWAS) have identified genetic loci that are associated with cIMT and have moderate heritability[7-11]. However, GWAS studies have

some limitations, including multiple testing burdens, modest effect size of SNPs on proportion of heritability, and difficulty identifying causal variants[12-14]. Further, GWAS studies need to utilize a significance threshold that is extremely high that consequently limits GWAS' ability to detect all the heritability explained by SNPs. The SNPs identified as genome-wide significant in metanalyses of GWAS can thus only explain a small fraction of cIMT heritability and the genetic contribution of cIMT development remains inconclusive because of possible overlapping genetic and environmental factors[7, 8, 15]. GWAS studies need to be further explored to understand the underlying mechanism behind these associations.

Gene expression profile studies and Transcriptome-wide association studies (TWAS) represent alternative approaches to GWAS for examining the genomics of a phenotype. They consist of testing clinical phenotypes for associations with the expression levels of genes across the genome[16]. Gene expression profiling using whole-blood for biomarker detection is commonly used in epidemiological research because it can be collected in a relatively non-invasive manner and, in the case of cIMT, it is involved in atherosclerosis development[17]. Even conclusions derived by use of TWAS using whole blood have limitations. Gene and mRNA expression is highly tissue specific, therefore, candidate genes may not reach our significance threshold after switching to a tissue with decreased mechanistic relevance to the phenotypic trait [18, 19]. There may be differential gene expression in more atherosclerotic-relevant tissue that may remain hidden by limiting the analysis to whole blood in expression profiling approaches. Finally, it is difficult to distinguish causal transcriptional effects and

those from reverse causation. For example, whether measured gene expression influences cIMT development and genes whose expression is influenced by changes in cIMT[17, 20].

The GTEx project obtained tissue-specific gene expression profiles from 54 different tissues and nearly 1000 donors to create a "reference transcriptome". The GTEx project serves as a large database of genetic associations with genotypes and mRNA expression that help identify genetic variants associated with gene expression levels, called expression quantitative trait loci (eQTLs)[21]. The reference transcriptomes from studies such as GTEx are used to train elastic net models of genetically regulated gene expression (GReX) levels and build a database of multiple-SNP prediction models. GReX is estimated by weighing reference markers for a gene and by the elastic-net-estimated effect sizes of those markers. These estimates are then stored to build the prediction models, PredictDB, which are publicly available to power methods such as S-PrediXcan. PrediXcan is a relatively novel approach that aims to address the limitations of both GWAS and TWAS. S-PrediXcan can integrate eQTL information to estimate the association of GReX in a range of tissues.[22]. S-PrediXcan does this by using the variance and covariance data of SNPs from the GTEx training set, the beta coefficients from GWAS summary statistics and weights from PredictDB to predict gene expression. The predicted gene expression is then correlated with the trait of interest using regression methods. Using S-PrediXcan can increase the power to detect associations in SNP-based association studies by using a SNP-aggregating approach that incorporates eQTL information while reducing the multiple testing burden that plagues traditional GWAS[22, 23]. In contrast to TWAS studies, S-PrediXcan predicted gene expression levels are a function of genetic variation alone and eliminate reverse causal

effects that would plague RNA sequencing methods while having the advantage of evaluating multiple tissues at a time. Other studies have already used PrediXcan (or the summary statistics version, S-PrediXcan) to find novel gene-trait associations in diverse datasets, expanding findings from TWAS and GWAS, however, this has not yet been done for cIMT[24-27] . In the present study we aim to complement TWAS with S-PrediXcan to better identify genes having genetically predicted expression levels associated with cIMT.

Public Health Significance

Cardiovascular disease is the leading cause of deaths in the developed world and is the underlying cause for approximately 50% of deaths[28]. We know genetic components play a role in the development of coronary heart disease, including atherosclerosis[7]. The development of atherosclerosis is usually preceded by subclinical atherosclerosis. Subclinical atherosclerosis can be detected from cIMT measurement by B-mode ultrasonography and may provide an independent assessment of coronary risk[29]. As a predictor of future coronary risk independent of other traditional risk factors, cIMT can be a useful tool to identify those at risk before the development of atherosclerosis, and eventually, more serious cardiovascular events such as myocardial infarction, stroke, or renal failure[1]. The way a gene is expressed is heritable[30] and can be considered as an intermediate phenotype between genetic variation and disease. PrediXcan can directly test the molecular mechanisms of gene expression by finding genes with expression levels that are correlated with cIMT, and by extension, coronary risk. By exploring the regulatory mechanisms that affect genetic variation at loci associated with cIMT, we can have a better understanding of the underlying

pathophysiology of CHD to guide future research. For example, genes that are found to be positively associated with cIMT using PrediXcan can be favorable targets for future drug development. Although GWAS has largely been successful in identifying countless genetic associations, they rely on large sample sizes to detect small effect sizes. PrediXcan can increase the power to detecting associations in SNP based GWAS studies and thus provides mechanistic insights that can potentially expand and improve gene-based tests.

Hypothesis, Research Question, Specific Aims or Objectives

This project aims to investigate associations between predicted gene expression levels and cIMT in order to uncover the underlying biological mechanisms in the development of atherosclerosis. There is an established relationship between the development of atherosclerosis and the genetic heritability of cIMT. The purpose of this study was to use S-PrediXcan to identify differential gene expression associated with cIMT. Secondly, we wanted to compare associations with predicted gene expression levels using S-PrediXcan to associations with measured gene expression levels.

METHODS

Genome-wide association study of cIMT

We used summary statistics from a meta-analysis of GWAS and cIMT of 71,128 individuals of European ancestry originating from 31 different studies from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. In the studies, cIMT was measured using high-resolution B-mode ultra-sonography[11]. cIMT was defined

as the mean of the maximum of several measurements from both the left and right carotid arteries, and included measurements of the near wall and, in some studies, the far wall. GWAS used in the meta-analysis used standardized protocols for phenotype ascertainment and statistical analyses. Each study used one of several genotyping arrays followed by genotype imputation using the 1000 Genomes Project phase 1 (March 2012) version 3 reference panel. Genomic positions were determined using the Human Genome Reference panel Build 37. Quality control was performed on the GWAS studies and data were excluded with regards to call rates that deviated from Hardy-Weinberg equilibrium, heterozygosity, sex discordance, cryptic relatedness, ethnicity, and non-autosomal SNPs[11].

Within each study, linear regression was then used to assess the association between each genetic variant and cIMT. Genetic variants associated with cIMT were analyzed using a linear regression model (or linear mixed effects models) as well as an additive genetic model of genotype dosage. Analyses were adjusted for age, sex, and various ancestry principal components. In addition, the summary estimates (regression coefficients and standard errors) from each study were then meta-analyzed with an inverse-variance weighted approach. Filters were applied for imputation quality (MACH[31] $r2 < 0.3$ and IMPUTE[32] info <0.4), a minor allele frequency less than 0.01 and for SNPs not present in at least four studies. The genome-wide significance threshold for SNPs in the meta-analysis was set at $p < 5.0 \times 10^{-8}$ [11].

GTEx

The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The data used for the analyses described in this manuscript were obtained from dbGaP accession number phs000424 using the GTEx Analysis V8 release. The GTEx project fund attempts to study the correlation between genetic variation and tissue-specific gene expression. Up to 54 different human tissues were collected from almost 1000 individuals, ranging from 4 in the kidney (medulla) to 706 in skeletal muscle. For our analysis, the minimum tissue-specific sample size was 73 for kidney cortex tissue. DNA and RNA were collected from multiple tissues for molecular assays including genome, exome and RNA sequencing, as well as analyzing data to identify eQTLs [21]. The GTEx project provides us with insight into the functional effects that the phenotype of interest has on genetic variation at the transcript level across 54 different tissue types.

S-PrediXcan

We applied S-PrediXcan to estimate the association of genetically determined gene expression levels with cIMT. When applying PrediXcan we use three different inputs: the study set, the expression training set, and the population reference set. Here, the study set is the summary statistics from GWAS on cIMT described above. The training set from the GTEx consortium consists of genotype data and normalized gene expression data that help train prediction models assign weights to each SNP as well as the predicted gene expression

of a given gene. In our analysis, the GTEx training set was used to compute the variances and covariances (linkage disequilibrium structure) of the markers used in the predicted expression levels. After S-PrediXcan was applied to these three input datasets, we filtered the tissue-specific results to include only genes with a cross-validation correlation > 0.1 and Benjamini-Hochberg correction q-value threshold of < 0.05 to adjust for multiple testing [33]. Using the midpoint of the start and end positions of each gene, we pruned the significant tissue-gene pairs by distance (±1 Mbp) to obtain only the most significant tissue gene pair at each locus.

CHARGE transcriptome-wide association study

Next, we compared the S-PrediXcan results from whole blood with a transcriptome-wide association study (TWAS) of cIMT from the CHARGE consortium[17]. The TWAS analyzed gene expression levels in whole blood using mRNA microarrays, in relation to cIMT from 5,647 individuals. Carotid intima media thickness was measured using B-mode ultrasonography in a similar protocol to the one described above [6, 17]. These associations of gene expression with cIMT were adjusted for age, sex, batch effects, cell counts, RNA quality, fasting, and smoking status.

Comparison of S-PrediXcan results with CHARGE TWAS

We computed pairwise correlations (Pearson's r correlation) of the beta coefficients and p-values between the genes in our TWAS and S-PrediXcan results. Finally, we stratified our pairwise correlation analysis based on quartiles of the S-PrediXcan model fit r squared,

which is a measure of how well the prediction models of S-PrediXcan fit the data, to see if there was a positive association between the strength of the model and correlation of beta coefficients of TWAS and S-PrediXcan results.

Ethics Statement

All studies from which the data were obtained were approved by appropriate research ethics committees or institutional review boards and conducted according to the Declaration of Helsinki developed by the World Medical Association. All participants signed informed consent prior to participation. The UTHSC Committee for Protection of Human Subjects approval for this study can be found in **Appendix B.**

RESULTS

Association analysis using S-PrediXcan

The number of genes that passed quality control and were included in the analysis ranged from 1,633 in kidney cortex tissue to 9,980 in tibial nerve tissue. The number of genes analyzed for each tissue is shown in **Appendix A: Supplementary Table 1** The median predicted performance r2 ranged from 0.064 (IQR: 0.03-0.142) for skeletal muscle tissue to 0.135 (IQR: 0.094-0.213) in kidney cortex tissue (**Appendix A: Supplementary Table 1**). After Benjamini Hochberg correction (q-value < 5%), 86 tissue-gene pairs were significantly associated with cIMT, involving 36 different genes and 33 different tissues. Limiting these results to a single lead tissue-gene pair at each locus, this corresponded to 22 loci that were significantly associated with cIMT. Descriptions of the associated loci are provided in **Appendix A: Supplementary Table 2 & 3**. Out of the 86 tissue-gene pairs

found to be significantly associated with cIMT, 21% (18/86) involved artery related tissues, more than what would be expected by chance (p-value 3.68 x 10^{-6}). In addition, 13 out of the 22 genes have not, to our knowledge, been previously associated with cIMT. A summary of these results are described in **Table 1.**

Table 1: Significant S-PrediXcan associated genes with cIMT across all tissues

Gene name	Most Significant Tissue	Effect Size	P-Value	B. H. P-Value	Novel (Yes/No)
RP1L1	Esophagus Muscularis	0.021	1.55E-10	4.3E-05	Yes
KIAA1462	Artery Tibial	0.039	2.85E-09	0.00018	No
TMEM170A	Breast Mammary Tissue	-0.066	8.06E-09	0.00022	No
CCDC71L	Artery Aorta	0.040	7.74E-09	0.00022	No
CBFA2T3	Testis	-0.155	1.41E-08	0.00032	No
CTD-2541M15.3	Brain Cerebellum	0.014	3.70E-08	0.00068	No
SH3BGRL3	Artery Tibial	-0.061	1.58E-07	0.0018	Yes
DNAH3	Skin (Sun Exposed) Lower Leg	-0.045	6.35E-07	0.0049	Yes
PRAG1	Adipose Visceral Omentum	-0.038	7.84E-07	0.0056	No
LOXL4	Thyroid	0.022	8.44E-07	0.0058	No
RAB6A	Artery Aorta	0.030	3.16E-06	0.016	Yes
SMIM22	Lung	0.061	3.63E-06	0.017	Yes
CAB39L	Artery Aorta	0.020	4.50E-06	0.020	No
PEX3	Esophagus Mucosa	0.052	4.63E-06	0.020	Yes
TLN2	Adipose Subcutaneous	-0.084	7.16E-06	0.029	Yes
DNAJC4	Colon Sigmoid	-0.048	7.36E-06	0.029	Yes
CCDC34	Artery Aorta	-0.055	9.01E-06	0.035	Yes
PHF10	Skin (Not Sun Exposed) Suprapubic	-0.045	1.01E-05	0.037	Yes
RP11-227D13.1	Artery Tibial	-0.022	1.24E-05	0.043	Yes
TCEA3	Adipose Visceral Omentum	-0.031	1.30E-05	0.044	Yes
CDH13	Artery Aorta	0.012	1.53E-05	0.049	No
SH2D4B	Spleen	0.015	1.54E-05	0.049	Yes

We restricted our analysis to include only the tissue-gene pairs that corresponded to arterial tissues (aortic, coronary, and tibial). We then pruned the 18 significant arterial tissue-

gene pairs by distance, resulting in 11 independent loci that were significantly associated with cIMT. 4 of these loci were novel and had not previously been associated with cIMT. The most significant association at each of these loci is described in **Table 2**.

Table 2: S-PrediXcan associated genes with carotid intima media thickness in artery-related tissues

Gene name	Tissue	Effect Size	P-Value	B. H. P-Value	Novel
CCDC71L	Artery Aorta	0.04	7.74E-09	0.00022	No
CD52	Artery Aorta	-0.04	2.70E-07	0.0029	Yes
CBFA2T3	Artery Aorta	0.047	4.00E-07	0.0037	No
KIAA1462	Artery Aorta	0.014	1.09E-06	0.0067	No
CTD-2541M15.3	Artery Tibial	0.01	3.06E-06	0.015	No
RAB6A	Artery Aorta	0.03	3.16E-06	0.016	Yes
BCAR1	Artery Aorta	0.018	4.01E-06	0.018	No
CAB39L	Artery Aorta	0.02	4.50E-06	0.02	No
CCDC34	Artery Aorta	-0.055	9.01E-06	0.035	Yes
RP11-227D13.1	Artery Tibial	-0.022	1.24E-05	0.043	Yes
CDH13	Artery Aorta	0.012	1.53E-05	0.049	No

Correlation of S-PrediXcan and TWAS

We examined the correlation of the regression coefficients (betas) in the CHARGE whole blood TWAS with those from the whole blood component of our S-PrediXcan study. The overall pairwise correlation between beta coefficients was low (Pearson's r: 0.024). When stratified by quartiles of model fit r-squared, the beta coefficients had a mixed upward trend (Pearson's r: Q1: 0.024; Q2: 0.034; Q3:0.011; Q4:0.059) with a dip in correlation observed in the third quantile. Pairwise correlation between p-values of TWAS and PrediXcan was also low (Pearson's r: -0.008). When we stratified by quartiles of model fit r-

squared, the p-values did not show a clear trend (Pearson's r: Q1: 0.009; Q2: -0.015; Q3: -0.004; Q4: -0.019).

We then explored the four novel genes found in arterial tissue of the S-PrediXcan results to see how their association with cIMT compared to those found in the CHARGE TWAS. None of the 4 novel genes found in our S-PrediXcan analysis was significant in the TWAS results (**Appendix A: Supplementary Table 4)**. In addition, only the *FBN1* showed a similar direction of effect as our S-PrediXcan analysis. When exploring the 3 significant genes found in the CHARGE TWAS study of cIMT in our S-PrediXcan analysis, none of the genes were significant when looking across all tissues. A summary of these findings is described in **Appendix A: Supplementary Table 5**.

DISCUSSION

S-PrediXcan

The aim of this study was to use S-PrediXcan to identify differential gene expression associated with cIMT. We found 86 tissue-gene pairs that were significantly associated with cIMT across 36 different tissues and 4 novel loci when restricting to arterial tissues. Results from our S-PrediXcan had low correlation with a previous CHARGE TWAS study. Using S-PrediXcan allowed us to discover new loci while at the same time confirming results from previous GWAS studies done on cIMT. Of the novel loci we identified in arterial tissues, all the genes we identified have plausible biological mechanisms through which they may affect cIMT.

Our marker near the RP11-227D13.1 transcript is within 290 Kbp of the *FBN1* gene, a gene that was not tested in arterial tissue for our analysis, whereas the RP11-227D13.1 transcript was. This could be to differences in model quality between gene arterial-tissue pairs of both the RP11-227D13.1 and *FBN1*, however, this remains unclear. The *FBN1* gene is part of a fibrillin family of glycoproteins found in connective tissue that serve as a structural component in microfibrils. The product of *FBN1* also provides the scaffolding for deposition and crosslinking of elastin, a major structural component of vessel walls that provides load bearing and anchoring functions, and is the most abundant protein in elastic arteries[34, 35]. In addition to the known biological role that *FBN1* plays in connective tissue, $ApoE^{-/-}$ $FBN1^{C1039G+/-}$ knockout mice are a commonly used animal model for atherosclerosis and are characterized by accelerated plaque development and plaque rupture[36]. Finally, the *FBN1* genotype has been associated with artery stiffening in individuals with coronary artery disease and may be involved in the development of atherosclerosis[37, 38]. The *FBN1* gene's involvement in atherosclerotic-prone animal models and its association with arterial stiffening in individuals with coronary artery disease supports the negative association seen in our data. We would expect a higher expression of the *FBN1* gene to be associated with more efficient arterial load bearing, decreased arterial stiffening, and thus, decreased intimal thickening.

CD52 is a small glycosylphosphatidylinositol (GPI)-anchored protein that is highly expressed in lymphatic cells, including $CD4^+$ T-Cells and B cells, as well as monocytes[39]. Research suggest that *CD52* acts as an anti-inflammatory molecule that works to suppress innate immunity cells and reduce the signaling responses by toll-like receptors (TLRs). TLRs

are thought to be critical signaling molecules in the development of atherosclerosis. This is because TLRs are a type of pattern recognition receptor molecules that can respond to non-microbial signals, such as intracellular contents from damaged cells, and thus initiate inflammatory signaling cascades [40]. These inflammatory responses increase the recruitment of neutrophils and macrophages, among other types of cells, to the smooth muscle cells of the intima. Macrophages are usually abundant near "fatty streaks", or a lesion within the smooth muscle cells of the intima. Fatty streaks are one of the earliest signs of atherosclerosis and precedes pathologic intimal thickening[5]. Given that *CD52* inhibits TLR activation and suppresses the inflammatory processes that may be involved in atherosclerotic development, it is no surprise that *CD52* showed a protective effect on cIMT in our study.

RAB6A, found in one of our novel loci, has similarly been implicated in inflammatory cascades that are involved in the development of atherosclerosis. *RAB6A* are trans-Golgi-localized GTPases that are important regulators of pro-inflammatory tumor necrosis factor (TNF) secretion in macrophages[41]. TNFs are part of a family of cytokines that coordinate inflammatory cascades to amplify inflammatory processes. Specifically, TNF promotes the inflammatory cascade within the arterial wall that has been shown to exacerbate endothelial cell injury[42, 43]. Endothelial cell injury allows circulating lipoproteins and monocytes to adhere to arterial wall in mechanisms known to precede the development of atherosclerosis and is consistent with the positive association with cIMT found in our results [5].

CCDC34 is a coiled-coil domain containing protein that was found to be associated with systolic blood pressure in a previous GWAS study[10]. Although blood pressure itself

has been linearly and positively associated with cIMT, it showed a protective effect on cIMT and thus more work is needed to understand the relationship between *CCDC34* and cIMT.

Disparity between S-PrediXcan and TWAS

To our knowledge, this is the first report of a direct comparison between the results of S-PrediXcan and TWAS in the same tissue. There are several possible explanations for why the S-PrediXcan and TWAS results were not strongly correlated. First, beta estimates from PrediXcan only represent the effect of the genetic component of gene expression on cIMT while the beta estimates from TWAS of measured gene expression levels with cIMT likely also include confounding and reverse causation.

Another issue could be that the gene expression data comes from GTEx and blood tissue procurement is pre-mortem for some donors and post-mortem for others. The donor's cause of death leading to tissue procurement can influence the quality of collected tissues [44] and may bias the results. This could contribute to the low correlation observed between the TWAS and S-PrediXcan results, especially considering we limited the comparison to whole blood. Whole blood is already a complicated tissue with multiple cell types and proteins. One study has shown that genes change expression in a tissue-specific manner when comparing pre-mortem donors versus post-mortem donors[45]. Some of these impacts on tissue transcriptomes are due to hypoxia which can trigger specific pathways in the blood [45].

Strengths and Limitations

S-PrediXcan has a lot of advantages compared to traditional GWAS and TWAS. S-PrediXcan has a much lower multiple-testing burden compared to traditional single-variant tests used in GWAS. Compared to TWAS studies looking at gene expression profiling, reverse causation is not a major concern since S-PrediXcan focuses on genetically determined gene expression which is not altered by our phenotype. For the same reason, confounding by dietary and lifestyle factors is also not a major concern with S-PrediXcan.

One of the limitations of our S-PrediXcan approach is linkage disequilibrium (LD) contamination. This would result in S-PrediXcan showing spurious associations with gene expression due to SNPs that are linked to nearby causal SNPs. The developers of S-PrediXcan used COLOC to try and limit issues with LD contamination in S-PrediXcan's prediction models. COLOC is a method to distinguish whether eQTLs and GWAS variants are colocalized or are separate but still linked by linkage disequilibrium[46]. COLOC assumes a single causal variant and consequently has reduced power when multiple causal variants are present; however, there is strong evidence that the architecture of gene expression is relatively sparse [47] and adding a polygenic component to the prediction models might limit its ability to account for LD contamination.

Another limitation is that the statistical power to associations between cIMT and gene expression is constrained by two factors: 1) the sample size of the GTEx training set, which can vary by tissue, and 2) the degree to which gene expression of each gene is genetically determined. Furthermore, since gene expression is correlated across tissues for many genes and given the differences in sample sizes between the tissues, we expect that in some cases

the most significant tissue is not the causal tissue. For example, we may get a better model of expression for a disease-relevant gene in a less disease-relevant tissue if the training set has a larger sample size for that tissue. For this reason, we performed an agnostic scanning of all tissues using S-PrediXcan in addition to a more targeted scan restricted to arterial tissues.

CONCLUSION

In conclusion, we were able to discover novel loci associated with cIMT using S-PrediXcan that harnesses gene expression data from multiple tissues. We discovered 4 novel loci associated with cIMT when limiting to arterial tissues, and 13 novel loci when looking across all tissues. We still do not know the exact mechanism by which the genes in these loci might contribute to the pathologic intimal thickening, but some of the genes found in these loci form part of large inflammatory and immune signaling cascades that tend to be involved in atherosclerotic development.

Although some loci found in GWAS studies overlapped with those found with S-PrediXcan, there was a disparity with the expression data from the CHARGE TWAS in whole blood. This, however, can be considered a strength of our study: our results are not restricted to one tissue type and neither confounding nor reverse causality is a major concern since S-PrediXcan looks at the genetic component of gene expression, and disease status does not alter genomic variation. We demonstrate that S-PrediXcan can identify novel genes, "replicate" those that have been previously discovered, all while providing insights into regulation of gene expression across a wide range of tissues.

APPENDICES

Appendix A: Supplementary Tables

Supplementary Table 1: Predicted Model Fit R^2 and number of genes across 48 tissues

Tissue	Predicted Performance R^2 Median (IQR)	Number of genes
Nerve (Tibial)	0.085 (0.039, 0.186)	9,980
Testis	0.1 (0.048, 0.21)	9,939
Thyroid	0.083 (0.037, 0.186)	9,630
Skin (Sun Exposed, Lower leg)	0.072 (0.034, 0.157)	9,267
Cells (Cultured fibroblasts)	0.09 (0.04, 0.194)	8,905
Skin (Not Sun Exposed, Suprapubic)	0.074 (0.035, 0.159)	8,627
Adipose Subcutaneous	0.072 (0.033, 0.154)	8,624
Artery Tibial	0.077 (0.036, 0.164)	8,594
Esophagus Mucosa	0.075 (0.035, 0.164)	8,493
Esophagus Muscularis	0.077 (0.036, 0.171)	8,199
Lung	0.066 (0.031, 0.142)	7,947
Artery Aorta	0.082 (0.039, 0.175)	7,576
Muscle Skeletal	0.064 (0.03, 0.142)	7,568
Adipose Visceral Omentum	0.067 (0.032, 0.143)	7,309
Whole Blood	0.07 (0.032, 0.159)	7,234
Brain Cerebellum	0.122 (0.061, 0.237)	6,773
Heart Atrial Appendage	0.078 (0.037, 0.16)	6,618
Breast Mammary Tissue	0.068 (0.032, 0.143)	6,436
Colon Transverse	0.076 (0.038, 0.157)	6,285
Esophagus Gastroesophageal Junction	0.084 (0.041, 0.172)	6,272
Colon Sigmoid	0.084 (0.04, 0.173)	6,143
Heart Left Ventricle	0.071 (0.035, 0.149)	5,996
Pancreas	0.092 (0.045, 0.191)	5,880
Spleen	0.108 (0.055, 0.209)	5,755
Brain Cerebellar Hemisphere	0.128 (0.067, 0.239)	5,733
Pituitary	0.093 (0.046, 0.186)	5,665
Brain Cortex	0.102 (0.054, 0.201)	5,477
Stomach	0.075 (0.038, 0.155)	5,137
Brain Caudate basal ganglia	0.1 (0.052, 0.196)	4,980
Brain Nucleus accumbens basal ganglia	0.097 (0.052, 0.187)	4,832
Adrenal Gland	0.096 (0.049, 0.185)	4,825
Brain Frontal Cortex (BA9)	0.104 (0.055, 0.195)	4,537
Prostate	0.09 (0.048, 0.174)	4,286
Artery Coronary	0.087 (0.046, 0.172)	4,032
Liver	0.091 (0.046, 0.182)	3,759
Brain Hippocampus	0.098 (0.053, 0.179)	3,677
Small Intestine (Terminal Ileum)	0.097 (0.055, 0.176)	3,659
Brain Hypothalamus	0.095 (0.052, 0.178)	3,637

Supplementary Table 1: Predicted Model Fit R2 and number of genes across 48 tissues (cont.)

Tissue	Predicted Performance R^2 Median (IQR)	Number of genes
Ovary	**0.105 (0.058, 0.196)**	**3,573**
Brain Anterior cingulate cortex (BA24)	0.109 (0.059, 0.201)	3,532
Brain Spinal cord cervical (C-1)	0.119 (0.068, 0.213)	3,240
Minor Salivary Gland	0.114 (0.065, 0.205)	2,904
Cells (EBV transformed lymphocytes)	0.118 (0.068, 0.207)	2,891
Brain Amygdala	0.108 (0.062, 0.193)	2,773
Vagina	0.099 (0.057, 0.171)	2,549
Brain Substantia nigra	0.113 (0.069, 0.195)	2,547
Uterus	0.109 (0.064, 0.194)	2,528
Kidney Cortex	0.135 (0.094, 0.213)	1,633

Supplementary Table 2: Descriptions of S-PrediXcan associated genes with carotid intima media thickness across all 48 tissues

Gene name	Tissue	Chr	Position	Effect Size	P-Value	Model Fit R	Number of SNPs in model	BH q value
TCEA3	Adipose Visceral Omentum	1	23402829	-0.031	1.30E-05	0.052	34	0.0444
CNKSR1	Colon Transverse	1	26183684	-0.039	6.52E-06	0.081	39	0.0273
SH3BGRL3	Artery Tibial	1	26280804	-0.061	1.58E-07	0.059	33	0.0018
SH3BGRL3	Thyroid	1	26280804	-0.042	1.01E-06	0.039	17	0.0064
SH3BGRL3	Stomach	1	26280804	-0.107	1.34E-06	0.043	4	0.0079
SH3BGRL3	Artery Aorta	1	26280804	-0.065	2.45E-06	0.089	26	0.0130
SH3BGRL3	Heart Atrial Appendage	1	26280804	-0.039	4.37E-06	0.070	15	0.0198
UBXN11	Testis	1	26299846	-0.021	2.95E-06	0.224	63	0.0153
UBXN11	Adipose Subcutaneous	1	26299846	0.044	3.31E-06	0.045	34	0.0161
UBXN11	Lung	1	26299846	0.031	8.46E-06	0.068	49	0.0330
UBXN11	Ovary	1	26299846	0.016	1.17E-05	0.077	68	0.0411
CD52	Artery Aorta	1	26319241	-0.040	2.70E-07	0.038	41	0.0029
CD52	Brain anterior cingulate cortex (BA24)	1	26319241	-0.010	2.99E-06	0.183	39	0.0153
CD52	Nerve tibial	1	26319241	-0.018	1.14E-05	0.064	24	0.0406
PEX3	Esophagus Mucosa	6	1.43E+08	0.052	4.63E-06	0.044	25	0.0203
PHF10	Skin Not Sun Exposed Suprapubic	6	1.7E+08	-0.045	1.01E-05	0.045	14	0.0374
CCDC71L	Artery Aorta	7	1.07E+08	0.040	7.74E-09	0.075	5	0.0002
CCDC71L	Esophagus Muscularis	7	1.07E+08	-0.056	8.10E-08	0.049	9	0.0013
PRKAR2B	Artery Aorta	7	1.07E+08	0.020	1.02E-06	0.092	14	0.0064
CTD-2541M15.3	Brain Cerebellum	8	6616401	0.014	3.70E-08	0.113	29	0.0007
CTD-2541M15.3	Skin Sun Exposed Lower leg	8	6616401	0.025	1.01E-07	0.051	18	0.0015
CTD-2541M15.3	Esophagus Muscularis	8	6616401	0.019	7.28E-07	0.069	15	0.0053
CTD-2541M15.3	Ovary	8	6616401	0.012	1.30E-06	0.166	30	0.0078

Supplementary Table 2: Descriptions of S-PrediXcan associated genes with carotid intima media thickness across all 48 tissues (cont.)

Gene name	Tissue	Chr	Position	Effect Size	P-Value	Model Fit R	Number of SNPs in model	BH q value
CTD-2541M15.3	Artery Tibial	8	6616401	0.010	3.06E-06	0.120	66	0.0154
CTD-2541M15.3	Nerve tibial	8	6616401	0.008	6.73E-06	0.230	37	0.0278
CTD-2541M15.3	Stomach	8	6616401	0.012	8.04E-06	0.049	28	0.0317
CTD-2541M15.3	Brain Amygdala	8	6616401	0.009	9.34E-06	0.160	48	0.0354
PRAG1	Adipose Visceral Omentum	8	8352088	-0.038	7.84E-07	0.036	13	0.0056
RP1L1	Esophagus Muscularis	8	10659268	0.021	1.55E-10	0.061	16	0.0000
RP1L1	Adipose Visceral Omentum	8	10659268	0.061	5.46E-10	0.042	4	0.0001
RP1L1	Nerve tibial	8	10659268	0.018	3.27E-09	0.132	25	0.0002
RP1L1	Skin Sun Exposed Lower leg	8	10659268	0.025	4.43E-09	0.083	14	0.0002
RP1L1	Lung	8	10659268	0.032	7.79E-09	0.050	14	0.0002
RP1L1	Esophagus Mucosa	8	10659268	0.012	1.24E-08	0.233	32	0.0003
RP1L1	Vagina	8	10659268	0.016	7.54E-08	0.134	30	0.0013
RP1L1	Breast Mammary Tissue	8	10659268	0.022	1.33E-07	0.077	12	0.0017
RP1L1	Adipose Subcutaneous	8	10659268	0.017	1.45E-07	0.117	14	0.0017
RP1L1	Thyroid	8	10659268	0.014	4.70E-07	0.046	61	0.0041
SOX7	Brain Cerebellar Hemisphere	8	10727140	-0.026	6.58E-09	0.060	29	0.0002
SOX7	Thyroid	8	10727140	0.021	1.32E-05	0.032	28	0.0444
MTMR9	Vagina	8	11306481	0.103	1.12E-09	0.066	13	0.0001
SLC35G5	Esophagus Gastroesophageal Junction	8	11331548	-0.028	1.11E-07	0.038	13	0.0015
FAM167A	Brain Cortex	8	11448692	-0.037	6.64E-07	0.032	24	0.0050
FAM167A	Brain Frontal Cortex (BA9)	8	11448692	-0.060	1.65E-06	0.079	7	0.0095
BLK	Brain Cerebellar Hemisphere	8	11525747	0.028	9.87E-06	0.054	8	0.0369
DEFB134	Pituitary	8	11994743	0.037	3.69E-06	0.031	7	0.0173
DEFB134	Brain Cerebellar Hemisphere	8	11994743	0.007	1.12E-05	0.399	28	0.0402

Supplementary Table 2: Descriptions of S-PrediXcan associated genes with carotid intima media thickness across all 48 tissues (cont.)

Gene name	Tissue	Chr	Position	Effect Size	P-Value	Model Fit R	Number of SNPs in model	BH q value
SVIL	Artery Tibial	10	29597060	0.018	1.03E-05	0.235	58	0.0376
SVIL	Artery Aorta	10	29597060	0.023	1.36E-05	0.212	13	0.0453
KIAA1462	Artery Tibial	10	30064149	0.039	2.85E-09	0.099	41	0.0002
KIAA1462	Artery Aorta	10	30064149	0.014	1.09E-06	0.318	32	0.0067
SH2D4B	Spleen	10	80592231	0.015	1.54E-05	0.058	17	0.0494
R3HCC1L	Pituitary	10	98189761	-0.014	6.21E-06	0.140	20	0.0264
LOXL4	Thyroid	10	98257942	0.022	8.44E-07	0.061	14	0.0058
CCDC34	Artery Aorta	11	27347031	-0.055	9.01E-06	0.029	5	0.0346
DNAJC4	Colon Sigmoid	11	64232282	-0.048	7.36E-06	0.022	31	0.0295
RAB6A	Artery Aorta	11	73718388	0.030	3.16E-06	0.147	43	0.0156
CAB39L	Artery Aorta	13	49376357	0.020	4.50E-06	0.099	39	0.0201
CAB39L	Liver	13	49376357	0.012	5.83E-06	0.144	63	0.0252
RP11-227D13.1	Artery Tibial	15	48648984	-0.022	1.24E-05	0.049	5	0.0427
RP11-227D13.1	Breast Mammary Tissue	15	48648984	-0.019	1.53E-05	0.107	3	0.0494
TLN2	Adipose Subcutaneous	15	62617579	-0.084	7.16E-06	0.040	5	0.0291
SMIM22	Lung	16	4792444	0.061	3.63E-06	0.016	7	0.0173
DNAH3	Skin Sun Exposed Lower leg	16	21046276	-0.045	6.35E-07	0.014	20	0.0049
BCAR1	Esophagus Mucosa	16	75248117	-0.013	4.35E-07	0.242	19	0.0039
BCAR1	Brain Cerebellum	16	75248117	-0.038	2.10E-06	0.048	12	0.0116
BCAR1	Artery Tibial	16	75248117	0.022	2.45E-06	0.151	44	0.0130
BCAR1	Artery Aorta	16	75248117	0.018	4.01E-06	0.237	79	0.0185
CFDP1	Brain Cerebellar Hemisphere	16	75363601	-0.047	9.90E-08	0.064	16	0.0015
CFDP1	Brain Hippocampus	16	75363601	-0.047	2.78E-07	0.049	15	0.0029

Supplementary Table 2: Descriptions of S-PrediXcan associated genes with carotid intima media thickness across all 48 tissues (cont.)

Gene name	Tissue	Chr	Position	Effect Size	P-Value	Model Fit R	Number of SNPs in model	BH q value
CFDP1	Brain Cerebellum	16	75363601	-0.026	3.38E-07	0.163	47	0.0032
CFDP1	Cells Cultured (fibroblasts)	16	75363601	-0.011	9.46E-07	0.403	44	0.0064
CFDP1	Muscle Skeletal	16	75363601	-0.034	2.03E-06	0.070	17	0.0114
TMEM170A	Breast Mammary Tissue	16	75454276	-0.066	8.06E-09	0.042	11	0.0002
TMEM170A	Muscle Skeletal	16	75454276	-0.097	3.02E-08	0.029	6	0.0006
TMEM170A	Cells Cultured (fibroblasts)	16	75454276	-0.028	1.31E-07	0.091	34	0.0017
TMEM170A	Esophagus Muscularis	16	75454276	-0.073	5.47E-07	0.049	11	0.0045
TMEM170A	Thyroid	16	75454276	-0.043	6.16E-07	0.057	5	0.0049
TMEM170A	Esophagus Gastroesophageal Junction	16	75454276	-0.055	9.71E-07	0.037	19	0.0064
CDH13	Artery Aorta	16	83213803	0.012	1.53E-05	0.321	24	0.0494
CBFA2T3	Testis	16	88926033	-0.155	1.41E-08	0.028	5	0.0003
CBFA2T3	Pancreas	16	88926033	0.016	2.76E-08	0.273	15	0.0006
CBFA2T3	Thyroid	16	88926033	0.066	2.95E-07	0.050	2	0.0029
CBFA2T3	Esophagus Gastroesophageal Junction	16	88926033	0.078	2.94E-07	0.039	7	0.0029
CBFA2T3	Artery Aorta	16	88926033	0.047	4.00E-07	0.037	12	0.0037
CBFA2T3	Heart Atrial Appendage	16	88926033	0.046	5.31E-07	0.065	4	0.0045

Supplementary Table 4: Association of the 4 novel genes from PrediXcan analysis S-PrediXcan significant (BH < 0.05) genes in the TWAS association analysis results

Probe ID	Gene name	Effect Size	Standard Error	P-Value
ILMN_2288784	*CCDC34*	-0.0586	0.0716	0.4129
ILMN_1657547	*CCDC34*	0.0477	0.0665	0.4728
ILMN_1803490	*CCDC34*	0.0064	0.0738	0.9309
ILMN_2208903	*CD52*	0.0498	0.0814	0.5405
ILMN_1783182	*FBN1*	-0.0616	0.0901	0.4941
ILMN_2377862	*RAB6A*	-0.0492	0.0803	0.5398
ILMN_1800871	*RAB6A*	0.0256	0.0723	0.7233

Supplementary Table 5: Association of the significant genes from the TWAS association analysis in S-PrediXcan significant (BH q-value < 0.05) genes across all tissues.

Gene name	Most Significant Tissue	Effect Size	P-Value	B. H. P-Value	Predicted Performance R^2
CEBPD	Artery Tibial	0.0088	0.021	0.980	0.013
METRNL	Artery Tibial	-0.0029	0.039	0.982	0.042
METRNL	Brain Caudate basal ganglia	0.0070	-0.066	0.925	0.074
METRNL	Brain Cortex	0.0074	0.040	0.956	0.075
METRNL	Brain Substantia nigra	-0.0009	-0.155	0.995	0.086
METRNL	Heart Left Ventricle	-0.0036	0.014	0.979	0.014
CEBPD	Heart Left Ventricle	0.0075	-0.061	0.988	0.018
METRNL	Lung	-0.0063	-0.045	0.973	0.017
METRNL	Pituitary	-0.0005	-0.038	0.998	0.094
METRNL	Skin Not Sun Exposed Suprapubic	0.0007	0.022	1.000	0.065
METRNL	Small Intestine Terminal Ileum	-0.0133	0.030	0.917	0.078
METRNL	Thyroid	0.0014	0.061	0.998	0.032

Appendix B: UTHSC Committee for Protection of Human Subjects Approval

UTHealth
The University of Texas
Health Science Center at Houston

Committee for the Protection of Human Subjects
6410 Fannin Street, Suite 1100
Houston, Texas 77030

Dr. Andy Castaneda
UT-H - SPH - Environ & Occup Health Science

January 28, 2020

HSC-SPH-20-0094 - *COMPARING GENE EXPRESSION VARIATION OF CAROTID INTIMA MEDIA THICKNESS IMPUTED FROM GWAS SUMMARY STATISTICS AGAINST GENE EXPRESSION FROM TRANCRIPTOME WIDE ASSOCIATION STUDIES*

The above named project is determined to qualify for exempt status according to 45 CFR 46.101(b)

b. federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

***CATEGORY #4** : Research, involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified directly or through identifiers linked to the subjects.*

CHANGES: Should you choose to make any changes to the protocol that would involve the inclusion of human subjects or identified data from humans, please submit the change via iRIS to the Committee for the Protection of Human Subjects for review.

INFORMED CONSENT DETERMINATION:
Waiver of Consent Granted

INFORMED CONSENT: When Informed consent is required, it must be obtained by the PI or designee(s), using the format and procedures approved by the CPHS. The PI is responsible to instruct the designee in the methods approved by the CPHS for the consent process. The individual obtaining informed consent must also sign the consent document. Please note that only copies of the stamped approved informed consent form can be used when obtaining consent.

HEALTH INSURANCE PORTABILITY and ACCOUNTABILITY ACT (HIPAA):

Exempt from HIPAA

STUDY CLOSURES: Upon completion of your project, submission of a study closure report is required. The study closure report should be submitted once all data has been collected and analyzed.

Should you have any questions, please contact the Office of Research Support Committees at 713-500-7943.

REFERENCES

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, Das SR, de Ferranti S, Despres JP, Fullerton HJ *et al*: **Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association**. *Circulation* 2016, **133**(4):e38-360.
2. **Arteriosclerosis/Artherosclerosis: Overview** [https://www.mayoclinic.org/diseases-conditions/arteriosclerosis-atherosclerosis/symptoms-causes/syc-20350569]
3. **Know the Difference Fact Sheet: Cardiovascular disease, Heart disease, and Coronary heart disease** [https://www.nhlbi.nih.gov/sites/default/files/media/docs/Fact_Sheet_Know_Diff_Design.508_pdf.pdf]
4. Stein JH, Korcarz CE, Hurst RT, Lonn E, Kendall CB, Mohler ER, Najjar SS, Rembold CM, Post WS: **Use of carotid ultrasound to identify subclinical vascular disease and evaluate cardiovascular disease risk: a consensus statement from the American Society of Echocardiography Carotid Intima-Media Thickness Task Force. Endorsed by the Society for Vascular Medicine**. *J Am Soc Echocardiogr* 2008, **21**(2):93-111; quiz 189-190.
5. Sakakura K, Nakano M, Otsuka F, Ladich E, Kolodgie FD, Virmani R: **Pathophysiology of Atherosclerosis Plaque Progression**. *Heart, Lung and Circulation* 2013, **22**(6):399-411.
6. Salonen JT, Salonen R: **Ultrasonographically assessed carotid morphology and the risk of coronary heart disease**. *Arteriosclerosis and thrombosis : a journal of vascular biology* 1991, **11**(5):1245-1249.
7. Bis JC, Kavousi M, Franceschini N, Isaacs A, Abecasis GR, Schminke U, Post WS, Smith AV, Cupples LA, Markus HS *et al*: **Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque**. *Nature Genetics* 2011, **43**(10):940-947.
8. Swan L, Birnie DH, Inglis G, Connell JM, Hillis WS: **The determination of carotid intima medial thickness in adults--a population-based twin study**. *Atherosclerosis* 2003, **166**(1):137-141.
9. Sayed-Tabatabaei FA, van Rijn MJ, Schut AF, Aulchenko YS, Croes EA, Zillikens MC, Pols HA, Witteman JC, Oostra BA, van Duijn CM: **Heritability of the function and structure of the arterial wall: findings of the Erasmus Rucphen Family (ERF) study**. *Stroke* 2005, **36**(11):2351-2356.
10. Kichaev G, Bhatia G, Loh P-R, Gazal S, Burch K, Freund MK, Schoech A, Pasaniuc B, Price AL: **Leveraging Polygenic Functional Enrichment to Improve GWAS Power**. *Am J Hum Genet* 2019, **104**(1):65-75.
11. Franceschini N, Giambartolomei C, de Vries PS, Finan C, Bis JC, Huntley RP, Lovering RC, Tajuddin SM, Winkler TW, Graff M *et al*: **GWAS and colocalization**

analyses implicate carotid intima-media thickness and carotid plaque loci in cardiovascular outcomes. *Nat Commun* 2018, **9**(1):5141.
12. Manolio TA: **Bringing genome-wide association findings into clinical use**. *Nat Rev Genet* 2013, **14**(8):549-558.
13. Dudbridge F, Gusnanto A: **Estimation of significance thresholds for genomewide association scans**. *Genet Epidemiol* 2008, **32**(3):227-234.
14. Altshuler D, Daly MJ, Lander ES: **Genetic mapping in human disease**. *Science* 2008, **322**(5903):881-888.
15. Rampersaud E, Bielak LF, Parsa A, Shen H, Post W, Ryan KA, Donnelly P, Rumberger JA, Sheedy PF, 2nd, Peyser PA *et al*: **The association of coronary artery calcification and carotid artery intima-media thickness with distinct, traditional coronary artery disease risk factors in asymptomatic adults**. *American journal of epidemiology* 2008, **168**(9):1016-1023.
16. Taurino C, Miller WH, McBride MW, McClure JD, Khanin R, Moreno MU, Dymott JA, Delles C, Dominiczak AF: **Gene expression profiling in whole blood of patients with coronary artery disease**. *Clin Sci (Lond)* 2010, **119**(8):335-343.
17. de Vries P: **Hemostasis and Cardiovascular Disease : a molecular epidemiology approach**. *Ph.D. thesis.* Erasmus University Rotterdam; 2016.
18. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K *et al*: **Opportunities and challenges for transcriptome-wide association studies**. *Nature Genetics* 2019, **51**(4):592-599.
19. Johnson BR, Atallah J, Plachetzki DC: **The importance of tissue specificity for RNA-seq: highlighting the errors of composite structure extractions**. *BMC Genomics* 2013, **14**(1):586.
20. Relton CL, Davey Smith G: **Epigenetic epidemiology of common complex disease: prospects for prediction, prevention, and treatment**. *PLoS Med* 2010, **7**(10):e1000356-e1000356.
21. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, Compton CC, DeLuca DS, Peter-Demchok J, Gelfand ET *et al*: **A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project**. *Biopreserv Biobank* 2015, **13**(5):311-319.
22. Gamazon ER, Wheeler HE, Shah KP, Mozaffari SV, Aquino-Michaels K, Carroll RJ, Eyler AE, Denny JC, Consortium GT, Nicolae DL *et al*: **A gene-based association method for mapping traits using reference transcriptome data**. *Nature Genetics* 2015, **47**:1091.
23. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL *et al*: **Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics**. *Nat Commun* 2018, **9**(1):1825-1825.
24. Pasman JA, Verweij KJH, Gerring Z, Stringer S, Sanchez-Roige S, Treur JL, Abdellaoui A, Nivard MG, Baselmans BML, Ong J-S *et al*: **Genome-wide association analysis of lifetime cannabis use (N=184,765) identifies new risk loci,**

genetic overlap with mental health, and a causal influence of schizophrenia on cannabis use. *bioRxiv* 2018:234294.
25. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, Torstenson ES, Kovesdy CP, Sun YV, Wilson OD *et al*: **Trans-ethnic association study of blood pressure determinants in over 750,000 individuals**. *Nature Genetics* 2019, **51**(1):51-62.
26. Sanchez-Roige S, Fontanillas P, Elson SL, Pandit A, Schmidt EM, Foerster JR, Abecasis GR, Gray JC, de Wit H, Davis LK *et al*: **Genome-wide association study of delay discounting in 23,217 adult research participants of European ancestry**. *Nature Neuroscience* 2018, **21**(1):16-18.
27. Petty LE, Highland HM, Gamazon ER, Hu H, Karhade M, Chen H-H, de Vries PS, Grove ML, Aguilar D, Bell GI *et al*: **Functionally oriented analysis of cardiometabolic traits in a trans-ethnic sample**. *Hum Mol Genet* 2019, **28**(7):1212-1224.
28. Ross R: **The pathogenesis of atherosclerosis: a perspective for the 1990s**. *Nature* 1993, **362**(6423):801-809.
29. Fuster V, Gotto AM: **Risk Reduction**. *Circulation* 2000, **102**(suppl_4):Iv-94-Iv-102.
30. Price AL, Helgason A, Thorleifsson G, McCarroll SA, Kong A, Stefansson K: **Single-Tissue and Cross-Tissue Heritability of Gene Expression Via Identity-by-Descent in Related or Unrelated Individuals (Heritability of Gene Expression across Tissues)**. *PLoS Genetics* 2011, **7**(2):e1001317.
31. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR: **Fast and accurate genotype imputation in genome-wide association studies through pre-phasing**. *Nat Genet* 2012, **44**(8):955-959.
32. Howie BN, Donnelly P, Marchini J: **A flexible and accurate genotype imputation method for the next generation of genome-wide association studies**. *PLoS Genet* 2009, **5**(6):e1000529.
33. Storey JD, Tibshirani R: **Statistical significance for genomewide studies**. *Proc Natl Acad Sci U S A* 2003, **100**(16):9440-9445.
34. Van Herck LJ, De Meyer RYG, Martinet EW, Van Hove HC, Foubert JK, Theunis GM, Apers GS, Bult GH, Vrints GC, Herman GA: **Impaired Fibrillin-1 Function Promotes Features of Plaque Instability in Apolipoprotein E–Deficient Mice**. *Circulation* 2009, **120**(24):2478-2487.
35. Kielty CM, Sherratt MJ, Shuttleworth CA: **Elastic fibres**. *Journal of Cell Science* 2002, **115**(14):2817-2828.
36. Emini Veseli B, Perrotta P, De Meyer GRA, Roth L, Van der Donckt C, Martinet W, De Meyer GRY: **Animal models of atherosclerosis**. *European Journal of Pharmacology* 2017, **816**:3-13.
37. Medley LT, Cole JT, Gatzka DC, Wang YSW, Dart MA, Kingwell AB: **Fibrillin-1 Genotype Is Associated With Aortic Stiffness and Disease Severity in Patients With Coronary Artery Disease**. *Circulation: Journal the of American Heart Association* 2002, **105**(7):810-815.

38. Dart AM, Kingwell BA: **Pulse pressure--a review of mechanisms and clinical relevance**. *J Am Coll Cardiol* 2001, **37**(4):975-984.
39. Rashidi M, Bandala-Sanchez E, Lawlor KE, Zhang Y, Neale AM, Vijayaraj SL, O'Donoghue R, Wentworth JM, Adams TE, Vince JE *et al*: **CD52 inhibits Toll-like receptor activation of NF-κB and triggers apoptosis to suppress inflammation**. *Cell Death & Differentiation* 2018, **25**(2):392-405.
40. Chen GY, Nuñez G: **Sterile inflammation: sensing and reacting to damage**. *Nature Reviews Immunology* 2010, **10**(12):826-837.
41. Micaroni M, Stanley AC, Khromykh T, Venturato J, Wong CXF, Lim JP, Marsh BJ, Storrie B, Gleeson PA, Stow JL: **Rab6a/a' are important Golgi regulators of pro-inflammatory TNF secretion in macrophages**. *PloS one* 2013, **8**(2):e57034-e57034.
42. McKellar GE, McCarey DW, Sattar N, McInnes IB: **Role for TNF in atherosclerosis? Lessons from autoimmune disease**. *Nature Reviews Cardiology* 2009, **6**(6):410-417.
43. Zhang H, Park Y, Wu J, Chen Xiu p, Lee S, Yang J, Dellsperger Kevin C, Zhang C: **Role of TNF-α in vascular dysfunction**. *Clinical Science* 2008, **116**(3):219-230.
44. Vennemann M, Koppelkamm A: **mRNA profiling in forensic genetics I: Possibilities and limitations**. *Forensic Science International* 2010, **203**(1):71-75.
45. Ferreira PG, Muñoz-Aguirre M, Reverter F, Sá Godinho CP, Sousa A, Amadoz A, Sodaei R, Hidalgo MR, Pervouchine D, Carbonell-Caballero J *et al*: **The effects of death and post-mortem cold ischemia on human tissue transcriptomes**. *Nat Commun* 2018, **9**(1):490.
46. Giambartolomei C, Vukcevic D, Schadt EE, Franke L, Hingorani AD, Wallace C, Plagnol V: **Bayesian test for colocalisation between pairs of genetic association studies using summary statistics**. *PLoS Genet* 2014, **10**(5):e1004383.
47. Wheeler HE, Shah KP, Brenner J, Garcia T, Aquino-Michaels K, Consortium GT, Cox NJ, Nicolae DL, Im HK: **Survey of the Heritability and Sparse Architecture of Gene Expression Traits across Human Tissues**. *PLoS genetics* 2016, **12**(11):e1006423-e1006423.