Lipoprotein(a) as a Genetic Risk Factor for Aortic Stenosis Across Ethnicities

Mary Hoekstra

Division of Experimental Medicine

McGill University

Montreal, Quebec, Canada

August, 2020

A thesis submitted to McGill University

in partial fulfillment of the requirements of the degree of

Master of Science in Experimental Medicine.

Table of Contents

Abstract

Background: Aortic stenosis (AS) is a heart valve disease that is becoming more prevalent with the aging population. Genome-wide association studies (GWAS) of AS in European-ancestry populations have identified the variant rs10455872 in *LPA* as associated with greater disease risk, an association that has been replicated in both African Americans and Hispanics. This association is mediated by elevated levels of lipoprotein(a) (Lp[a]), which are primarily determined by variation at the *LPA* locus. The identification of additional risk loci for Lp(a) levels could reveal new targets for Lp(a)-lowering therapies.

Apart from *LPA*, several other risk loci have been identified for AS; however, a genetic risk score (GRS) has never been created to assess the contribution of multiple genetic variants to this disease. Given that Lp(a) is a risk factor for AS with strong genetic determinants, a GRS for Lp(a) may also be effective at explaining variance in AS.

Methods: A large-scale GWAS for Lp(a) was performed in White British Individuals from the UK Biobank (N=293,274). Approximately 93,095,623 variants were tested for association with natural log-transformed Lp(a) levels using linear regression models adjusted for age, sex, genotype batch, and 20 principal components of genetic ancestry. Independent variants reaching genome-wide significance ($P \leq 5 \times 10^{-8}$) were tested for association in a meta-analysis of two European-ancestry cohorts, as well as additional ethnic groups in the UK Biobank. Results from the Lp(a) GWAS and a previously-performed AS GWAS were used to develop genetic risk scores (GRSs) in Non-Hispanic Whites (n=55,192), African Americans (n=1,917), and Hispanics (n=3,482) from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, where the

p-value thresholds for the scores were selected based on maximal association with AS in these populations. The proportion of variance explained in AS by the GRSs was assessed in each ethnicity, with and without the most significant variant, rs10455872. **Results:** A total of 131 independent variants were associated with Lp(a) at genome-wide significance, validating previous associations at *LPA*, *APOE*, and *CETP,* and identifying a novel association at *APOH.* The variant rs8178824 at *APOH*, encoding beta2-glycoprotein I (β2GPI), was associated with increased Lp(a) levels (β [95% CI] (ln nmol/L), 0.064 [0.047, 0.081]; $P = 2.8 \times 10^{-13}$). This association was replicated in a meta-analysis of 5,465 European-ancestry individuals from the Framingham Offspring Study and Multi-Ethnic Study of Atherosclerosis (β [95% CI] (ln mg/dL), 0.16 [0.044, 0.28]; $P = 0.0071$) but failed to replicate in other ethnicities from the UK Biobank (all $P>0.05$). Genetic risk scores constructed from the Lp(a) GWAS contained 263, 1,291, and 11,217 variants when maximized for association in GERA Whites, African Americans, and Hispanics, respectively, while GRSs constructed from an AS GWAS in the same manner contained 3, 10, and 496 variants, respectively. The best-fit (by ethnicity) Lp(a)-GRSs were significantly associated with AS in each ethnicity, but explained a small proportion of variance in AS (Whites: R^2=0.13%, $P= 6.2 \times 10^{-9}$; African Americans: R^2=0.56%, P=0.045; Hispanics: R^2=0.31%, P=0.045). The best-fit (by ethnicity) AS-GRSs explained only marginally more phenotypic variance than the Lp(a)-GRSs (Whites: R^2=0.16%, $P=1.1 \times 10^{-10}$; African Americans: R^2=0.63%, P=0.033; Hispanics: R^2=0.39%, P=0.024). When rs10455872 was removed from the Lp(a)-GRSs, the variance explained decreased to 0.076% in Whites, 0.25% in African Americans, and 0.26% in Hispanics. Similarly, excluding rs10455872 from the AS-GRSs decreased

the variance explained in Whites (0.063%), African Americans (0.13%), and Hispanics (0.37%).

Conclusions: A large-scale GWAS of Lp(a) levels revealed *APOH* as a novel locus for Lp(a) in individuals of European ancestry, highlighting β2GPI as a potential therapeutic target. An Lp(a)-GRS derived from this GWAS was associated with AS to nearly the same extent as an AS-GRS when assessed in Whites, African Americans, and Hispanics. Both types of GRSs explained only a small proportion of variance in AS, and the *LPA* variant rs10455872 accounted for a considerable fraction of this in Whites and African Americans, suggesting that the genetic etiology of AS may be predominantly explained by variation at the *LPA* locus.

Résumé

Contexte: La sténose aortique (SA) est une maladie des valves cardiaques de plus en plus répandue en raison du vieillissement de la population. Des études d'associations pangénomiques (GWAS) de la SA dans des populations d'ascendance européenne ont permis d'identifier la variation rs10455872 du locus *LPA* comme étant associée à un plus grand risque de maladie, une association qui a été par la suite reproduite autant chez les Afro-Américains que chez les Hispaniques. Cette association est médiée par des niveaux élevés de lipoprotéine(a) (Lp[a]), qui sont déterminés principalement par la variation au locus *LPA*. L'identification de loci de risque additionnels pour les niveaux de Lp(a) pourrait permettre d'identifier de nouvelles cibles pour le développement de thérapies visant à diminuer les niveaux de Lp(a).

Outre le locus *LPA*, plusieurs autres loci de risque ont été identifiés pour la SA. Cependant, jamais un score de risque polygénique (GRS) n'avait été créé pour mesurer la contribution combinée de multiples variations génétiques à cette maladie. Étant donné que la Lp(a) est un facteur de risque génétique pour la SA, un GRS pour la Lp(a) pourrait aussi être utile pour expliquer la variation de l'étiologie génétique de la SA.

Méthodes: Nous avons réalisé un GWAS à grande échelle chez des Britanniques blancs de la cohorte UK Biobank (n=293 274 individus). Nous avons testé environ 93 095 623 variations pour évaluer leur association avec les niveaux de Lp(a) transformés en log en utilisant un modèle de régression linéaire ajustée pour l'âge, le sexe, le lot de génotypage et 20 composantes principales d'ascendance génétique. Les variations atteignant une signification au niveau du génome ($P \leq 5 \times 10^{-8}$) ont été testées pour la

réplication dans une méta-analyse de deux autres cohortes d'ascendance européenne et d'autres groupes ethniques de la cohorte UK Biobank.

Les données générées par le GWAS pour la Lp(a) et par un GWAS pour la SA ont été utilisées pour développer des GRS chez les Blancs non hispaniques (n=55192), les Afro-Américains (n=1917) et les Hispaniques (n=3482) qui font partie de la cohorte GERA (Genetic Epidemiology Research on Adult Health and Aging), où le seuil de valeur *P* des scores a été sélectionné pour maximiser l'association avec la SA dans chaque groupe ethnique. La proportion de la variance expliquée de la SA par chaque GRS a été évaluée avec et sans la variation la plus significative, soit rs10455872.

Résultats: Au total, 131 variations indépendantes ont été associées de façon statistiquement significative au niveau pangénomique à Lp(a), validant ainsi des associations précédemment établies avec *LPA*, *APOE* et *CETP*, et identifiant une nouvelle association avec *APOH*. La variation rs8178824 sur *APOH*, qui code pour la bêta2-glycoprotéine I (β2GPI), est associée avec une augmentation des niveaux de Lp(a) (β [95% CI] (ln nmol/L), 0.064 [0.047, 0.081]; $P = 2.8 \times 10^{-13}$). Cette association a été reproduite dans une méta-analyse de 5465 individus d'ascendance européenne des cohortes Framingham Offspring Study et MESA (Multi-Ethnic Study of Atherosclerosis) (β [95% CI] (ln mg/dL), 0.16 [0.044, 0.28]; $P = 0.0071$), mais l'association n'a pu être reproduite chez les autres groupes ethniques de la cohorte UK Biobank (tous sont à $P>0.05$).

Les GRS construits à partir du GWAS de Lp(a) et maximisés pour l'association chez les blancs de la cohorte, les Afro-Américains et les Hispaniques de la cohorte GERA comprenaient respectivement 263, 1,291 et 11,217 variants tandis que les GRS

construits par la même méthode, à partir du GWAS de SA, comprenaient respectivement 3,10, et 496 variants. Les GRS-Lp(a) optimaux étaient tous associés significativement avec la SA mais expliquaient seulement une petite portion la variance de la SA (Blancs : R^2=0.13%, P= 6.2 × 10^{-9}; Afro-Américains: R^2=0.56%, P=0.045; Hispaniques: R^2=0.31%, P=0.045). Les GRS-SA optimaux expliquaient marginalement plus de variance phénotypique que les GRS-Lp(a) (Blancs : R^2=0.16%, P= 1.1 × 10^{-10}; Afro-Américains: R^2=0.63%, P=0.033; Hispaniques: R^2=0.39%, P=0.024). Lorsque nous avons retirés rs10455872 des GRS-Lp(a) optimaux, la variance expliquée a diminué de 0.076% chez les Blancs, 0.25% chez les Afro-Américains et de 0.26% chez les Hispaniques. Pareillement, si la variation rs10455872 est retirée des GRS-SA optimaux, la variance expliquée diminue respectivement de 0.063%, 0.13% et 0.37%.

Conclusions: Un GWAS à grande échelle des niveaux de Lp(a) a permis de révéler *APOH* comme un nouveau locus pour Lp(a) chez des individus d'ascendance européenne, révélant β2GPI comme une cible thérapeutique potentielle. Un GRS-Lp(a) dérivé de ce GWAS a également été associé à la SA presque autant qu'un GRS-SA lorsqu'étudié chez les Blancs, les Afro-Américains et les Hispaniques. Les deux types de GRS expliquent seulement une faible proportion de la variance de la SA, la variation *LPA* rs10455872 représentant une fraction considérable de la variance chez les Blancs et les Afro-Américains. Ces résultats suggèrent que l'étiologie génétique de la SA peut être expliquée principalement par des variations au niveau du locus *LPA* et appuient le rôle de la Lp(a) en tant que facteur de risque pour la SA dans toutes les groupes ethniques.

Acknowledgements

I would like to express my deepest thanks to my supervisor, Dr. James C. Engert, for his attention and support throughout my degree. I greatly appreciated the frequent one-on-one meetings, the continuous feedback, and the impromptu genetics lessons. Your dedication to your students does not go unnoticed. As well, your intellectual curiosity and enthusiasm for science are truly inspiring.

I am also indebted to my supervisor, Dr. George Thanassoulis, for his brilliant insights and interesting analysis ideas. With your help, I was able to connect to other researchers and strengthen the impact of my research. In addition, your comments and questions always helped me to think about the broader clinical applications of my work.

I am extremely grateful to Hao Yu Chen for her endless guidance and support; for helping me to navigate the world of presentations, conferences, and publications; for sharing code and resources that saved me time and headaches; for answering my emails at all hours of the day and night; for training me to think critically about my results; and finally, for caring about my health and mental wellbeing during particularly busy times. You are an incredible role model in the lab and otherwise.

Many thanks to Line Dufresne for her statistical wisdom and guidance with analysis; to Katia Desbiens for her prompt help with IT issues; to Maedeh Mohammadnetaaj for her assistance with administrative tasks; and to Hannah Burr, Thomas Marsh, and Yoshihiko (Tiger) Nagai for their friendship and support.

Finally, I would like to express my gratitude to Dr. Simon Gravel, Dr. Benjamin Smith, and Dr. Suhad Ali for their valuable feedback and suggestions during my committee meetings.

Contribution of Authors

Chapter 2: Genome-Wide Association Study Highlights *APOH* as a Novel Locus for Lipoprotein(a) Levels (Manuscript*)

Hoekstra: Main analysis of Lp(a) GWAS in the UK Biobank, meta-analysis of Framingham Offspring Study (FOS) and the Multi-Ethnic Study of Atherosclerosis (MESA), drafting of manuscript

Chen: Cleaning of phenotype data in the UK Biobank, running of Lp(a) GWAS in the UK Biobank, critical revision of manuscript

Rong: Replication analysis in the FOS

Yao & Guo: Replication analysis in the MESA

Dufresne: Critical revision of manuscript

Engert, Thanassoulis, Larson, Rotter: Supervisory support, critical revision of manuscript

Post, Vasan, Tsimikas: Administrative support, critical revision of manuscript

Tsai: Collection of Lp(a) data in the Multi-Ethnic Study of Atherosclerosis

*submitted to *Arteriosclerosis, Thrombosis, and Vascular Biology (ATVB)* as a Brief Report.

Chapter 4: Assessing the Performance of Genetic Risk Scores for Aortic Stenosis Across Ethnicities (Manuscript*)

Hoekstra: Quality control of summary statistics for AS and Lp(a) GWAS in the UK Biobank, imputation of genotype data for Hispanics in the Genetic Epidemiology

Research on Adult Health and Aging (GERA) cohort, creation of genetic risk scores, statistical analysis, drafting of manuscript

Chen: Cleaning of phenotype data in the UK Biobank, running of Lp(a) GWAS in the UK Biobank, critical revision of manuscript

Dufresne: Quality control of genotype data in the GERA cohort, cleaning of phenotype data in the GERA cohort, critical revision of manuscript

Burr: Imputation of genotype data for Whites in the GERA cohort

Ambikkumar: Imputation of genotype data for African Americans in the GERA cohort

Whitmer: Administrative support

Munter & Cairns: Running of AS GWAS in the UK Biobank

Lathrop: Holding of UK Biobank data for AS GWAS

Engert & Thanassoulis: Concept and design, supervisory support, critical revision of manuscript

* to be submitted to *Circulation: Genomic and Precision Medicine.*

Chapter 1: General Introduction

Aortic stenosis: a common heart valve disease with a poor prognosis

Aortic stenosis (AS) is a heart condition involving the progressive thickening of the aortic valve due to calcium buildup. As the aortic valve stiffens, blood flow through the aorta is restricted, causing angina, syncope, and dyspnea[1]. The left ventricle must work harder to pump blood through the smaller opening, leading to enlargement of the ventricle and heart failure. Because AS is a progressive disease, many individuals don't experience noticeable symptoms until the valve is severely calcified and blood flow through the heart is greatly reduced.

AS is the most common acquired valvular heart disease in the developed world, affecting 12.4% of individuals 75 years or older[2]. AS may also occur in younger people born with a bicuspid aortic valve, whose valves calcify sooner[1]. The prevalence of AS increases with age and is expected to more than double by 2040 due to the aging population[3]. Accordingly, healthcare costs associated with AS are also expected to increase, with current costs in the United States estimated at $1.3 billion a year[4].

While AS was once thought to be a degenerative condition that was inevitable with aging, an increasing body of evidence suggests otherwise[5]. The development of AS is now thought to be an active process involving lipid metabolism and inflammation, much like atherosclerosis[6]. Indeed, there is substantial overlap between risk factors for AS and coronary artery disease (CAD), including older age, male sex, elevated blood pressure, adiposity, dyslipidemia, and cigarette smoking[7]. Emerging evidence also suggests that AS may be influenced by family history, with one study showing that

individuals who had a sibling with AS were at least three times more likely to develop it themselves[8].

Like CAD, the association between elevated low-density lipoprotein cholesterol (LDL-C) and AS has been noted in many observational and retrospective studies[9–12]. However, three separate randomized trials have failed to show that lipid-lowering can halt the progress of AS[13–15]. Since all clinical trials were performed in severe cases of AS, it is still possible that lowering LDL-C could have benefits earlier in the disease process[16]. Indeed, a Mendelian randomization analysis performed in several cohorts showed that having a genetic predisposition to elevated LDL-C increased the odds for aortic valve calcification (AVC) and incident AS[17]. Unlike observational studies, Mendelian randomization provides evidence for potentially causal associations[18], warranting further investigation into LDL-C as a preventative target for AS.

Apart from statins, other therapies tested to treat AS include antihypertensive medication and drugs targeting phosphate and calcium metabolism[19]. None of these therapies have demonstrated strong benefit in stopping or slowing disease progression. As a result, the only definitive treatment option for AS is aortic valve replacement via transcatheter procedures or conventional open-heart surgery[20]. Without valve replacement, half of patients with severe AS will die within one year, while nearly 90% of patients will die within 5 years[4]. Research aimed at understanding the causes of AS is needed to identify new avenues for prevention and treatment.

Genetic approaches for understanding disease

One way to better understand the basis of complex traits and diseases like AS is by performing genome-wide association studies (GWAS). This approach involves genotyping a dense set of markers, called single nucleotide polymorphisms (SNPs), in a large number of people and assessing each SNP's association with a given trait[21,22]. To account for the large number of tests performed, the standard p-value threshold indicating significant evidence of association in a GWAS is $P \leq 5.0 \times 10^{-8}$[21]. Compared to the linkage analysis used in the past, GWAS are much more powerful for identifying common genetic variants that have modest effects on disease[21,23,24]. In addition, GWAS are hypothesis-free and do not require any previous evidence of genes affecting disease, representing an unbiased approach for discovery.

Other genetic approaches for understanding disease include next-generation sequencing techniques, such as whole exome sequencing (WES) and whole genome sequencing (WGS). These methods involve sequencing all of the protein-coding regions of the genome, or in the case of WGS, the entire genome. In contrast to GWAS, which focus on variants with moderate-to-high frequency, WES and WGS can identify rare variants with larger effect sizes, making them well-suited for the study of Mendelian disorders[25,26]. Accordingly, WES and WGS have been valuable for identifying mutations in familial supravalvular AS, a form of AS that develops before birth[27,28]. Given that non-congenital AS has many similarities to CAD, a complex disease with many risk loci[29], GWAS may be a more appropriate method for understanding its genetic basis.

One common issue encountered in GWAS is population stratification. In populations that are geographically isolated, genetic variants can become more or less

frequent due to genetic drift or as an adaptation to the environment[30,31]. Similarly, isolated populations can develop different patterns of linkage disequilibrium (LD), which refers to the correlation of genotypes between nearby SNPs. When populations with different genetic backgrounds are included in the same GWAS, differences in allele frequencies between cases and controls due to differences in ancestry can lead to false-positive associations or masking of true associations[32–34]. In a simulation study that included both Han Chinese individuals and Caucasian individuals in the same sample, the false positive rate in a dataset with only one disease susceptibility locus was over 10%[35].

To reduce systemic bias resulting from population stratification, GWAS often include individuals from the same ethnic background. However, genetic admixture within a population may still lead to confounding[36]. Genetic admixture occurs when previously isolated populations interbreed, resulting in offspring whose genomes are mosaics of chromosomal segments with different ancestry[31]. These proportions of ancestry can vary widely between individuals; in one study of African Americans, proportions of European ancestry were estimated to range between 0% and 72%[37]. In a classic example of confounding by ancestry, Knowler *et al.* identified a false association between the Gm haplotype Gm3;5,13,14 and type 2 diabetes in American Indians that resulted from the association of this haplotype with European ancestry[38].

Several strategies have been proposed to correct for population stratification[39,40], including the genomic control method and the principal component (PC) method. The genomic control method involves adjusting for a single factor (λ) that captures systemic bias in the test statistic[41]. While this strategy is effective at addressing inflation globally,

it does not account for differences that are specific to each SNP[42]. The PC method involves performing principal component analysis on several thousand unlinked genetic markers in a GWAS dataset[43]. The resulting PCs capture variation in genetic ancestry and can be used as covariates in the GWAS (Figure 1). When several adjustment methods for population stratification were compared, adjusting for PCs proved to be the most effective method for reducing false positives in populations with varying degrees of stratification[35].

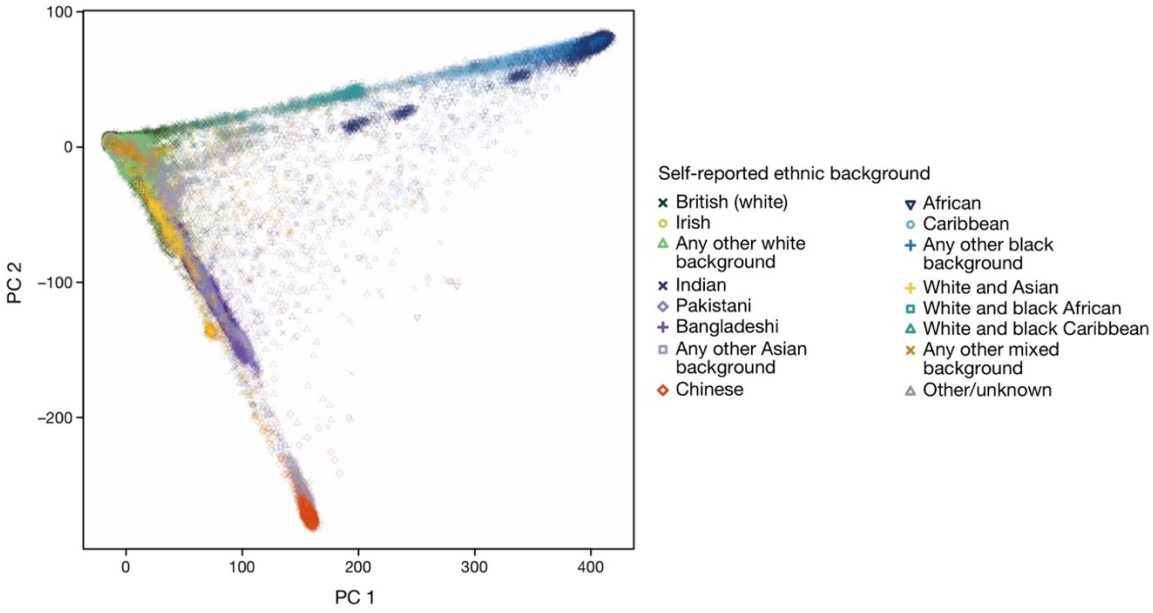

Figure 1: Principal components (PCs) for 488,377 individuals from the UK Biobank, where colour and shape represent the self-reported ethnic background of each individual. The first two PCs separate individuals with African, European, and Asian ancestry. Modified from Bycroft et al., 2018.

Another factor influencing the success of GWAS is the selection of SNPs to be genotyped. Genotyping arrays that take LD into account tend to achieve the greatest coverage, since more genetic variation can be captured by fewer SNPs[44]. Patterns of LD can also be used to increase coverage through a process known as genotype

imputation, where genotypes are predicted for variants that were not directly genotyped. These genotypes can be used to increase the number of variants tested in a GWAS, increasing the power of a study and facilitating meta-analysis of studies which used different genotyping arrays[45]. To ensure that imputation is accurate, the study sample must have LD patterns similar to the population(s) used to create the reference panel. For admixed populations, panels including samples from all ancestral populations leads to the highest accuracy[46,47].

In addition to improving our understanding of disease, GWAS can also be used to create genetic risk scores (GRSs). A GRS aims to summarize the cumulative risk of a set of genetic variants into a single variable. This variable can then be used to assess the contribution of genetic variation to disease and/or to predict the likelihood of an individual developing disease[48]. The predictive accuracy of a GRS is often assessed by measuring the area under a receiver operator characteristic (ROC) curve (AUC), which generally varies between 0.5 (a model with predictive power no better than chance) and 1 (a perfect model)[49]. As an alternative to measuring the AUC, the proportion of trait variability explained by the GRS can be calculated using the multiple R^2 from linear regression, or a pseudo-R^2 for binary traits[50].

One simple approach to creating a GRS involves selecting all variants reaching some threshold of significance in a GWAS and summing the risk alleles in a weighted fashion. This method is known as clumping (or pruning) and thresholding and can be implemented using the software PRSice or PRSice-2[51]. Other methods for creating GRSs include penalized regression (eg. lassosum[52]) and Bayesian approaches (eg. LDPred[53]). When these three approaches were compared for run time, memory usage,

and predictive power, PRSice-2 was more efficient than lassosum and LDPred at all sample sizes tested and had comparable predictive accuracy[51]. In addition, the clumping and thresholding approach used by PRSice-2 typically includes fewer variants in the GRSs, allowing for easier interpretation of the results and potentially greater portability to other populations.

Genetic studies of AS and Lipoprotein(a)

In 2013, the first GWAS for aortic valve disease was performed, identifying rs10455872 in the *LPA* gene as associated with both AVC and incident AS[54]. This variant, which was previously associated with CAD[55], has since been robustly associated with AS in numerous studies[56–59]. Recently, a transcriptome-wide association study (TWAS) of AS identified *PALMD* as an additional risk locus, with the lead variant increasing odds of disease by 29%[60]. Another study simultaneously reported associations at *PALMD* and *TEX41*[61]. Finally, a meta-GWAS and TWAS for AS demonstrated additional associations at *IL6*, *ALPL*, and *NAV1*[62]. Of all risk loci identified for AS, *LPA* is estimated to have the largest effect, with each minor allele of rs10455872 increasing risk by 66%[58].

The *LPA* gene codes for apolipoprotein(a), which, when bound to an LDL-like particle, forms lipoprotein(a) (Lp[a]). Genetic variation in the *LPA* gene, determined by a variable number of kringle IV type 2 repeats and other sequence variation, is estimated to explain over 90% of variation in plasma Lp(a) levels in European populations[63]. The link between *LPA* and AS has therefore implicated elevated plasma Lp(a) levels in the development of AS, an association that is thought to be causal[54,57]. As a result, Lp(a)-

lowering therapies are currently being tested as a treatment for both AS and CAD[64,65]. While these trials have shown promising results, the identification of additional loci for Lp(a) could reveal new targets for Lp(a)-lowering therapies and contribute to our knowledge of Lp(a) metabolism.

Although several GWAS have been performed for AS, to our knowledge, no GRSs have been developed for AS. In contrast, over 15 GRSs have been published for CAD in the last 5 years[66]. The creation of a GRS for AS could be challenging for a number of reasons; AS has a late onset and is not highly prevalent in cohorts of middle-aged adults. In addition, individuals with less severe AS are often asymptomatic and are not diagnosed. Finally, AS is a binary trait, leading to less statistical power in association studies. Given that Lp(a) is a continuous trait and levels are relatively independent of age[67], a GRS for Lp(a) may be more effective at explaining risk of AS than a GRS for AS.

The discovery of Lp(a) a risk factor for AS has highlighted the importance of genetic studies for contributing to our understanding of AS etiology. However, all genome-wide discovery studies for AS have been performed in populations of European descent. This lack of ethnic diversity is part of a larger research trend that sees a disproportionate focus on individuals of European ancestry[68–71]. As a result, findings from genetics studies may not be generalizable to broader populations. In AS research, only one study has attempted to replicate a previous association in non-European populations[54]. Studies assessing the transferability of previous results to other ethnicities are urgently needed to improve our understanding of AS in all populations.

Objectives

The main objectives of this thesis are: (1) to determine if Lp(a) levels are influenced by loci other than *LPA,* and (2) to determine whether genetic associations with Lp(a) can explain variance in AS across ethnic groups. To achieve these goals, a large-scale GWAS of Lp(a) was performed in White British individuals from the UK Biobank, with attempted replication in other ethnicities from the same cohort. Next, GRSs for Lp(a) were derived from the GWAS and fit for maximal association with AS across ethnicities in an American cohort. The performance of the Lp(a)-GRSs was compared to that of AS-GRSs derived from the same cohort, and the contribution of the *LPA* variant rs10455872 to both types of scores was assessed.

Genome-wide association study highlights *APOH* as a novel locus for lipoprotein(a) levels

Running title: *APOH* and lipoprotein(a) levels

Mary Hoekstra[1,2], Hao Yu Chen[1,2], Jian Rong[3], Line Dufresne[2], Jie Yao[4], Xiuqing Guo[4], Michael Y. Tsai[5], Sotirios Tsimikas[6], Wendy S. Post[7], Ramachandran S. Vasan[3], Jerome I. Rotter[4], Martin G. Larson[3], George Thanassoulis[1,2*], James C. Engert[1,2,8*]

1. Division of Experimental Medicine, McGill University, Montreal, Quebec;
2. Preventive and Genomic Cardiology, McGill University Health Centre and Research Institute, Montreal, Quebec;
3. Boston University's and NHLBI's Framingham Heart Study, Boston, Massachusetts;
4. The Institute for Translational Genomics and Population Sciences, Department of Pediatrics, The Lundquist Institute for Biomedical Innovation at Harbor-UCLA Medical Center, Torrance, California;
5. Department of Laboratory Medicine and Pathology, University of Minnesota, Minneapolis, Minnesota;
6. Division of Cardiovascular Medicine, Sulpizio Cardiovascular Center, University of California San Diego, La Jolla, California;
7. Division of Cardiology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland;

8. Department of Human Genetics, McGill University, Montreal, Quebec.

*co-senior authors.

Corresponding authors:

James C. Engert, PhD

McGill University Health Centre

1001 Decarie Blvd., Room EM1.2218

Montreal, Quebec, Canada H4A 3J1

Phone: 1-514-934-1934 ext. 35325 | e-mail: jamie.engert@mcgill.ca

George Thanassoulis, MD, MSc

McGill University Health Centre

1001 Decarie Blvd., Room D05.5120

Montreal, Quebec, Canada H4A 3J1

Phone: 1-514-934-1934 ext. 35465 | e-mail: george.thanassoulis@mcgill.ca

Keywords: Lipoprotein(a), Lipids, Lipoproteins, GWAS

Subject codes: Lipids and Cholesterol, Genetics, Genetic Association Studies

Word count: 5176

Total number of figures and tables: 1 figure,1 table

TOC category: Clinical and population studies

TOC subcategory: Arteriosclerosis, Thrombosis, and Vascular Biology

Abstract

Objective: Lipoprotein(a) (Lp(a]) is an independent risk factor for cardiovascular diseases and plasma levels are primarily determined by variation at the *LPA* locus. We performed a genome-wide association study (GWAS) in the UK Biobank to determine whether additional loci influence Lp(a) levels.

Approach and Results: We included 293,274 White British individuals in the discovery analysis. Approximately 93,095,623 variants were tested for association with natural log-transformed Lp(a) levels using linear regression models adjusted for age, sex, genotype batch, and 20 principal components of genetic ancestry. After quality control, 131 independent variants were associated at genome-wide significance *($P \leq 5 \times 10^{-8}$)*. In addition to validating previous associations at *LPA*, *APOE*, and *CETP*, we identified a novel variant at the *APOH* locus, encoding beta2-glycoprotein I (β2GPI). The *APOH* variant rs8178824 was associated with increased Lp(a) levels (β [95% CI] (ln nmol/L), 0.064 [0.047, 0.081]; $P = 2.8 \times 10^{-13}$) and demonstrated a stronger effect after adjustment for variation at the *LPA* locus (β [95% CI] (ln nmol/L), 0.089 [0.076, 0.10]; $P < 1.0 \times 10^{-25}$). This association was replicated in a meta-analysis of 5,465 European-ancestry individuals from the Framingham Offspring Study and Multi-Ethnic Study of Atherosclerosis (β [95% CI] (ln mg/dL), 0.16 [0.044, 0.28]; $P = 0.0071$).

Conclusions: In a large-scale GWAS of Lp(a) levels, we identified *APOH* as a novel locus for Lp(a) in individuals of European ancestry. Additional studies are needed to determine the precise role of β2GPI in influencing Lp(a) levels as well as its potential as a therapeutic target.

Abbreviations

Lp(a): Lipoprotein(a)

GWAS: Genome-wide association study

Apo(a): Apolipoprotein(a)

LDL: Low-density lipoprotein

β2GPI: Beta2-glycoprotein I

MESA: Multi-Ethnic Study of Atherosclerosis

Introduction

Lipoprotein(a) (Lp[a]) is an independent risk factor for both coronary artery disease and aortic valve stenosis[1,2]. Lp(a) consists of a low-density lipoprotein (LDL)-like particle covalently bound to the glycoprotein apolipoprotein(a) (apo[a]). Levels of Lp(a) are primarily controlled by the size of the apo(a) protein, with smaller apo(a) isoforms leading to higher concentrations of plasma Lp(a). This size polymorphism is caused by a variable number of kringle IV type 2 repeats in the *LPA* gene. Together with other sequence variation in *LPA*, these kringle IV type 2 repeats are estimated to explain more than 90% of variability in Lp(a) concentration in individuals of European ancestry[3,4].

Lp(a) plasma concentrations vary widely between populations, with African-ancestry individuals having 2-3 fold higher levels than European-ancestry individuals[5,6]. The distribution of Lp(a) is right-skewed across populations, with most individuals having very low levels[7]. While the precise physiological functions of Lp(a) are still unclear, there is evidence that it has proatherogenic and proinflammatory properties[8]. In

pathophysiological studies, Lp(a) or apo(a) have been detected in both the lesioned intima of human arteries[9–12] and in aortic valve lesions[13,14].

Several genome-wide association studies (GWAS) of Lp(a) have been performed, highlighting *LPA* as the major genetic determinant of Lp(a) levels[4,7,15–20]. However, these studies have been limited by small sample sizes (N<15,000), sparse genotyping arrays, or a focus on founder populations. In this study, we aimed to identify novel loci for Lp(a) by performing a GWAS in nearly 300,000 individuals from the UK Biobank. The findings could provide further insights into the regulation and clearance of Lp(a) particles and highlight novel targets for Lp(a)-lowering therapies.

Materials and Methods

Study Population

The UK Biobank study recruited over 500,000 individuals aged 40-69 years from 22 recruitment centers across the United Kingdom between 2006 and 2010. Participants provided blood samples for DNA extraction and biomarker analysis and completed a series of questionnaires, as previously described[21]. UK Biobank received ethical approval from the North West Multi-Centre Research Ethics Committee and all participants provided written informed consent. All relevant internal review boards approved this study. Only genetically-confirmed White British individuals were included in the discovery analysis to reduce confounding by ancestry, where White British ancestry was determined using a combination of self-reported ethnicity and results from a principal component analysis[21].

Phenotyping

Lp(a) (nmol/L) was measured using an immunoturbidimetric analysis on a Randox AU5800. Measurements were taken at the initial assessment visit (2006-2010) or the first repeat assessment visit (2012-2013). Measurements that returned an error from the analyzer or were outside of the reportable range (3.80-189 nmol/L) were excluded (n=91,426). Additional phenotypes are described in the Supplemental Material.

Genotyping

Genotyping was performed using the Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants, while the remaining 450,000 participants were genotyped using the Affymetrix UK Biobank Axiom array. Quality control and imputation were performed centrally by the UK Biobank as described previously[21]. Briefly, genetic markers were tested for batch effects, plate effects, departure from Hardy-Weinberg equilibrium, sex effects, array effects, and discordance across control replicates; markers that failed at least one test in a given batch had their genotype calls set to missing. Imputation was performed using only markers present on both the UK BiLEVE and UK Biobank Axiom arrays, and markers that failed quality control in more than one batch, had a >5% missing rate, or had a minor allele frequency <0.0001 were removed. Samples with unusually high heterozygosity or >5% missing rate were excluded from analysis.

GWAS for Lp(a)

Associations of 93,095,623 genetic variants with natural log-transformed Lp(a) were tested in linear regression models assuming additive genetic effects in PLINK version

2.0[22,23]. All models were adjusted for age, sex, genotype batch, and 20 principal components of ancestry. Variants with minor allele frequency < 0.01 or imputation quality score < 0.3 were removed (n=83,266,620). Clumping was performed on variants reaching genome-wide significance ($P \leq 5 \times 10^{-8}$) with PLINK version 1.9[22,24]; index variants were chosen greedily starting with the lowest p-value, and variants less than 1Mb away from an index variant with an r^2 of 0.01 were assigned to that index variant's clump. The most significant independent variants in each locus (lead variants) were queried for previous associations using PhenoScanner (accessed 02/04/2020)[25,26]. In a conditional analysis, all lead variants were tested for association with Lp(a) after additional adjustment for assessment center.

Conditional analysis for the *LPA* locus

A weighted *LPA*-region genetic risk score was created using all independent genome-wide significant variants in the *LPA* region. Lead variants outside the *LPA* region were tested for association with Lp(a) after adjusting for age, sex, genotype batch, 20 principal components, and the *LPA*-region genetic risk score.

Replication in Other Populations

The lead variant in each locus was tested for association with Lp(a) in other ethnic groups from the UK Biobank containing at least 1,000 unrelated individuals (South Asians, Black Africans, and Black Caribbeans). Lp(a) (nmol/L) was natural log transformed and models were adjusted for age, sex, genotype batch, and 20 principal components. Lead variants were also assessed in a fixed-effects meta-analysis of self-

reported White individuals from the Multi-Ethnic Study of Atherosclerosis (MESA) and the Framingham Offspring Study. The variant rs1065853 was not available in these cohorts so rs7412 was used as a proxy (linkage disequilibrium r^2=0.99 in the UK Biobank). Cohort descriptions and model details are provided in the Supplemental Material.

Statistical Analyses

Two-sided p-values ≤ 5×10^{-8} were considered significant in the GWAS and two-sided p-values ≤ 0.05 were considered significant in all other analyses. Proportion of variance explained was calculated for independent significant variants in the *LPA* region independently, when modeled together, and when combined as a weighted genetic risk score. For data availability, please see the Major Resources Table in the Supplemental Material.

Results

A total of 293,274 individuals with Lp(a) measurements were included in the study. Demographic characteristics of the UK Biobank Lp(a) subset are presented in Table I in the Supplemental Material, with individuals stratified by Lp(a) levels. Individuals with Lp(a) levels greater than or equal to the median were more likely to have coronary artery disease and aortic stenosis but were less likely to be female or diabetic (all *P*<0.001). These individuals also had higher levels of LDL cholesterol (corrected and uncorrected for Lp[a]) and high-density lipoprotein cholesterol (all *P*<0.001).

Following quality control, 9,829,003 variants with a minor allele frequency > 0.01 remained for further analysis. The association of these variants with Lp(a) showed no substantial inflation in the test statistics (genomic inflation factor [λ] = 1.03, Figure 1). After clumping, we identified 131 independent variants associated with Lp(a) at the genome-wide significance level of $P \leq 5 \times 10^{-8}$. The most significant variant, rs10455872 in *LPA*, explained 29% of variance in Lp(a) levels (Table II in Supplemental Material). There were 126 other independent variants in the *LPA* region, explaining an additional 20% of variance.

Outside the *LPA* region, we identified variants in four loci (Table 1). Variant rs1065853 on chromosome 19, located downstream of *APOE,* was associated with decreased Lp(a) levels (β [95% CI] (ln nmol/L), -0.11 [-0.12, -0.10]; $P < 1.0 \times 10^{-25}$), as was variant rs247617 on chromosome 16, located upstream of *CETP* (β [95% CI] (ln nmol/L), -0.023 [-0.030, -0.017]; $P = 1.0 \times 10^{-13}$). On chromosome 17, rs8178824 in *APOH* was associated with an increase in Lp(a) (β [95% CI] (ln nmol/L), 0.064 [0.047, 0.081]; $P = 2.8 \times 10^{-13}$). Finally, variant rs826128 on chromosome 2, located in the long non-coding RNA *AC093639.1*, was associated with decreased Lp(a) levels (β [95% CI] (ln nmol/L), -0.039 [-0.053, -0.026]; $P = 5.9 \times 10^{-9}$). The median Lp(a) level by genotype is shown for all lead variants in Table III of the Supplemental Material. Additional adjustment for assessment center did not materially change the results (Table IV in Supplemental Material).

The lead variant in each locus was evaluated for association with Lp(a) in the following populations from the UK Biobank: South Asian (n=6,101), Black African (n=2,510), and Black Caribbean (n=3,207). Consistent with results in White British

individuals, rs10455872 was associated with increased Lp(a) in South Asians (β [95% CI] (ln nmol/L), 1.01 [0.82,1.19]; $P < 1.0 \times 10^{-25}$) and Black Caribbeans (β [95% CI] (ln nmol/L), 0.81 [0.59, 1.03]; $P = 1.4 \times 10^{-12}$). This variant showed no association in Black Africans, likely owing to its low frequency in this population (minor allele frequency = 6.3×10^{-4}). The variant rs1065853 near *APOE* was significantly associated with Lp(a) across ethnic groups, with Black African individuals demonstrating the largest decrease in levels per minor allele (T) (β [95% CI] (ln nmol/L), -0.28 [-0.34, -0.21]; $P = 9.4 \times 10^{-16}$). The lead variants in *CETP*, *APOH*, and *AC093639.1* were not significantly associated in South Asians, Black Africans, or Black Caribbeans (Table V in Supplemental Material).

The lead variants were also tested for association with Lp(a) in a meta-analysis of 5,465 European-ancestry individuals from the Multi-Ethnic Study of Atherosclerosis and the Framingham Offspring Study. Both rs10455872 in *LPA* and rs8178824 in *APOH* were significantly associated with increased Lp(a) levels in the meta-analysis (rs10455872: β [95% CI] (ln mg/dL), 2.1 [2.0, 2.2]; $P < 1.0 \times 10^{-25}$ and rs8178824: β [95% CI] (ln mg/dL), 0.16 [0.044, 0.28]; $P = 0.0071$). The lead variants in *APOE*, *CETP*, and *AC093639.1* showed no significant effects (Table VI in Supplemental Material).

The weighted *LPA*-region genetic risk score contained 127 variants and explained 44% of the variance in Lp(a) levels. After adjusting for this score, variants in *APOE*, *CETP*, and *APOH* showed stronger effects on Lp(a) (Table VII in Supplemental Material). Conversely, the variant rs826128 on chromosome 2 showed a decreased effect and no longer reached genome-wide significance (β [95% CI] (ln nmol/L), -0.016 [-0.026, -0.0058]; $P = 0.0019$).

Discussion

We performed a GWAS for plasma Lp(a) levels in 293,274 White British individuals from the UK Biobank. We confirmed the association of loci in the *LPA* region with Lp(a) levels, as well as *APOE* and *CETP*. In addition, we identified *APOH* as a novel risk locus and replicated this association in a meta-analysis of two independent cohorts.

As expected, our association study identified many significant variants in the *LPA* gene and the surrounding region. Despite imposing a stringent r^2 threshold (≤ 0.01), 127 variants were independently associated with Lp(a). Together, the top 4 variants explained 40% of the variance in Lp(a) levels, while the remaining 123 explained an additional 9%. Consistent with previous work[1,20,27], the variant rs10455872 was the most strongly associated with Lp(a), explaining 29% of variation in Lp(a) levels alone.

Outside of the *LPA* region, we identified variants at four loci, two of which have been previously associated with Lp(a) or Lp(a)-cholesterol: *APOE* and *CETP*. Our lead variant near the *APOE* locus, rs1065853, is in high linkage disequilibrium with the apoE2-defining variant rs7412 (r^2=0.99*)*, which has been previously associated with decreased Lp(a)[4,19,28,29]. Relative to the apoE3 and apoE4 isoforms, apoE2 has a lower affinity for LDL receptors and LDL receptor-related protein I, potentially leading to less competition for Lp(a) binding and greater clearance of Lp(a)[30]. Upstream of the *CETP* locus, the variant rs247617 was also associated with decreased Lp(a) levels. This finding is consistent with clinical studies showing that inhibition of cholesteryl ester transfer protein, the product of *CETP*, decreases Lp(a) levels[31–33]. This variant is also in high linkage disequilibrium with rs247616 (r^2=0.99), which has been previously associated with Lp(a)-cholesterol[28].

Apart from *APOE* and *CETP*, no other loci outside the *LPA* region have been associated with Lp(a) levels at the genome-wide significance level. Here, we identify rs8178824 in *APOH* as significantly associated with increased Lp(a) and provide independent replication. Relative to rs10455872, the effect size of rs8178824 is small, with individuals homozygous for the minor allele having a median Lp(a) level only 4.7 nmol/L higher than individuals with two major alleles. However, this effect is comparable to those seen for *APOE* and *CETP*, where the difference in homozygous genotype classes is 10.5 nmol/L and 1.9 nmol/L, respectively. As demonstrated previously with treatments targeting *CETP*, which produced reductions in Lp(a) of more than 30%[31–33], therapeutic targeting of the *APOH* locus could have a more substantial effect on Lp(a) levels than the effect of a single variant.

The *APOH* locus encodes beta2-glycoprotein I (β2GPI), a single chain plasma protein with a high affinity for negatively charged surfaces[34]. Recently, β2GPI has been shown to interact with proprotein convertase subtilisin/kexin-9 (PCSK9)[35], whose inhibition leads to reductions in LDL cholesterol[36]. This evidence is supported by previous studies demonstrating that genetic variation in *APOH* is associated with decreased levels of LDL cholesterol[37–40] and peak particle diameter[41]. In vitro, β2GPI has also been shown to bind to Lp(a), both through the phospholipids of the LDL component and through the kringle IV-domain of apo(a)[42]. Given that apo(a) is a major site for the accumulation of negatively charged oxidized phospholipids[43], the interaction of β2GPI and Lp(a) may be primarily mediated through binding of β2GPI to these phospholipids.

The variant we identified in *APOH* is in perfect linkage disequilibrium with rs1801689 (r^2=1.0). Interestingly, the amino acid change caused by rs1801689 (Cys325Gly, also known as Cys306Gly) has been shown to alter the binding capacity of β2GPI for phospholipids[44]. This change may reduce β2GPI's affinity for oxidized phospholipids on apo(a), thereby allowing more free molecules of β2GPI and Lp(a) to circulate in the plasma. Indeed, rs1801689 has also been previously associated with increased levels of plasma β2GPI[45,46]. The potential role of β2GPI in lipid metabolism is further supported by the observation that it can accelerate triglyceride clearance in rats[47]. Future studies should investigate whether the presence of β2GPI similarly affects Lp(a) clearance or affects its pathogenicity through other mechanisms.

This study has several strengths and limitations. The UK Biobank discovery sample was larger than any previous Lp(a) GWAS, and thus had more power to detect novel associations. In addition, we were able to replicate our novel finding in *APOH* in a meta-analysis of two other European-ancestry cohorts. However, the *APOH* variant showed no association with Lp(a) in other ethnicities from the UK Biobank. The lack of replication observed for this variant and others may reflect reduced power due to smaller sample sizes, different allele frequencies, or different patterns of linkage disequilbrium[48]; nonetheless, additional studies in larger non-European cohorts are warranted. Another limitation of our study is the high percentage of individuals missing Lp(a) measurements (>20%) in the UK Biobank due to the assay's limited reportable range (3.80-189 nmol/L); our results may therefore not apply to individuals with very high levels of Lp(a).

In summary, we have performed a large-scale GWAS of Lp(a) levels, validating previous loci and identifying *APOH* as a novel locus. Our findings provide further insight into the regulation of Lp(a) levels and highlight β2GPI as a potential therapeutic target in individuals with elevated Lp(a).

Acknowledgements

This research has been conducted using the UK Biobank Resource under application number 41025, with supporting data from the Framingham Heart Study of the National Heart Lung and Blood Institute (NHLBI) of the National Institutes of Health and Boston University School of Medicine, as well as the Multi-Ethnic Study of Atherosclerosis (MESA). We thank the investigators, staff, and participants of these cohorts for their valuable contributions. A full list of participating MESA and institutions can be found at http://www.mesa-nhlbi.org.

Sources of Funding

Ms Hoekstra was supported by an award from the Canadian Institutes of Health Research (CIHR) for the duration of the study. Ms Chen has received grants from the McGill University Faculty of Medicine and McGill University Health Centre Foundation. Dr Thanassoulis has received research support from the CIHR, the Heart and Stroke Foundation of Canada, the NHLBI (grant R01 HL128550), the Doggone Foundation, the Courtois Foundation, and the Satoko Shibata and Richard Ingram Foundation. Dr Thanassoulis receives salary support from the Fonds de Recherche Santé–Québec.

MESA and the MESA SHARe project are conducted and supported by the NHLBI in collaboration with MESA investigators. Support for MESA is provided by contracts 75N92020D00001, HHSN268201500003I, N01-HC-95159, 75N92020D00005, N01-HC-95160, 75N92020D00002, N01-HC-95161, 75N92020D00003, N01-HC-95162, 75N92020D00006, N01-HC-95163, 75N92020D00004, N01-HC-95164, 75N92020D00007, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, UL1-TR-000040, UL1-TR-001079, UL1-TR-001420. Also supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The funding for SHARe genotyping was provided by NHLBI contract N02-HL-64278. Genotyping was performed at Affymetrix (Santa Clara, California, USA) and the Broad Institute of Harvard and MIT (Boston, Massachusetts, USA) using the Affymetrix Genome-Wide Human SNP Array 6.0. The Framingham Heart Study acknowledges the support of contracts NO1-HC-25195, HHSN268201500001I and 75N92019D00031 from the NHLBI and grant supplement R01 HL092577-06S1 for this research.

Disclosures

Dr Thanassoulis has participated in advisory boards for Amgen, Sanofi/Regeneron, Ionis, HLS Therapeutics, and Servier Canada and has received research grants from Ionis and Servier for research outside the scope of this work. Dr Tsimikas is a co-inventor and receives royalties from patents owned by UCSD on oxidation-specific

antibodies and of biomarkers related to oxidized lipoproteins and is a co-founder and has an equity interest in Oxitope, Inc and its affiliates ("Oxitope") as well as in Kleanthi Diagnostics, LLC ("Kleanthi"). Although these relationships have been identified for conflict of interest management based on the overall scope of the project and its potential benefit to Oxitope and Kleanthi, the research findings included in this particular publication may not necessarily relate to the interests of Oxitope and Kleanthi. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies.

References

1. Clarke R, Peden JF, Hopewell JC, Kyriakou T, Goel A, Heath SC, Parish S, Barlera S, Franzosi MG, Rust S, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med*. 2009;361:2518-2528.
2. Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith A V., Peloso GM, Kerr KF, Pechlivanis S, Budoff MJ, Harris TB, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. *N Engl J Med*. 2013;368:503-512.
3. Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest*. 1992;90:52-60.
4. Mack S, Coassin S, Rueedi R, Yousri NA, Seppälä I, Gieger C, Schönherr S, Forer L, Erhart G, Marques-Vidal P, et al. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res*. 2017;58:1834-1844.

5. Kraft HG. Frequency distributions of apolipoprotein(a) Kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: The distribution of null alleles is non-random. *Eur J Hum Genet*. 1996;4:74-87.
6. Cobbaert C, Kesteloot H. Serum lipoprotein(a) levels in racially different populations. *Am J Epidemiol*. 1992;136:441-449.
7. Kronenberg F, Utermann G. Lipoprotein(a): Resurrected by genetics. *J Intern Med*. 2013;273:6-30.
8. Orsó E, Schmitz G. Lipoprotein(a) and its role in inflammation, atherosclerosis and malignancies. *Clin Res Cardiol Suppl*. 2017;12:31-37.
9. Rath M, Niendorf A, Reblin T, Dietel M, Krebber HJ, Beisiegel U. Detection and quantification of lipoprotein(a) in the arterial wall of 107 coronary bypass patients. *Arteriosclerosis*. 1989;9:579-592.
10. Niendorf A, Rath M, Wolf K, Peters S, Arps H, Beisiegel U, Dietel M. Morphological detection and quantification of lipoprotein(a) deposition in atheromatous lesions of human aorta and coronary arteries. *Virchows Arch A Pathol Anat Histopathol*. 1990;417:105-111.
11. Hoff HF, Neil JO, Yashiro A. Partial characterization of lipoproteins containing apo [a] in human atherosclerotic lesions. *J Lipid Res*. 1993;34:789-798.
12. Reblin T, Meyer N, Labeur C, Henne-Bruns D, Beisiegel U. Extraction of lipoprotein(a), apo B, and apo E from fresh human arterial wall and atherosclerotic plaques. *Atherosclerosis*. 1995;113:179-188.

13. O'Brien KD, Reichenbach DD, Marcovina SM, Kuusisto J, Alpers CE, Otto CM. Apolipoproteins B, (a), and E accumulate in the morphologically early lesion of "degenerative" valvular aortic stenosis. *Arterioscler Thromb Vasc Biol*. 1996;16:523-532.

14. Torzewski M, Ravandi A, Yeang C, Edel A, Bhindi R, Kath S, Twardowski L, Schmid J, Yang X, Franke UFW, et al. Lipoprotein(a)-Associated Molecules Are Prominent Components in Plasma and Valve Leaflets in Calcific Aortic Valve Stenosis. *JACC Basic to Transl Sci*. 2017;2:229-240.

15. Liu Y, Ma H, Zhu Q, Zhang B, Yan H, Li H, Meng J, Lai W, Li L, Yu D, et al. A genome-wide association study on lipoprotein (a) levels and coronary artery disease severity in a Chinese population. *J Lipid Res*. 2019;60:1440-1448.

16. Lu W, Cheng YC, Chen K, Wang H, Gerhard GS, Still CD, Chu X, Yang R, Parihar A, O'Connell JR, et al. Evidence for several independent genetic variants affecting lipoprotein (a) cholesterol levels. *Hum Mol Genet*. 2015;24:2390-2400.

17. Ober C, Nord AS, Thompson EE, Pan L, Tan Z, Cusanovich D, Sun Y, Nicolae R, Edelstein C, Schneider DH, et al. Genome-wide association study of plasma lipoprotein(a) levels identifies multiple genes on chromosome 6q. *J Lipid Res*. 2009;50:798-806.

18. Qi Q, Workalemahu T, Zhang C, Hu FB, Qi L. Genetic variants, plasma lipoprotein(a) levels, and risk of cardiovascular morbidity and mortality among two prospective cohorts of type 2 diabetes. *Eur Heart J*. 2012;33:325-334.

19. Li J, Lange LA, Sabourin J, Duan Q, Valdar W, Willis MS, Li Y, Wilson JG, Lange EM. Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study. *J Hum Genet*. 2015;60:755-761.

20. Zabaneh D, Kumari M, Sandhu M, Wareham N, Wainwright N, Papamarkou T, Hopewell J, Clarke R, Li KW, Palmen J, et al. Meta analysis of candidate gene variants outside the LPA locus with Lp(a) plasma levels in 14,500 participants of six White European cohorts. *Atherosclerosis*. 2011;217:447-451.

21. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, Motyer A, Vukcevic D, Delaneau O, O'Connell J, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203-209.

22. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.

23. Purcell S, Chang C. PLINK 2.0. www.cog-genomics.org/plink/2.0/

24. Purcell S, Chang C. PLINK 1.9. www.cog-genomics.org/plink/1.9/

25. Staley JR, Blackshaw J, Kamat MA, Ellis S, Surendran P, Sun BB, Paul DS, Freitag D, Burgess S, Danesh J, et al. PhenoScanner: A database of human genotype-phenotype associations. *Bioinformatics*. 2016;32:3207-3209.

26. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, Butterworth AS, Staley JR. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics*. 2019;35:4851-4853.

27. Ronald J, Rajagopalan R, Cerrato F, Nord AS, Hatsukami T, Kohler T, Marcovina S, Heagerty P, Jarvik GP. Genetic variation in LPAL2, LPA, and PLG predicts plasma lipoprotein(a) level and carotid artery disease risk. *Stroke*. 2011;42:2-9.

28. Zekavat SM, Ruotsalainen S, Handsaker RE, Alver M, Bloom J, Poterba T, Seed C, Ernst J, Chaffin M, Engreitz J, et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat Commun*. 2018;9:1-14.

29. Moriarty PM, Varvel SA, Gordts PLSM, McConnell JP, Tsimikas S. Lipoprotein(a) Mass Levels Increase Significantly According to APOE Genotype: An Analysis of 431 239 Patients. *Arterioscler Thromb Vasc Biol*. 2017;37:580-588.

30. Phillips MC. Apolipoprotein e isoforms and lipoprotein metabolism. *IUBMB Life*. 2014;66:616-623.

31. Bloomfield D, Carlson GL, Sapre A, Tribble D, McKenney JM, Littlejohn TW, Sisk CMC, Mitchel Y, Pasternak RC. Efficacy and safety of the cholesteryl ester transfer protein inhibitor anacetrapib as monotherapy and coadministered with atorvastatin in dyslipidemic patients. *Am Heart J*. 2009;157:352-360.

32. Cannon CP, Shah S, Dansky HM, Davidson M, Brinton EA, Gotto AM, Stepanavage M, Liu SX, Gibbons P, Ashraf TB, et al. Safety of anacetrapib in patients with or at high risk for coronary heart disease. *N Engl J Med*. 2010;363:2406-2415.

33. Thomas T, Zhou H, Karmally W, Ramakrishnan R, Holleran S, Liu Y, Jumes P, Wagner JA, Hubbard B, Previs SF, et al. CETP (Cholesteryl Ester Transfer Protein) Inhibition with Anacetrapib Decreases Production of Lipoprotein(a) in

Mildly Hypercholesterolemic Subjects. *Arterioscler Thromb Vasc Biol*. 2017;37:1770-1775.

34. Miyakis S, Giannakopoulos B, Krilis SA. Beta 2 glycoprotein I-function in health and disease. *Thromb Res*. 2004;114:335-346.

35. Melendez QM, Wooten CJ, Krishnaji ST, Knagge K, Kirchner D, Lopez D. Identification of Novel Proteins Interacting with Proprotein Convertase Subtilisin/Kexin 9. 2020;3:1-27.

36. Horton JD, Cohen JC, Hobbs HH. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem Sci*. 2007;32:71-77.

37. Leduc MS, Shimmin LC, Klos KLE, Hanis C, Boerwinkle E, Hixson JE. Comprehensive evaluation of apolipoprotein H gene (APOH) variation identifies novel associations with measures of lipid metabolism in GENOA. *J Lipid Res*. 2008;49:2648-2656.

38. Asselbergs FW, Guo Y, Van Iperen EPA, Sivapalaratnam S, Tragante V, Lanktree MB, Lange LA, Almoguera B, Appelman YE, Barnard J, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet*. 2012;91:823-838.

39. Elbers CC, Guo Y, Tragante V, van Iperen EPA, Lanktree MB, Castillo BA, Chen F, Yanek LR, Wojczynski MK, Li YR, et al. Gene-Centric Meta-Analysis of Lipid Traits in African, East Asian and Hispanic Populations. *PLoS One*. 2012;7:e50198.

40. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, Ganna A, Chen J, Buchkovich ML, Mora S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274-1283.

41. Bossé Y, Feitosa MF, Després JP, Lamarche B, Rice T, Rao DC, Bouchard C, Pérusse L, Vohl MC. Detection of a major gene effect for LDL peak particle diameter and association with apolipoprotein H gene haplotype. *Atherosclerosis*. 2005;182:231-239.

42. Köchl S, Fresser F, Lobentanz E, Baier G, Utermann G. Novel interaction of apolipoprotein(a) with β-2 glycoprotein I mediated by the kringle IV domain. *Blood*. 1997;90:1482-1489.

43. Leibundgut G, Scipione C, Yin H, Schneider M, Boffa MB, Green S, Yang X, Dennis E, Witztum JL, Koschinsky ML, et al. Determinants of binding of oxidized phospholipids on apolipoprotein (a) and lipoprotein (a). *J Lipid Res*. 2013;54:2815-2830.

44. Sanghera DK, Wagenknecht DR, McIntyre JA, Kamboh MI. Identification of structural mutations in the fifth domain of apolipoprotein H (β2-glycoprotein I) which affect phospholipid binding. *Hum Mol Genet*. 1997;6:311-316.

45. Kamboh MI, Manzi S, Mehdi H, Fitzgerald S, Sanghera DK, Kuller LH, Atson CE. Genetic variation in apolipoprotein H (β2-glycoprotein I) affects the occurrence of antiphospholipid antibodies and apolipoprotein H concentrations in systemic lupus erythematosus. *Lupus*. 1999;8:742-750.

46. Mather KA, Thalamuthu A, Oldmeadow C, Song F, Armstrong NJ, Poljak A, Holliday EG, McEvoy M, Kwok JB, Assareh AA, et al. Genome-wide significant

results identified for plasma apolipoprotein H levels in middle-aged and older adults. *Sci Rep*. 2016;6:23675.

47. Wurm H, Beubler E, Polz E, Holasek A, Kostner G. Studies on the possible function of β2-glycoprotein-I: Influence in the triglyceride metabolism in the rat. *Metabolism*. 1982;31:484-486.

48. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in g lobally diverse populations. *Nat Rev Genet*. 2019;20:520-535.

Highlights

- We have performed a large-scale genome-wide association study of Lp(a) levels.
- We confirmed the association of variants in *LPA*, *APOE*, and *CETP* with Lp(a).
- We identified *APOH* as a novel risk locus for Lp(a), highlighting β2GPI as a determinant of Lp(a) levels and a potential therapeutic target.

Tables

Table 1: Association of lead variants with Lp(a).

Variant	CHR	Position	Genes in Locus	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*	Variants in Locus*
rs10455872	6	161010118	*LPA, ZDHHC14, SNX9*†	G (0.076)	1.7 [1.7,1.7]	$<1.0 \times 10^{-25}$	127
rs1065853	19	45413233	*APOE*	T (0.080)	-0.11 [-0.12, -0.10]	$<1.0 \times 10^{-25}$	1
rs247617	16	56990716	*CETP*	A (0.32)	-0.023 [-0.030, -0.017]	1.0×10^{-13}	1
rs8178824	17	64224775	*APOH*	T (0.030)	0.064 [0.047,0.081]	2.8×10^{-13}	1
rs826128	2	184797074	*AC093639.1*	A (0.054)	-0.039 [-0.053, -0.026]	5.9×10^{-9}	1

*Number of independent ($r^2 < 0.01$), genome-wide significant variants.

†Additional genes (+/- 3.5 MB from *LPA*): *TULP4, SYTL3, EZR, RP1-155D22.1, RSPH3, RP1-111C20.4, FNDC1, RP11-125D12.1, RP11-125D12.2, RP3-393E18.1, SOD2, ACAT2, PNLDC1, RP1-249F5.3, IGF2R, SLC22A1, SLC22A2, SLC22A3, LPAL2, PLG, RP11-235G24.1, RP11-235G24.2, RP11-235G24.3, RP3-428L16.1, MAP3K4, AGPAT4, PARK2.*

SUPPLEMENTAL MATERIAL

Supplemental Methods

Phenotype Definitions in the UK Biobank

Sex was coded as concordant genetic sex and self-reported gender. Age refers to the age of the participant on the day they attended an Assessment Centre. Presence of diabetes was determined using self-report of diagnosis by a doctor. Cases of coronary artery disease were determined by diagnosis of *International Classification of Diseases, 9th Revision (ICD-9)* 410, 410.0, 410.1, 410.2, 410.3, 410.4, 410.5, 410.6, 410.7, 410.8, 410.9, 411.9, 412, 412.9, 414.0, 414.1, 414.8, 414.9, or 429.7 or diagnosis of *ICD-10* I21, I21.0, I21.1, I21.2, I21.3, I21.4, I21.9, I22, I22.0, I22.1, I22.8, I22.9, I23, I23.0, I23.1, I23.2, I23.3, I23.6, I23.8, I24.1, I25.0, I25.1, I25.2, I25.3, I25.4, I25.5, I25.6, I25.8, I25.9, I51.0, I51.3 in the hospital inpatient records or death records, or OPCS4 procedure codes K40.1, K40.2, K40.3, K40.4, K40.8, K40.9, K41.1, K41.2, K41.3, K41.4, K41.8, K41.9, K42.1, K42.2, K42.3, K42.4, K42.8, K42.9, K43.1, K43.2, K43.3, K43.4, K43.8, K43.9, K44.1, K44.2, K44.8, K44.9, K45.1, K45.2, K45.3, K45.4, K45.5, K45.6, K45.8, K45.9, K46.1, K46.2, K46.3, K46.4, K46.5, K46.8, K46.9, K49.1, K49.2, K49.3, K49.4, K49.8, K49.9, K50.1, K50.2, K50.3, K50.4, K50.8, K50.9, K75.1, K75.2, K75.3, K75.4, K75.8, K75.9 in hospital inpatient records. Low-density lipoprotein cholesterol (LDL-C) in nmol/L was measured by enzymatic protective selection analysis on a Beckman Coulter AU5800. Corrected LDL-C was calculated by converting both LDL-C and lipoprotein(a) (Lp[a]) into mg/dL and subtracting 30% of the Lp(a) value from the LDL-C value, as previously described[1]. High-density lipoprotein cholesterol (nmol/L) was measured by enzyme immunoinhibition analysis on a Beckman Coulter AU5800.

Systolic and diastolic blood pressure (mmHg) were each measured automatically using an Omron device. Body mass index (kg/m^2) was calculated using weight and height measurements from the initial assessment visit.

Replication Cohorts

The Multi-Ethnic Study of Atherosclerosis (MESA) recruited 6,814 individuals of diverse ancestry from 6 communities in the United States between 2000 and 2002, as previously described[2]. Individuals were 45 to 84 years old at the baseline exam and free of clinical cardiovascular disease. Blood samples were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 and imputation was performed using the 1000 Genomes Phase 3 reference panel[3]. Lp(a) in mg/dL was measured at baseline by Health Diagnostics Laboratory (Richmond, VA) using a latex-enhanced turbidimetric immunoassay (Denka Seiken, Tokyo, Japan) that controls for the heterogeneous sizes of apolipoprotein(a). Our analysis included 2,456 unrelated individuals of European ancestry. Associations between genetic variants and natural log-transformed Lp(a) were tested in linear regression models adjusted for age, sex, recruitment site, and two principal components of ancestry.

The Framingham Offspring Study (FOS) is a longitudinal population-based study that recruited 5,124 of the original Framingham Heart Study participants' offspring and their spouses, as previously described[4]. During the 5th examination cycle (1991-1995), participants underwent a medical history, physical examination, and had blood drawn for plasma lipid and lipoprotein measurements. Lp(a) was measured in mg/dL using an

immunoturbidimetric assay from Wako Chemicals USA (Richmond, VA). Genotyping was performed on the Affymetrix GeneChip Human Mapping 500K Array and 50K Human Gene Focused Panel and genotypes were imputed using the Haplotype Reference Consortium panel[5]. Our analysis included 3,009 unrelated individuals of European ancestry. Associations between genetic variants and natural log-transformed Lp(a) were tested using linear regression models adjusted for age, sex, and 10 principal components of ancestry.

Supplemental Figures

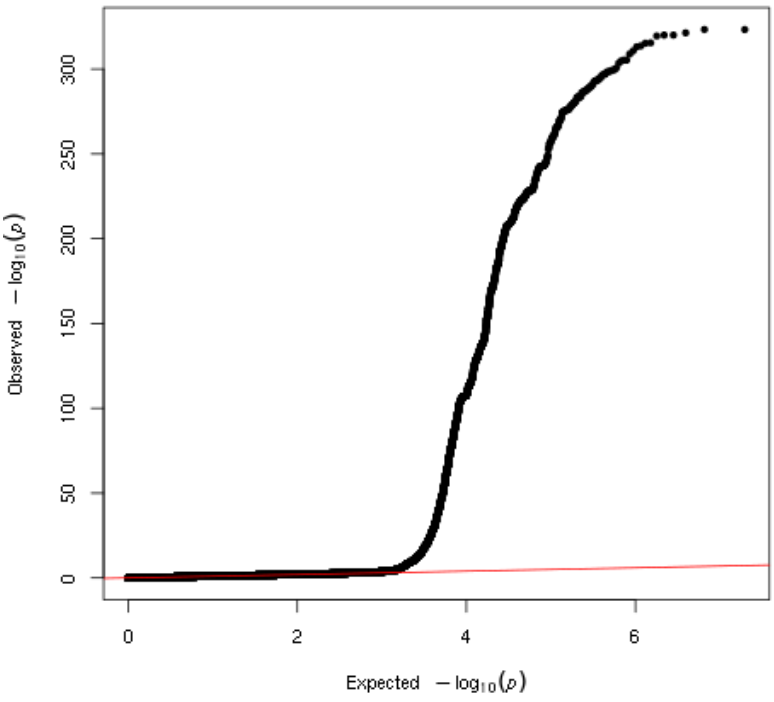

Figure I: QQ-plot for the GWAS of natural log-transformed Lp(a) adjusted for age, sex, genotype batch, and 20 principal components. Associations with $P < 4.9 \times 10^{-324}$ are not shown due to limitations in plotting software.

Supplemental Tables

Table I: Population characteristics of the UK Biobank cohort with Lp(a) measurements.

	All	**< Lp(a) median**	**>= Lp(a) median**	***P***
N	293,274	146,416	146,858	NA
Male (%)	135,220 (46.1)	69,918 (47.8)	65,302 (44.5)	<0.001
Age, y (mean (SD))	57.9 (8.2)	57.7 (8.2)	58.2 (8.1)	<0.001
BMI, kg/m^2 (mean (SD))	27.4 (4.7)	27.4 (4.8)	27.4 (4.7)	0.73
CAD (%)	22,391 (7.6)	10,172 (7.0)	12,219 (8.3)	<0.001
AS (%)	1722 (0.6)	740 (0.5)	982 (0.7)	<0.001
Diabetes (%)	14,340 (4.9)	7,710 (5.3)	6,630 (4.5)	<0.001
Ever smoked (%)	176,845 (60.5)	88,317 (60.5)	88,528 (60.5)	0.92
SBP, mm Hg (mean (SD))	138.6 (18.7)	138.5 (18.6)	138.6 (18.7)	0.10
DBP, mm Hg (mean (SD))	82.0 (10.1)	82.0 (10.2)	82.0 (10.1)	0.073
HDL-C, mmol/L (mean (SD))	1.45 (0.38)	1.45 (0.38)	1.46 (0.38)	<0.001
LDL-C, mmol/L (mean (SD))	3.57 (0.87)	3.50 (0.85)	3.64 (0.87)	<0.001
Corrected LDL-C, nmol/L (mean (SD))	3.53 (0.86)	3.49 (0.85)	3.57 (0.87)	<0.001
Lp(a), nmol/L (median [IQR])	20.1 [9.3, 60.2]	9.3 [6.1, 13.3]	60.1 [33.8, 119.0]	<0.001

Abbreviations: BMI, body mass index; CAD, coronary artery disease; AS, aortic stenosis; SBP, systolic blood pressure; DBP, diastolic blood pressure; HLD-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; Lp(a), lipoprotein(a); SD, standard deviation; IQR, inter-quartile range.

Table II: Variance explained for the top 10 independent variants in the *LPA* region, modeled independently and cumulatively.

Variant	Rank	Variant R^2	Model R^2
rs10455872	1	0.29	0.29
rs73596816	2	0.046	0.36
rs150415123	3	0.026	0.37
rs544366796	4	0.023	0.40
rs140570886*	5	0.015	0.42
rs78439586	6	0.014	0.42
rs41269133	7	0.013	0.43
rs77009508	8	0.012	0.43
6:160489092_TGG_T	9	0.012	0.44
rs528521448	10	0.011	0.44

Abbreviations: R^2, proportion of variance explained.

*rs140570886 is in high linkage disequilibrium with rs3798220 (r^2=0.81 in LDlink[6]), which was not present in the dataset.

Table III: Median Lp(a) level within each genotypic class for the lead variants.

Variant	Locus	Minor Allele (Freq)	Median Lp(a) level (nmol/L)		
			Homozygous Major	Heterozygous	Homozygous Minor
rs10455872	*LPA*	G (0.076)	16.1	127	148
rs1065853	*APOE*	T (0.080)	20.8	16.8	10.3
rs247617	*CETP*	A (0.32)	20.7	19.9	18.8
rs8178824	*APOH*	T (0.030)	20.0	22.6	24.7
rs826128	*AC093639.1*	A (0.054)	20.2	19.1	17.4

Table IV: Association of lead variants with natural log-transformed Lp(a) after adjusting for age, sex, genotype batch, 20 principal components, and assessment center.

Variant	Locus	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*
rs10455872	*LPA*	G (0.076)	1.7 [1.7, 1.7]	$<1.0 \times 10^{-25}$
rs1065853	*APOE*	T (0.080)	-0.11 [-0.12, -0.10]	$<1.0 \times 10^{-25}$
rs247617	*CETP*	A (0.32)	-0.023 [-0.030, -0.017]	1.1×10^{-13}
rs8178824	*APOH*	T (0.030)	0.064 [0.047, 0.081]	2.9×10^{-13}
rs826128	*AC093639.1*	A (0.054)	-0.040 [-0.053, -0.0026]	5.4×10^{-9}

Abbreviations: CI, confidence interval.

Table V: Association of lead variants with natural log-transformed Lp(a) in 6,101 South Asians, 2,510 Black Africans, 3,207 Black Caribbeans, and 293,274 White British from the UK Biobank.

Variant	Locus	Group	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*
rs10455872	*LPA*	SA	G (0.011)	1.01 [0.82, 1.19]	$<1.0 \times 10^{-25}$
		BA	G (0.00063)	0.68 [-0.24, 1.6]	0.15
		BC	G (0.010)	0.81 [0.59, 1.03]	1.4×10^{-12}
		WB	G (0.076)	1.7 [1.7, 1.7]	$<1.0 \times 10^{-25}$
rs1065853	*APOE*	SA	T (0.045)	-0.17 [-0.25, -0.081]	1.3×10^{-4}
		BA	T (0.12)	-0.28 [-0.34, -0.21]	9.4×10^{-16}
		BC	T (0.12)	-0.21 [-0.27, -0.15]	1.2×10^{-11}
		WB	T (0.080)	-0.11 [-0.12, -0.10]	$<1.0 \times 10^{-25}$
rs247617	*CETP*	SA	A (0.34)	-0.0046 [-0.041, 0.032]	0.81
		BA	A (0.26)	-0.020 [-0.70, 0.031]	0.45

		BC	A (0.25)	0.0086 [-0.038, 0.055]	0.72
		WB	A (0.32)	-0.023 [-0.030, -0.017]	1.0×10^{-13}
rs8178824	*APOH*	SA	T (0.015)	0.051 [-0.095, 0.20]	0.49
		BA	T (0.0031)	0.22 [-0.18, 0.61]	0.28
		BC	T (0.0047)	-0.14 [-0.44, 0.16]	0.35
		WB	T (0.030)	0.064 [0.047, 0.081]	2.8×10^{-13}
rs826128	*AC093639.1*	SA	A (0.065)	0.023 [-0.049, 0.096]	0.53
		BA	A (0.13)	-0.00066 [-0.068, 0.067]	0.98
		BC	A (0.12)	-0.01 [-0.073, 0.052]	0.75
		WB	A (0.054)	-0.039 [-0.053, -0.026]	5.9×10^{-9}

Abbreviations: SA, South Asians; BA, Black Africans; BC, Black Caribbeans; WB, White British; CI, confidence interval.

Table VI: Association of lead variants with natural log-transformed Lp(a) in 3,009 White individuals from the FOS cohort, 2,456 White individuals from the MESA cohort, and a meta-analysis of these two cohorts.

Variant	**Locus**	**Minor Allele (Freq*)**	**FOS**		**MESA**		**Meta-analysis**	
			β [95% CI] (ln mg/dL)	***P***	**β [95% CI] (ln mg/dL)**	***P***	**β [95% CI] (ln mg/dL)**	***P***
rs10455872	*LPA*	G (0.066)	2.3 [2.2, 2.4]	$<1.0 \times 10^{-25}$	1.6 [1.5, 1.8]	$<1.0 \times 10^{-25}$	2.1 [2.0, 2.2]	$<1.0 \times 10^{-25}$
rs7412†	*APOE*	T (0.077)	0.0014 [-0.12, 0.13]	0.99	-0.15 [-0.29, -0.0084]	0.038	-0.065 [-0.16, 0.028]	0.17
rs247617	*CETP*	A (0.32)	-0.026 [-0.082, 0.030]	0.54	0.018 [-0.043, 0.078]	0.57	-0.0057 [-0.047, 0.035]	0.79
rs8178824	*APOH*	T (0.035)	0.19 [0.049, 0.34]	0.082	0.11 [-0.089, 0.31]	0.28	0.16 [0.044, 0.28]	0.0071
rs826128	*AC093639.1*	A (0.049)	0.050 [-0.072, 0.17]	0.58	-0.047 [-0.17, 0.074]	0.45	0.0013 [-0.084, 0.087]	0.98

Abbreviations: CI, confidence interval; FOS, Framingham Offspring Study; MESA, Multi-Ethnic Study of Atherosclerosis.
*Allele frequency in the Framingham Offspring Study.
†rs7412 used as a proxy for rs1065853 (linkage disequilibrium $r^2 = 0.99$ in the UK Biobank).

Table VII: Association of non-*LPA* lead variants with natural log-transformed Lp(a) after adjustment for age, sex, genotype batch, 20 principal components, and the *LPA*-region genetic risk score.

Variant	**Locus**	**Minor Allele (Freq)**	**β [95% CI] (ln nmol/L)**	***P***
rs1065853	*APOE*	T (0.080)	-0.17 [-0.17, -0.16]	$<1.0 \times 10^{-25}$
rs247617	*CETP*	A (0.32)	-0.029 [-0.034, -0.024]	$<1.0 \times 10^{-25}$
rs8178824	*APOH*	T (0.030)	0.089 [0.076, 0.10]	$<1.0 \times 10^{-25}$
rs826128	*AC093639.1*	A (0.054)	-0.016 [-0.026, -0.0058]	0.0019

Abbreviations: CI, confidence interval.

Major Resources Table

Data & Code Availability

Description	**Source / Repository**	**Persistent ID / URL**
UK Biobank genetic & phenotypic data	http://www.ukbiobank.ac.uk/using-the-resource/	NA
MESA genetic & phenotypic data	dbGaP Study Accession: phs000209.v13.p3	https://www.mesa-nhlbi.org/default.aspx
FOS genetic & phenotypic data	dbGaP Study Accession: phs000007.v30.p11	https://framinghamheartstudy.org
Analysis code	Available upon request	NA

Abbreviations: MESA, Multi-Ethnic Study of Atherosclerosis; FOS, Framingham Offspring Study.

References

1. Viney NJ, van Capelleveen JC, Geary RS, Xia S, Tami JA, Yu RZ, Marcovina SM, Hughes SG, Graham MJ, Crooke RM, et al. Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. *Lancet*. 2016;388:2239-2253.
2. Bild DE, Bluemke DA, Burke GL, Detrano R, Diez Roux A V., Folsom AR, Greenland P, Jacobs DR, Kronmal R, Liu K, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol*. 2002; 156:871-881.
3. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, Clark AG, Donnelly P, Eichler EE, Flicek P, et al. A global reference for human genetic variation. *Nature*. 2015;26:68-74.
4. Kannel WB, Feinleib M, MeNAMARA PM, Garrison RJ, Castelli WP. An Investigation of Coronary Heart Disease in Families: The Framingham Offspring Study. *Am J Epidemiol*. 2017;185:1103-1108.
5. McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, Kang HM, Fuchsberger C, Danecek P, Sharp K, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279-1283.
6. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555-3557.

Chapter 3: Transition

In the previous chapter, a large-scale GWAS for Lp(a) was performed in White British individuals from the UK Biobank. The association of variants in previously discovered loci were validated and *APOH* was identified as a novel risk locus. When the lead variants were assessed in other ethnicities from the UK Biobank, *LPA* and *APOE* were the only loci to demonstrate significant associations. While it is well established that *LPA*'s rs10455872 plays a role in the development of aortic stenosis (AS) in populations of European descent, the extent to which other genetic determinants of Lp(a) explain AS is unknown. In addition, it is not clear how large a role rs10455872 may play in explaining AS in non-European ethnicities.

In the next chapter, genetic risk scores (GRSs) will be used to assess the explanatory value of genetic associations with Lp(a) for aortic stenosis (AS) in several ethnicities. Using the variants and effect sizes from the Lp(a) GWAS in White British, GRSs will be fit for maximal association with AS in non-Hispanic Whites, African Americans, and Hispanics from an American cohort (the GERA cohort) using PRSice-2. The performance of the best-fit (by ethnicity) Lp(a)-GRSs will be compared to AS-GRSs developed using a previously performed AS GWAS. In addition, the contribution of rs10455872 to AS will be assessed within each ethnicity, both as part of the risk scores and independently. The results of this study will provide insights into the transferability of genetic associations with AS and Lp(a) in European-ancestry populations to other ethnicities, while verifying the value of Lp(a) as a risk factor for AS.

Assessing the Performance of Genetic Risk Scores for Aortic Stenosis Across Ethnicities

Mary Hoekstra, BCmp[1,2]; Hao Yu Chen, MSc[1,2]; Hannah Burr, BSc[2]; Athithan Ambikkumar, BSc[2]; Line Dufresne, MSc[2]; Rachel A. Whitmer, PhD[3]; Hans Markus Munter, PhD[4]; Benjamin J. Cairns, PhD[5-7], Mark Lathrop, PhD[4,8]; George Thanassoulis, MD[1,2*], James C. Engert, PhD[1,2,8*]

1. Division of Experimental Medicine, McGill University, Montreal, Quebec;
2. Preventive and Genomic Cardiology, McGill University Health Centre and Research Institute, Montreal, Canada;
3. Department of Public Health Sciences, University of California, Davis, California;
4. McGill University and Genome Quebec Innovation Centre, Montreal, Quebec;
5. MRC (Medical Research Council) Population Health Research Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom;
6. Clinical Trial Service Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom;
7. Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford, United Kingdom;
8. Department of Human Genetics, McGill University, Montreal, Quebec.

*co-senior authors.

Corresponding authors:

James C. Engert, PhD

McGill University Health Centre

1001 Decarie Blvd., Room EM1.2218

Montreal, Quebec, Canada H4A 3J1

Phone: 1-514-934-1934 ext. 35325 | e-mail: jamie.engert@mcgill.ca

George Thanassoulis, MD, MSc

McGill University Health Centre

1001 Decarie Blvd., Room D05.5120

Montreal, Quebec, Canada H4A 3J1

Phone: 1-514-934-1934 ext. 35465 | e-mail: george.thanassoulis@mcgill.ca

Word Count: 5004

Journal Subject Terms: Genetics; Epidemiology and Big Data; Genetic, Association Studies; Lipids and Cholesterol; Race and Ethnicity; Valvular Heart Disease; Stenosis.

Abstract

Background: Aortic stenosis (AS) is becoming increasingly common as the population ages. Genome-wide association studies (GWAS) have identified *LPA* as a risk locus for AS, highlighting lipoprotein(a) (Lp[a]) as an important risk factor. We sought to determine whether genetic risk scores (GRSs) for AS and Lp(a) can explain variance in AS across ethnicities.

Methods: Using GWAS performed in the UK Biobank, we derived best-fit GRSs for AS and Lp(a) in unrelated non-Hispanic Whites (n=55,192), African Americans (n=1,917), and Hispanics (n=3,4582) from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. The variance explained in AS was assessed for each GRS, with and without the *LPA* variant rs10455872.

Results: Depending on ethnicity, the best-fit AS-GRS contained between 3 and 496 variants, while the best-fit Lp(a)-GRS contained between 263 and 11,217 variants. The best-fit (by ethnicity) AS-GRSs were all significantly associated with AS but explained only a small proportion of phenotypic variance (Whites: R^2=0.16%, P=1.1 × 10^{-10}; African Americans: R^2=0.63%, P=0.033 & Hispanics: R^2=0.39%, P=0.024). The best-fit Lp(a)-GRSs were also significantly associated with AS and had comparable performances. Removing rs10455872 from the GRSs decreased the variance explained substantially in Whites and African Americans, and modestly in Hispanics.

Conclusions: UK Biobank-derived GRSs for AS and Lp(a) are associated with AS in Whites, African Americans, and Hispanics when the p-value threshold is selected in these ethnicities. The *LPA* variant rs10455872 accounts for a considerable proportion of

the variance explained by both types of score in Whites and African Americans, emphasizing genetic variation at *LPA* as a risk factor for AS in these ethnicities.

Keywords: aortic stenosis, lipoprotein(a), genetic risk scores.

Introduction

Aortic stenosis (AS) is the most common acquired valve disorder in developed countries and its prevalence is expected to increase considerably with the aging population[1,2]. Currently, there are no medical therapies to stop or slow the progression of AS. Without an aortic valve replacement, half of patients with severe AS will die within 1 year[3].

In recent years, genome-wide association studies have helped identify the genetic etiology of AS development. Variants in the *LPA* locus have been strongly associated with both aortic valve calcium, a subclinical precursor to AS, and established valve disease[4–8]. The most strongly associated variant, rs10455872, has been shown to increase odds of AS by over 60%[7]. These associations are mediated by elevated levels of lipoprotein(a) [Lp(a)], which is now recognized as an important risk factor for AS and other cardiovascular diseases[9].

Since the discovery that genetic variants in *LPA* contribute to AS susceptibility, several additional loci have been implicated in this disease, each with small to moderate effects[10–13]. These findings are consistent with a polygenic model of inheritance, where a trait is determined by the combined action of many genes. Genetic risk scores (GRSs), also known as polygenic risk scores, aim to capture the polygenic nature of traits by combining the cumulative risk of many genetic variants into a single variable.

Recently, GRSs have been created to assess the extent to which cumulative genetic variation contributes to complex diseases[14–16]. It is unknown to what extent a GRS can explain variance in AS.

Creating a GRS that captures significant variation in AS may be challenging for several reasons. The prevalence of AS is significantly higher in individuals aged 75 years or older[17], making it less frequent in cohorts of predominantly young to middle-aged individuals. In addition, asymptomatic individuals and those with less severe disease may not diagnosed[18], further reducing statistical power in association studies. Lp(a) levels, on the other hand, can be measured as a continuous variable in all individuals and are relatively constant across the lifetime[19]. Given that studies of Lp(a) are more highly powered, a GRS for Lp(a) may better explain an individual's risk for AS.

Another challenge hindering the development of a GRS for AS (as well as many other diseases) is ensuring that it can accurately predict risk in diverse populations[20]. Despite being a significant source of morbidity across ethnicities[21], our knowledge of the genetic determinants of AS in non-Europeans is limited. A study of aortic valve calcium in European-ancestry individuals found that the lead variant, rs10544872, was also associated with valve calcium in African Americans and Hispanics[4]. Other than this replication, no studies have assessed whether genetic associations for AS in European populations are transferable to other ethnicities.

In this study, we explore the extent to which genetic associations with AS and Lp(a) in a population of European ancestry can explain variation in AS in non-Hispanic Whites, African Americans, and Hispanics from the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. Using PRSice-2 to perform p-value

thresholding, we find the best-fit GRSs in each ethnicity and compare the performances of these scores to the performance of the *LPA* variant rs10455872. Our findings provide insights into the transferability of genetic associations with AS to non-European populations, as well as the role of Lp(a) in AS development across ethnicities.

Methods

Study Populations

The UK Biobank study recruited over 500,000 individuals aged 40-69 years from 22 recruitment centers across the United Kingdom between 2006 and 2010[22]. Participants provided blood samples for DNA extraction and biomarker analysis and completed a series of questionnaires. The UK Biobank received ethical approval from the North West Multi-Centre Research Ethics Committee and all participants provided written, informed consent. Cases of AS were determined using *International Classification of Diseases, 10th Revision* (*ICD-10*) diagnosis codes for AS from electronic health records (*ICD-10* I35.0 or 135.2); all remaining participants were designated controls. Lp(a) (nmol/L) was measured using an immunoturbidimetric analysis on a Randox AU5800. Measurements were taken at the initial assessment visit (2006-2010) or the first repeat assessment visit (2012-2013). Measurements that returned an error from the analyzer or were outside of the reportable range (3.80-189 nmol/L) were excluded.

The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort includes 110,266 members of the Kaiser Permanente Medical Care Plan in the Northern California Region (dbGaP Study Accession: phs000674.v3.p3). All participants answered a detailed survey, provided saliva samples for extraction of DNA, and gave

broad consent for the use of their data. The present study included only unrelated participants aged 55 years or older of self-reported White, African American, or Hispanic race/ethnicity. AS status was determined using electronic health record data from January 1, 1996, to December 31, 2015, inclusive. Cases were classified using the *International Classification of Diseases, 9th Revision* (ICD-9) diagnosis code for AS (424.1) or a procedure code for aortic valve replacement; all other individuals were classified as controls. Individuals with congenital valvular disease (*ICD-9* codes 746-747) were excluded from the analysis. Demographic characteristics are described in the Supplemental Material.

Genotyping and Imputation

For the UK Biobank samples, genotyping was performed using the Affymetrix UK BiLEVE Axiom array on an initial 50,000 participants, while the remaining 450,000 participants were genotyped using the Affymetrix UK Biobank Axiom array. Quality control and imputation were performed centrally by the UK Biobank as previously described[22].

In the GERA cohort, genome-wide genotyping was performed using ethnicity-specific Axiom Genotyping Solution (Affymetrix) arrays that have been described elsewhere[23]. Genotype data was imputed on the Michigan Imputation Server[24] using the Haplotype Reference Consortium (version r1.1)[25] for White individuals, the Consortium on Asthma among African-ancestry Populations in the Americas reference panel[26] for African Americans, and the 1000 Genomes Phase 3 reference panel[27] for Hispanic individuals.

GWAS for AS and Lp(a) in the UK Biobank

Only unrelated, genetically-confirmed White British individuals were included in the association studies. The AS GWAS included 214,947 individuals aged 55 years or older (n=1,399 AS cases); variants were assessed for association with AS using logistic regression models adjusted for age, age^2, sex, and recruitment centre. The Lp(a) GWAS included 293,274 individuals of all ages and linear regression models were adjusted for age, sex, and genotype batch. Both GWAS were run with PLINK 2.0[28,29] with adjustment for 20 principal components of ancestry.

Risk Score Construction

Genetic risk scores were constructed using GWAS for AS and Lp(a) in the UK Biobank. To be included in the genetic risk scores, variants had to be well-imputed in the UK Biobank (imputation quality score >= 0.90) with a minor allele frequency >= 1%. Variants also had to be present in the data for Whites, African Americans, and Hispanics of the GERA cohort and had to possess the same two unambiguous alleles across ethnicities. For each risk score, the overlapping sets of variants were clumped with PLINK (version 1.9)[30] to retain only independent variants with P<0.05 in the AS GWAS and Lp(a) GWAS (r^2=0.1, window size=1Mb). The independent sets of variants were analyzed to find the optimal p-value threshold for AS in each ethnicity of GERA using PRSice version 2.2.13[31]. All p-value thresholds between 5.0×10^{-8} and 0.05 were tested at intervals of 5.0×10^{-7}, for a total of 31,859 thresholds. After determining the optimal p-value thresholds using PRSice-2, the scores were computed in PLINK 2.0. For each individual, the GRS was calculated by multiplying the number of risk alleles

per variant (represented as dosages) with the regression coefficient (beta in Lp(a) GWAS and log odds ratio in AS GWAS), summed over all variants.

Association of Risk Scores with AS in GERA

The best-fit AS-GRS and Lp(a)-GRS in each ethnicity was standardized and tested for association with AS, unadjusted and adjusted for age, age^2, and sex. The best-fit AS-GRS and Lp(a)-GRS in Whites was also tested for association with AS in African Americans and Hispanics. In a sensitivity analysis, the variant rs10455872 was removed from the GRSs and associations with AS were re-tested. The variance explained was calculated for the risk scores, with and without rs10455872. The variant rs10455872 was also tested independently for association with AS after standardization. All association tests were performed using logistic regression in R[32], with two-sided p-values <0.05 considered significant.

Results

A total of 55,192 Whites, 1,917 African Americans, and 3,482 Hispanics were included in the study. Demographic characteristics of each ethnicity are presented in Table 1. Compared to Whites, African Americans and Hispanics had a larger body mass index and a higher prevalence of dyslipidemia, coronary artery disease, and diabetes (all *P*<0.001). African Americans had a higher prevalence of hypertension than Whites and Hispanics, while Whites had the highest prevalence of AS (all *P*<0.001).

In total, 37,570 variants were eligible for inclusion in the AS-GRSs and 39,018 variants were eligible for inclusion into the Lp(a)-GRSs (Figure 1). The best-fit AS-GRS

in Whites was found at a p-value threshold of 1.6×10^{-6} (n=10 variants) and was associated with an 11% increase in odds for AS (odds ratio [OR] per SD [95% CI], 1.11 [1.08, 1.15]; $P=1.1 \times 10^{-10}$). This score was not significantly associated with AS in African Americans or Hispanics (Table 1 in Supplemental Material). The optimal p-value thresholds for the best-fit AS-GRSs in African Americans and Hispanics were 5.0×10^{-8} (n=3 variants) and 3.4×10^{-4} (n=496 variants), respectively; per SD, these GRSs were associated with a 25% increase in odds for AS in African Americans (1.25 [1.02, 1.53]; $P=0.033$), and a 20% increase in odds for AS in Hispanics (1.20 [1.02,1.40], $P=0.024$).

The best-fit Lp(a)-GRS in Whites was found at a p-value threshold of 6.6×10^{-4} and included 1,291 variants. This score was associated with an 11% increase in odds for AS in Whites (OR per SD [95% CI], 1.10 [1.07,1.14]; $P=6.2 \times 10^{-9}$), but was not significantly associated with AS in African Americans or Hispanics (Table 1 in Supplemental Material). The best-fit Lp(a)-GRSs in African Americans and Hispanics were found at p-value thresholds of 1.1×10^{-6} (n=263 variants) and 0.010 (n=11,217 variants), and per SD, were associated with a 24% increase in odds for AS (1.24 [1.00,1.52]; $P=0.045$) and a 17% increase in odds for AS (1.17 [1.00,1.37]; $P=0.045$), respectively. The association results for all risk scores are summarized in Figure 2. Adjusting for age, age^2, and sex did not materially change the results (Figure 1 in Supplemental Material).

When the variant rs10455872 was excluded from the AS-GRSs, the variance explained (R^2) decreased from 0.16% to 0.063% in Whites, from 0.63% to 0.13% in African Americans, and from 0.39% to 0.37% in Hispanics. When rs10455872 was excluded from the Lp(a)-GRSs, the variance explained decreased from 0.13% to

0.076% in Whites, from 0.56% to 0.25% in African Americans, and from 0.31% to 0.26% in Hispanics. When tested independently, each SD increase of rs10455872 was associated with a 10% increase in odds for AS in Whites (OR per SD [95% CI], 1.10 [1.07, 1.14]; $P=3.9 \times 10^{-10}$) and explained 0.14% of phenotypic variance. In African Americans, rs10455872 was associated with a 25% increase in odds for AS (1.25 [1.09, 1.43]; $P=0.0010$) and explained 1.2% of phenotypic variance. The variant rs10455872 had no clear effect on AS in Hispanics (0.99 [0.84, 1.17]; $P=0.92$; $R^2<0.001$). A comparison of the variance explained by the best-fit (by ethnicity) GRSs and rs10455872 is shown in Figure 3.

Discussion

In this study, we assessed the explanatory value of genetic associations with AS and Lp(a) in the UK Biobank for AS in other ethnicities. Using PRSice-2, we created GRSs based on GWAS in White British and fit for maximal association with AS in Whites, African Americans, and Hispanics from the GERA cohort. All best-fit (by ethnicity) GRSs were significantly associated with AS, but explained only a small proportion of phenotypic variance. The variant rs10455872 accounted for a large proportion of variance explained in the best-fit GRSs for Whites and African Americans, but not Hispanics.

To our knowledge, this is the first study to assess the explanatory value of a GRS for AS. To create our GRSs, we selected only independent variants and used PRSice-2 to perform p-value thresholding. The best p-value thresholds were selected based on the maximal association with AS within the testing populations. As a result, the GRSs

are over-fitted by nature and reflect the best possible performance of European-derived genetic associations in these populations. While clumping and thresholding is one way to derive GRSs, there are several other methods including penalized regression (eg. lassosum) and Bayesian approaches (eg. LDPred). When tested on simulated data, PRSice-2 had comparable performance to these methods[31,33]. In addition, the typically fewer number of variants included using the clumping and thresholding method make the GRSs more clinically relevant and easier to assess in other populations.

In addition to creating GRSs based on associations with AS, we also created GRSs based on genetic associations with Lp(a). Both types of GRSs were derived in White British from the UK Biobank and were significantly associated with AS across ethnicities in GERA when the p-value threshold was selected in each ethnicity. However, the number of variants in the best-fit scores varied greatly by ethnicity. For example, the best-fit Lp(a)-GRS in African Americans contained 263 variants while the best-fit Lp(a)-GRS in Hispanics contained over 11,000. Furthermore, when the Lp(a)-GRS created in Whites (n=1,291 variants) was tested in African Americans and Hispanics, it had non-significant and attenuated effects on AS compared to the best-fit (by ethnicity) scores. These results are consistent with previous studies showing that GRSs developed in European populations perform less well in non-European ethnicities[34,35]. While European associations with AS may have some predictive value in admixed populations, GRSs are likely not transferable without ethnicity-specific optimization.

As mentioned, the GRSs in this study were derived using a clumping and thresholding method. In previous studies, when this method was used to develop risk

scores for coronary artery disease, the p-value threshold for the optimal GRS was 0.001 in one study[36] and 0.05 in another study[37]. The liberal p-value thresholds for these scores is consistent with the notion that coronary artery disease is a polygenic trait, with genetic risk determined by the cumulative risk of a large number of variants with small effects. In contrast, the p-value threshold for the best-fit AS-GRS in Whites from our study was 1.6×10^{-6}. This more stringent threshold may suggest that compared to coronary artery disease, risk of AS may be less polygenic and influenced by a smaller number of variants.

Despite being created based on maximum association with AS, the GRSs in our study explained less than 1.0% of phenotypic variance. By comparison, risk scores recently developed for type 2 diabetes and coronary artery disease explained 2.9% and 4.0% of phenotypic variance, respectively[37]. When rs10455872 was excluded from the AS- and Lp(a) GRSs in Whites and African Americans, the variance explained was reduced by more than half. On its own, the variance explained by rs10455872 was comparable to that of the GRSs in Whites (0.14%) and was even higher than the GRSs in African Americans (1.2%). These findings further support the notion that the genetic architecture of AS may be oligogenic with strong influence from variants in *LPA*.

While rs10455872 was an important component of the GRSs in Whites and African Americans, the variance explained only decreased slightly in Hispanics when this variant was removed. In addition, rs10455872 alone showed no significant association with AS among Hispanics. This finding is in contrast to a previous study in the Multi-Ethnic Study of Atherosclerosis (MESA), where rs10455872 was significantly associated with aortic valve calcium in Hispanics[4]. However, another study in MESA

showed that Lp(a) levels themselves were not significantly associated with aortic valve calcium in Hispanics[38]. Follow-up studies are needed to confirm the role of Lp(a) in disease development in this population.

This study had several strengths and limitations. This was the first study to broadly assess the transferability of genetic associations with AS in a European-ancestry population to African Americans and Hispanics. Our GRSs were derived using associations in the UK Biobank, a large and well-powered cohort. To ensure that our risk scores had the best chance of transferring to these populations, we selected only variants that were common across ethnicities to be included in the risk scores. As a result, ethnicity-specific risk variants may have been excluded from the risk scores, resulting in false negatives. In addition, our sample sizes for African Americans and Hispanics were smaller than for Whites, decreasing statistical power and likely hindering transferability to these populations. Finally, AS is a late-onset disease, further hindering the performance of a GRS in a cohort of predominantly middle-aged adults (i.e. individuals with high genetic risk may not have developed the disease yet).

In summary, we created GRSs for AS using genetic associations with AS and Lp(a) in the UK Biobank and fitted for association with AS in Whites, African Americans, and Hispanics from the GERA cohort. Our results demonstrate that European-derived GRSs have some explanatory value in African Americans and Hispanics when the p-value thresholds are selected in these populations, but are not significantly associated otherwise. In addition, our findings suggest that AS risk may be greatly explained by variation in *LPA*, highlighting the importance of plasma Lp(a) as a risk factor for AS.

Acknowledgments

This research has been conducted using the UK Biobank Resource under applications 24281 and 41025. We thank the participants, staff, and investigators of the UK Biobank and GERA cohorts for their contributions.

Funding Sources

Ms Hoekstra was supported by an award from the Canadian Institutes of Health Research (CIHR) for the duration of the study. Ms Chen has received grants from the McGill University Faculty of Medicine and McGill University Health Centre Foundation. Dr Thanassoulis has received research support from the CIHR, the Heart and Stroke Foundation of Canada, the NHLBI (grant R01 HL128550), the Doggone Foundation, the Courtois Foundation, and the Satoko Shibata and Richard Ingram Foundation. Dr Thanassoulis receives salary support from the Fonds de Recherche Santé–Québec.

Data came from a grant, the Resource for Genetic Epidemiology Research in Adult Health and Aging (RC2 AG033067; Schaefer and Risch, PIs) awarded to the Kaiser Permanente Research Program on Genes, Environment, and Health (RPGEH) and the UCSF Institute for Human Genetics. The RPGEH was supported by grants from the Robert Wood Johnson Foundation, the Wayne and Gladys Valley Foundation, the Ellison Medical Foundation, Kaiser Permanente Northern California, and the Kaiser Permanente National and Northern California Community Benefit Programs.

Disclosures

Dr Thanassoulis has participated in advisory boards for Amgen, Sanofi/Regeneron, Ionis, HLS Therapeutics, and Servier Canada and has received research grants from Ionis and Servier for research outside the scope of this work.

References

1. Iung B, Baron G, Butchart EG, Delahaye F, Gohlke-Bärwolf C, Levang OW, et al. A prospective survey of patients with valvular heart disease in Europe: The Euro Heart Survey on valvular heart disease. *Eur Heart J*. 2003;24:1231–1243.
2. Danielsen R, Aspelund T, Harris TB, Gudnason V. The prevalence of aortic stenosis in the elderly in Iceland and predictions for the coming decades: The AGES-Reykjavík study. *Int J Cardiol*. 2014;176:916–922.
3. Clark MA, Arnold S V., Duhay FG, Thompson AK, Keyes MJ, Svensson LG, et al. Five-year clinical and economic outcomes among patients with medically managed severe aortic stenosis: Results from a medicare claims analysis. *Circ Cardiovasc Qual Outcomes*. 2012;5:697–704.
4. Thanassoulis G, Campbell CY, Owens DS, Smith JG, Smith A V., Peloso GM, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. *N Engl J Med*. 2013;368:503–512.
5. Kamstrup PR, Tybjærg-Hansen A, Nordestgaard BG. Elevated lipoprotein(a) and risk of aortic valve stenosis in the general population. *J Am Coll Cardiol*. 2014;63:470–477.
6. Arsenault BJ, Boekholdt SM, Dubé MP, Rhéaume É, Wareham NJ, Khaw KT, et

al. Lipoprotein(a) levels, genotype, and incident aortic valve stenosis a prospective mendelian randomization study and replication in a case-control cohort. *Circ Cardiovasc Genet*. 2014;7:304–310.

7. Cairns BJ, Coffey S, Travis RC, Prendergast B, Green J, Engert JC, et al. A replicated, genome-wide significant association of aortic stenosis with a genetic variant for lipoprotein(a): Meta-analysis of published and novel data. *Circulation*. 2017;135:1181–1183.
8. Chen HY, Dufresne L, Burr H, Ambikkumar A, Ranatunga DK, Whitmer RA, et al. Observational and Genetic Associations of Cardiovascular Risk Factors with Aortic Stenosis. *Atheroscler Suppl*. 2018;32:74.
9. Nordestgaard BG, Langsted A. Lipoprotein (a) as a cause of cardiovascular disease: Insights from epidemiology, genetics, and biology. *J Lipid Res*. 2016;7:1953–1975.
10. Helgadottir A, Thorleifsson G, Gretarsdottir S, Stefansson OA, Tragante V, Thorolfsdottir RB, et al. Genome-wide analysis yields new loci associating with aortic valve stenosis. *Nat Commun*. 2018;9.
11. Thériault S, Gaudreault N, Lamontagne M, Rosa M, Boulanger MC, Messika-Zeitoun D, et al. A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. *Nat Commun*. 2018;9:6–13.
12. Thériault S, Dina C, Messika-zeitoun D, Scouarnec S Le, Capoulade R, Gaudreault N, et al. Genetic association analyses highlight IL6, ALPL, and NAV1 as three new susceptibility genes underlying calcific aortic valve stenosis. *BioRxiv*. 2019:1–35.

13. Chen HY, Cairns BJ, Small AM, Burr HA, Ambikkumar A, Martinsson A, et al. Association of FADS1/2 Locus Variants and Polyunsaturated Fatty Acids with Aortic Stenosis. *JAMA Cardiol*. 2020:1–9.

14. Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460:748–752.

15. Demirkan A, Penninx BWJH, Hek K, Wray NR, Amin N, Aulchenko YS, et al. Genetic risk profiles for depression and anxiety in adult and elderly cohorts. *Mol Psychiatry*. 2011;16:773–783.

16. Ganna A, Magnusson PKE, Pedersen NL, De Faire U, Reilly M, Ärnlöv J, et al. Multilocus genetic risk scores for coronary heart disease prediction. *Arterioscler Thromb Vasc Biol*. 2013;33:2267–2272.

17. Osnabrugge RLJ, Mylotte D, Head SJ, Van Mieghem NM, Nkomo VT, Lereun CM, et al. Aortic stenosis in the elderly: Disease prevalence and number of candidates for transcatheter aortic valve replacement: A meta-analysis and modeling study. *J Am Coll Cardiol*. 2013;62:1002–1012.

18. Généreux P, Stone GW, O'Gara PT, Marquis-Gravel G, Redfors B, Giustino G, et al. Natural History, Diagnostic Approaches, and Therapeutic Strategies for Patients with Asymptomatic Severe Aortic Stenosis. *J Am Coll Cardiol*. 2016;67:2263–2288.

19. Enkhmaa B, Anuurad E, Berglund L. Lipoprotein (a): Impact by ethnicity and environmental and medical conditions. *J Lipid Res*. 2016;57:1111–1125.

20. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of

current polygenic risk scores may exacerbate health disparities. *Nat Genet*. 2019;51:584–591.

21. Beydoun HA, Beydoun MA, Liang H, Dore GA, Shaked D, Zonderman AB, et al. Sex, Race, and Socioeconomic Disparities in Patients With Aortic Stenosis (from a Nationwide Inpatient Sample). *Am J Cardiol*. 2016;118:860–865.
22. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562:203–209.
23. Hoffmann TJ, Zhan Y, Kvale MN, Hesselson SE, Gollub J, Iribarren C, et al. Design and coverage of high throughput genotyping arrays optimized for individuals of East Asian, African American, and Latino race/ethnicity using imputation and a novel hybrid SNP selection algorithm. *Genomics*. 2011;98:422–430.
24. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–1287.
25. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48:1443–1448.
26. Mathias RA, Taub MA, Gignoux CR, Fu W, Musharoff S, O'Connor TD, et al. A continuum of admixture in the Western Hemisphere revealed by the African Diaspora genome. *Nat Commun*. 2016;7:12522.
27. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et

al. A global reference for human genetic variation. *Nature*. 2015;26:68–74.

28. Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
29. Purcell S, Chang C. PLINK 2.0. https://www.cog-genomics.org/plink/2.0/
30. Purcell S, Chang C. PLINK 1.9. https://www.cog-genomics.org/plink/1.9/
31. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8:giz082.
32. R Development Core Team R. R: A Language and Environment for Statistical Computing. 2011.
33. Ge T, Chen CY, Ni Y, Feng YCA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun*. 2019;10:1776.
34. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet*. 2017;100:635–649.
35. Grinde KE, Qi Q, Thornton TA, Liu S, Shadyab AH, Chan KHK, et al. Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet Epidemiol*. 2019;43:50–62.
36. Goldstein BA, Yang L, Salfati E, Assimes TL. Contemporary Considerations for Constructing a Genetic Risk Score: An Empirical Approach. *Genet Epidemiol*. 2015;39:439–445.
37. Khera A V., Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk

equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224.

38. Cao J, Steffen BT, Budoff M, Post WS, Thanassoulis G, Kestenbaum B, et al. Lipoprotein(a) levels are associated with subclinical calcific aortic valve disease in white and black individuals: The multi-ethnic study of atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2016;36:1003–1009.

Tables

Table 1. Characteristics of individuals 55 years and older in the GERA cohort.

	Whites	**African Americans**	**Hispanics**	***P***
N	55,192	1,917	3,482	N/A
Male (%)	23,687 (42.9)	844 (44.0)	1,543 (44.3)	0.18
Age, years (mean (SD))	67.9 (8.6)	66.2 (8.0)	65.9 (7.9)	<0.001
BMI, kg/m^2 (mean (SD))	26.9 (5.1)	29.1 (5.6)	28.1 (5.2)	<0.001
Dyslipidemia (%)	31,344 (56.8)	1,255 (65.5)	2,180 (62.6)	<0.001
Coronary artery disease (%)	12,895 (23.4)	493 (25.7)	804 (23.1)	0.051
Hypertension (%)	22,005 (39.9)	1,139 (59.4)	1,396 (40.1)	<0.001
Ever smoked (%)	26,374 (50.0)	911 (50.9)	1,524 (46.3)	<0.001
Diabetes (%)	5,960 (10.8)	424 (22.1)	697 (20.0)	<0.001
Aortic valve stenosis (%)	3,469 (6.3)	86 (4.5)	159 (4.6)	<0.001

Abbreviations: BMI, body mass index; GERA, Genetic Epidemiology Research on Adult Health and Aging.

Figures

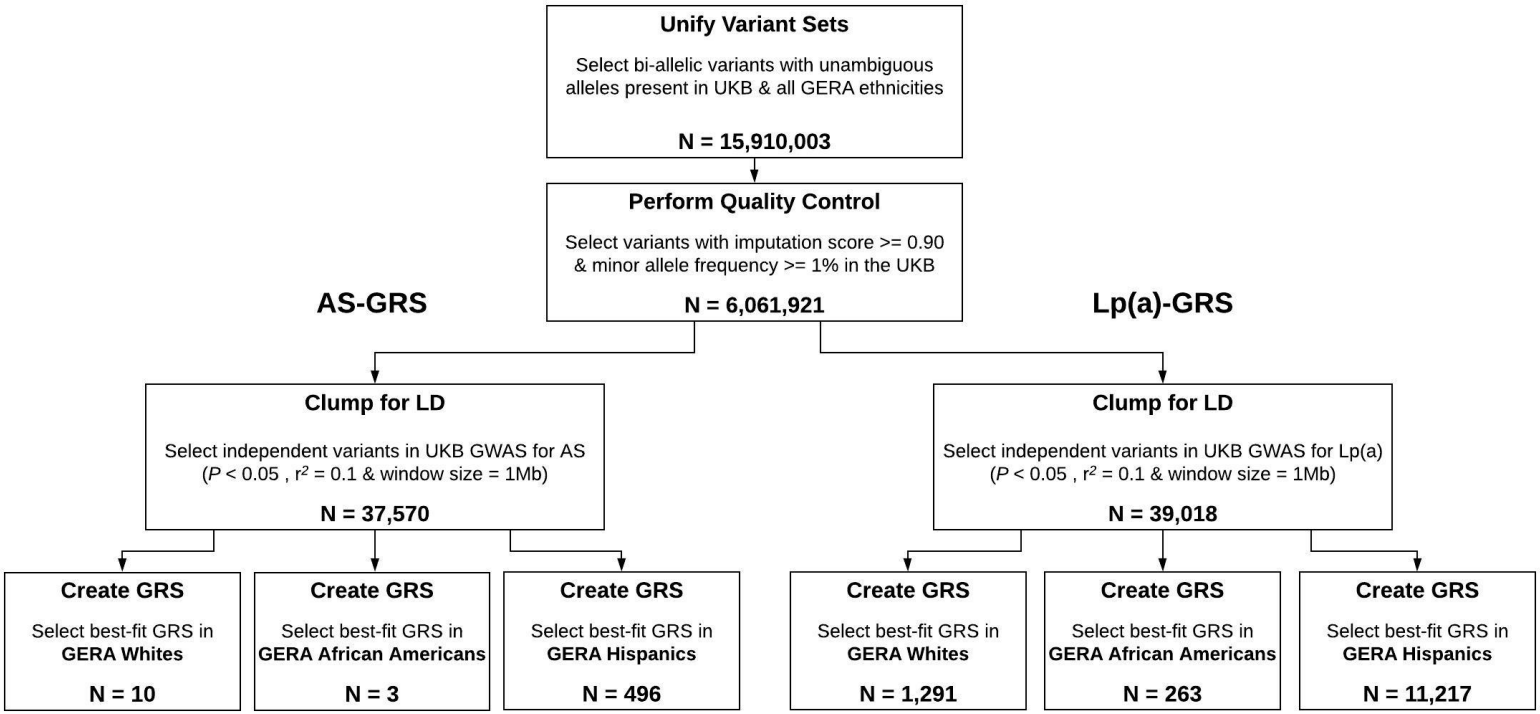

Figure 1. Variant selection process for the AS-GRSs and Lp(a)-GRSs created using GWAS in the UK Biobank and fitted for association with AS in Whites, African Americans, and Hispanics from the GERA cohort. N refers to the number of variants. Abbreviations: UKB, UK Biobank; GERA, Genetic Epidemiology Research on Adult Health and Aging; AS, aortic stenosis; Lp(a), lipoprotein(a); LD, linkage disequilibrium; GRS, genetic risk score; GWAS, genome-wide association study.

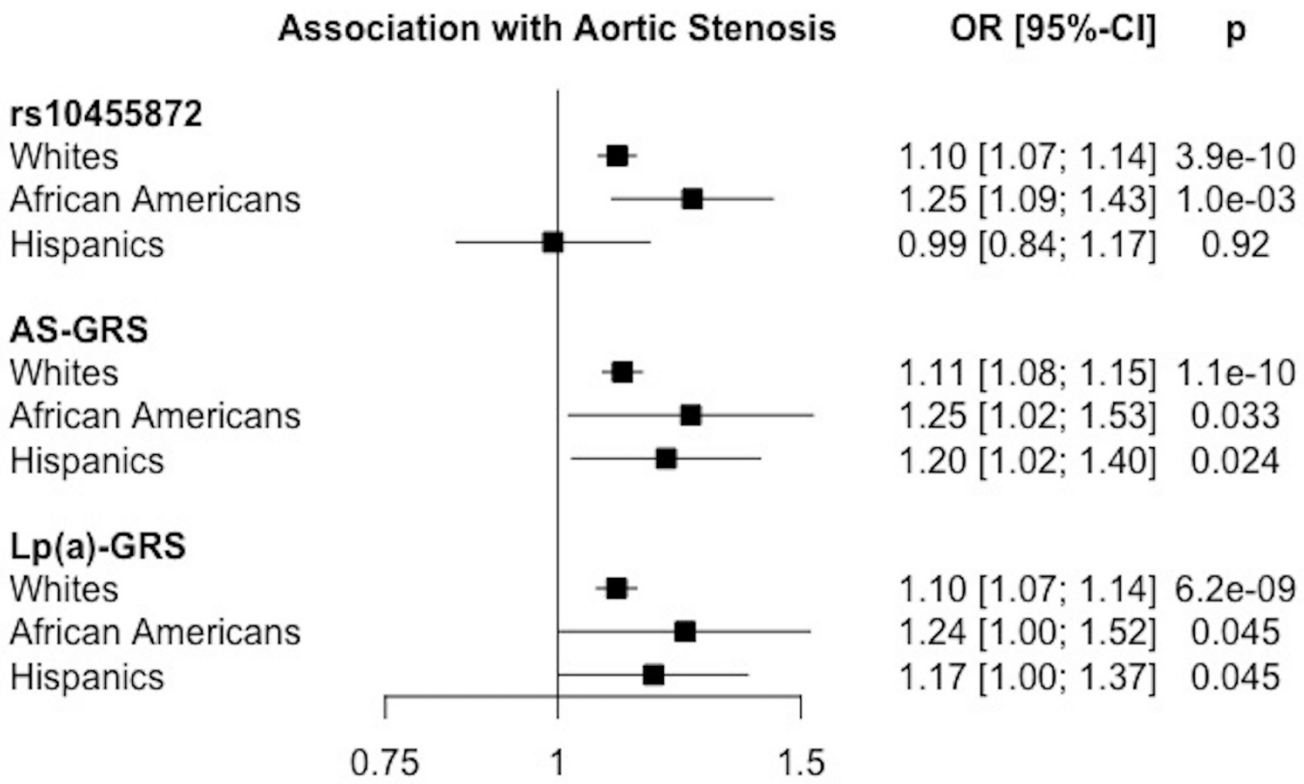

Figure 2: Association of rs10455872 and the best-fit (by ethnicity) AS-GRSs and Lp(a)-GRSs in Whites, African Americans, and Hispanics from the GERA cohort. Odds ratio (OR) is per standard deviation. Abbreviations: CI, confidence interval; AS, aortic stenosis; Lp(a), lipoprotein(a); GRS, genetic risk score; GERA, Genetic Epidemiology Research on Adult Health and Aging.

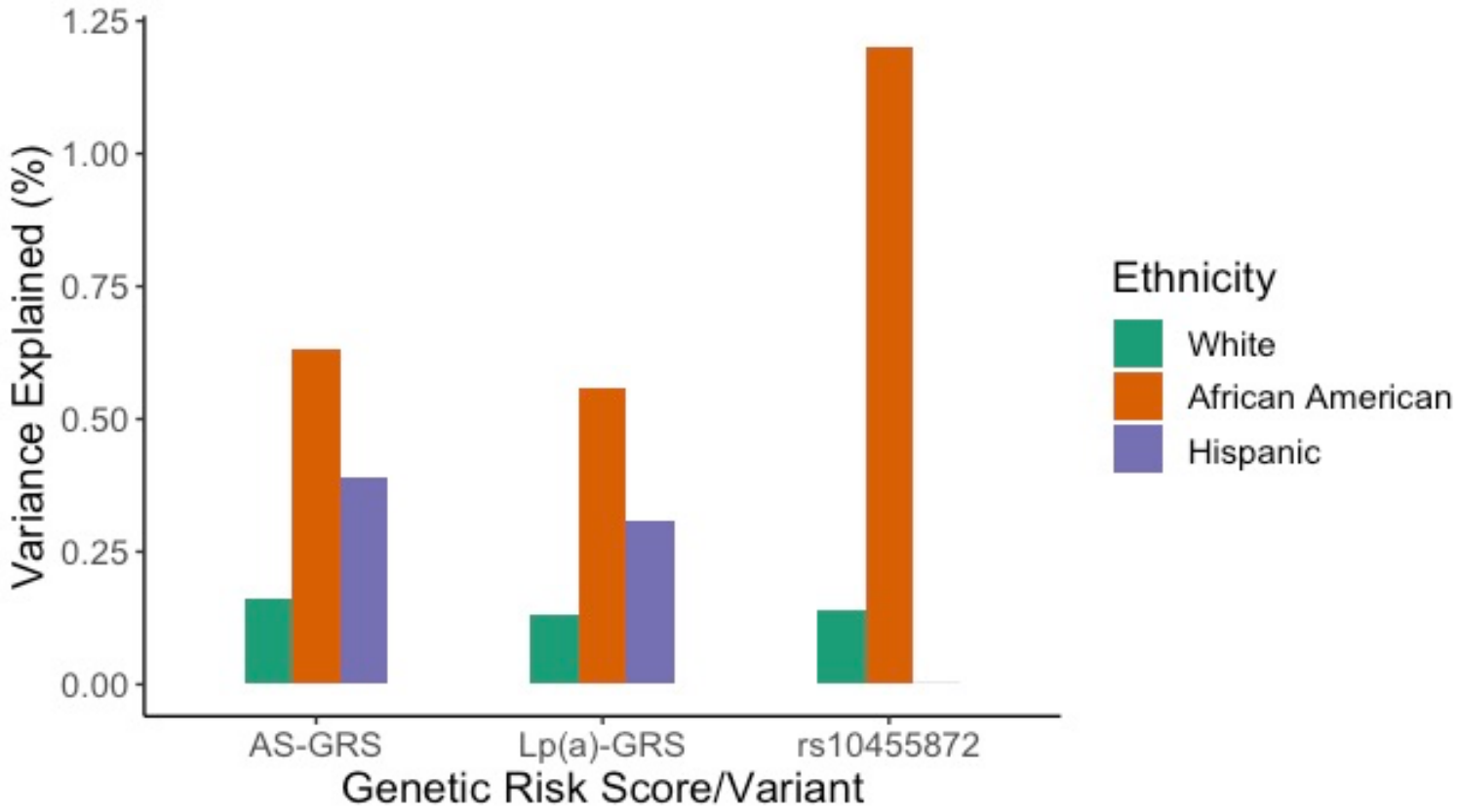

Figure 3: Proportion of variance explained in AS by rs10455872 and the best-fit (by ethnicity) AS-GRSs and Lp(a)-GRSs in Whites, African Americans, and Hispanics from the GERA cohort. Abbreviations: AS, aortic stenosis; Lp(a), lipoprotein(a); GRS, genetic risk score; GERA, Genetic Epidemiology Research on Adult Health and Aging.

SUPPLEMENTAL MATERIAL

Supplemental Methods

Variable Descriptions

In the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, age and sex were recorded from the medical chart, with ages greater than 90 years rounded down to 90 to enhance privacy (n=412). Hypertension, smoking, and diabetes were self-reported in a questionnaire and body mass index was calculated from self-reported height and weight. Dyslipidemia was defined as two or more diagnoses of lipid metabolism disorders in the electronic health records (*International Classification of Diseases, Ninth Revision [ICD-9]* 272) and one or more statin prescriptions in the Kaiser Permanente prescriptions database. Coronary artery disease was defined as a diagnosis of myocardial infarction or coronary artery disease (*ICD-9* 410-414), a procedure code for coronary artery bypass surgery, or percutaneous coronary intervention in the electronic health records.

Supplemental Tables

Table 1. Association of the best-fit AS-GRS and Lp(a)-GRS in GERA Whites across ethnicities in the GERA cohort.

Ethnicity	**AS-GRS**		**Lp(a)-GRS**	
	OR [95% CI]	***P***	**OR [95% CI]**	***P***
White	1.11 [1.08,1.15]	1.1×10^{-10}	1.10 [1.07,1.14]	6.2×10^{-9}
African American	1.00 [0.80,1.23]	1.00	1.20 [0.97,1.48]	0.092
Hispanic	1.11 [0.95,1.27]	0.17	1.11 [0.95,1.29]	0.20

Abbreviations: AS, aortic stenosis; Lp(a), lipoprotein(a); GRS, genetic risk score; OR, odds ratio; CI, confidence interval; GERA, Genetic Epidemiology Research on Adult Health and Aging.

Supplemental Figures

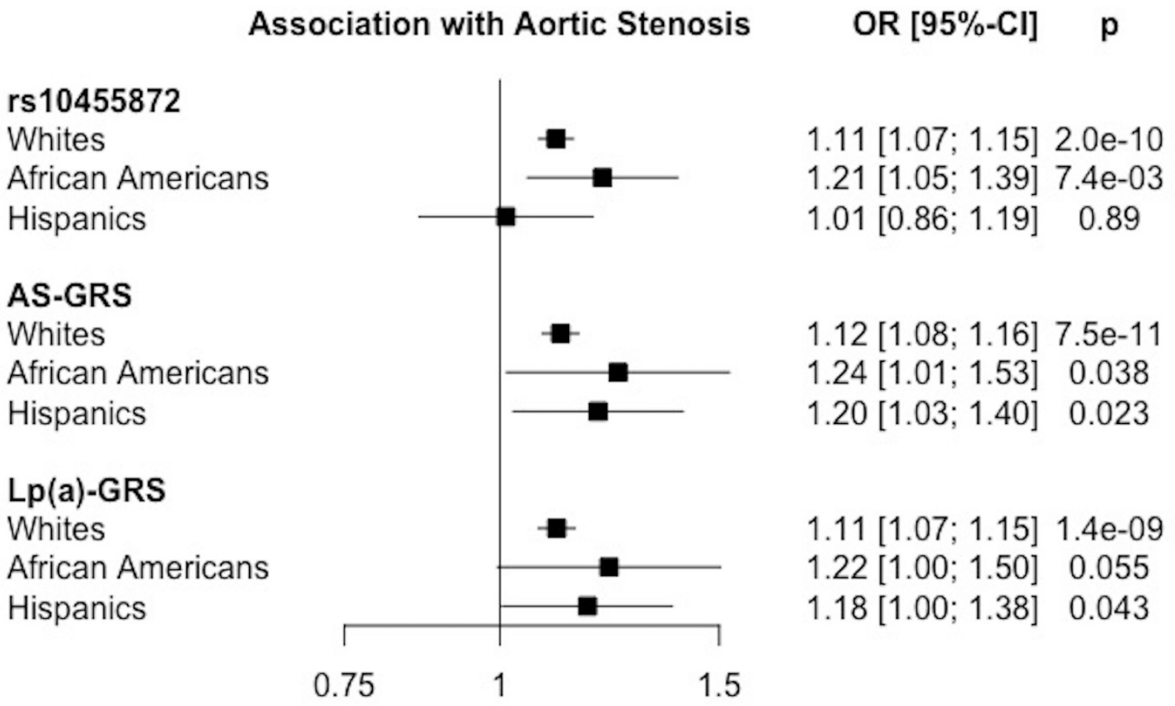

Figure 2: Association of rs10455872 and the best-fit (by ethnicity) AS-GRSs and Lp(a)-GRSs with AS in Whites, African Americans, and Hispanics from the GERA cohort after adjustment for age, age^2, and sex. Abbreviations: AS, aortic stenosis; Lp(a), lipoprotein(a); GRS, genetic risk score; OR, odds ratio; CI, confidence interval; GERA, Genetic Epidemiology Research on Adult Health and Aging.

Chapter 5: General Discussion

Summary

In this thesis, a large-scale GWAS of Lp(a) was performed in White British individuals from the UK Biobank, validating previously identified loci, including *LPA*, *APOE*, and *CETP*, and identifying *APOH* as a novel risk locus. The Lp(a) GWAS was used to create GRSs in Whites, African Americans, and Hispanics from the GERA cohort, where the p-value threshold for the GRSs was selected based on clumping and thresholding to maximize association with AS in these populations. GRSs were derived from an GWAS for AS in the UK Biobank in a similar manner. The best-fit (by ethnicity) Lp(a)- and AS-GRSs were significantly associated with AS in each ethnicity, with the Lp(a)-GRSs explaining a comparable proportion of variance to the AS-GRSs. However, the variance explained by all scores was small, and the variant rs10455872 in *LPA* accounted for a considerable proportion in Whites and African Americans.

Genetic associations with Lp(a)

Consistent with previous GWAS of Lp(a)[55,72,73], the variant rs10455872 had the largest effect on Lp(a) levels. White British individuals carrying two minor alleles for rs10455872 had a median Lp(a) level of 148 nmol/L, compared to the cohort median of 20.1 nmol/L. In addition, it was demonstrated for the first time that rs10455872 is associated with Lp(a) in South Asians; other studies in this population have failed to assess this variant due to limited sample sizes and a lower risk allele frequency in this population[74]. As shown in other studies of admixed populations with African ancestry[75,76], rs10455872

was also significantly associated in Black Caribbeans. This variant was not significantly associated in Black Africans, likely due to its near absence in this population.

The variant rs1065853 at *APOE* had the next largest effect in White British, with individuals homozygous for the minor allele having levels approximately 10 nmol/L lower than the median. Like rs10455872, the variant rs1065853 in *APOE* (or a SNP in high LD, the apoE2-defining rs7412) has been previously associated with Lp(a) in populations of African ancestry[77,78], but has never been investigated in a South Asian population. In the present study, the variant rs1065853 was significantly associated with Lp(a) in Black Caribbeans, Black Africans, and South Asians from the UK Biobank, making it the only lead variant to replicate in all ethnicities tested. The specific mechanism underlying this association requires further investigation, but may involve the apoE2 isoform having reduced affinity for LDL receptors, resulting in a greater number of receptors available for Lp(a)[79,80].

In addition to *LPA* and *APOE*, a new locus for Lp(a) was identified: *APOH*. The variant rs8178824 in *APOH* was significantly associated with increased Lp(a) levels, an association that was replicated in a meta-analysis of two independent European-ancestry cohorts. Previously, in vitro studies have shown that β2GPI, the protein product of *APOH*, interacts directly with the apo(a) component of Lp(a)[81]. Since β2GPI is known to have a high affinity for negatively charged phospholipids[82], this interaction may be mediated by the binding of β2GPI to oxidized phospholipids, which accumulate on apo(a)[83]. A study demonstrating that β2GPI accelerates triglyceride clearance in rats[84] supports the hypothesis that β2GPI may play a role in the clearance of Lp(a) from plasma.

Neither the variant in *APOH,* nor the variants in *CETP* and *AC093639.1,* were significantly associated in the other ethnicities from the UK Biobank. Relative to *LPA* and *APOE*, these variants had only small effects on Lp(a) in White British; thus, the sample sizes of the other ethnicities may not have been large enough to detect an effect. Other potential reasons for the differing effects between populations include gene-gene or gene-environment interactions, whereby the effect of a gene is influenced by other genes (epistasis)[85] or by environmental exposures, such as diet or climate[86]. Alternatively, genetic heterogeneity may be caused by differences in LD structure between the populations[87], which can occur when the identified variant in the GWAS is not causal and tags different variants in different populations (Figure 1)[31].

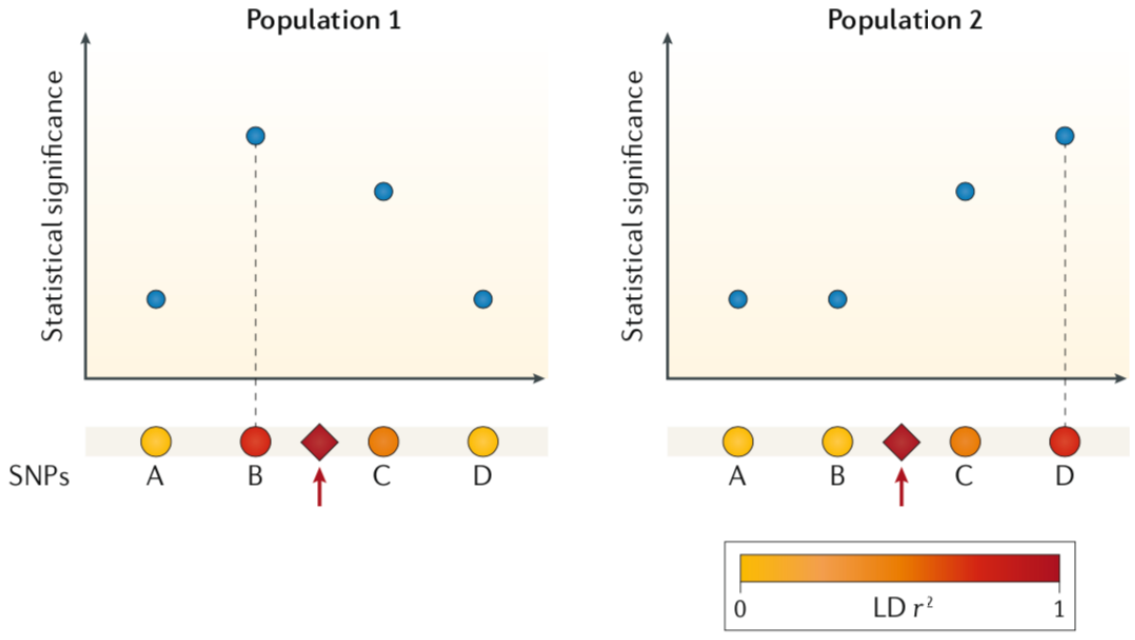

Figure 3: Differences in linkage disequilibrium (LD) between an unobserved causal variant (the diamond) and observed variants (the circles) between two populations. Modified from Gurdasani et al., 2019.

The lack of reproducibility observed for these variants may also be due to false positive associations in the discovery study. However, the loci *CETP* and *APOH* have been previously linked to Lp(a), either through previous association studies[78], clinical studies[88–90], or in vitro studies[81], supporting the validity of these associations. In addition, the *APOH* variant was significantly associated in a meta-analysis of independent European-ancestry cohorts, further strengthening the evidence for a true association at this locus. The variant in *AC093639.1*, on the other hand, did not replicate in other ethnicities, showed no association in the meta-analysis, and has not been previously connected to Lp(a), suggesting a potential false positive association.

Lp(a) as a risk factor for aortic stenosis

The Lp(a) GWAS and a GWAS for AS were used to create GRSs for AS, where the p-value thresholds for the GRSs were selected based on maximal association with AS in Whites, African Americans and Hispanics of the GERA cohort. It was demonstrated, for the first time, that using the same SNP selection procedure, GRSs for Lp(a) explain nearly the same proportion of variance in AS as GRSs for AS, suggesting an important role for Lp(a) in AS development. Interestingly, when the Lp(a)- and AS-GRSs fit for association in Whites were applied to African Americans, the Lp(a)-GRS showed a strong effect (OR=1.20), while the AS-GRS showed no effect (OR=1.00). While neither effect was statistically significant and the confidence intervals were large, this trend suggests that an Lp(a)-GRS could be more robust to transferring to African Americans when creating an ethnicity-specific GRS for AS isn't possible.

The effects of the Lp(a)- and AS-GRS were largest in African Americans, despite these scores containing fewer variants than the best-fit scores in Whites and Hispanics. The variant rs10455872 also demonstrated the largest effect on AS in this population; each SD was associated with a 25% increase in odds for AS, corresponding to more than a 4-fold increase in the odds per risk allele. When rs10455872 was previously tested for association with AVC in African Americans of the MESA cohort, it similarly demonstrated a large effect, with each risk allele more than tripling the odds for AVC[54]. As was demonstrated in Black Caribbeans from the UK Biobank, this variant is also associated with increased Lp(a) levels in populations of African ancestry, suggesting that the association with AS/AVC is likely mediated by Lp(a), as it is in Whites.

Interestingly, while levels of Lp(a) are several times higher in African Americans than other populations[91,92], these increased levels don't seem to translate to a higher risk of AS[93]. In fact, in the present study, the prevalence of AS was significantly lower in African Americans (4.5%) than Whites (6.3%) from the GERA cohort. This association has been noted in several other studies[94–98], and persists after adjustment for traditional cardiovascular risk factors. One possible explanation for this trend is referral bias, where African Americans are less likely to be referred for echocardiography and therefore less likely to be diagnosed with AS[99]. One large study assessed this possibility by comparing the prevalence of severe mitral regurgitation, another condition requiring an echocardiogram-based diagnosis, between African Americans and Whites; no differences were found, suggesting that referral bias did not play a large role[96]. Other potential explanations for the lower prevalence of AS in African Americans include lower socioeconomic status and reduced access to healthcare in this population[99]. It also

remains possible that African Americans are protected from their elevated Lp(a) levels or have a different etiology of disease, but this requires further investigation.

The relationship between Lp(a) and AS in Hispanics is even less clear. The best-fit Lp(a)-GRS in Hispanics was significantly associated with AS and, like the score in African Americans, explained a larger proportion of phenotypic variance than the score in Whites. However, the lead variant rs10455872 showed no association with AS in this population, despite this variant showing association with AVC in a previous study[54]. Studies of Lp(a) levels in Hispanics have also had conflicting results. One study of 316 Mexican Americans showed that Lp(a) levels were significantly lower in this population than in Whites (n=242)[100], while another study showed that Hispanic men (n=126) had significantly higher levels of Lp(a) than White men (n=222)[101]. In the MESA cohort, median Lp(a) levels were nearly identical in Hispanics (n=1,044) and Whites (n=1,677)[102], yet Lp(a) levels in Hispanics were not significantly associated with AVC[93].

The somewhat contrasting results between studies of Hispanics may reflect the complex admixture in this population. Hispanics typically possess European, African, and Native American ancestry, the proportions of which vary widely between Hispanic subgroups[103–106]. In the MESA cohort, where most Hispanics self-identify as having origins in either Central America, Cuba, the Dominican Republic, Mexico, Puerto Rico, or South America, there are large differences in ancestry proportions between the subgroups[107]. For example, genomes within the Caribbean subgroups primarily reflect admixture between European and African ancestry, while those in the Mexican subgroup primarily reflect admixture of European and Native American ancestry (Figure

2). An analysis of Hispanics stratified by subgroup could contribute to our understanding of why rs10455872 has demonstrated conflicting results between Hispanic cohorts.

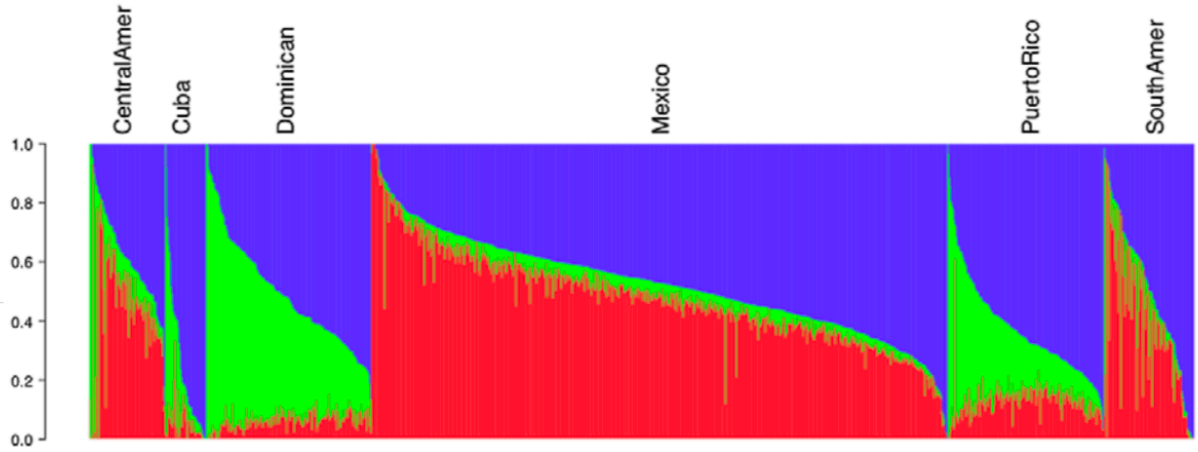

Figure 4: Estimated proportions of ancestry in 1,374 Hispanic individuals from the MESA cohort with self-identified origins in Central America, Cuba, Dominican Republic, Mexico, Puerto Rice, and South America, where red represents Native American ancestry, green represents African ancestry, and blue represents European ancestry. Modified from Manichaikul et al., 2012.

Genetic risk scores for aortic stenosis

The GRSs in the present study were developed to assess the role of genetic variation in disease. However, an increasing number of GRSs are being developed to predict the risk of common complex diseases[108–110]. Given that an individual's genetic profile is determined at birth, GRSs may theoretically be used to predict future disease risk in unaffected individuals, or to predict future disease severity in early stage disease[111–113]. However, a major obstacle preventing the clinical implementation of GRSs is ensuring that they are equally applicable to all ethnic groups. Currently, the majority of GRSs are being developed using GWAS of European-ancestry individuals[114]; therefore, the

variants and their effect sizes are specific to these populations. As a result, the GRSs perform less well when applied to other ethnicities[115,116].

The present study demonstrated that selecting ethnicity-specific p-value thresholds for GRSs can improve their performance, even when the variants are selected from a European GWAS. However, several other methods have been proposed for improving the accuracy of GRSs in non-European populations. In a study assessing the transferability of GRSs from Europeans to Hispanics, the optimal GRSs most often included variants selected from a European GWAS, but with weights selected from a meta-analysis of the European GWAS and a Hispanic GWAS[117]. Similarly, the "multi-ethnic" approach proposed by Marquez-Luna *et al*. involves combining weights from a European population and the target population in a linear combination[118]. One method designed specifically for admixed individuals involves creating "partial" GRSs for each region of the genome corresponding to local ancestry at that region and combining them into one score[119]. All of these methods aim to leverage the large sample sizes of European populations while incorporating ethnic-specific information to improve prediction accuracy in the target population.

Future directions

As the body of evidence supporting Lp(a)'s role in both AS and CAD increases, research into Lp(a)-lowering therapies is becoming increasingly important. Several therapies have been shown to modestly lower Lp(a), including niacin[120], mipomersen[121], cholesteryl ester transfer protein (CETP) inhibitors[88–90] and proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors[122–125]. Currently, more specific Lp(a) therapies

are being tested that target apo(a) mRNA in the liver. In clinical trials, these antisense oligonucleotide therapies have demonstrated potent and dose-dependent reductions in Lp(a) levels[64,65].

The present study identified *APOH* as a novel locus for Lp(a), pointing to β2GPI, the product of the *APOH* gene, as a potential therapeutic target. Like *PCSK9*, mutations in *APOH* also result in reductions of LDL-C[126–130]. In fact, β2GPI was recently identified as a potential endogenous inhibitor of PCKS9[131], suggesting that upregulating the action of *APOH* may produce similar results to PCSK9 inhibitors. The combined action of lowering LDL-C and Lp(a) would make this type of drug potentially useful for the prevention of AS[16]. Future work should investigate the mechanism by which β2GPI alters levels of Lp(a) and LDL-C, and whether targeting of the *APOH* locus or β2GPI could produce a clinically useful effect.

In addition to providing novel insights into the determinants of Lp(a), this study attempted to extend these findings to other ethnic groups. However, to truly understand ethnic differences in associations with Lp(a) and AS, discovery studies must be performed in non-European populations. For many diseases and traits, a major barrier to this pursuit is limited sample sizes. Large sample sizes are highly valued in genomic research, which has resulted in cohorts like the UK Biobank becoming the new standard for discovery GWAS[71]. However, opportunities to study non-European populations are increasing with the creation of large and diverse cohorts, such as the Million Veteran Program, where non-European individuals make up nearly 30% of the ~800,000 participants[132]; the Trans-Omics for Precision Medicine (TOPMed) Program, where 60% of ~155,000 participants are of non-European ancestry[133]; and the Population

Architecture using Genomics and Epidemiology (PAGE) Consortium, whose next phase (PAGE II) will genotype approximately 50,000 individuals from non-European cohorts[134].

Another challenge of performing GWAS in non-European populations is a lack of adequate reference panels for genetic imputation. If the markers for a reference panel are selected based on allele frequencies and LD patterns in European populations, performing imputation in other populations will result in reduced accuracy and therefore reduced statistical power in a GWAS[135]. Earlier available panels include the Human Genome Diversity Project[136], the HapMap Consortium[137], and the 1000 Genomes Project (1000G)[138], all of which were created using diverse but small samples[139]. The Human Reference Consortium (HRC) created a much larger reference panel but included predominantly individuals of European ancestry[140]. Recently, the TOPMed panel was released, which was created using over 60,000 deeply sequenced genomes from a diverse sample including ~50% non-European individuals[141]. Compared to the HRC and 1000G panels, the TOPMed panel has been shown to greatly increase the imputation quality of low-frequency variants in admixed populations like African Americans[141].

Another barrier to performing discovery analyses in non-Europeans is genetic admixture[142]. As previously discussed, ancestry can be a confounding factor in GWAS, making admixed populations more challenging to study[31]. However, several methods have been developed to detect and adjust for stratification due to admixture[143], including: genomic control methods[41,144]; the inference of genetic ancestry through model-based programs such as STRUCTURE[145] and ADMIXTURE[146]; adjusting for principal components (PCs)[43]; or using mixed model programs to perform GWAS, such

as TASSEL[147] and EMMAX[148]. In some cases, genetic admixture can actually be leveraged to discover new risk loci[149]; in a technique known as admixture mapping, ancestry analysis in admixed individuals can help identify regions in the genome with excess ancestry among disease cases[150].

In addition to facilitating the discovery of ancestry-linked risk loci, performing genetic studies in diverse populations may lead to the discovery of variants that are simply too rare to detect in Europeans[151]. For example, a study in African Americans identified nonsense variants in *PCSK9* that are common in this population (~2%), but rare (<0.1%) in individuals of European descent[152]. These variants were shown to be associated with remarkable reductions in LDL-C and an accompanying reduction in the risk for CAD[153]. Findings from these studies have not only enhanced our understanding of the genetic determinants of LDL-C, but have encouraged the development of LDL-C-lowering therapies (PCSK9 inhibitors) that stand to benefit the broader population[154]. Genetic discovery studies of Lp(a) and AS in diverse populations may achieve similar successes.

Conclusion

In conclusion, a large-scale GWAS in the UK Biobank identified several loci associated with Lp(a) levels in White British individuals, including the novel locus *APOH*, encoding β2GPI. Only the lead variants in *LPA* and *APOE* were significantly associated in smaller sample sizes of Black Africans, Black Caribbeans, and South Asians from this cohort. Genetic risk scores derived from the Lp(a) GWAS explained a small but statistically significant proportion of variance in AS in Whites, African Americans, and Hispanics

from the GERA cohort when the p-value thresholds for the scores were selected in each ethnicity. These Lp(a)-GRSs performed nearly as well as AS-GRSs derived from the same cohort, emphasizing the importance of Lp(a) as a risk factor for AS.

Future studies should focus on performing GWAS for AS and Lp(a) in diverse populations, while using appropriate techniques to reduce confounding by ancestry. Alternatively, a trans-ethnic meta-GWAS involving both European populations and other ethnicities could improve statistical power and strengthen the ability to identify causal variants[155]. Once summary statistics are available for ethnic-specific or trans-ethnic GWAS, the performance of GRSs for AS should be assessed when variants and/or weights are selected from these GWAS. Future studies should also investigate the role of Lp(a) in the development of AS across ethnicities, with a particular focus on Hispanics. Together, results from work in this field will improve our understanding of the etiology of AS and other cardiovascular diseases across ethnicities while highlighting new targets for much-needed therapies.

Chapter 6: Master Reference List

1. Yutzey KE, Demer LL, Body SC, et al. Calcific aortic valve disease: A consensus summary from the alliance of investigators on calcific aortic valve disease. *Arterioscler Thromb Vasc Biol*. 2014. doi:10.1161/ATVBAHA.114.302523
2. Osnabrugge RLJ, Mylotte D, Head SJ, et al. Aortic stenosis in the elderly: Disease prevalence and number of candidates for transcatheter aortic valve replacement: A meta-analysis and modeling study. *J Am Coll Cardiol*. 2013;62(11):1002-1012. doi:10.1016/j.jacc.2013.05.015
3. Danielsen R, Aspelund T, Harris TB, Gudnason V. The prevalence of aortic stenosis in the elderly in Iceland and predictions for the coming decades: The AGES-Reykjavík study. *Int J Cardiol*. 2014;176(3):916-922. doi:10.1016/j.ijcard.2014.08.053
4. Clark MA, Arnold S V., Duhay FG, et al. Five-year clinical and economic outcomes among patients with medically managed severe aortic stenosis: Results from a medicare claims analysis. *Circ Cardiovasc Qual Outcomes*. 2012;5(5):697-704. doi:10.1161/CIRCOUTCOMES.112.966002
5. Chan KL. Is aortic stenosis a preventable disease? *J Am Coll Cardiol*. 2003. doi:10.1016/S0735-1097(03)00786-1
6. Goody PR, Hosen MR, Christmann D, et al. Aortic valve stenosis: From basic mechanisms to novel therapeutic targets. *Arterioscler Thromb Vasc Biol*. 2020. doi:10.1161/ATVBAHA.119.313067
7. Chen HY, Engert JC, Thanassoulis G. Risk factors for valvular calcification. *Curr Opin Endocrinol Diabetes Obes*. 2019. doi:10.1097/MED.0000000000000471

8. Martinsson A, Li X, Zöller B, et al. Familial Aggregation of Aortic Valvular Stenosis: A Nationwide Study of Sibling Risk. *Circ Cardiovasc Genet*. 2017. doi:10.1161/CIRCGENETICS.117.001742

9. Stewart BF, Siscovick D, Lind BK, et al. Clinical factors associated with calcific aortic valve disease. *J Am Coll Cardiol*. 1997. doi:10.1016/S0735-1097(96)00563-3

10. Novaro GM, Tiong IY, Pearce GL, Lauer MS, Sprecher DL, Griffin BP. Effect of hydroxymethylglutaryl coenzyme A reductase inhibitors on the progression of calcific aortic stenosis. *Circulation*. 2001. doi:10.1161/hc4301.098249

11. Bellamy MF, Pellikka PA, Klarich KW, Tajik AJ, Enriquez-Sarano M. Association of cholesterol levels, hydroxymethylglutaryl coenzyme-A reductase inhibitor treatment, and progression of aortic stenosis in the community. *J Am Coll Cardiol*. 2002. doi:10.1016/S0735-1097(02)02496-8

12. Rosenhek R, Rader F, Loho N, et al. Statins but not angiotensin-converting enzyme inhibitors delay progression of aortic stenosis. *Circulation*. 2004. doi:10.1161/01.CIR.0000140723.15274.53

13. Cowell SJ, Newby DE, Prescott RJ, et al. A randomized trial of intensive lipid-lowering therapy in calcific aortic stenosis. *N Engl J Med*. 2005. doi:10.1056/NEJMoa043876

14. Rossebø AB, Pedersen TR, Boman K, et al. Intensive lipid lowering with simvastatin and ezetimibe in aortic stenosis. *N Engl J Med*. 2008. doi:10.1056/NEJMoa0804602

15. Chan KL, Teo K, Dumesnil JG, Ni A, Tam J. Effect of lipid lowering with

rosuvastatin on progression of aortic stenosis: Results of the aortic stenosis progression observation: Measuring effects of rosuvastatin (Astronomer) trial. *Circulation*. 2010. doi:10.1161/CIRCULATIONAHA.109.900027

16. Thanassoulis G. Preventing Aortic Stenosis by Changing the Way We Think about an Old Disease. *JAMA Cardiol*. 2020. doi:10.1001/jamacardio.2020.0722
17. Smith JG, Luk K, Schulz CA, et al. Association of low-density lipoprotein cholesterol - Related genetic variants with aortic valve calcium and incident aortic stenosis. *JAMA - J Am Med Assoc*. 2014. doi:10.1001/jama.2014.13959
18. Smith GD, Ebrahim S. “Mendelian randomization”: Can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003. doi:10.1093/ije/dyg070
19. Marquis-Gravel G, Redfors B, Leon MB, Généreux P. Medical Treatment of Aortic Stenosis. *Circulation*. 2016. doi:10.1161/CIRCULATIONAHA.116.023997
20. Nishimura RA, Otto CM, Bonow RO, et al. 2014 AHA/ACC Guideline for the Management of Patients With Valvular Heart Disease: Executive Summary. *Circulation*. 2014. doi:10.1161/cir.0000000000000029
21. Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science (80-)*. 1996. doi:10.1126/science.273.5281.1516
22. Visscher PM, Wray NR, Zhang Q, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. *Am J Hum Genet*. 2017. doi:10.1016/j.ajhg.2017.06.005
23. Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet*. 2001. doi:10.1038/35052543

24. Hirschhorn JN, Daly MJ. Genome-wide association studies for common diseases and complex traits. *Nat Rev Genet*. 2005. doi:10.1038/nrg1521

25. Ku CS, Naidoo N, Pawitan Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet*. 2011. doi:10.1007/s00439-011-0964-2

26. Petersen BS, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017. doi:10.1186/s12863-017-0479-5

27. Hayano S, Okuno Y, Tsutsumi M, et al. Frequent intragenic microdeletions of elastin in familial supravalvular aortic stenosis. *Int J Cardiol*. 2019. doi:10.1016/j.ijcard.2018.09.032

28. Pons L, Bouvagnet P, Bakloul M, et al. Supravalvular aortic stenosis caused by a familial chromosome 7 inversion disrupting the eln gene uncovered by whole-genome sequencing. *Mol Syndromol*. 2019. doi:10.1159/000500215

29. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. *Circ Res*. 2016. doi:10.1161/CIRCRESAHA.115.306566

30. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;26(7571):68-74. doi:10.1038/nature15393

31. Gurdasani D, Barroso I, Zeggini E, Sandhu MS. Genomics of disease risk in globally diverse populations. *Nat Rev Genet*. 2019. doi:10.1038/s41576-019-0144-0

32. Freedman ML, Reich D, Penney KL, et al. Assessing the impact of population stratification on genetic association studies. *Nat Genet*. 2004. doi:10.1038/ng1333

33. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population

structure on large genetic association studies. *Nat Genet*. 2004. doi:10.1038/ng1337

34. Hellwege JN, Keaton JM, Giri A, Gao X, Velez Edwards DR, Edwards TL. Population Stratification in Genetic Association Studies. *Curr Protoc Hum Genet*. 2017. doi:10.1002/cphg.48
35. Bouaziz M, Ambroise C, Guedj M. Accounting for population stratification in practice: A comparison of the main strategies dedicated to genome-wide association studies. *PLoS One*. 2011. doi:10.1371/journal.pone.0028845
36. Liu J, Lewinger JP, Gilliland FD, Gauderman WJ, Conti D V. Confounding and heterogeneity in genetic association studies with admixed populations. *Am J Epidemiol*. 2013;177(4):351-360. doi:10.1093/aje/kws234
37. Zakharia F, Basu A, Absher D, et al. Characterizing the admixed African ancestry of African Americans. *Genome Biol*. 2009. doi:10.1186/gb-2009-10-12-r141
38. Knowler WC, Williams RC, Pettitt DJ, Steinberg AG. Gm(3;5,13,14) and type 2 diabetes mellitus: An association in American Indians with genetic admixture. *Am J Hum Genet*. 1988.
39. Wu C, Dewan A, Hoh J, Wang Z. A Comparison of Association Methods Correcting for Population Stratification in Case-Control Studies. *Ann Hum Genet*. 2011. doi:10.1111/j.1469-1809.2010.00639.x
40. Edwards TL, Gao X. Methods for detecting and correcting for population stratification. *Curr Protoc Hum Genet*. 2012. doi:10.1002/0471142905.hg0122s73
41. Devlin B, Roeder K. Genomic control for association studies. *Biometrics*. 1999. doi:10.1111/j.0006-341X.1999.00997.x

42. Hoggart CJ, Parra EJ, Shriver MD, et al. Control of confounding of genetic associations in stratified populations. *Am J Hum Genet*. 2003. doi:10.1086/375613

43. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006. doi:10.1038/ng1847

44. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nat Rev Genet*. 2008. doi:10.1038/nrg2344

45. Marchini J, Howie B. Genotype imputation for genome-wide association studies. *Nat Rev Genet*. 2010. doi:10.1038/nrg2796

46. Huang L, Li Y, Singleton AB, et al. Genotype-imputation accuracy across worldwide human populations. *Am J Hum Genet*. 2008. doi:10.1016/j.ajhg.2009.01.013

47. Huang GH, Tseng YC. Genotype imputation accuracy with different reference panels in admixed populations. In: *BMC Proceedings*. ; 2014. doi:10.1186/1753-6561-8-S1-S64

48. Igo RP, Kinzy TG, Cooke Bailey JN. Genetic Risk Scores. *Curr Protoc Hum Genet*. 2019. doi:10.1002/cphg.95

49. Janssens ACJW, Moonesinghe R, Yang Q, Steyerberg EW, Van Duijn CM, Khoury MJ. The impact of genotype frequencies on the clinical validity of genomic profiling for predicting common chronic diseases. *Genet Med*. 2007. doi:10.1097/GIM.0b013e31812eece0

50. Menard S. Coefficients of determination for multiple logistic regression analysis. *Am Stat*. 2000. doi:10.1080/00031305.2000.10474502

51. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience*. 2019;8(7):giz082. doi:10.1093/gigascience/giz082

52. Mak TSH, Porsch RM, Choi SW, Zhou X, Sham PC. Polygenic scores via penalized regression on summary statistics. *Genet Epidemiol*. 2017. doi:10.1002/gepi.22050

53. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015. doi:10.1016/j.ajhg.2015.09.001

54. Thanassoulis G, Campbell CY, Owens DS, et al. Genetic Associations with Valvular Calcification and Aortic Stenosis. *N Engl J Med*. 2013;368:503-512. doi:10.1056/NEJMoa1109034

55. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med*. 2009. doi:10.1056/NEJMoa0902604

56. Arsenault BJ, Boekholdt SM, Dubé MP, et al. Lipoprotein(a) levels, genotype, and incident aortic valve stenosis a prospective mendelian randomization study and replication in a case-control cohort. *Circ Cardiovasc Genet*. 2014;7(3):304-310. doi:10.1161/CIRCGENETICS.113.000400

57. Kamstrup PR, Tybjærg-Hansen A, Nordestgaard BG. Elevated lipoprotein(a) and risk of aortic valve stenosis in the general population. *J Am Coll Cardiol*. 2014;63(5):470-477. doi:10.1016/j.jacc.2013.09.038

58. Cairns BJ, Coffey S, Travis RC, et al. A replicated, genome-wide significant association of aortic stenosis with a genetic variant for lipoprotein(a): Meta-analysis of published and novel data. *Circulation*. 2017;135(12):1181-1183. doi:10.1161/CIRCULATIONAHA.116.026103

59. Chen HY, Dufresne L, Burr H, et al. Association of LPA variants with aortic stenosis a large-scale study using diagnostic and procedural codes from electronic health records. *JAMA Cardiol*. 2018;3(1):18-23. doi:10.1001/jamacardio.2017.4266

60. Thériault S, Gaudreault N, Lamontagne M, et al. A transcriptome-wide association study identifies PALMD as a susceptibility gene for calcific aortic valve stenosis. *Nat Commun*. 2018;9(1):6-13. doi:10.1038/s41467-018-03260-6

61. Helgadottir A, Thorleifsson G, Gretarsdottir S, et al. Genome-wide analysis yields new loci associating with aortic valve stenosis. *Nat Commun*. 2018;9(1). doi:10.1038/s41467-018-03252-6

62. Thériault S, Dina C, Messika-zeitoun D, et al. Genetic association analyses highlight IL6, ALPL, and NAV1 as three new susceptibility genes underlying calcific aortic valve stenosis. *bioRxiv*. 2019:1-35.

63. Boerwinkle E, Leffert CC, Lin J, Lackner C, Chiesa G, Hobbs HH. Apolipoprotein(a) gene accounts for greater than 90% of the variation in plasma lipoprotein(a) concentrations. *J Clin Invest*. 1992;90(1):52-60. doi:10.1172/JCI115855

64. Tsimikas S, Viney NJ, Hughes SG, et al. Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study.

Lancet. 2015. doi:10.1016/S0140-6736(15)61252-1

65. Viney NJ, van Capelleveen JC, Geary RS, et al. Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. *Lancet*. 2016. doi:10.1016/S0140-6736(16)31009-1
66. Assimes TL, Salfati EL, Del Gobbo LC. Leveraging information from genetic risk scores of coronary atherosclerosis. *Curr Opin Lipidol*. 2017. doi:10.1097/MOL.0000000000000400
67. Enkhmaa B, Anuurad E, Berglund L. Lipoprotein (a): Impact by ethnicity and environmental and medical conditions. *J Lipid Res*. 2016;57(7):1111-1125. doi:10.1194/jlr.R051904
68. Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet*. 2009. doi:10.1016/j.tig.2009.09.012
69. Bustamante CD, De La Vega FM, Burchard EG. Genomics for the world. *Nature*. 2011. doi:10.1038/475163a
70. Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016. doi:10.1038/538161a
71. Bentley AR, Callier S, Rotimi CN. Diversity and inclusion in genomic research: why the uneven progress? *J Community Genet*. 2017;8(4):255-266. doi:10.1007/s12687-017-0316-6
72. Zabaneh D, Kumari M, Sandhu M, et al. Meta analysis of candidate gene variants outside the LPA locus with Lp(a) plasma levels in 14,500 participants of six White European cohorts. *Atherosclerosis*. 2011.

doi:10.1016/j.atherosclerosis.2011.04.015

73. Ronald J, Rajagopalan R, Cerrato F, et al. Genetic variation in LPAL2, LPA, and PLG predicts plasma lipoprotein(a) level and carotid artery disease risk. *Stroke*. 2011. doi:10.1161/STROKEAHA.110.591230

74. Lanktree MB, Anand SS, Yusuf S, Hegele RA. Comprehensive analysis of genomic variation in the LPA locus and its relationship to plasma lipoprotein(a) in south asians, chinese, and european caucasians. *Circ Cardiovasc Genet*. 2010. doi:10.1161/CIRCGENETICS.109.907642

75. Deo RC, Wilson JG, Xing C, et al. Single-nucleotide polymorphisms in LPA explain most of the Ancestry-specific variation in Lp(a) levels in African Americans. *PLoS One*. 2011. doi:10.1371/journal.pone.0014581

76. Cardoso-Saldaña G, Fragoso JM, Lale-Farjat S, et al. The rs10455872-g allele of the lpa gene is associated with high lipoprotein(A) levels and increased aortic valve calcium in a mexican adult population. *Genet Mol Biol*. 2019. doi:10.1590/1678-4685-gmb-2017-0371

77. Li J, Lange LA, Sabourin J, et al. Genome- and exome-wide association study of serum lipoprotein (a) in the Jackson Heart Study. *J Hum Genet*. 2015;60(12):755-761. doi:10.1038/jhg.2015.107

78. Zekavat SM, Ruotsalainen S, Handsaker RE, et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat Commun*. 2018. doi:10.1038/s41467-018-04668-w

79. Mack S, Coassin S, Rueedi R, et al. A genome-wide association meta-analysis on lipoprotein (a) concentrations adjusted for apolipoprotein (a) isoforms. *J Lipid Res*.

2017. doi:10.1194/jlr.M076232

80. Moriarty PM, Varvel SA, Gordts PLSM, McConnell JP, Tsimikas S. Lipoprotein(a) Mass Levels Increase Significantly According to APOE Genotype: An Analysis of 431 239 Patients. *Arterioscler Thromb Vasc Biol*. 2017. doi:10.1161/ATVBAHA.116.308704

81. Köchl S, Fresser F, Lobentanz E, Baier G, Utermann G. Novel interaction of apolipoprotein(a) with β-2 glycoprotein I mediated by the kringle IV domain. *Blood*. 1997. doi:10.1182/blood.v90.4.1482.1482_1482_1489

82. Miyakis S, Giannakopoulos B, Krilis SA. Beta 2 glycoprotein I-function in health and disease. *Thromb Res*. 2004. doi:10.1016/j.thromres.2004.07.017

83. Leibundgut G, Scipione C, Yin H, et al. Determinants of binding of oxidized phospholipids on apolipoprotein (a) and lipoprotein (a). *J Lipid Res*. 2013. doi:10.1194/jlr.M040733

84. Wurm H, Beubler E, Polz E, Holasek A, Kostner G. Studies on the possible function of β2-glycoprotein-I: Influence in the triglyceride metabolism in the rat. *Metabolism*. 1982. doi:10.1016/0026-0495(82)90238-4

85. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet*. 2009. doi:10.1038/nrg2579

86. Hunter DJ. Gene-environment interactions in human diseases. *Nat Rev Genet*. 2005. doi:10.1038/nrg1578

87. Marigorta UM, Navarro A. High Trans-ethnic Replicability of GWAS Results Implies Common Causal Variants. *PLoS Genet*. 2013. doi:10.1371/journal.pgen.1003566

88. Bloomfield D, Carlson GL, Sapre A, et al. Efficacy and safety of the cholesteryl ester transfer protein inhibitor anacetrapib as monotherapy and coadministered with atorvastatin in dyslipidemic patients. *Am Heart J*. 2009. doi:10.1016/j.ahj.2008.09.022

89. Cannon CP, Shah S, Dansky HM, et al. Safety of anacetrapib in patients with or at high risk for coronary heart disease. *N Engl J Med*. 2010. doi:10.1056/NEJMoa1009744

90. Thomas T, Zhou H, Karmally W, et al. CETP (Cholesteryl Ester Transfer Protein) Inhibition with Anacetrapib Decreases Production of Lipoprotein(a) in Mildly Hypercholesterolemic Subjects. *Arterioscler Thromb Vasc Biol*. 2017. doi:10.1161/ATVBAHA.117.309549

91. Cobbaert C, Kesteloot H. Serum lipoprotein(a) levels in racially different populations. *Am J Epidemiol*. 1992. doi:10.1093/oxfordjournals.aje.a116517

92. Kraft HG. Frequency distributions of apolipoprotein(a) Kringle IV repeat alleles and their effects on lipoprotein(a) levels in Caucasian, Asian, and African populations: The distribution of null alleles is non-random. *Eur J Hum Genet*. 1996. doi:10.1159/000472175

93. Cao J, Steffen BT, Budoff M, et al. Lipoprotein(a) levels are associated with subclinical calcific aortic valve disease in white and black individuals: The multi-ethnic study of atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2016;36(5):1003-1009. doi:10.1161/ATVBAHA.115.306683

94. Novaro GM, Katz R, Aviles RJ, et al. Clinical Factors, But Not C-Reactive Protein, Predict Progression of Calcific Aortic-Valve Disease. The Cardiovascular Health

Study. *J Am Coll Cardiol*. 2007. doi:10.1016/j.jacc.2007.07.064

95. Novaro GM, Houghtaling PL, Gillinov AM, Blackstone EH, Asher CR. Prevalence of mitral valve prolapse and congenital bicuspid aortic valves in black and white patients undergoing cardiac valve operations. *Am J Cardiol*. 2013. doi:10.1016/j.amjcard.2012.11.051

96. Patel DK, Green KD, Fudim M, Harrell FE, Wang TJ, Robbins MA. Racial differences in the prevalence of severe aortic stenosis. *J Am Heart Assoc*. 2014. doi:10.1161/JAHA.114.000879

97. Beydoun HA, Beydoun MA, Liang H, et al. Sex, Race, and Socioeconomic Disparities in Patients With Aortic Stenosis (from a Nationwide Inpatient Sample). *Am J Cardiol*. 2016. doi:10.1016/j.amjcard.2016.06.039

98. Alqahtani F, Aljohani S, Amin AH, et al. Effect of Race on the Incidence of Aortic Stenosis and Outcomes of Aortic Valve Replacement in the United States. *Mayo Clin Proc*. 2018. doi:10.1016/j.mayocp.2017.11.026

99. Batchelor W, Anwaruddin S, Ross L, et al. Aortic Valve Stenosis Treatment Disparities in the Underserved: JACC Council Perspectives. *J Am Coll Cardiol*. 2019. doi:10.1016/j.jacc.2019.08.1035

100. Haffner SM, Gruber KK, Morales PA, et al. Lipoprotein(a) concentrations in Mexican Americans and non-hispanic whites: The San Antonio heart study. *Am J Epidemiol*. 1992. doi:10.1093/oxfordjournals.aje.a116571

101. Kamboh M I., Rewers M, Aston CE, Hamman RF. Plasma Apolipoprotein A-I, Apolipoprotein B, and Lipoprotein(a) Concentrations in Normoglycemic Hispanics and Non-Hispanic Whites from the San Luis Valley, Colorado. *Am J Epidemiol*.

1997. doi:10.1093/oxfordjournals.aje.a009229

102. Guan W, Cao J, Steffen BT, et al. Race is a key variable in assigning lipoprotein(a) cutoff values for coronary heart disease risk assessment: The multi-ethnic study of atherosclerosis. *Arterioscler Thromb Vasc Biol*. 2015. doi:10.1161/ATVBAHA.114.304785

103. Bertoni B, Budowle B, Sans M, Barton SA, Chakraborty R. Admixture in Hispanics: Distribution of ancestral population contributions in the continental United States. *Hum Biol*. 2003. doi:10.1353/hub.2003.0016

104. Bonilla C, Parra EJ, Pfaff CL, et al. Admixture in the Hispanics of the San Luis Valley, Colorado, and its implications for complex trait gene mapping. *Ann Hum Genet*. 2004. doi:10.1046/j.1529-8817.2003.00084.x

105. Halder I, Yang BZ, Kranzler HR, Stein MB, Shriver MD, Gelernter J. Measurement of admixture proportions and description of admixture structure in different U.S. populations. *Hum Mutat*. 2009. doi:10.1002/humu.21045

106. Bryc K, Velez C, Karafet T, et al. Genome-wide patterns of population structure and admixture among Hispanic/Latino populations. *Proc Natl Acad Sci U S A*. 2010;107(SUPPL. 2):8954-8961. doi:10.1073/pnas.0914618107

107. Manichaikul A, Palmas W, Rodriguez CJ, et al. Population Structure of Hispanics in the United States : The Multi-Ethnic Study of Atherosclerosis. 2012;8(4). doi:10.1371/journal.pgen.1002640

108. Khera A V., Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50(9):1219-1224. doi:10.1038/s41588-018-0183-z

109. Mavaddat N, Michailidou K, Dennis J, et al. Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes. *Am J Hum Genet*. 2019. doi:10.1016/j.ajhg.2018.11.002

110. Khera A V., Chaffin M, Wade KH, et al. Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell*. 2019. doi:10.1016/j.cell.2019.03.028

111. Torkamani A, Wineinger NE, Topol EJ. The personal and clinical utility of polygenic risk scores. *Nat Rev Genet*. 2018. doi:10.1038/s41576-018-0018-x

112. Lambert SA, Abraham G, Inouye M. Towards clinical utility of polygenic risk scores. *Hum Mol Genet*. 2019. doi:10.1093/hmg/ddz187

113. Lewis CM, Vassos E. Polygenic risk scores: From research tools to clinical instruments. *Genome Med*. 2020. doi:10.1186/s13073-020-00742-5

114. Duncan L, Shen H, Gelaye B, et al. Analysis of polygenic risk score usage and performance in diverse human populations. *Nat Commun*. 2019. doi:10.1038/s41467-019-11112-0

115. Martin AR, Gignoux CR, Walters RK, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet*. 2017;100(4):635-649. doi:10.1016/j.ajhg.2017.03.004

116. Kim MS, Patel KP, Teng AK, Berens AJ, Lachance J. Genetic disease risks can be misestimated across global populations. *Genome Biol*. 2018. doi:10.1186/s13059-018-1561-7

117. Grinde KE, Qi Q, Thornton TA, et al. Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genet Epidemiol*. 2019;43(1):50-62.

doi:10.1002/gepi.22166

118. Márquez-Luna C, Loh PR, Price AL, et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet Epidemiol*. 2017. doi:10.1002/gepi.22083

119. Marnetto D, Pärna K, Läll K, et al. Ancestry deconvolution and partial polygenic score can improve susceptibility predictions in recently admixed individuals. *Nat Commun*. 2020. doi:10.1038/s41467-020-15464-w

120. Sahebkar A, Reiner Ž, Simental-Mendía LE, Ferretti G, Cicero AFG. Effect of extended-release niacin on plasma lipoprotein(a) levels: A systematic review and meta-analysis of randomized placebo-controlled trials. *Metabolism*. 2016. doi:10.1016/j.metabol.2016.08.007

121. Santos RD, Raal FJ, Catapano AL, Witztum JL, Steinhagen-Thiessen E, Tsimikas S. Mipomersen, an Antisense Oligonucleotide to Apolipoprotein B-100, Reduces Lipoprotein(a) in Various Populations with Hypercholesterolemia: Results of 4 Phase III Trials. *Arterioscler Thromb Vasc Biol*. 2015. doi:10.1161/ATVBAHA.114.304549

122. Desai NR, Kohli P, Giugliano RP, et al. AMG145, a monoclonal antibody against proprotein convertase subtilisin kexin type 9, significantly reduces lipoprotein(a) in hypercholesterolemic patients receiving statin therapy: An analysis from the LDL-C assessment with proprotein convertase subtilisi. *Circulation*. 2013. doi:10.1161/CIRCULATIONAHA.113.001969

123. Raal FJ, Honarpour N, Blom DJ, et al. Inhibition of PCSK9 with evolocumab in homozygous familial hypercholesterolaemia (TESLA Part B): A randomised,

double-blind, placebo-controlled trial. *Lancet*. 2015. doi:10.1016/S0140-6736(14)61374-X

124. Gaudet D, Watts GF, Robinson JG, et al. Effect of Alirocumab on Lipoprotein(a) Over ≥1.5 Years (from the Phase 3 ODYSSEY Program). *Am J Cardiol*. 2017. doi:10.1016/j.amjcard.2016.09.010
125. O'Donoghue ML, Fazio S, Giugliano RP, et al. Lipoprotein(a), PCSK9 Inhibition, and Cardiovascular Risk. *Circulation*. 2019. doi:10.1161/circulationaha.118.037184
126. Leduc MS, Shimmin LC, Klos KLE, Hanis C, Boerwinkle E, Hixson JE. Comprehensive evaluation of apolipoprotein H gene (APOH) variation identifies novel associations with measures of lipid metabolism in GENOA. *J Lipid Res*. 2008. doi:10.1194/jlr.M800155-JLR200
127. Asselbergs FW, Guo Y, Van Iperen EPA, et al. Large-scale gene-centric meta-analysis across 32 studies identifies multiple lipid loci. *Am J Hum Genet*. 2012. doi:10.1016/j.ajhg.2012.08.032
128. Elbers CC, Guo Y, Tragante V, et al. Gene-Centric Meta-Analysis of Lipid Traits in African, East Asian and Hispanic Populations. *PLoS One*. 2012. doi:10.1371/journal.pone.0050198
129. Helske S, Kupari M, Lindstedt KA, Kovanen PT. Aortic valve stenosis: An active atheroinflammatory process. *Curr Opin Lipidol*. 2007. doi:10.1097/MOL.0b013e3282a66099
130. Willer CJ, Schmidt EM, Sengupta S, et al. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013. doi:10.1038/ng.2797

131. Melendez QM, Wooten CJ, Krishnaji ST, Knagge K, Kirchner D, Lopez D. Identification of Novel Proteins Interacting with Proprotein Convertase Subtilisin/Kexin 9. 2020;3(1):1-27. doi:10.31531/2581-4745.1000123.

132. Gaziano JM, Concato J, Brophy M, et al. Million Veteran Program: A mega-biobank to study genetic influences on health and disease. *J Clin Epidemiol*. 2016. doi:10.1016/j.jclinepi.2015.09.016

133. National Heart, Lung and BI. NHLBI Trans-Omics for Precision Medicine. https://www.nhlbiwgs.org. Published 2020. Accessed June 7, 2020.

134. Institute NHGR. Population Architecture Using Genomics and Epidemiology (PAGE) Consortium. https://www.genome.gov/Funded-Programs-Projects/Population-Architecture-Using-Genomics-and-Epidemiology. Published 2019. Accessed June 7, 2020.

135. Rosenberg NA, Huang L, Jewett EM, Szpiech ZA. Genome-wide association studies in diverse populations. *Nat Publ Gr*. 2010;11(5):356-366. doi:10.1038/nrg2760

136. Cavalli-Sforza LL. The human genome diversity project: Past, present and future. *Nat Rev Genet*. 2005. doi:10.1038/nrg1596

137. Altshuler DM, Gibbs RA, Peltonen L, et al. Integrating common and rare genetic variation in diverse human populations. *Nature*. 2010. doi:10.1038/nature09298

138. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015. doi:10.1038/nature15394

139. Vergara C, Parker MM, Franco L, et al. Genotype imputation performance of three reference panels using African ancestry individuals. *Hum Genet*. 2018.

doi:10.1007/s00439-018-1881-4

140. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016. doi:10.1038/ng.3643

141. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *bioRxiv*. 2019. doi:10.1101/563866

142. Medina-Gomez C, Felix JF, Estrada K, et al. Challenges in conducting genome-wide association studies in highly admixed multi-ethnic populations: the Generation R Study. *Eur J Epidemiol*. 2015. doi:10.1007/s10654-015-9998-4

143. Price AL, Zaitlen NA, Reich D, Patterson N. New approaches to population stratification in genome-wide association studies. *Nat Rev Genet*. 2010. doi:10.1038/nrg2813

144. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet*. 1999. doi:10.1086/302449

145. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000.

146. Alexander DH, Novembre J. Fast Model-Based Estimation of Ancestry in Unrelated Individuals. 2009:1655-1664. doi:10.1101/gr.094052.109.vidual

147. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*. 2007. doi:10.1093/bioinformatics/btm308

148. Kang HM, Sul JH, Service SK, et al. Variance component model to account for sample structure in genome-wide association studies. *Nat Genet*. 2010.

doi:10.1038/ng.548

149. McKeigue PM. Mapping genes that underlie ethnic differences in disease risk: Methods for detecting linkage in admixed populations, by conditioning on parental admixture. *Am J Hum Genet*. 1998. doi:10.1086/301908

150. Winkler CA, Nelson GW, Smith MW. Admixture Mapping Comes of Age. *Annu Rev Genomics Hum Genet*. 2010. doi:10.1146/annurev-genom-082509-141523

151. Pulit SL, Voight B, de Bakker PIW. Multiethnic genetic association studies improve power for locus discovery. *PLoS One*. 2010. doi:10.1371/journal.pone.0012600

152. Cohen J, Pertsemlidis A, Kotowski IK, Graham R, Garcia CK, Hobbs HH. Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in PCSK9. *Nat Genet*. 2005. doi:10.1038/ng1509

153. Cohen JC, Boerwinkle E, Mosley TH, Hobbs HH. Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med*. 2006. doi:10.1056/NEJMoa054013

154. Horton JD, Cohen JC, Hobbs HH. Molecular biology of PCSK9: its role in LDL metabolism. *Trends Biochem Sci*. 2007. doi:10.1016/j.tibs.2006.12.008

155. Li YR, Keating BJ. Trans-ethnic genome-wide association studies: Advantages and challenges of mapping in diverse populations. *Genome Med*. 2014. doi:10.1186/s13073-014-0091-5