Association of Haplotypes with Lipoprotein(a) Concentrations and Cardiovascular Disease Risk Across Multiple Ethnicities

Thomas Marsh

Human Genetics

McGill University, Montreal

September 20th, 2021

A thesis submitted to McGill University in partial

fulfillment of the requirements of the degree of

Master of Science in Human Genetics

Abstract:

Introduction:

Lipoprotein(a) (Lp(a)) is a particle that circulates in plasma and is associated with several cardiovascular diseases (CVDs). This particle is composed of a low-density lipoprotein-like moiety and an apolipoprotein(a) (apo(a)) glycoprotein coded for by the *LPA* gene. Shorter *LPA* isoforms are associated with greater apo(a) concentrations and increased CVD risk. *LPA* isoform size is determined by a copy number variant known as the kringle IV type 2 repeats (KIV_2). The average number of KIV_2 repeats and Lp(a) levels are known to vary widely between different ethnicities. Haplotypes, which are collections of variants on the same chromosome, may be useful to better understand the link between ethnicity and Lp(a) concentrations.

Methods:

Putative haplotypes that span the KIV_2 repeats in the *LPA* gene were identified using unrelated individuals from the UK Biobank (UKB) cohort. Haplotypes were constructed from 36 common SNPs (minor allele frequency (MAF) > 0.30) using White British individuals and then were investigated in three other ethnicities: South Asians, Chinese, and Blacks. They were tested for their association with Lp(a) concentrations, and for coronary artery disease (CAD) and aortic stenosis (AS). Replication analyses were completed for Lp(a) concentrations in the Multi-Ethnic Study of Atherosclerosis (MESA), and for CVD in the Genetic Epidemiology Research in Adult Health and Aging (GERA) cohorts. After the initial haplotypic analyses, three variants previously associated with Lp(a) levels and CVD were added to the *LPA* haplotypes to assess their impact on associations. The variant rs10455872, an intronic variant, and rs3798220, a missense variant, are two well-characterized variants associated with Lp(a), CVD, and KIV_2 repeat length and rs140570886 is in high linkage disequilibrium with rs3798220. A subset of

four variants out of the 36 identified in the UKB were used to calculate haplotype frequencies from populations present in the Human Genome Diversity Project (HGDP), including Sub-Saharan Africans, South Asians, East Asians, West Asians, and Native Americans.

Results:

Two haplotypes were identified in the *LPA* region that span the KIV_2 repeats in UKB White British that account for >99% of the haplotypes. These two *LPA* haplotypes are associated with opposing effects on Lp(a) levels. These same two haplotypes were also identified, at similar frequencies, in Chinese and South Asian individuals and were again associated with Lp(a) levels. Black individuals have a much more heterogenous haplotypic composition. In UKB White British, after inclusion of the *LPA* variants rs10455872, rs3798220, and rs140570886, it became clear that the observed associations with the original haplotypes were largely due to the presence of these variants on subsets of the major haplotype. Rs3798220 was found to have a significant association with Lp(a) concentrations only when found on a haplotype with rs140570886. Similar results were found in MESA.

The *LPA* haplotypes, with the three SNPs added, were associated with CAD and AS odds in both UKB White British and GERA European samples, in the expected direction, though power was reduced for these diseases.

The *LPA* haplotype frequencies are comparable between UKB and HGDP for all non-African ethnicities. All but one of the haplotypes in Sub-Saharan Africans are also found in UKB Blacks, as well as MESA and GERA African Americans.

Conclusions:

This study has identified KIV_2 repeat spanning *LPA* haplotypes that are associated with Lp(a) concentrations and are highly conserved in many populations. However, the association of these haplotypes with Lp(a) and CVD are driven by other variants that are in linkage disequilibrium with the haplotypes, including rs10455872. When assessing the associations between rs3798220 and rs140570886 with Lp(a) in the context of *LPA* haplotypes it is clear that rs3798220, despite being an amino acid changing variant, is not a causal variant. Future work should incorporate direct sequencing of the KIV_2 repeats using third-generation sequencing to better assess the association between *LPA* variation and Lp(a) levels.

Résumé:

Introduction:

La lipoprotéine(a) [Lp(a)] est une particule qui circule dans le plasma et qui est associée à plusieurs maladies cardio-vasculaires (MCVs). Cette particule est composée d'une partie similaire aux lipoprotéines de faible densité, ainsi que d'une molécule d'apolipoprotéine(a) [apo(a)], une glycoprotéine codée par le gène *LPA*. Des isoformes plus courts de *LPA* sont associés à de plus grandes concentrations d'apo(a), ainsi qu'à un risque accru de MCVs. La taille de l'isoforme de *LPA* est déterminée par un variant du nombre de copies connu sous le nom de répétitions de Kringle IV de type 2 (KIV_2). Le nombre moyen de répétitions KIV_2 et les niveaux de Lp(a) varient beaucoup entre les différentes ethnicités. Les haplotypes, qui consistent en des collections de variants sur le même chromosome, pourraient être utiles afin de mieux comprendre le lien entre l'ethnicité et les concentrations de Lp(a).

Méthodes:

Des haplotypes qui couvrent les répétitions KIV_2 dans le gène *LPA* ont été identifiés en utilisant des individus non-apparentés provenant de la cohorte UK Biobank (UKB). Les haplotypes ont été construits à partir de 36 polymorphismes d'un seul nucléotide (PSNs) communs (fréquence de l'allèle mineur (FAM) > 0.30) en utilisant des individus britanniques blancs et ont ensuite été étudiés dans trois autres ethnicités: les Sud-Asiatiques, les Chinois et les Noirs. Ils ont été testés pour leur association avec les concentrations de Lp(a), la maladie coronarienne (MC), ainsi que la sténose aortique (SA). Des analyses de réplication ont été accomplies pour des concentrations de Lp (a) dans l'étude multiethnique de l'athérosclérose (MESA), et pour la CVD dans les cohortes de recherche en épidémiologie génétique dans la

santé et le vieillissement adultes (GERA). Après les analyses haplotypiques initiales, trois variants auparavant associés avec les niveaux de Lp(a) et les MCVs ont été ajoutés aux haplotypes de *LPA* afin d'évaluer leur impact sur ces associations. Les variants rs10455872 et rs3798220, respectivement un variant intronique, ainsi qu'un variant faux-sens, sont deux variants bien caractérisés qui sont associés avec les niveaux de Lp(a), les MCVs, ainsi que la longueur des répétitions KIV_2. Le variant rs140570886 est en fort déséquilibre de liaison avec le variant rs3798220.

Sur les 36 polymorphismes initiaux identifiés dans la cohorte UKB, un sous-groupe de quatre variants a été utilisé dans le Human Genome Diversity Project (HGDP) afin de calculer les fréquences des haplotypes provenant de différentes populations, parmi lesquelles on retrouve les Africains subsahariens, les Sud-Asiatiques, les Asiatiques de l'Est, les Asiatiques occidentaux et les Autochtones.

Résultats:

Chez les individus britanniques blancs provenant de la cohorte UKB, deux haplotypes qui comptent pour >99% de tous les haplotypes ont été identifiés dans la région du gène *LPA* qui couvre les répétitions KIV_2. Ces deux haplotypes de *LPA* sont associés à des effets opposés sur les niveaux de Lp(a). Ces mêmes haplotypes ont également été identifiés avec des fréquences similaires chez des Chinois et des Sud-Asiatiques et ont une fois de plus été associés avec les niveaux de Lp(a) chez ces individus. Les individus de race noire ont une composition haplotypique beaucoup plus hétérogène. Chez les individus britanniques blancs provenant de la cohorte UKB, l'inclusion des trois variants clés de *LPA* (rs10455872, rs3798220 et rs140570886) dans les analyses haplotypiques a révélé que les associations avec les haplotypes originaux étaient principalement dues à la présence de ces variants dans des sous-groupes de l'haplotype majeur. Le variant rs3798220 possède une association significative avec les concentrations de Lp(a) qui n’est présente que lorsqu’il se trouve sur un haplotype avec le variant rs140570886. Des résultats similaires ont été observés dans la cohorte MESA.

En incluant les trois PSNs clés, les haplotypes de *LPA* sont associés avec un risque réduit de MC et de SA autant dans les échantillons de blancs britanniques provenant de la cohorte UKB que ceux d’Européens issus de GERA. Même si la puissance statistique était réduite pour ces maladies, ces associations étaient dans la direction attendue.

Les fréquences des haplotypes de *LPA* sont comparables entre les cohortes UKB et HGDP pour les ethnicités non-africaines. Tous les haplotypes chez les Africains subsahariens sauf un sont également retrouvés chez les Noirs provenant de la cohorte UKB, ainsi que chez les Afro-Américains issus de MESA et GERA.

Conclusions:

Cette étude a identifié des haplotypes de *LPA* couvrant les répétitions KIV_2 qui sont associés avec les concentrations de Lp(a) et qui sont hautement conservés dans plusieurs populations. Cependant, l’association de ces haplotypes avec les niveaux de Lp(a) et les MCVs est induite par d’autres variants en déséquilibre de liaison avec les haplotypes en question, y compris le variant rs10455872. En évaluant l’association entre les variants rs3798220 et rs140570886, ainsi que les niveaux de Lp(a) dans le contexte des différents haplotypes, il semble que ce variant n’est pas causatif malgré le fait qu’il change la séquence d’acides aminés. Les travaux futurs devraient incorporer le séquençage direct des répétitions KIV_2 en utilisant le séquençage de troisième génération afin de mieux évaluer l’association entre le gène *LPA* et les niveaux de Lp(a).

Table of Contents:

Abstract..2

Résumé..5

List of Abbreviations..9

List of Figures..10

List of Tables..11

Acknowledgements...13

Format of the Thesis...15

Contribution of Authors...16

Chapter 1: Introduction...17

Chapter 2: Association of *LPA* haplotypes with Lipoprotein(a) and Cardiovascular Diseases...38

Chapter 3: General Discussion..93

Chapter 4: Conclusions and Future Directions...99

Chapter 5: References: Master Reference List...101

Appendix: Ethics Certificates...115

List of Abbreviations:

AS: Aortic Stenosis

Apo(a): Apolipoprotein(a)

CVD: Cardiovascular Disease

CAD: Coronary Artery Disease

GERA: Genetic Epidemiology Research on Adult Health and Aging

GWAS: Genome-Wide Association Study

HGDP: Human Genome Diversity Project

KIV$_2$: Kringle IV Type 2

Lp(a): Lipoprotein(a)

LD: Linkage Disequilibrium

MAF: Minor Allele Frequency

MESA: Multi-Ethnic Study of Atherosclerosis

OR: Odds Ratio

SNP: Single Nucleotide Polymorphism

UKB: UK Biobank

List of Figures:

Chapter 1:

Figure 1: Lipoprotein(a) Particle Diagram..19

Figure 2. *LPA* Gene Diagram..19

Figure 3. Human Genome Diversity Project Population Distribution...................................31

Chapter 2:

Figure 1. Average Lp(a) Concentrations (nmol/L) by Genotype in UKB White British Individuals..80

Figure 2. Average Haplotypic Lp(a) Concentrations (nmol/L) in UKB White British Individuals..81

List of Tables:

Chapter 2:

Table 1: Association of *LPA* Haplotypes with Lp(a) Concentrations in Select UKB Ethnicities..........65

Table 2. Association of *LPA* Haplotypes with Lp(a) Concentrations in UKB Blacks..........66

Table 3. Association of *LPA* Haplotypes with Lp(a) Concentrations in Select MESA Ethnicities..........67

Table 4. Association of *LPA* Haplotypes with Lp(a) Concentrations in MESA African Americans..........68

Table 5. UKB *LPA* Haplotypes with Variants-Associations with Lp(a)..........69–70

Table 6. MESA *LPA* Haplotypes with Variants-Associations with Lp(a)..........71

Table 7. UKB White British *LPA* Region Lp(a) GWAS Hits Genotypic-Missingness (Mean Homozygote Minor Lp(a) Concentrations ≥ 70 nmol/L)..........72

Table 8. UKB White British *LPA* Region Lp(a) GWAS Variants Censored Lp(a) Concentrations..........73

Table 9. UKB White British *LPA* Null Variants Censored Lp(a) Concentrations..........74

Table 10. Table 10. UKB Black *LPA* Region Null Variants Missingness..........74

Table 11. UKB South Asian *LPA* Region Null Variants Missingness..........75

Table 12. UKB White British *LPA* Haplotypes Haplotypic Associations with Coronary Artery Disease..........75

Table 13. UKB White British *LPA* Haplotypes Haplotypic Associations with Aortic Stenosis..........76

Table 14. GERA European *LPA* Haplotypes Haplotypic Associations with Coronary Artery Disease..........77

Table 15. GERA European *LPA* Haplotypes Haplotypic Associations with Aortic Stenosis..........78

Table 16. HGDP *LPA* Region Haplotype Frequencies..........79

Table S1. Sample Sizes for Analyzed HGDP Populations..........88–89

Table S2. *LPA* Region Haplotype SNPs..........90

Table S3. UKB White British Lp(a) GWAS *LPA* Haplotypes..........91–92

Table S4. Haplotype Frequencies in Individuals of African Descent..........92

Acknowledgements:

I owe much thanks to Dr. Engert for his continued help and support throughout the last two years. Thanks to his patience and guidance, I was able to learn critical skills, including programming, scientific writing, and data analysis. Despite having a busy schedule, Dr. Engert always put in the necessary time to support my project and to discuss my work. It has been rewarding to work for such an excellent and highly motivated supervisor. I am now ready to start a PhD thanks to Dr. Engert’s mentorship.

I would like to thank Dr. Thanassoulis for his feedback and support throughout my master’s degree. There are many times when Dr. Thanassoulis would challenge and critique me, which allowed me to greatly improve upon my academic skills. His friendly demeanor allowed very easy communication. His expertise in medicine and cardiology was indispensable for my research.

I owe much gratitude to Hao Yu Chen for her contributions to my work. Hao Yu was frequently available to mentor me despite her extremely busy schedule. She greatly improved my writing skills through her very detailed critiques of my works. I would not have developed my programming and statistics skills so extensively without her clear explanations.

Thanks to Line Dufresnse who helped me many times throughout my thesis. She helped improve my presentation skills and made my PowerPoint presentations look more professional. She noticed even the smallest details of my work in her critiques, which taught me to be very thorough in my research.

I would like to thank Yoshihiko Nagai for all his help with server and computer-related issues. He is also a great friend who has helped me out and supported me since he first entered the Engert-Thanassoulis lab.

I would like to thank previous lab members Mary Hoekstra and Hannah Burr for aid in programming and data analysis problems. In addition, I would like to thank them for making the student office an enjoyable place to work.

I would like to thank Katia Desbiens for her help with computer issues. I would like to thank Maedeh Mohammadnetaaj for her weekly lab meeting reminders and help with bureaucratic issues. I want to thank Justin Miron for translating my thesis abstract and for his support. In addition, I would like to thank Guenievre Grondin for her help and understanding in bureaucratic issues.

Format of the Thesis:

This is a manuscript-based thesis that was completed under the supervision of Dr. James Engert and Dr. George Thanassoulis. This manuscript identifies associations between haplotypes in the *LPA* region and lipoprotein(a) concentrations, as well as cardiovascular disease, across multiple ethnicities.

The first chapter is an introduction to research on Lp(a). The second chapter contains a manuscript to be submitted to a journal. The third chapter is a more general discussion of the results.

Contribution of Authors:

Marsh, T: Analysis of the UK Biobank (UKB), Genetic Epidemiology Research in Adult Health and Aging (GERA), Multiethnic Study of Atherosclerosis (MESA), and Human Genome Diversity Project (HGDP) cohorts; imputation, annotation, and quality control of GERA cohort; and drafting of the manuscript

Chen, HY: Revision of the manuscript

Dufresne, L: Revision of the manuscript

Nagai, Y: Revision of the manuscript

Engert, JC: Revision of the manuscript

Thanassoulis, G: Revision of the manuscript

Chapter 1: Introduction:

History:

Lipoprotein(a) (Lp(a)) was first discovered by Karl Berg, in 1963, through injection of human serum into rabbits and then retrieving lipoprotein-related antigens.[1] The “(a)” in Lp(a) refers to “antigen”, as Lp(a) was discovered as an antigen that bound low-density lipoprotein (LDL).[2] A decade later, Berg also identified the first association between Lp(a) and a form of cardiovascular disease (CVD), coronary artery disease (CAD).[3] Berg initially reported that Lp(a) was found in a minority of individuals, but it was soon discovered, in 1977, that Lp(a) is present in everyone, at varying concentrations.[4] By the 1980s, Lp(a) was confirmed as an important CVD risk factor.[2,5,6]

Despite the novelty of Lp(a), research on this particle was slow to advance until, in 1987, the *LPA* gene that codes for the apolipoprotein(a) (apo(a)) component of Lp(a) was sequenced.[7] The *LPA* gene was found to resemble the *PLG* gene and contain several repeats of a domain similar in sequence to the *PLG* kringle 4 domain.[7] *LPA* was also found to contain a kringle 5-like domain and a protease domain, domains which are also found in *PLG*.[7] Another key discovery was that the sequence required to gain protease function was mutated, resulting in an inactive protease domain.[7] This discovery was followed by a plethora of new Lp(a)-related research over the next 8 years.[2,8-12]

There was a decline in research on Lp(a) in the early 1990s because of negative results for associations of Lp(a) with CVD.[13,14] These negative results were driven by poor practices such as improper storage of Lp(a) samples and use of non-standardized assays of Lp(a).[15] In 2004, a group of researchers performed a study similar to one of the previous ones,[14] and found that elevated Lp(a) concentrations led to an increase in angina pectoris.[16] The results of this study helped spur the modern interest in Lp(a).[17] Several years later, there were three landmark studies that solidified Lp(a) as a causal risk factor for CVD.[18-20] These studies found that Lp(a) concentrations and *LPA* variation are robustly associated with coronary artery disease,[18,19] stroke,[19] and myocardial infarction.[20] In 2013, a study identified elevated Lp(a) levels as having a strong association with aortic valve calcification and aortic stenosis (AS).[21] Lp(a) was found to be causative in the development of AS soon after.[22]

Structure, Metabolism, and Function of Lp(a):

Lp(a) is a plasma particle that is composed of an LDL component that is covalently linked by a disulfide bond to an apo(a) glycoprotein (see Figure 1).[15] The apo(a) is usually found bound to Lp(a) in plasma. Apo(a) is coded for by the *LPA* gene located on chromosome 6q25.3–6q26 (see Figure 2). The *LPA* gene is homologous to the gene immediately upstream that codes for plasminogen (*PLG*).[7] The plasminogen protein contains several domains, including an amino-terminal domain, five kringle domains (K1–5) and a protease domain. Apo(a), however, is composed of 10 Kringle 4 (KIV) domains that are each homologous to the K4 domain, a KV domain that is homologous to the K5 domain, and an inactive protease domain. Apo(a) lacks domains that are homologous to K1–3. Each of the KIV domains in apo(a) have unique amino acid sequences.[23] All of the KIV domains in apo(a) are found in a single copy except for the KIV type 2 (KIV_2) domain, originally found to be present in anywhere from three to 42 copies,[24] which can be found repeated from two[25] to over 50 times.[26] Most of the KIV_2 repeats have the same amino acid sequence as other KIV_2 repeats.[7] Almost all individuals are heterozygous for KIV_2 repeat count.[27]

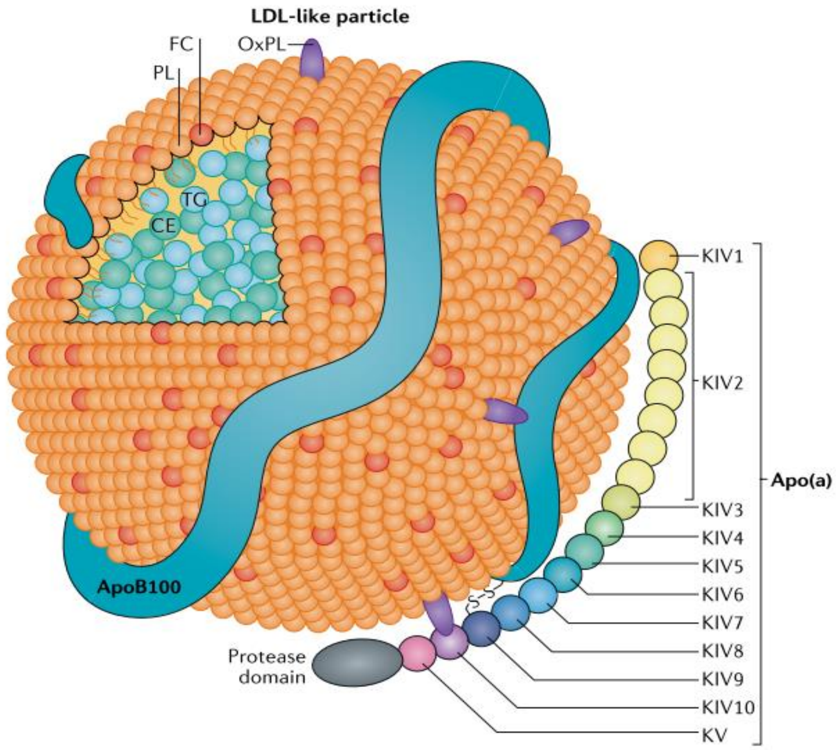


Figure 1: Lipoprotein(a) Particle Diagram. This figure shows the structure of the lipoprotein(a) particle. Diagram taken from Boffa et al., 2019, *Cardiology*.[15]

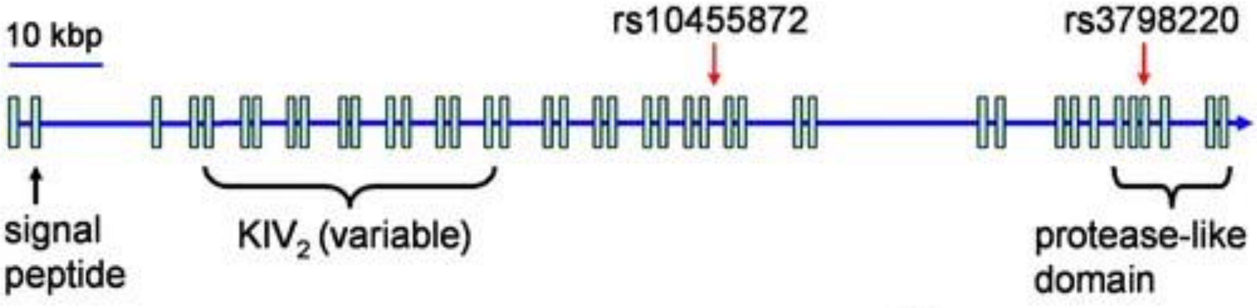


Figure 2. *LPA* Gene Diagram. This figure shows the *LPA* gene, as well as key domains and variants found within the gene. Diagram taken from Li et al., 2011, *Circulation*.[28]

The kringle domains are 114 amino acid[27] ligand binding domains that have three loops that are each stabilized by a disulfide bond.[29] The KIV domains in apo(a) each have different amino acid structures, and have two exons with an intervening intron;[30] the exons are 160 and 182 bp in length, respectively, while the intervening intron is ~4kb in length.[31] The KV domain also contains two exons.[30] Kringle domains tend to be found in molecules with proteolytic activity in humans, such as plasminogen, prothrombin, urokinase, tissue plasminogen activator, and some coagulation factors.[7] However, Lp(a) is not the only kringle-containing particle without proteolytic activity; hepatocyte growth factor is a kringle-containing protein without protease activity.[32] Hepatocyte growth factor also contains two exons per kringle domain, an inactive protease domain, and likely also evolved from the *PLG* gene.[32] However, apo(a) is the only protein to contain a copy number variant containing a kringle domain.[33] The lysine-binding sites on the kringles in plasminogen are used to interact with target molecules and to maintain the structure of plasminogen.[34] In apo(a) the lysine binding sites of $KIV_{1–4}$ and KIV_9 have been inactivated and thus do not function in binding Lp(a) substrates.[35] However, these domains do still engage in other types of intermolecular interactions that may contribute to Lp(a) functionality.[35] Interestingly, the linker domain between each kringle in apo(a) is glycosylated, while this modification is absent in other kringle-domain containing proteins such as plasminogen.[30]

Differences exist in the binding capacity of different kringle domains in Lp(a), as well as other kringle domain-containing proteins.[15] Some kringle repeats have lysine binding sites that are critical to their function of binding.[15] During synthesis of Lp(a), the lysine binding domains in KIV_5–KIV_8 are responsible for non-covalent interactions between apolipoprotein B (apoB) and apo(a) before disulfide bond formation.[36] A critical lysine binding site is also found in KIV_{10}, which is involved in Lp(a) functionality and is likely the major driver of the pathogenic effects of Lp(a).[15] A unique non-lysine binding site is found in the KIV_9 domain in apo(a) and is responsible for the disulfide bond between apo(a) and the apoB component of the LDL-like moiety of Lp(a).[37] The function of the KIV_2 repeats is largely unknown, and is driven by non-lysine binding sites, but these repeats are strongly associated with Lp(a) concentrations: the fewer KIV_2 repeats present, the shorter the *LPA* gene, and the higher Lp(a) concentrations.[38]

Although the kringle domains are relatively well understood, much remains to be learned about Lp(a). It is known that the majority of Lp(a) is synthesized in the liver by primary hepatocytes.[15] However, whether the assembly of the particle occurs intracellularly or extracellularly is controversial. Some studies indicate that Lp(a) is assembled intracellularly,[39] whereas others indicate that Lp(a) is assembled extracellularly,[37] or that both of these cellular regions are sites of Lp(a) assembly.[40] If assembled extracellularly, it has been proposed that Lp(a) is formed on the primary hepatocyte cell surface[41] or formed through interactions between LDL and apo(a) in the plasma.[15] In HepG2 cells, Lp(a) was not detected intracellularly, but was found to occur in blood plasma.[37] However, an *in vivo* stable-isotope kinetic study indicated the Lp(a) was synthesized intracellularly because Lp(a)-apoB was synthesized at similar rates to apo(a), but at different rates compared to LDL.[39] The differences in study results is likely due to differing methodologies being used, including differing Lp(a) purification protocols.[39] One problem with research indicating extracellular assembly of Lp(a) is that they use model systems that have important differences compared to humans since Lp(a) is not found naturally in non-primate animals.[39] Issues have also been identified in research that supports intracellular Lp(a) assembly, including small sample sizes of blood donors.[39]

The site and mechanism by which Lp(a) is cleared and catabolized is also poorly understood.[15] It is known that Lp(a) is catabolized at a similar rate between different isoforms of the Lp(a) particle.[38] Before catabolism, most Lp(a) is transferred to the liver, with a small percentage being transported to the kidneys[42] which likely play a role in Lp(a) catabolism.[43] Plasma Lp(a) concentrations decrease after blood passes through the kidneys, although this finding was identified only in individuals with chronic diseases.[44] The same study also showed that there is a small loss of Lp(a) in the kidneys, with about 1.5 mg of Lp(a) lost per dL of plasma, and there was no indication that catabolism occurs in the kidneys.[44] However, apo(a) is found in human urine, indicating that the kidneys participate to some extent in clearance of Lp(a) from the body, even if they do not play a role in catabolism.[33]

Several receptors are putatively associated with Lp(a) clearance: LDL receptors, plasminogen receptors, toll-like receptors, scavenger receptor class B member 1, and lectins.[15] One study found, using HepG2 cells, that increasing low-density lipoprotein receptor (LDLR) expression reduced Lp(a) concentrations and that the LDLR is involved, to some extent, in the uptake of Lp(a).[45] Other studies indicate that LDLR does not play a role in Lp(a) clearance, as statin administration, which increases LDLR production by decreasing intracellular cholesterol synthesis, does not impact Lp(a) concentrations.[46-49] The inconsistencies between studies is likely because PCSK9, one of the targets of LDLR increasing therapy, targets other receptors besides LDLR, such as LRP1[50] and apolipoprotein E receptor 2.[51] Lp(a) does not bind PCSK9 *in vitro*, indicating the PCSK9 effects are not modulated directly through Lp(a).[45] Another receptor associated with Lp(a) clearance is the plasminogen receptor, as plasminogen appears to competitively inhibit Lp(a) clearance.[52] One study found that both the LDLR and plasminogen receptor are important in Lp(a) clearance and degradation.[52] The LDLR was found to be calcium-dependent and capable of endocytosing Lp(a) particles, while the plasminogen receptor was calcium-independent and could endocytose both apo(a) and Lp(a). Once internalized into a cell through clathrin-coat mediated endocytosis, Lp(a) is transported to lysosomes for degradation.[45] The predominant mechanism of Lp(a) uptake differs between different cell types.[52] Another consideration is that the non-covalent interactions between LDL and Lp(a) may affect how Lp(a) is cleared.[53]

Regardless of the mechanisms of Lp(a) synthesis, catabolism and clearance, Lp(a) levels are largely regulated at the level of synthesis.[54] The length of the *LPA* gene is determined by the number of KIV_2 repeats.[15] The longer the *LPA* gene, the longer the resulting apo(a) isoform, resulting in a longer Lp(a) particle, which results in lower plasma concentrations of Lp(a).[38,55-57] The putative reason that longer *LPA* gene lengths are associated with lower Lp(a) concentrations is that longer isoforms of Lp(a) require more post-translational modifications in the endoplasmic reticulum, which slows their secretion rate[55] as Lp(a) undergoes many post-translational modifications.[58] Lp(a) synthesis rates are also controlled by transcription factors binding to the 5'-UTR of the *LPA* gene.[59] For example, a study in baboon hepatocytes found a binding site for hepatocyte nuclear factor 1α in the promoter region of *LPA*.[59] These promoter regions show activity that is hepatocyte-specific, indicating they are important in hepatic *LPA* expression.[59,60]

Despite the general association between *LPA* gene length and Lp(a) concentrations, exceptions to this rule have been identified.[15] The same apo(a) isoform can be associated with substantially different Lp(a) concentrations across different individuals[27] and across different ethnicities.[30] Several reasons for this difference have been proposed.[15] One explanation is a variant located in an intron of the KIV_2 repeats found more commonly in short apo(a) isoforms that interferes with proper splicing, resulting in longer isoforms of Lp(a) than expected based on the number of KIV_2 repeats.[61] Variants may also be also found in genes outside of *LPA* that regulate the activity of Lp(a) that can result in unusual *LPA* transcription levels.[15,62,63]

The function of Lp(a) remains a mystery. Lp(a) may have no required function in the body as homozygotes for *LPA* null alleles show no obvious clinical issues nor increased risk of death but rather, a benefit of decreased odds of CVD.[64] Regardless, since Lp(a) remains expressed in humans, there may be a residual function that benefits humans. There have been several proposed functions, but none have been confirmed. Lp(a) has been found to be the predominant carrier of oxidized phospholipids (OxPLs) in the plasma, carrying at least 85% of plasma phosphocholine-containing OxPLs.[65] The OxPLs bind both the LDL-component and protein components of Lp(a).[65] Apo(a) binds to OxPLs through a lysine binding site located in the KIV_{10} domain.[66] Interestingly, the capacity for Lp(a) to carry OxPLs decreases with increasing length which may explain the pathology associated with shorter *LPA* gene lengths.[65] However, is has also been proposed that, since OxPLs contribute to atherosclerotic CVD, Lp(a) particles may be critical to the detoxification of OxPLs by binding OxPLs released by apoptotic or necrotic cells.[65] While these functions have been indicated for OxPLs that contain phosphocholine head groups, one study found that Lp(a) is a general transporter of OxPLs, regardless of the OxPL headgroup.[66]

Another potential function of Lp(a) is to promote wound healing.[30] Lp(a) has been identified in high concentrations at wound sites after coagulation.[67] It is believed that Lp(a) transports cholesterol to be used for wound healing by traveling through the plasma and binding endothelial and smooth muscle cells at sites of injury.[30,68] Lp(a) also activates pro-inflammatory and pro-thrombotic molecules that stimulate the wound healing response.[69] Lp(a) inhibits the conversion of plasminogen to plasmin, which would allow wounds to remain plugged longer.[70] Lp(a) can bind both fibrin and plasminogen inhibiting enzymes, which allows it to further inhibit fibrin lysis.[69] Lp(a) can also bind to molecules specifically involved in wound healing, such as histidine-rich glycoprotein, which supports the hypothesis of Lp(a) being involved in such processes.[69]

In the field of Lp(a) research there are unfortunately inconsistencies due to different Lp(a) measurement methods. Many previous studies have measured the mass per unit volume of Lp(a), which is problematic as the mass of Lp(a) can vary considerably between individuals due to the variable mass of several different components of Lp(a).[71] More contemporary work measures Lp(a) concentrations using molar units to evade the issue of variable mass. Further, two isoform-insensitive methods are becoming more popular for measuring Lp(a) concentrations: ELISA and mass spectrometry. Another issue with previous Lp(a) measurements is that, in methods that use immunochemistry to identify Lp(a), antibodies directed at apo(a) have a high probability of being non-specific because of the high sequence homology between kringle domains in apo(a), especially within the KIV_2 repeats. Even if an antibody is supposed to bind uniquely to the Lp(a) molecule, many have been understudied for precise Lp(a) binding locations. Compounding the problem is that the specific calibrators used for Lp(a) measurement are unknown, and some researchers fail to mention the assay used to measure Lp(a) concentrations. Many issues related to Lp(a) will likely be solved in the coming years as isoform-insensitive assays and improved Lp(a) measurements become predominant in Lp(a) research.

Evolution of *LPA*:

The *LPA* gene evolved from duplication of the *PLG* gene, with subsequent deletion and mutation.[72] The *LPA* gene arose around 33 million years ago, but the KIV_2 repeats emerged more recently in human evolution.[73] In humans only, it has been proposed that further mutations occurred that resulted in restoration of a lysine binding site in KIV_{10} of apo(a), thus allowing it to bind OxPLs.[66] Because it arose after the evolutionary split of the Old and New World Monkey groups *LPA* is found only in Old World Monkeys[73] including bonobo, chimpanzee, gorilla, orangutan, baboon, cynomolgus, and rhesus monkeys.[66] All of these species have a protease domain that is inactivated, as in humans, as well as a single KIV domain, unlike in humans.[66] Differences exist in the sequence and genetic structure of *LPA* between primates. The KV domain is absent in macaques, baboons,[30] and cynomolgus monkeys, but is present in other Old World Monkeys.[66] Also two mutations are present in the key protease triad in the protease domain of *LPA* in rhesus monkeys that are absent in humans.[74] There are other various mutations to the lysine binding site in the KIV domain in all Old World Monkeys, except orangutans.[66] The human apo(a) also appears to have gained a function that is absent from other primates with the *LPA* gene: the ability to carry OxPLs in the plasma.[66]

The *LPA* gene has evolved twice during evolution: once in hedgehogs and once in primates.[72] The hedgehog version of *LPA* emerged much earlier than the primate version, around 86 million years ago.[73] Both types of *LPA* are expressed in the liver.[72] The *LPA* gene differs between primates and hedgehogs, with the hedgehog *LPA* containing only sequences that are paralogous to the *PLG* K3 domain and the primate *LPA* containing sequences paralogous to the *PLG* K4, K5, and protease domains.[72] Further, the KIII is found in multiple copies in hedgehog *LPA* while the KIV domain is found in multiple copies in the primate *LPA*.[72] As in primates, the *LPA* gene product in hedgehogs forms covalent bonds with the apoB component of LDL via an unpaired cysteine found in only one of its repeated domains.[72] In addition, the kringle domain repeat polymorphism also arose many years after the initial *LPA* gene formation.[73]

The apo(a) glycoprotein likely serves, or served, an adaptive function as it has evolved twice during mammalian evolution.[73] It is possible that it has instead become maladaptive over time, which explains its association with CVDs. Hedgehogs could provide an animal model of elevated Lp(a) that will allow a better understanding of the function of apo(a).

Ethnic Differences in Lp(a) Concentrations:

Lp(a) levels vary widely across individuals, with levels varying from lower than 1 mg/dL to greater than 100 mg/dL.[7] In addition to differences between individuals, differences are present in the average Lp(a) concentrations between different ethnicities.[75] Black individuals tend to have the highest Lp(a) concentrations, with Chinese individuals tending to have the lowest Lp(a) concentrations.[76] There are even differences in average Lp(a) concentrations between different subpopulations of Europeans. For example, Finnish individuals have some of the lowest average Lp(a) concentrations compared to other European ethnicities.[77] There seems to be a similar increase in CVD risk for each unit increase in Lp(a) concentration across ethnicities despite the difference in average Lp(a) concentrations between ethnicities.[76]

It is likely that there is a genetic basis for the ethnic differences in Lp(a) concentrations, as research indicates no significant effect of environmental factors across ethnicities.[78] The genetic basis for the differences in Lp(a) concentrations across ethnicities is poorly understood, though research indicates that the differences are primarily due to genetic variation in the *LPA* region.[77,79-82] One reason for this lack of understanding is the low sample sizes that are found in cohorts of non-European ethnicities, though this has not prevented researchers from attempting to study Lp(a) across multiple ethnicities. One study found that shorter *LPA* gene lengths are associated with increased Lp(a) concentrations across several ethnicities, including Africans, Chinese, Arabs, Europeans, Latinos, South Asians, and Southeast Asians.[82] That study also showed that there is a difference in the average Lp(a) isoform size between different ethnicities, which explains at least some of the inter-ethnicity variability in Lp(a) concentrations. Chinese and South Asian individuals were found to have the highest average KIV_2 repeat count, while Africans had the lowest average KIV_2 repeat count. Similar results have been found in other studies that used a variety of methodologies such as whole genome sequencing, qPCR, and ELISA.[77,79,80]

Although it is compelling to think of the differences in Lp(a) concentrations between ethnicities to be largely due to KIV_2 repeat count, there can be differing average Lp(a) concentrations between Europeans and Blacks with similar apo(a) isoforms.[80] This data indicates that, although KIV_2 repeats are critical in determining Lp(a) concentrations, the differing Lp(a) concentrations between ethnicities may be explained, at least in part, by other genetic variation. Some key variants are predicted to have severe consequences on Lp(a) concentrations, but show very different frequencies across ethnicities.[31] Some of these variants are common in one ethnicity, but completely absent in another. For example, it has been found that null alleles (alleles that result in undetectable levels of Lp(a)) are much more common in Europeans than in Blacks.[80] Regardless of the number of KIV_2 repeats, Lp(a) concentrations will be lower if a null allele is present. Another study found that individuals of African descent have three times fewer Lp(a) decreasing alleles in the *LPA* gene, but three times higher Lp(a) increasing alleles in the 5'-UTR region of *LPA* than Europeans.[25] Europeans have higher average Lp(a) concentrations than several Asian populations and the variant rs10455872, which is associated with increased Lp(a) concentrations and decreased Lp(a) isoform size, is found commonly in Europeans, but uncommonly in Asian populations.[79] These results indicate that rs10455872 could be a causative factor in the difference in Lp(a) concentrations between Europeans and Asians. However, rs10455872 may be a marker for variation in KIV_2 length rather than functional because it is correlated with KIV_2 repeat count.[18] Regardless, a study found that both single nucleotide polymorphisms (SNPs) and KIV_2 repeats contribute independently to Lp(a) concentrations in Europeans, South Asians, and Chinese.[79] Another example of a key variant associated with Lp(a) concentrations which shows differences in frequency across ethnicities is rs3798220; it is very common in Latinos, common in Europeans, and non-existent in Africans.[83] [25] However, both KIV_2 repeats and SNPs may both be independently important, as indicated by previous research.[77,84]

There are also several possible functional explanations for differing Lp(a) concentrations between certain ethnicities. One study found that African Americans have a variant more often than Europeans that increases promoter activity and thus apo(a) production.[85] These results indicate that the ethnic differences in Lp(a) are in part due to differences in transcription rate. This same study found that a variant associated with decreased Lp(a) concentrations in the region coding for the KIV_8 domain, which is critical for non-covalent bonding during Lp(a) synthesis, is found more often in Europeans than African Americans. Such a variant likely decreases Lp(a) levels by interfering with proper synthesis of the Lp(a) particle. Studies have found ethnic differences in other Lp(a)-related phenotypes. For example, Black individuals have high levels of oxidized phospholipids (OxPLs) compared to European and Latino individuals.[78] OxPLs are negatively associated with KIV_2 repeat count and positively associated with Lp(a) concentrations in Europeans, Latinos, and Blacks.[78] This means that OxPL carrying capacity or availability could be another reason for variation in Lp(a) concentrations across ethnicities. It is likely that multiple mechanisms explain ethnic differences in Lp(a) concentrations.

The distributions of *LPA* gene length have been shown to differ between European, Latino, and Black individuals, with Blacks having a normal distribution, and Europeans and Latinos having a bimodal distribution.[78] Further, the Lp(a) distribution in blacks shows minimal skewing, while Europeans and Latinos show a defined skew-to-the-right.[78] The genetic basis for this phenomenon is also poorly understood.

An understanding of the source of ethnic differences in the *LPA* region can be gained by examining cohorts with multiple ancestral populations such as the Human Genome Diversity Project (HGDP).[86] One can observe the diversity of populations included in this cohort in Figure 3. This cohort was recently sequenced which allows for further study of the conservation of genetic sequences across ethnicities.[87]

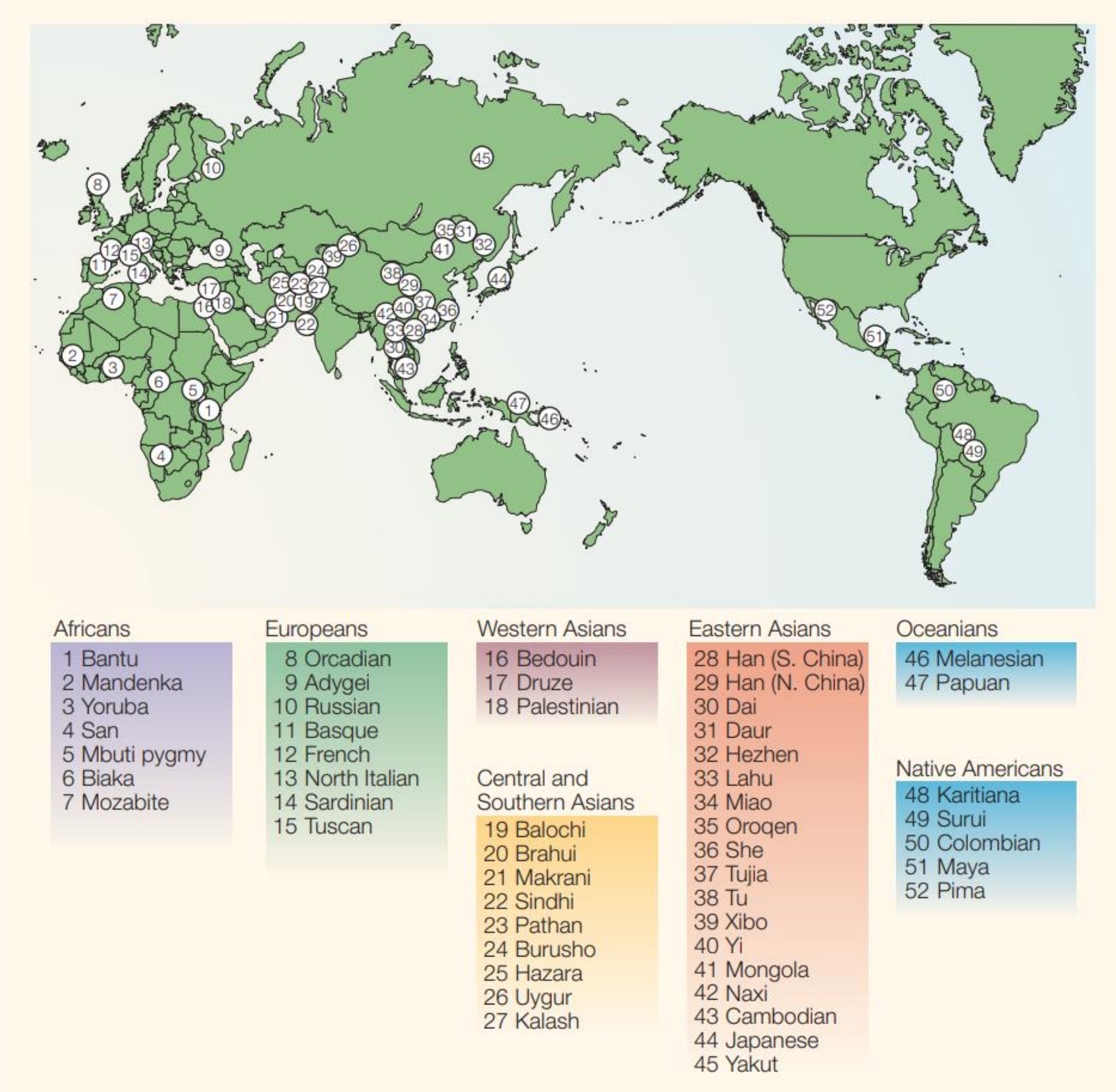


Figure 3. Human Genome Diversity Project Population Distribution. This figure shows the geographical location of each population recruited for the Human Genome Diversity Project. Figure taken from Cavalli-Sforza, 2005, Nature Reviews: Genetics.[88]

Haplotypes and Lp(a):

Haplotypes are collections of genetic variants that are found on the same chromosome.[89] They are a potential tool for better understanding genetic associations with disease. Haplotypes can be used to identify areas of recombination,[89] which can be useful when selecting potential clinical diagnostic variants. Haplotypes are already used in the clinic when patients have their *HLA* region genotyped to assess their compatibility with donor organs.[89] The closer the match between a donor and receiver haplotype, the lower the probability that an organ will be rejected.[89] In individuals with sickle cell disease, haplotypes of the ß-globin cluster have been used to identify severity of the disease.[90] The ß-globin haplotypes have also been shown to differ in frequency across multiple ethnicities.[90] However, besides these two clinical uses of haplotypes, there is currently minimal use of haplotypes in clinical settings.[89] Haplotypes are generally unused in clinics because other, more valid alternatives such as single variants can be used.[89] Few studies have used haplotypes to study Lp(a) concentrations, and even fewer include both haplotypes and multiple ethnicities. Despite the poverty of research using haplotypes, haplotypes are a potentially important tool for Lp(a) research. This idea is corroborated by the fact that pharmacogenomics researchers have used haplotypic analysis to show that specific haplotypes, but not the variants that make up these haplotypes, affect the response of asthmatics to ß-agonists.[91]

There has been some research assessing Lp(a) concentrations in relation to haplotypes. Two *LPA* region haplotypes, one of which spanned the KIV_2 repeats, have been identified in European, South Asian and Chinese samples.[79] One can predict that haplotype structure in Africans is different than other ethnicities, which may explain previous results showing that they have the higher average Lp(a) concentrations across populations. Another study found a haplotype that spans the *SLC22A3*, *LPAL2* and *LPA* genes was associated with Lp(a) concentrations.[92]

As previously stated, some variants have opposing effects in different ethnicities. It is unlikely that there is a key biological difference associated with the same variant between different ethnicities, so haplotypic variation is proposed to explain these differences. One study found that the variant rs19287380, which is located in an intron of *LPAL2* located adjacent to the *LPA* gene, is associated with increased Lp(a) concentrations in Africans, but decreased Lp(a) concentrations in Finns.[77] Interestingly, the authors found that this same variant had a different haplotypic background between the two ethnicities, but they did not run a haplotype-based analysis to confirm a haplotype-based association. There are other genes with association with *LPA* yet to be investigated for their haplotypic association with Lp(a) concentrations.[77,93,94] The APOH gene was recently identified as a locus strongly associated with Lp(a) concentrations in a recent GWAS, and thus will likely be an important target of haplotype analyses.[93]

Haplotypes have also been used to show associations between the *LPA* region and CVD. A four SNP haplotype that spanned the *SLC22A3*, *LPAL2* and *LPA* genes was found to be associated with CAD.[92] The haplotypes in this region that were associated with the highest Lp(a) levels were also associated with increased myocardial infarction and CAD risk. These haplotypes showed a stronger association between the *LPA* region and CAD than any of single SNP analyses in the *LPA* region. An overlapping haplotype has also been found to be associated with carotid artery disease.[84] There were nine SNPs in this overlapping haplotype and together these SNPs explained ~30% of variance in Lp(a) concentrations. However, this study found that the haplotypic analysis did not add additional information compared to individual SNP analysis. The haplotypes built in these studies were relatively small (2–9 SNPs), so it is possible that larger haplotypes would allow for more information that would allow better assessment of haplotypic associations with Lp(a) levels.

Lp(a) and Cardiovascular Disease:

Two CVDs that have been shown to be associated with Lp(a) concentrations are CAD and AS.[95] AS is a heart valve disease in which calcification, fibrosis, inflammation, neoangeogenesis, and lipid accumulation result in stiffening and thickening of the aortic valve. AS is the third most common CVD in North America. The diseases result in left ventricular hypertrophy, as well as diastolic and systolic dysfunction. There are no accepted pharmacological treatments for AS. The only treatment options for AS are transcatheter aortic valve replacement, surgical aortic valve replacement, or lipoprotein apheresis. The mechanism by which Lp(a) contributes to AS development and progression are not well understood, but Lp(a) likely delivers pro-inflammatory OxPLs to the aortic valve after endothelial damage.[95]

Several stages for the pathology of Lp(a) have been proposed. First, endothelial damage occurs in the endothelial layers of the aortic valves.[96] Then, Lp(a), with bound OxPLs, binds to the damaged endothelium and deposits in the aortic valves.[97] Over time, Lp(a) and other lipids accumulate in the aortic valve.[96] The Lp(a) and oxidized lipids can be taken up by macrophages which converts them to pro-inflammatory foam cells.[98] The OxPLs bound to Lp(a) also contribute to the atherosclerotic process by activating various immune cell activities,[15] as well as by activating calcification pathways.[99] OxPLs activate these calcifying pathways through activation of the NF-κB pathway which downstream leads to the production of the pro-osteogenic factor BMP2.[99] The OxPLs have pathogenic effects because Lp(a) also transports lipoprotein-associated phospholipase A_2 and autotaxin, which together convert OxPL into a pathogenic molecule.[99]

CAD is the most common CVD in North America.[95] The prevalence of CAD varies across populations.[100] Symptoms of CAD include angina and myocardial infarction.[100] CAD is caused by the build up of plaques in the coronary arteries, which supply blood to the heart, eventually resulting in occlusion of the arteries and thus insufficient blood flow to the heart.[100] Myocardial infarction resulting from coronary artery blockage can be treated with percutaneous coronary intervention to open the blocked coronary arteries, with stent insertion to keep the coronary arteries open.[100] Patients can also take antiplatelet agents, such as aspirin, to prevent myocardial infarction or cardiovascular emergencies.[100] It is likely that Lp(a) contributes to CAD by delivering pro-inflammatory OxPLs to the arterial intima after endothelial damage, similar to AS.[101] The main difference between the involvement of Lp(a) in AS and CAD is that, in CAD, Lp(a) activates proliferation and migration of smooth muscle cells[98] and that activation of calcification pathways occurs later in CAD.[102] Also, the activation of apoptosis in macrophages by OxPLs can lead to the necrosis and destabilization of plaques in coronary arteries which can lead to thromboembolic disease.[103]

There have been associations found between Lp(a) concentrations and CVD across multiple ethnicities. However, few studies have assessed CVD in non-European ethnicities and the results are not always consistent. There are differences in the rates of CVD between different ethnicities, but the increased risk for each unit increase in Lp(a) appears to be similar across ethnicities.[76] There may be differences, however, in the association between Lp(a) and certain CVDs across ethnicities. Lp(a) concentrations were found to be associated with atherosclerosis risk in both Europeans and Africans in one study,[77] but another study found no association between Lp(a) and myocardial infarction in African individuals.[82] One study found that an *LPA* gene risk score predicted CAD, stroke, and myocardial infarction risk in both Europeans and

African Americans.[77] These results indicate that more research is required to better understand the association between *LPA* and CVD across ethnicities.

Some treatments are being assessed for their utility in treating CVD by lowering Lp(a) concentrations. The most promising treatment is TQJ230, as it can lower Lp(a) mRNA levels by ~90%.[104] The HORIZON phase III clinical trial is being completed to assess the utility of this antisense RNA in lowering CVD disease risk (NCT03887520). Ongoing clinical trials are assessing the utility of an siRNA treatment, called AMG 890, in CVD patients with very high plasma Lp(a) concentrations (NCT04270760). The effectiveness of these RNA treatments is increased due to an N-acetyl galactosamine molecule that is attached to the RNA molecules to target them specifically to the liver.[105] Antisense and siRNA treatments work by inactivating and/or cleaving apo(a) mRNA, so that it cannot be translated into the apo(a) glycoprotein.[106] Clinical trials have also demonstrated that PCSK9 inhibitors can lower Lp(a) concentrations and decrease the risk of CVD.[107] PCSK9 inhibitors lower Lp(a) concentrations by up to 30%,[108] but also lower the concentration of LDL cholesterol and other substances;[109] this means that the benefits of PCSK9 inhibitors may not be due to the Lp(a) lowering effects. It will be seen in the coming years if Lp(a) is a useful target of treatment. Regardless, it is important to understand the genetic association of *LPA* with CVD so that diagnostic variation can be identified and so that the mechanism by which Lp(a) associates with CVD is understood.

Hypotheses:

Variants in the *LPA* gene demonstrate a strong association with Lp(a) and CVD in European and non-European populations, but the genetic basis for associations in non-Europeans are poorly understood. It is hypothesized that haplotypes in the *LPA* region will be associated with both Lp(a) concentrations and CVD risk. It is also hypothesized that the *LPA* haplotypes will be highly conserved across all assessed ethnicities, except those of African descent. Most importantly, it is hypothesized that SNP haplotypes in *LPA* can be markers of KIV_2 repeat count.

Objectives:

The objectives of the present research are to (1) identify associations between *LPA* haplotypes and Lp(a) in multiple ethnicities, (2) assess the conservation of *LPA* haplotypes across multiple ethnicities, and (3) identify the association between *LPA* haplotypes and CVD in multiple ethnicities. The first objectives were achieved by performing haplotypic analyses using linear regression on individuals of four different ethnicities from UK Biobank: White British, South Asian, Chinese, and Black; the same analyses were repeated in the available ethnicities of the Multi-Ethnic study of atherosclerosis (MESA) cohort. Three common variants were then included in the haplotypes to elaborate on the association between *LPA* haplotypes and Lp(a) concentrations: rs3798220, rs10455872, and rs140570886. The conservation of the haplotypes was assessed using haplotype frequency analyses with individuals from 36 different populations selected from HGDP. CVD associations were assessed using logistic regression analyses in seven different ethnicities, from UK Biobank and the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort, with CAD and AS outcomes. There must be linkage disequilibrium (LD) between kringles and SNPs, and if we can assess KIV_2 repeat number with SNPs this would improve testing in the future since the KIV_2 repeats are difficult to sequence.

Association of *LPA* Haplotypes with Lipoprotein(a) Concentrations Across Multiple Ethnicities

Authors:

Thomas W. Marsh[1,2], Hao Yu Chen[2,3], Line Dufresne[2], Yoshihiko Nagai[2], George Thanassoulis[2,3], and James C. Engert[1,2,3]

1. Department of Human Genetics, McGill University, Montreal, Quebec
2. McGill University Health Center Research Institute, Montreal, Quebec
3. Division of Experimental Medicine, McGill University, Montreal, Quebec

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), Haplotype, Ethnicity, Kringle Repeats, Copy Number Variation

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

Word count: 6,476

Total number of figures and tables: 2 figures, 20 tables

TOC category: Clinical and population studies

TOC subcategory: Arteriosclerosis, Thrombosis, and Vascular Biology

Abstract:

Introduction:

The plasma particle lipoprotein(a) contributes to several cardiovascular diseases. Shorter isoforms of apolipoprotein(a), coded for by the *LPA* gene, are associated with greater lipoprotein(a) concentrations and increased disease risk. *LPA* isoform size, determined by the kringle IV type 2 (KIV_2) repeats, and lipoprotein(a) levels vary between ethnicities. Single nucleotide polymorphism (SNP) haplotypes may be in linkage disequilibrium with the KIV_2 repeats and improve the identification of at-risk individuals.

Methods:

Common haplotypes that span *LPA* were tested for their association with lipoprotein(a) concentrations and cardiovascular diseases in several ethnicities of the UK Biobank (UKB) (White British, South Asians, Chinese, and Blacks). Replication was assessed in the Multiethnic Study of Atherosclerosis (MESA) and the Genetic Epidemiology in Adult Health and Aging (GERA) cohorts. Additional haplotypes were created with previously identified variants. The ancestral origins of haplotypes were assessed with the Human Genome Diversity Project (HGDP).

Results:

Two haplotypes were identified in the *LPA* region that span the KIV_2 repeats and accounted for >98% of the haplotypes in the White British, Chinese, and South Asian individuals of the UKB. These same haplotypes were also found at high frequency in European and Latino individuals in MESA and GERA. These two “yin yang” haplotypes are associated with opposing effects on lipoprotein(a) levels in most ethnicities. However, in individuals of European descent,

the haplotype that increases lipoprotein(a) levels has an effect that is due entirely to the presence of two common variants: rs10455872 and rs140570886. The yin yang haplotypes were also associated with coronary artery disease and aortic stenosis across several ethnicities in GERA and UKB. Similar haplotype frequencies were observed for all non-African ethnicities in the HGDP. Blacks had a more heterogenous haplotypic composition, but showed similar haplotype frequencies across all cohorts.

Conclusions:

Two *LPA* haplotypes are associated with lipoprotein(a) concentrations, as well as cardiovascular diseases, and are highly conserved in non-African origin populations. The observed effect of the haplotypes in Europeans and some other ethnicities is driven by known SNP variants and potentially KIV$_2$ repeat number.

Abbreviations:

AS: Aortic Stenosis
Apo(a): Apolipoprotein(a)
CVD: Cardiovascular Disease
CAD: Coronary Artery Disease
GERA: Genetic Epidemiology Research on Adult Health and Aging
GWAS: Genome-Wide Association Study
HGDP: Human Genome Diversity Project
KIV$_2$: Kringle IV Type 2
LD: Linkage Disequilibrium
Lp(a): Lipoprotein(a)
MESA: Multi-Ethnic Study of Atherosclerosis

MAF: Minor Allele Frequency

OR: Odds Ratio

SNP: Single Nucleotide Polymorphism

UKB: UK Biobank

Introduction:

Lipoprotein(a) (Lp(a)) is a plasma particle associated with several cardiovascular diseases (CVDs) including coronary artery disease (CAD), myocardial infarction, and aortic stenosis (AS).[1] Lp(a) is composed of an LDL-like particle, which contains apolipoprotein B-100, and an apolipoprotein(a) (apo(a)) glycoprotein.[1] The apo(a) length is determined primarily by the number of kringle IV Type 2 (KIV_2) repeats in the *LPA* gene that codes for apo(a)[2] and fewer repeats lead to higher Lp(a) plasma concentrations [3] and increased CVD risk.[4-6] *LPA* single nucleotide polymorphisms (SNPs) are associated with plasma Lp(a) levels[4,7-10] and apo(a) length.[2,11]

While the association between *LPA* variation and Lp(a) concentrations are well studied in individuals of European descent, fewer studies have examined associations in other ethnicities. The available data indicate that KIV_2 count, SNP allele frequency, and Lp(a) concentrations all differ across ethnicities, with Chinese individuals tending to have the longest average *LPA* length and lowest average Lp(a) concentrations, while Africans tend to have the shortest average *LPA* length and highest average Lp(a) concentrations.[12,13] Interestingly, Black individuals have higher Lp(a) levels than White individuals, even when individuals between these two ethnicities have similar *LPA* KIV_2 repeat counts.[13-15] Despite this variability, it has been demonstrated that Lp(a) isoform length is negatively correlated with Lp(a) concentrations across several ethnicities.[11,12,16,17]

Studies also indicate that there may be inter-ethnic differences in CVD risk.[12,18,19] One study found that Arab and African individuals did not show an association between high Lp(a) concentrations and myocardial infarction, in contrast to Europeans, Chinese, South Asians, Southeast Asians, and Latin Americans.[12] In addition, *LPA* length has a weaker association with

Lp(a) concentrations in Black individuals.[20] Another study found that, despite differences in Lp(a) concentrations across ethnicities, the increased risk of atherosclerotic CVD per unit increase in Lp(a) is the same across ethnicities.[19]

While the *LPA* variant rs6415084 (a variant that is common across all assessed ethnicities) was found to be associated with increased Lp(a) concentrations across multiple ethnicities, including Europeans, South Asians, and Chinese,[11] other SNPs strongly associated with Lp(a) in Europeans, such as rs10455872 and rs3798220,[2] have markedly lower allele frequencies in other ethnicities.[11,20] However, the rs3798220 minor allele is ten times more frequent in Latino individuals (minor allele frequency (MAF) = 0.42) than Europeans (MAF = 0.04) and has an opposing effect on *LPA* length between Europeans and Latinos.[20] Thus, the association between SNPs and Lp(a) concentrations is nuanced, indicating possible linkage disequilibrium (LD) with KIV_2 repeats or varying biological function or both. Variants also exist that are strongly associated with Lp(a) concentrations, but that have not been assessed across different ethnicities. Recently, we identified 127 independent variants in the *LPA* region that explained about half the variation in circulating levels of Lp(a).[10]

Haplotypes may be useful to improve our understanding of the association between *LPA* region variation and Lp(a) concentrations. Previous research has identified a haplotype block that spans the *LPA* region that is highly conserved across ethnicities.[11] Another study has demonstrated that the same SNP can have opposing effects on Lp(a) concentrations in different ethnicities, indicating possible haplotypic modulation of SNP effects.[16] SNP haplotypes may also serve as proxies for KIV_2 repeat length, which are difficult to assess with most current technologies.[21]

The aims of the present study are to identify haplotypes in the *LPA* gene that are associated with Lp(a) concentrations and CVD, and assess the frequency and effect of these haplotypes in multiple ethnicities.

Methods:

Cohorts:

UK Biobank (UKB): The UKB is a cohort of around 500,000 individuals between the ages of 37–82 recruited from 2006–2010 by 22 different centers across the UK.[22] The cohort has been described previously.[22] Briefly, a variety of phenotypic data is available, including ethnicity, disease status, and plasma Lp(a) concentrations. Lp(a) was measured using a Randox immunoturbidimetric assay. Importantly, 24.0% of the Lp(a) values were not available because measured plasma Lp(a) concentrations <3.8 nmol/L and >189 nmol/L were not reported. We created a censored Lp(a) concentration variable that replaces low and high missing values with the threshold value for missing Lp(a) concentrations. Four ethnicities from the UKB were analyzed: South Asians (n = 6,081), Chinese (n = 1,254), Blacks (n = 5,712), and White British (n = 291,547).

Multi-Ethnic Study of Atherosclerosis (MESA): The MESA cohort was established to study CVD progression in individuals with subclinical CVD and has been previously described.[23] Briefly, it is a collection of almost 7,000 individuals, aged 45–84 years, recruited during the years 2000–2002, from six different states across the USA. A diverse set of phenotypic data was collected, including plasma Lp(a) measured using a turbidimetric immunoassay, as well as DNA. All four MESA ethnicities were analyzed: Europeans (n = 1,578), Chinese (n = 777), African Americans (n = 1,123), and Latinos (n = 934).

Genetic Epidemiology in Adult Health and Aging (GERA) cohort: This cohort is comprised of 103,006 individuals aged 18 years and older that were members of the Kaiser Permanente North California healthcare system.[24] Imputed whole genome genotyping was available, and the present study analyzed unrelated European individuals who were aged 55 years or older (n = 55,192) for CAD and AS.

Human Genome Diversity Project (HGDP): The HGDP is a cohort of 1,064 individuals that includes 51 different populations from 60 different geographical regions.[25] Whole genome genotyping was performed on DNA from lymphoblastoid cell lines. Six composite ethnicities were included in the present analyses: Europeans (n = 160), South Asians (n = 125), West Asians (n = 146), East Asians (n = 142), Native Americans (n = 108), and Sub-Saharan Africans (n = 121). All available ethnicities were included in their respective composite ethnicity, except for East Asians, which included only Chinese, Japanese, and Korean individuals. The list of populations included is provided in Table S1.

LPA Region:

The GRCh37/hg19 build of the human genome was used for all analyses except for those involving the GERA cohort, for which GRCh38/hg38 was used. However, all positions presented in this paper reference the GRCh37/hg19 build. For the identification of common *LPA* haplotypes in UKB White British individuals, SNPs were included if they were within 250 kb of the *LPA* gene (chr6:160,952,515–161,087,407). The KIV_2 region spans from 161,032,565 to 161,067,901 on chromosome 6.[16]

Haplotype Construction and Analysis in the UKB:

In UKB White British individuals, two common haplotypes in the *LPA* region were identified using PLINK 1.07[26] based on two criteria: all SNPs had an MAF that differed from the average MAF by less than 0.05, and had an $r^2 \geq 0.95$ with the other variants. The haplotypes were assessed for association with Lp(a) using multiple linear regression adjusted for age, age^2, sex, genotype batch, and 20 principal components as implemented in PLINK 1.07. The statistical significance threshold for prioritizing haplotypes was set at $p \leq 0.05$. The resulting haplotypes, composed of 36 SNPs, were tested for association with Lp(a) in UKB South Asian, Chinese, and Black individuals using the same parameters.

Haplotype Association with Lp(a) in the MESA:

Thirty-five of the 36 UKB SNPs in the *LPA* haplotype were found in the MESA cohort. Multiple linear regression adjusting for sex and age was performed on unrelated individuals with genotyping success ≥90%.

Incorporation of Additional SNPs in Haplotype Association Tests with Lp(a):

Additional haplotypes were constructed by adding additional SNPs to the common haplotypes and performing association analysis on all four UKB ethnicities. Again, covariates for the multiple linear regression were age, age^2, sex, genotype batch, and 20 principal components.

These analyses were also performed on MESA European, African American, and Latino samples. Chinese individuals from MESA were excluded from analyses because all the additional variants were absent. Covariates for the multiple linear regression were age and sex.

Aortic Stenosis and Coronary Artery Disease in the UKB:

The association of the *LPA* region haplotypes with CAD (cases = 32,867) and AS (cases = 3,443) in White British (n = 239,987) individuals was assessed using multiple logistic regression adjusted for sex, age, age^2, genotype batch, and 20 principal components. CAD was defined as the presence of an International Classification of Diseases Tenth Revision (ICD-10) code for atherosclerotic heart disease of native coronary artery (*I25.1*). AS was defined as the presence of an ICD-10 code for nonrheumatic aortic valve disorder (*I35*). Covariates were the same as for the Lp(a) analyses. Unrelated individuals aged 55 and older were included. Associations with CAD in Chinese were not assessed due to case counts <100. Associations with AS in Blacks, South Asians, and Chinese were not assessed due to case counts <100.

Coronary Artery Disease and Aortic Stenosis in the GERA Cohort:

LPA haplotypic associations were tested against CAD (13,245 cases), and AS (3,469 cases) in GERA individuals of European ancestry using multiple logistic regression on unrelated individuals aged 55 and older adjusted for age, age^2, sex, and ten European-specific principal components. CAD was defined as the presence of an International Classification of Diseases Ninth Revision (ICD-9) code for coronary atherosclerosis of native coronary artery (*414.0*). AS was defined as the presence of an ICD-9 code for aortic valve disorder (*424.1*) or a procedure code for aortic valve replacement in the electronic health record. Individuals with an ICD-9 code for congenital valvular disease (*746–747*) were excluded.

Incorporation of Additional SNPs in Haplotype Associations Tests with Lp(a):

Mean Lp(a) concentrations were calculated for haplotypes containing 11 different Lp(a) genome-wide significant *LPA* region variants: rs57175128, rs114284743, rs558685607, rs544366796, rs10455872, rs140570886 (a variant in high LD with rs3798220 ($r^2 = 0.74$)), rs73596816, rs369686024, rs528521448, rs150776685, and rs193210914.[10] These variants were selected because individuals who were homozygous for the minor alleles had average Lp(a) concentrations ≥ 70 nmol/L and because they were highly independent ($r^2 \leq 0.01$).[10] The average Lp(a) concentration for each haplotype was calculated and then plotted to assess trends. The missingness of Lp(a) concentrations in individuals who were carriers for each variant was also assessed. Only UKB White British were included because of the rarity of homozygous individuals in other ethnicities (no homozygous individuals for ≥5 variants). Lp(a) concentrations were calculated for all available haplotypes.

Several null variants have been identified in the *LPA* region: rs201306475,[27] rs139145675,[28] rs41267811,[29] rs143431368,[30] rs41267813,[29] rs41272114,[31] rs201297680,[27] rs41259144,[28] and rs1623955.[27] These variants were used to construct haplotypes with the 36 common SNPs. The UKB did not report Lp(a) measurements <3.80 nmol/L but did identify these samples. Thus, individuals with Lp(a) measurements <3.80 nmol/L were assessed and compared to individuals who had the major allele for all null variants. For these analyses, all UKB White British (n = 384,669), Blacks (n = 7,261), and South Asians (n = 7,319) were included. Chinese participants were not analyzed because none of the null variants were found in Chinese participants except for rs41272114 which was found in only three individuals.

Censored Lp(a) Concentrations in the UKB White British:

We created a censored Lp(a) variable by setting the out-of-range Lp(a) values to the threshold values [3.80 nmol/L and 186.00 nmol/L]. This new variable was analyzed for association with the *LPA* genome-wide significant variants and the null variants.

Haplotype Frequencies in HGDP:

Of the original 36 SNPs that defined the common haplotypes in the UKB, four variants were found in HGDP: rs9365179 (upstream of the kringle repeats), rs10945682, rs1740428, and rs1321196 (downstream of the kringle repeats). These were used to construct and assess haplotype frequencies across the six HGDP composite populations.

Results:

Two common *LPA* haplotypes were identified in the *LPA* gene that span the KIV_2 repeats. The haplotypes were composed of 36 SNPs (Table S2). The variants were in high LD with each other ($r^2 = 0.97–1.00$), and thus had nearly the same MAF (MAF range = 0.36–0.37).

The major and minor haplotypes composed of the 36 SNPs accounted for >98% of haplotypes for White British, South Asian, and Chinese individuals. These haplotypes only comprised 42.1% of haplotypes in the Black individuals indicating greater haplotypic heterogeneity, with eight additional haplotypes having a frequency of at least 0.01. While the major haplotype in the other ethnicities was the most common haplotype in Black individuals, the minor haplotype was only the fifth most common.

The major haplotype was associated with increased Lp(a) levels in White British (ß = 0.22, $p < 1 \times 10^{-315}$), South Asian (ß = 0.28, $p = 1.28 \times 10^{-51}$), and Chinese (ß = 0.24, $p = 2.55 \times 10^{-9}$) individuals (Table 1). In the *LPA* region of Blacks, a conditional haplotype association analysis was performed with the major haplotype as the reference; the minor haplotype still showed a negative association with Lp(a) levels (ß = -0.26, omnibus $p = 6.24 \times 10^{-47}$; Table 2). The major haplotype was not statistically significantly associated with Lp(a) levels in Blacks (ß = 0.02, p = 0.13).

The association between the two *LPA* haplotypes and Lp(a) concentrations was weaker in MESA Europeans (major haplotype: ß = 0.07, p = 0.06; minor haplotype: ß = -0.10, p = 0.01) but had a concordant direction of effect as seen in the UKB White British (Table 3). In MESA Chinese, the major haplotype had a positive association (ß = 0.28, $p = 7.56 \times 10^{-7}$) with Lp(a) levels, and the minor haplotype had a negative association (ß = -0.27, $p = 1.76 \times 10^{-6}$) with Lp(a) levels. In MESA African Americans (Table 4), both the major (ß = -0.13, $p = 4.52 \times 10^{-4}$) and minor (ß = -0.48, p = 1.36×10^{-13}) haplotypes had a negative association with Lp(a) levels. In MESA Latinos, both the major (ß = -0.20, p = 8.83×10^{-5}) and minor (ß = -0.13, p = 0.03) haplotypes also had a negative association with Lp(a) levels. Similar to UKB Blacks, more haplotype heterogeneity was present in MESA African Americans (with the major and minor haplotypes having a combined frequency of 0.48); ten haplotypes had a frequency above 0.01. Latinos had a joint major (haplotype frequency = 0.63) and minor (haplotype frequency = 0.25) haplotype frequency of 0.88 and three additional haplotypes had a frequency above 0.01.

When additional variants were included in the haplotypes, these variants were clearly responsible for the effect of the major haplotype in UKB White British (Table 5) because they were most often found on the major haplotypic background (rs10455872, 99.84% on the major haplotypic background; rs140570886, 99.52%; rs3798220, 70.61%), Rs10455872 had a large effect on Lp(a) (ß = 1.72, $p < 1 \times 10^{-315}$) when on the major but not the minor haplotypic background (ß = 0.04, p = 0.79), though this was a rare haplotype. Rs3798220 had a negative effect or no effect on Lp(a) when found on haplotypes without rs140570886. Rs140570886, on the other hand, had a large effect regardless of the presence of rs3798220 and regardless of haplotypic background.

When the three variants were included in the *LPA* haplotypes in the UKB, no difference was observed in the direction or magnitude of effect of the major + 3 SNPs (major haplotype plus the common allele of all three additional SNPs) or minor + 3 SNPs (minor haplotype plus the common allele of all three additional SNPs) haplotypes, compared to the major and minor haplotypes, in UKB South Asians, Chinese, and Blacks (Table 5). Rs10455872 was also associated with a large positive effect on Lp(a) concentrations in South Asians (ß = 1.04, p = 1.39×10^{-24}) and Blacks (ß = 0.80, p = 9.93×10^{-13}). The MAF was too low to assess rs10455872 in Chinese individuals (MAF = 4.00 × 10^{-4}). Rs3798220 and rs140570886 together were not significantly associated with Lp(a) levels in Blacks, were associated with decreased Lp(a) concentrations in Chinese (ß = -0.41, p = 0.01), and were associated with increased Lp(a) concentrations in South Asians (ß = 0.60, p = 0.02). Rs140570886 was not statistically significantly associated with Lp(a) concentrations in Blacks, Chinese, and South Asians, but the sample size for individuals with rs140570886 without rs3798220 was small in these ethnicities (<9 people). Rs3798220 was always found with rs140570886 in Blacks. Also, rs3978220 was not significantly associated with Lp(a) levels in Chinese nor South Asians.

In MESA Europeans, the major + 3 SNPs haplotype had an opposite effect on Lp(a) levels (ß = -0.17, p = 7.46 × 10^{-6}) compared to the major haplotype (Table 6). The minor + 3 SNPs haplotype effect showed the same direction of effect as the minor haplotype in MESA Europeans (ß = -0.11, p = 4.71 × 10^{-3}), similar to UKB White British. As in UKB White British, rs10455872 was associated with a large increase in Lp(a) concentrations (ß = 1.36, p = 1.09 × 10^{-37}), rs3798220 and rs140570886 together were associated with a large increase in Lp(a) levels (ß = 1.84, p = 6.48 × 10^{-22}), and rs3798220 alone was not statistically significantly associated with Lp(a) levels. Similar to UKB Blacks, in MESA African Americans (Table 6), rs10455872 was associated with a large increase in Lp(a) concentrations (ß = 1.03, p = 6.84 × 10^{-6}), and rs3798220 and rs140570886 together were not associated with Lp(a) levels. In MESA Latinos (Table 6), there was no effect of the major +3 SNPs and minor + 3 SNPs haplotypes, unlike the negative effect of the major and minor haplotypes. Rs10455872 was associated with a large increase in Lp(a) concentrations in Latinos (ß = 1.64, p = 2.12 × 10^{-18}), but rs3798220 and rs140570886 together were associated with a moderate decrease in Lp(a) levels (ß = -0.41, p = 2.05 × 10^{-9}), in contrast to the other ethnicities. Haplotypes bearing rs140570886 or rs3798220 alone had no association with Lp(a) concentrations in Latinos. Additionally, rs3798220 and rs140570886 were much more common in Latinos (MAF = 0.18–0.19) compared to Europeans (MAF = 0.01) and African Americans (MAF = $8.01–8.92 \times 10^{-3}$). None of the three variants were found in MESA Chinese individuals.

In UKB White British individuals, the minor + 3 SNPs haplotype had no effect on Lp(a) concentrations when using the major + 3 SNPs haplotype as a reference (ß = -0.04, omnibus $p < 1 \times 10^{-315}$; Table 5). However, UKB South Asians (ß = -0.27, omnibus $p = 2.11 \times 10^{-72}$), Chinese (ß = -0.27, omnibus $p = 7.05 \times 10^{-9}$), and Blacks (ß = -0.26, omnibus $p = 1.75 \times 10^{-54}$) demonstrated effects similar to the minor + 3 SNPs haplotype.

In MESA Europeans, conditional haplotype analyses with the major + 3 SNPs haplotype used as a reference showed no difference in the effect of the major + 3 SNPs and minor + 3 SNPs haplotypes (ß = 5.38×10^{-3}, omnibus $p = 1.73 \times 10^{-64}$; Table 6). Similar to Europeans, the minor + 3 SNPs haplotype in MESA Latinos had no effect when using the major + 3 SNPs haplotype as a reference (ß = -0.05, omnibus $p = 1.33 \times 10^{-24}$). Similar to UKB Blacks, MESA African Americans (ß = -0.34, omnibus $p = 1.83 \times 10^{-33}$) maintained the same effect size as the one-versus-all haplotypic analyses. MESA Chinese individuals did not have any of the three variants so conditional haplotype analyses with the major + 3 SNPS haplotypes as the reference were not performed.

Eleven haplotypes containing Lp(a) genome-wide significant variants showed a trend of increasing Lp(a) concentrations with each additional copy of the haplotype (Figure 1). These variants were found on independent haplotypes ~91% of the time. Most variants showed a similar increase in Lp(a) concentrations for each additional copy of the minor allele. The plots appeared linear overall, although some variants, such as rs10455872, rs140570886, and rs544366796, showed a less linear trend.

The haplotypes that were less linear may be due to the fact that high Lp(a) values (Lp(a) > 189 nmol/L) were removed from the UKB dataset. The absence of this data would cause high effect variants to have an underestimated effect on Lp(a) concentrations, especially in homozygote minor individuals. Thus, the missingness of the Lp(a) genome-wide significant variants was also assessed (Table 7). The three variants associated with less linear trend lines each had missingness rates that were above 50% and higher than any of the other variants. Of the 65 individuals homozygous for rs140570886, 62 of them had missing Lp(a) values (95.4%) and for rs10455872, the missingness rate was 78%.

To test the idea that higher missingness is indicative of higher Lp(a) concentrations in UKB White British, censored Lp(a) values were analyzed using the 11 Lp(a) genome-wide significant variants (Table 8). The results corroborated the idea that the effect of key variants in the *LPA* region were underestimated in analyses that excluded the UKB missing Lp(a) data.

The 11 Lp(a) genome-wide significant variants, although independent at an r^2 of 0.01, had effects that were dependent on other variants (Figure 2). The variants rs57175128, rs1507766825, rs193210914, rs558685607, and rs114284743 showed a much lower effect when accounting for the respective other variants. However, rs558685607 appeared to have an effect when found in two copies. To extend these results, haplotypes were constructed with all 11 Lp(a) genome-wide significant variants, plus a marker for the *LPA* haplotypes that span the KIV_2 repeats (Table S3). Rs57175128, rs1507766825, rs193210914, rs558685607, and rs114284743 did not have effects when found without other variants on a haplotype. These variants were only associated with a strong increase in Lp(a) concentrations when rs10455872 was also present. The other six variants appeared to show effects that were independent of all the other assessed Lp(a) genome-wide significant variants. However, rs73596816 had effects that were dependent on haplotype, indicating it was simply in LD with the actual causative variation. The five variants that were dependent on rs10455872 all had low effect sizes according to a conditional haplotype analysis (ß ≤ 0.18; Table 7). In fact, the conditional haplotype analyses showed that, except for rs150556685, the effect size of these five variants was ≤ 0.06. The other six Lp(a) genome-wide significant variants each had higher effects on Lp(a) concentrations (ß ≥ 0.80).

Although the five rs10455872-dependent variants were independent based on an r^2 threshold of 0.01, the D' prime gave different information (Table 11). Each of these variants had a D' with rs10455872 of at least 0.16 and at most 0.65. Interestingly, the variants that were haplotypically independent from rs10455872 had a high D' with rs10455872 of at least 0.58, and at most 1.

Censored Lp(a) values were also assessed in UKB White British individuals using *LPA* null variants (Table 9). All null variants had a strong reduction in average Lp(a) concentrations of 25.41 nmol/L when one copy of the respective null allele was present (range of mean Lp(a) in heterozygotes = 6.17–28.67). We saw very low average Lp(a) concentrations in individuals homozygous for rs143431368 (3.80 nmol/L), rs41272114 (3.85 nmol/L), and rs41259144 (3.82 nmol/L). Three of the haplotypes containing null variants assessed in the UKB White British, rs143431368, rs41272114, and rs41259144, were found to be homozygous in some individuals (Table 9); the rest were only present as heterozygotes. Rs143431368 was the only variant with 100% missingness, but also had the smallest homozygous minor individual count (n = 4).

Only individuals who were heterozygous for the minor allele of the null variants rs139145675, rs143431368, rs41272114, rs41259144, and rs1623955 were assessed for missingness in UKB Blacks because no homozygous individuals were observed, and all other null variants were found in ≤1 individual. Rs143431368 and rs41259144 had the highest missingness rates in heterozygote individuals (Table 10), two of the same variants that were identified in White British individuals. In contrast, rs41272114 had a much lower missingness rate (missingness = 10.71%), than in the White British (missingness = 41.73%).

In UKB South Asians, only the variants rs139145675, rs143431368, rs41272114, and rs41259144 were assessed for missingness because all other null variants had ≤1 individual with the minor allele. Only rs41272114 was homozygous minor in ≥1 individual. The heterozygotes for rs143431368 had a similar missing rate to homozygote major individuals (Table 11). Heterozygotes for rs41272114 (missingness = 30.33%) and rs41259144 (missingness = 41.18%) showed higher missingness rates than homozygote major individuals, but not as high as the missingness rates found in the UKB White British. However, 88.9% of individuals who were homozygous for rs41272114 were missing Lp(a) concentration measurement values, similar to White British individuals. The missingness rates in South Asians were generally lower than those found in White British, but higher than those found in Blacks. The null variants were not assessed in the UKB Chinese as only three individuals were carriers.

In UKB White British participants, the haplotypes associated with an increase in Lp(a) concentrations were also associated with increased CAD (OR = 1.04, p = 7.74×10^{-6}) and AS (odds ratio (OR) = 1.06, p = 0.03) odds (Tables 12 and 13).

Consistent with the Lp(a) results, haplotypes containing rs10455872 were associated with a large increase in CAD (OR = 1.33, p = 2.85×10^{-80}) and AS (OR = 1.55, p = 1.11×10^{-29}). Rs3798220 was not statistically significantly associated with CAD, nor AS, except when found on the same haplotype as rs140570886. Rs140570886 was associated with increased odds of

CAD (OR = 1.54, p = 1.14 × 10^{-41}) and AS (OR = 1.82, p = 3.91 × 10^{-14}) when found with rs3798220, also consistent with the Lp(a) results. When found alone on the major haplotypic background, rs140570886 was only found to be associated with increased odds of CAD (OR = 1.41, p = 7.82 × 10^{-3}). Similar results were seen in the GERA Europeans, in whom there were significant effects of the haplotypes for both CAD (Table 14) and AS (Table 15). Rs10455872 was associated with increased odds of CAD (OR = 1.22, p = 1.23 × 10^{-11}) and AS (OR = 1.37 p = 4.05 × 10^{-11}), just as in UKB White British. Rs3798220 and rs140570886 together were associated with increased odds of CAD (OR = 1.31, p = 2.78 × 10^{-5}) and AS (OR = 1.40, p = 8.75 × 10^{-4}), just as in UKB White British. Rs3798220 and rs140570886 were not associated with CAD nor AS when analyzed in isolation. However, rs140570886 did show an OR concordant with that found in UKB White British individuals (OR = 1.37, p = 0.09). Rs3798220 had non-statistically significant opposing effects on Lp(a) between the major and minor haplotype, just like in White British individuals.

In UKB Blacks (n = 1,973, cases = 222), none of the haplotypes showed a significant association with CAD; however, the OR for the major + rs10455872 haplotype was large (OR = 2.39, p = 0.20). In UKB South Asians (n = 2,710, cases = 673), none of the haplotypes of interest showed a statistically significant association with CAD. Instead of the major + rs10455872 haplotype (OR = 0.78, p = 0.53, haplotype count = 44) showing a large OR, it was the minor + rs10455872 haplotype that showed a large effect on CAD (OR = 5.38, p = 0.21, haplotype count = 4). Also, the major + 3 SNPs (OR = 1.08, p = 0.24) and minor + 3 SNPs (OR = 0.92, p = 0.23) haplotypes showed opposing effects on CAD.

The frequencies of the two common *LPA* haplotypes across the six HGDP populations indicated that these haplotypes were highly conserved (Table 16). For all ethnicities, except Sub-

Saharan Africans, these two haplotypes composed >99% of haplotypes, with haplotypic frequencies very similar to those of the other cohorts. In Sub-Saharan Africans, these two haplotypes comprised <92% of haplotypes and six haplotypes were observed, with the minor haplotype being the most common. South Asians and Native Americans had the least variation, with only the major and minor haplotypes being present.

Because only four variants of the original 36 were available for analysis in HGDP, the observed minor haplotype in Sub-Saharan Africans could represent two other haplotypes identified in the other African-descent cohorts. There were four additional haplotypes in Sub-Saharan Africans, and three of them were also found in Black individuals of the other cohorts (Table S4). The fourth, which was not seen in the other cohorts, was found in only one individual from the Biaka people. Importantly, all of the GERA and MESA African American haplotypes were found in UKB Blacks.

Discussion:

Two highly conserved haplotypes in the *LPA* region were associated with opposing effects on Lp(a) concentrations in individuals of European descent. These haplotypes were also associated with opposing effects on CAD and AS in European individuals in a direction that is consistent with higher Lp(a) levels being associated with increased CVD risk. These results were replicated in other ethnicities except for Latinos and Africans, consistent with non-African ethnicities having significantly lower diversity than African individuals.[32] Latinos likely showed differing results because they are a highly admixed population composed of alleles from Africans, Europeans, and Native Americans.[33] When additional variants were included in the

LPA haplotype, it became clear that the effect on Lp(a) concentrations was the result of the accumulated effect of rare variants that arose on the major haplotype.

The haplotypes that had the strongest associations with Lp(a) concentrations in UKB White British contained the risk alleles of rs10455872 and rs140570886. Based on our analyses, these two variants had minimal impact on the effect size of both the major and minor haplotypes in the other UKB ethnicities: South Asians, Chinese, and Blacks. This means that rs10455872 and rs140570886 are likely markers for another type of genetic variation, such as the KIV_2 repeats. However, these results could also be due to the low power of the non-European samples. Previous research has found that rs10455872 and rs3798220 are both associated with fewer KIV_2 repeats.[2,5] Haplotypes provide specific LD information that could be more precise for predicting *LPA* gene length. Future large-scale studies should include a larger number of SNPs to identify haplotypes that can be used as KIV_2 proxies in each ethnicity.

Our replication analyses in the MESA cohort corroborated the results from the UKB, and extended them to Latinos. The MESA Europeans demonstrated haplotype associations with Lp(a) concentrations comparable to the associations found in the UKB White British, although the effect sizes were smaller. The effect sizes and frequencies were concordant between the UKB and MESA Chinese individuals. The major haplotype was not associated with Lp(a) in UKB Blacks but was negatively associated with Lp(a) in MESA African Americans. The minor haplotype showed a similar effect size between UKB Blacks and MESA African Americans. Additionally, UKB Blacks, and MESA and GERA African Americans all had a greater number of *LPA* haplotypes. In Latinos, both common haplotypes showed a negative association with Lp(a). This result could be because Latinos are an admixed population of Europeans, Africans, and Native Americans.[33] However, the major + 3 SNPs and minor + 3 SNPs haplotypes had the same effect on Lp(a) when using the major + 3 SNPs haplotype as a reference, similar to MESA Europeans and UKB White British. Taken together, these results indicate that the *LPA* haplotypes are associated with Lp(a) across multiple ethnicities.

The haplotype frequencies in HGDP indicated that the major and minor haplotypes are the predominate haplotypes in the *LPA* region across many populations, with Europeans and South Asians having haplotype frequencies very similar to the UKB. HGDP Sub-Saharan Africans also had similar haplotype frequencies to UKB Blacks for all perfectly overlapping haplotypes. Because fewer SNPs defined it, the minor haplotype had a higher frequency in the HGDP, but this haplotype would probably resolve into multiple specific haplotypes in the other cohorts. Although Black individuals showed more *LPA* haplotypic diversity than other ethnicities, the haplotypic composition was similar across individuals of African descent in GERA, MESA, and the UKB. Most of the HGDP Sub-Saharan African haplotypes were found in UKB Blacks. Thus, while substantially different from non-African populations, individual cohorts of African descent individuals were highly similar. One haplotype found in an HGDP Biaka individual was not found in any other individual of African descent across all cohorts. This haplotype may be restricted to certain populations.

The differing haplotype frequencies and effect sizes in different ethnicities could indicate that previously identified genetic markers of high Lp(a) are not causative, but instead markers of other variation, such as the KIV_2 repeats. For example, rs3798220 is a highly studied missense variant with varying frequency across ethnicities.[2,4,34-37] It was speculated that rs3798220 may have a mechanism independent of KIV_2 repeat count because it causes an amino acid change,[37] but our work confirms that a variant in LD, rs140570886, has a greater effect and is associated with Lp(a) concentrations regardless of haplotypic background while rs3798220 is actually associated with a small decrease, or no effect at all, in Lp(a) levels when rs140570886 is not present on the same haplotype. In addition, our results demonstrate that rs3798220 is much more common in Latinos than other ethnicities and is also associated with lower Lp(a) concentrations. This is consistent with a previous report.[20] Thus, rs140570886 is likely a marker for the KIV_2 repeats, but only in Europeans. These results again demonstrate the importance of using haplotypes for *LPA* research.

Haplotypes composed of additional genome-wide significant variants[10] demonstrated similar linear trends of increasing Lp(a) concentrations with each additional copy of the respective minor alleles in UKB White British. Some of the strongest variants, including rs10455872 and rs140570886, showed a less linear trend than other variants, but this was likely due to the censoring of Lp(a) measurements higher than 189 nmol/L because individuals with these variants had high missingness rates. The linear trends may indicate that these variants are associated with very specific KIV_2 counts, such that having two copies of the minor allele doubles the KIV_2 count. However, when accounting for rs10455872, half of the Lp(a) genome-wide significant variants no longer showed a significant nor linear association with Lp(a). Because these variants likely only showed an effect because of rs10455872 and also showed a linear effect, this is evidence that rs10455872 has a linear effect as well, further indicating an association with specific KIV_2 counts. Because all the Lp(a) genome-wide significant variants were independent at an r^2 of 0.01, the D' statistic may be more important than the r^2 statistic.

Three null alleles (rs143431368,[27] rs41272114,[38] and rs41259144[39]) were each associated with very low censored Lp(a) measurements. Heterozygous individuals for any of the null variants had on average a halving of their Lp(a) concentration. Individuals homozygous for null variants in UKB White British individuals, Rs143431368, rs41272114, and rs41259144, had high missingness rates (88.1%–100.0%), likely due to Lp(a) concentrations <3.8 nmol/L. Although the censored Lp(a) values are more useful than missing values, future studies would benefit from the raw Lp(a) measurement values. Also, these variants may be useful to create pseudo-haploid models of Lp(a) concentrations, which would allow direct assessment of specific *LPA* haplotypes.[27]

The major + 3 SNPs haplotype was associated with decreased CAD and AS odds in both GERA Europeans and UKB White British. The minor + 3 SNPs haplotype was associated with decreased CAD and AS odds only in UKB White British. One possibility is that the three variants included in the haplotype analyses accounted for essentially all of the effect associated with the minor haplotype. This idea is strengthened by the fact that there were three times more CAD cases in GERA Europeans than AS in UKB White British, indicating that power was not an issue.

Rs140570886 was associated with CAD, but not AS, in both GERA Europeans and the UKB White British; this effect was found only when rs140570886 was on the major haplotypic background. However, it appears that rs140570886 was associated with CAD regardless of haplotypic background in White British individuals, but this association did not reach statistical significance. Most likely, there was insufficient power to identify an association between rs140570886 and AS because there were many fewer AS cases than CAD cases in both the GERA and UKB. A previous study found that rs140570886 showed a strong association with CAD[40] and the effect was modulated through interactions with other *LPA* variants, specifically rs9458001 and rs1800769. Although that study did not use haplotypes, interactions between variants could identify important haplotypic backgrounds. A haplotype analysis could have been more informative as one can account for multiple variants in a single analysis.

A limitation of our work is the reduced power to detect effects in the much smaller non-European populations in all the cohorts. In the future, other cohorts with larger sample sizes for non-European ethnic groups should be analyzed. For example, the Million Veteran Program (MVP)[41] has at least ten times more individuals of African descent than the UKB, so the MVP may improve the study of individuals of African descent.[42] Another consideration for increasing power in future work will be to increase the number of cases for certain diseases/phenotypes, such as. Despite being unable to show an association between AS and the *LPA* haplotypes in non-European ethnicities, we believe that these haplotypes are associated with CAD, AS, and other CVDs, due to their consistent association with Lp(a) levels across many ethnicities. In addition, the effect sizes for CAD in GERA African Americans and East Asians were more extreme than those found in Europeans. Another limitation of our work is that all haplotype analyses were performed using the expectation maximization algorithm as implemented in PLINK 1.07. The specific approach used in PLINK 1.07 may have some issues, such as lower accuracy and imputation efficacy, that were identified in a publication comparing haplotyping software.[43] However, we confirmed the presence of haplotypes using 1000 Genomes Europeans in LDLink, which uses haplotypes created with Bayesian algorithms, machine learning models, and probability models.[44,45] But given the high haplotype frequencies and conservation of the two *LPA* haplotypes, imputation accuracy is likely not a problem. Another limitation of the present study is that the haplotypes only flanked the KIV_2 region but did not include SNPs occurring within them. A notable strength of the present work is the use of haplotypic analyses for the *LPA* region, which allowed more nuanced assessment of genetic associations in *LPA* than single SNP analyses. Another key strength of this paper is the inclusion of multiple ethnicities. There were 367,974 individuals from 11 different ethnicities included in the present study, with

36 sub-ethnicities composing the 6 HGDP ethnicities, which allowed for reliable assessment of the conservation of the *LPA* haplotypes, as well as replication. An additional strength of our work was the assessment of censored Lp(a) concentrations in place of missing Lp(a) values. This is important as missing UKB Lp(a) concentrations may affect analysis outcomes, as there were ~93,850 individuals who had missing Lp(a) concentrations. There is a high probability that the lack of these values skew UKB Lp(a) effect sizes and significance values.

This work has used a haplotype analysis of the *LPA* region to identify two highly conserved haplotypes associated with Lp(a) concentrations and CVD. Such analyses allowed us to demonstrate that SNPs are likely just markers for other causal variation, potentially the KIV_2 repeats. Future efforts should use larger cohorts to better identify potentially diagnostic haplotypes in a clinical setting, to understand human evolution, and to assess if *LPA* haplotypes are associated with specific KIV_2 repeat numbers.

Tables and Figures:

Table 1. Association of *LPA* Haplotypes with Lp(a) Concentrations in Select UKB Ethnicities

Ethnicity	**Haplotype**	**Haplotype Frequency**	**Beta**	**P-Value**
W. British	GTCGTTGTCGCATGCCTATGCGTGGTCAAGCTTTGT	0.61	0.22	$< 1 \times 10^{-315}$
	TCGCCAAGTATGCATGCGCATCGAACGTGATCGGAC	0.38	-0.24	$< 1 \times 10^{-315}$
Chinese	GTCGTTGTCGCATGCCTATGCGTGGTCAAGCTTTGT	0.60	0.24	2.55×10^{-9}
	TCGCCAAGTATGCATGCGCATCGAACGTGATCGGAC	0.39	-0.24	4.10×10^{-9}
S. Asian	GTCGTTGTCGCATGCCTATGCGTGGTCAAGCTTTGT	0.59	0.28	1.28×10^{-51}
	TCGCCAAGTATGCATGCGCATCGAACGTGATCGGAC	0.39	-0.28	2.49×10^{-54}

The major haplotype, colored in black, and the minor haplotype, colored in red, are shown for each ethnicity. Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs9456552 to rs1321196. Only haplotypes that had a frequency of at least 0.01 are included in the table. Betas are calculated for the haplotype versus all others. Covariates included were age, age^2, sex, genotype batch, and 20 principal components. Analyses included 291,547 White British, 1,254 Chinese, and 6,081 South Asian individuals. Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 2. Association of *LPA* Haplotypes with Lp(a) Concentrations in UKB Blacks

Haplotype Name	Haplotype	Haplotype Frequency	Beta	P-Value
Major	GTCGTTGTCGCATGCCTATGCGTGGTCAAGCTTTGT	0.36	0.02	0.13
Haplotype 1	TCGCTAAGTATGCACCCGCACCGAACGTGATCGGAC	0.19	0.01	0.70
Haplotype 2	TCGCTAAGTATGCACCCGCATCGAACGTGATCGGAC	0.18	0.15	1.49×10^{-13}
Haplotype 3	GTCGTTGGCGCATGCCTGTGCGTGGTCAAGCCTTGT	0.06	0.15	1.35×10^{-5}
Minor	TCGCCAAGTATGCATGCGCATCGAACGTGATCGGAC	0.06	-0.28	2.51×10^{-16}
Haplotype 4	TCGCCAAGTATGCATGCGCACCGAACGTGATCGGAT	0.04	-0.35	2.31×10^{-17}
Haplotype 5	GTCGTTGGCGCATGCCTGTGCGTGATCAAGTCTTGT	0.03	0.05	0.24
Haplotype 6	GTCGTTGGCGCGTGCCTGTACGTGATCAAGTCTTAT	0.02	-0.18	1.08×10^{-3}
Haplotype 7	GTCGTTGGCGCGTGCCTGTGCGTGATCAAGTCTTAT	0.02	-0.24	9.29×10^{-6}
Haplotype 8	GTCGTTGGCGCATGCCTGTACGTGGTCAAGCCTTGT	0.01	0.22	0.02

Major and minor are according to UKB non-Black haplotypes. The haplotype in row one is the major haplotype and the haplotype in row 5 is the minor haplotype. The haplotypes span the *LPA* gene from variants rs9456552 to rs1321196. Only haplotypes with a frequency of at least 0.01 are included. Betas are calculated for the haplotype versus all others. Covariates included were age, age^2, sex, genotype batch, and 20 principal components. Analyses included 5,712 Black individuals. Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 3. Association of *LPA* Haplotypes with Lp(a) Concentrations in Select MESA Ethnicities

Ethnicity	Haplotype	Haplotype Frequency	Haplotype Count	Beta	P-Value
European	Major	0.62	1,991	0.07	0.06
	Minor	0.37	1,185	-0.10	1.26×10^{-2}
Latino	Major	0.63	1,182	-0.20	8.83×10^{-5}
	Minor	0.25	468	-0.13	0.03
Chinese	Major	0.57	887	0.28	7.56×10^{-7}
	Minor	0.41	638	-0.27	1.76×10^{-6}

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs9456552 to rs1321196. Only haplotypes with a frequency of at least 0.01 are included in the table. Betas are calculated for the haplotype versus all others. Covariates included were age and sex. Analyses included 1,578 European, 934 Latino, and 777 Chinese individuals. Lp(a): Lipoprotein(a); MESA: Multi-Ethnic Study of Atherosclerosis; UKB: UK Biobank.

Table 4. Association of *LPA* Haplotypes with Lp(a) Concentrations in MESA African Americans

Haplotype	Haplotype Frequency	Haplotype Count	Beta	P-Value
Major	0.39	866	-0.13	4.52×10^{-4}
Haplotype 2	0.16	368	0.41	9.31×10^{-16}
Haplotype 1	0.15	345	0.12	0.02
Minor	0.09	209	-0.48	1.36×10^{-13}
Haplotype 3	0.04	96	-0.09	0.35
Haplotype 4	0.04	80	-0.34	3.41×10^{-4}
Haplotype 5	0.04	79	0.44	3.58×10^{-6}
Haplotype 6	0.02	52	-0.16	0.22
Haplotype 7	0.02	35	-0.18	0.24

Major and minor are according to UKB non-Black haplotypes. Haplotype numbering is according to UKB Blacks. The haplotypes span the *LPA* gene from rs9456552 to rs1321196. Only haplotypes with a frequency of at least 0.01 are included in the table. Betas are calculated for the haplotype versus all others. Covariates included were age and sex. Analyses included 1,123 African American individuals. Lp(a): Lipoprotein(a); MESA: Multi-Ethnic Study of Atherosclerosis; UKB: UK Biobank.

Table 5. UKB *LPA* Haplotypes with Variants-Associations with Lp(a)

Ethnicity	Haplotype	Haplotype Frequency	Beta	Conditional Beta	P-Value
White British					
	Major + 3 SNP Common Alleles	0.54	-0.24	**Ref.**	$< 1 \times 10^{-315}$
	Minor + 3 SNP Common Alleles	0.38	-0.24	-0.04	$< 1 \times 10^{-315}$
	Major + rs10455872	0.07	1.72	1.72	$< 1 \times 10^{-315}$
	Major + rs3798220 + rs140570886	4.61×10^{-3}	1.38	1.56	$< 1 \times 10^{-315}$
	Minor + rs3798220	2.36×10^{-3}	-0.35	-0.26	3.42×10^{-27}
	Major + rs3798220	1.06×10^{-3}	-0.04	0.18	0.53
	Major + rs140570886	4.72×10^{-4}	1.54	1.75	7.85×10^{-109}
	Minor + rs10455872	1.09×10^{-4}	0.04	-0.09	0.79
	Minor + rs140570886	2.47×10^{-5}	1.01	0.15	0.01
South Asian					
	Major + 3 SNP Common Alleles	0.59	0.24	**Ref.**	2.78×10^{-39}
	Minor + 3 SNP Common Alleles	0.39	-0.29	-0.27	7.09×10^{-56}
	Major + rs10455872	0.01	1.04	0.91	1.39×10^{-24}
	Major + rs3798220	1.65×10^{-3}	0.09	-0.27	0.74
	Major + rs3789220 + rs140570886	1.50×10^{-3}	0.60	0.55	0.02
	Minor + rs3798220	1.20×10^{-3}	-0.05	0.04	0.89
	Major + rs140570886	5.01×10^{-4}	0.11	0.11	0.79
Chinese					
	Major + 3 SNP Common Alleles	0.52	0.26	**Ref.**	3.48×10^{-10}
	Minor + 3 SNP Common Alleles	0.38	-0.24	-0.27	8.17×10^{-9}
	Major + rs3798220	0.06	0.02	-0.08	0.79
	Major + rs3798220 + rs140570886	0.02	-0.41	-0.50	0.01
	Minor + rs3798220	0.01	-0.25	-0.25	0.30
	Major + rs140570886	3×10^{-3}	0.17	0.09	0.65
Black					
	Major + 3 SNP Common Alleles	0.36	8.00×10^{-3}	**Ref.**	0.62
	Haplotype 2 + 3 SNP Common Alleles	0.18	0.15	0.12	6.17×10^{-14}
	Haplotype 1 + 3 SNP Common Alleles	0.18	0.02	7.50×10^{-3}	0.43
	Minor + 3 SNP Common Alleles	0.06	-0.28	-0.26	5.27×10^{-17}
	Haplotype 3 + 3 SNP Common Alleles	0.06	0.15	0.12	6.19×10^{-6}
	Haplotype 4 + 3 SNP Common Alleles	0.04	-0.34	-0.34	2.66×10^{-16}
	Haplotype 5 + 3 SNP Common Alleles	0.03	0.05	0.05	0.26
	Haplotype 6 + 3 SNP Common Alleles	0.02	-0.17	-0.16	1.84×10^{-3}
	Haplotype 7 + 3 SNP Common Alleles	0.02	-0.29	-0.22	5.87×10^{-7}
	Haplotype 8 + 3 SNP Common Alleles	0.01	0.24	0.098	8.34×10^{-3}
	Major + rs10455872	4.93×10^{-3}	0.80	0.73	9.93×10^{-13}
	Major + rs3798220 + rs140570886	1.38×10^{-3}	-0.19	-0.33	0.39
	Major + rs140570886	2.78×10^{-4}	-0.72	-0.81	0.12

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Variants included in haplotypes are rs3798220,

(*Table 5* cont'd) rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the variants. Covariates included were age, age^2, sex, genotype batch, and 20 principal components. Betas are calculated for the haplotype versus all others. Conditional betas are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term "Ref.". Analyses included 290,819 White British, 6,081 South Asian, 1,254 Chinese, and 5,712 Black individuals. Lp(a): Lipoprotein(a); SNP: Single Nucleotide Polymorphism; UKB: UK Biobank.

Table 6. MESA *LPA* Haplotypes with Variants-Associations with Lp(a)

Ethnicity	Haplotype	Haplotype Frequency	Beta	Conditional Beta	P-Value
European					
	Major + 3 SNP Common Alleles	0.56	-0.17	**Ref.**	7.46×10^{-6}
	Minor + 3 SNP Common Alleles	0.37	-0.11	4.57×10^{-3}	4.71×10^{-3}
	Major + rs10455872	0.04	1.36	1.39	1.09×10^{-37}
	Major + rs3798220 + rs140570886	0.01	1.84	1.87	6.48×10^{-22}
	Minor + rs3798220	2.21×10^{-3}	-0.73	-0.57	0.10
African American					
	Major + 3 SNP Common Alleles	0.37	-0.11	**Ref.**	0.03
	Minor + 3 SNP Common Alleles	0.09	-0.44	-0.30	6.48×10^{-8}
	Major + rs10455872	0.01	1.03	1.03	6.84×10^{-6}
	Major + rs3798220 + rs140570886	7.91×10^{-3}	-0.18	-0.1	0.50
Latino					
	Major + 3 SNP Common Alleles	0.42	-0.03	**Ref.**	0.57
	Minor + 3 SNP Common Alleles	0.25	-0.09	-0.10	0.14
	Major + rs3798220 + rs140570886	0.16	-0.41	-0.30	2.05×10^{-9}
	Major + rs10455872	0.02	1.64	1.64	2.12×10^{-18}
	Minor + rs3798220	0.02	-0.11	-0.20	0.64
	Major + rs140570886	0.01	-0.13	-0.20	0.54
	Minor + rs3798220	6.05×10^{-3}	0.07	0.40	0.88

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Variants included in haplotype are rs3798220, rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the variants. Betas are calculated for the haplotype versus all others. Conditional betas are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term “Ref.”. Covariates included were age and sex. Analyses included 1,619 European, 1,123 African American, and 933 Latino individuals. Lp(a): Lipoprotein(a); MESA: Multi-Ethnic Study of Atherosclerosis; SNP: Single Nucleotide Polymorphism; UKB: UK Biobank.

Table 7. UKB White British *LPA* Region Lp(a) GWAS Hits Genotypic Missingness

rsID	Homozygote Minor Count	Homozygote Minor: % Missing Lp(a)	Heterozygote: % Missing Lp(a)	Conditional Beta
rs57175128	72	36.11	26.11	0.03
rs114284743	2	0.00	26.77	0.06
rs558685607	13	30.77	23.70	0.02
rs544366796	37	51.35	33.48	1.58
rs10455872	2492	78.25	29.02	1.82
rs140570886	65	95.38	74.80	1.78
rs73596816	411	33.33	19.62	1.17
rs369686024	17	11.76	12.91	0.80
rs528521448	16	18.75	18.22	1.32
rs150776685	22	22.73	21.96	1.81
rs193210914	25	48.00	30.73	0.03

The variants in this table were prioritized in an Lp(a) GWAS. They are located in the *LPA* region, and had a mean homozygote minor Lp(a) concentration ≥ 70 nmol/L in UKB White British individuals. The Conditional Beta column specifies the effect size of each variant in the context of a conditional haplotype analysis that contained all 11 Lp(a) GWAS variants plus rs3798220. Missingness analyses included 384,669 UKB White British individuals. Conditional haplotype analyses included 288,632 UKB White British individuals. GWAS: Genome-Wide Association Study; Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 8. UKB White British *LPA* Region Lp(a) GWAS Variants Censored Lp(a) concentrations

rsID	Minor Allele Counts	Homozygote Minor Mean Lp(a)	Heterozygote Mean Lp(a)	D’ with Rs10455872	R^2 with Rs10455872
rs57175128	144	99.31	72.00	0.31	0.01
rs114284743	4	158.05	41.16	0.65	1.05×10^{-4}
rs558685607	26	93.97	55.00	0.16	$1.97 \times 10^{--3}$
rs544366796	74	137.39	122.24	0.58	2.96×10^{-4}
rs10455872	4,984	161.89	124.44	Reference	Reference
rs140570886	130	165.97	155.85	1	1.33×10^{-3}
rs73596816	822	110.82	79.87	0.98	2.92×10^{-3}
rs369686024	34	64.11	54.54	1	6.75×10^{-4}
rs528521448	32	104.46	78.32	0.96	5.75×10^{-4}
rs150776685	44	93.35	64.93	0.21	4.95×10^{-3}
rs193210914	50	106.18	71.01	0.25	5.49×10^{-3}

The variants in this table were prioritized in an Lp(a) GWAS. They are located in the *LPA* region, and had a mean homozygote minor Lp(a) concentration ≥ 70 nmol/L in UKB White British individuals. Analyses included 384,669 UKB White British individuals. GWAS: Genome-Wide Association Study; Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 9. UKB White British *LPA* Null Variants Censored Lp(a) Concentrations

rsID	Minor Allele Count	Homozygote Minor Mean Lp(a)	Heterozygote Mean Lp(a)	Homozygote Minor: % Missing Lp(a)	Heterozygote: % Missing Lp(a)
rs201306475	0	NA	27.30	NA	44.44
rs139145675	0	NA	26.33	NA	31.58
rs41267811	0	NA	28.26	NA	49.47
rs143431368	8	3.80	27.90	100.00	39.80
rs41267813	0	NA	31.73	NA	24.84
rs41272114	1,192	3.85	27.58	98.32	41.73
rs201297680	0	NA	9.96	NA	60.00
rs41259144	86	3.82	27.60	97.67	40.49
rs1623955	0	NA	21.20	NA	50.00

Variants included in this table are null variants in UKB. All non-percentage values refer to censored Lp(a) concentrations. Individuals without these variants had an average Lp(a) concentration of 53.04 nmol/L. Calculations included 384,669 White British individuals. Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 10. UKB Black *LPA* Region Null Variants Missingness

rsID	Minor Allele Counts	Heterozygote Missing: % Missing Lp(a)
rs139145675	112.34	16.67
rs143431368	14.00	28.57
rs41272114	56.00	10.71
rs41259144	14.00	21.43
rs1623955	25.27	14.29

Variants included in this table are null variants in UKB.

Calculations included 7,261 Black individuals.

Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 11. UKB South Asian *LPA* Region Null Variants Missingness

rsID	Minor Allele Counts	Homozygote Minor: % Missing Lp(a)	Heterozygote Missing: % Missing Lp(a)
rs139145675	29.99	NA	7.69
rs143431368	7.00	NA	14.29
rs41272114	473.00	88.89	30.33
rs41259144	16.99	NA	41.18

Variants included in this table are null variants in UKB. Calculations included 7,319 South Asian individuals. Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table 12. UKB White British *LPA* Haplotypes Haplotypic Associations with Coronary Artery Disease

Haplotype	Haplotype Frequency	Haplotype Count	Odds Ratio	Conditional Odds Ratio	P-Value
Major + 3 SNP Common Alleles	0.53	260,703	0.92	Ref.	1.12×10^{-21}
Minor + 3 SNP Common Alleles	0.36	175,925	0.96	1.01	5.04×10^{-6}
Major + rs10455872	0.08	38,713	1.33	1.35	2.85×10^{-80}
Major + rs3798220 + rs140570886	0.01	7,056	1.54	1.60	1.14×10^{-41}
Minor + rs3798220	2.73×10^{-3}	1,337	1.12	1.14	0.19
Major + rs3798220	1.04×10^{-3}	509	1.19	1.16	0.28
Major + rs140570886	9.29×10^{-4}	455	1.41	1.50	7.82×10^{-3}

Major and minor are according to UKB non-Black haplotypes. Variants included in haplotype are rs3798220, rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the three variants. Odds ratios are calculated for the haplotype versus all others. Conditional odds ratios are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term "Ref.". Covariates included were age, age^2, sex, genotype batch, and 20 principal components. Analyses included 247,532 White British individuals (32,867 cases). SNP: Single Nucleotide Polymorphism; UKB: UK Biobank

Table 13. UKB White British *LPA* Haplotypes Haplotypic Associations with Aortic Stenosis

Haplotype	Haplotype Frequency	Haplotype Count	Odds Ratio	Conditional Odds Ratio	P-Value
Major + 3 SNP Common Alleles	0.53	259,875	0.87	Ref.	5.30×10^{-9}
Minor + 3 SNP Common Alleles	0.36	175,367	0.93	1.03	8.34×10^{-3}
Major + rs10455872	0.08	38,590	1.55	1.59	1.11×10^{-29}
Major + rs3798220 + rs140570886	0.01	7,034	1.82	1.92	3.91×10^{-14}
Minor + rs3798220	2.73×10^{-3}	1,333	0.81	0.79	0.44
Major + rs3798220	1.04×10^{-3}	508	1.75	1.86	0.13
Major + rs140570886	9.30×10^{-4}	454	1.09	1.16	0.82

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Variants included in haplotype are rs3798220, rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the three variants. Covariates included were age, age^2, sex, genotype batch, and 20 principal components. Odds ratios are calculated as one haplotype versus all. Conditional odds ratios are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term “Ref.”. Analyses included 239,987 White British individuals (3,443 cases). SNP: Single Nucleotide Polymorphism; UKB: UK Biobank.

Table 14. GERA European *LPA* Haplotypes Haplotypic Associations with Coronary Artery Disease

Haplotype	Haplotype Frequency	Haplotype Count	Odds Ratio	Conditional Odds Ratio	P-Value
Major + 3 SNP Common Alleles	0.54	59,925	0.94	Ref.	2.75×10^{-5}
Minor + 3 SNP Common Alleles	0.36	39,950	0.99	1.03	0.52
Major + rs10455872	0.06	7,140	1.22	1.24	1.23×10^{-11}
Major + rs3798220 + rs140570886	0.01	1,390	1.31	1.37	2.78×10^{-5}
Minor + rs3798220	2.73×10^{-3}	301	0.77	0.79	0.11
Major + rs140570886	1.52×10^{-3}	167	1.37	1.39	0.09
Major + rs3798220	1.14×10^{-3}	125	0.98	1.10	0.95

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Variants included in haplotype are rs3798220, rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the three variants. Odds ratios are calculated as one haplotype versus all. Conditional odds ratios are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term “Ref.”. Covariates included were sex, age, age^2, and ten European-specific principal components. Analyses included 55,180 individuals (12,883 cases). GERA: Genetic Epidemiology Research on Adult Health and Aging; SNP: Single Nucleotide Polymorphism; UKB: UK Biobank.

Table 15. GERA European *LPA* Haplotypes Haplotypic Associations with Aortic Stenosis

Haplotype	Haplotype Frequency	Haplotype Count	Odds Ratio	Conditional Odds Ratio	P-Value
Major + 3 SNP Common Alleles	0.54	59,925	0.92	Ref.	1.49×10^{-3}
Minor + 3 SNP Common Alleles	0.36	39,950	0.98	1.03	0.45
Major + rs10455872	0.06	7,140	1.37	1.39	4.05×10^{-11}
Major + rs3798220 + rs140570886	0.01	1,390	1.40	1.47	8.75×10^{-4}
Minor + rs3798220	2.73×10^{-3}	301	0.76	0.80	0.36
Major + rs140570886	1.52×10^{-3}	167	1.08	1.14	0.81
Major + rs3798220	1.14×10^{-3}	125	1.82	1.51	0.13

Major and minor are according to UKB non-Black haplotypes. The haplotypes span the *LPA* gene from rs3798220 to rs1321196. Variants included in haplotype are rs3798220, rs10455872, and rs140570886, and in this table refer to the presence of the rare allele. Haplotypes included in the table have a frequency of at least 0.01 or contain one of the minor alleles of the three variants. Odds ratios are calculated as one haplotype versus all. Conditional odds ratios are calculated with the major + 3 SNPs haplotype as the reference, as indicated by the term “Ref.”. Covariates included were sex, age, age^2, and ten European-specific principal components. Analyses included 55,180 European individuals (3,469 cases). GERA: Genetic Epidemiology Research on Adult Health and Aging; SNP: Single Nucleotide Polymorphism: UKB: UK Biobank.

Table 16. HGDP *LPA* Region Haplotype Frequencies

	Haplotype	European	South Asian	West Asian	East Asian	Native American	Sub-Saharan African
N		160	125	146	142	108	121
Major	CGCA	0.61	0.59	0.64	0.53	0.76	0.34
Minor	TATG	0.38	0.41	0.35	0.47	0.24	0.57
% Yin Yang		>99%	100%	>99%	100%	100%	<92%

The four variants in the HGDP haplotypes were drawn from the 36 variants that compose the major and minor haplotypes that are predominant in the *LPA* region in non-Black ethnicities of the UKB. Variants included in the haplotype are rs9365179, rs10945682, rs1740428, and rs1321196. HGDP: Human Genome Diversity Project; UKB: UK Biobank.

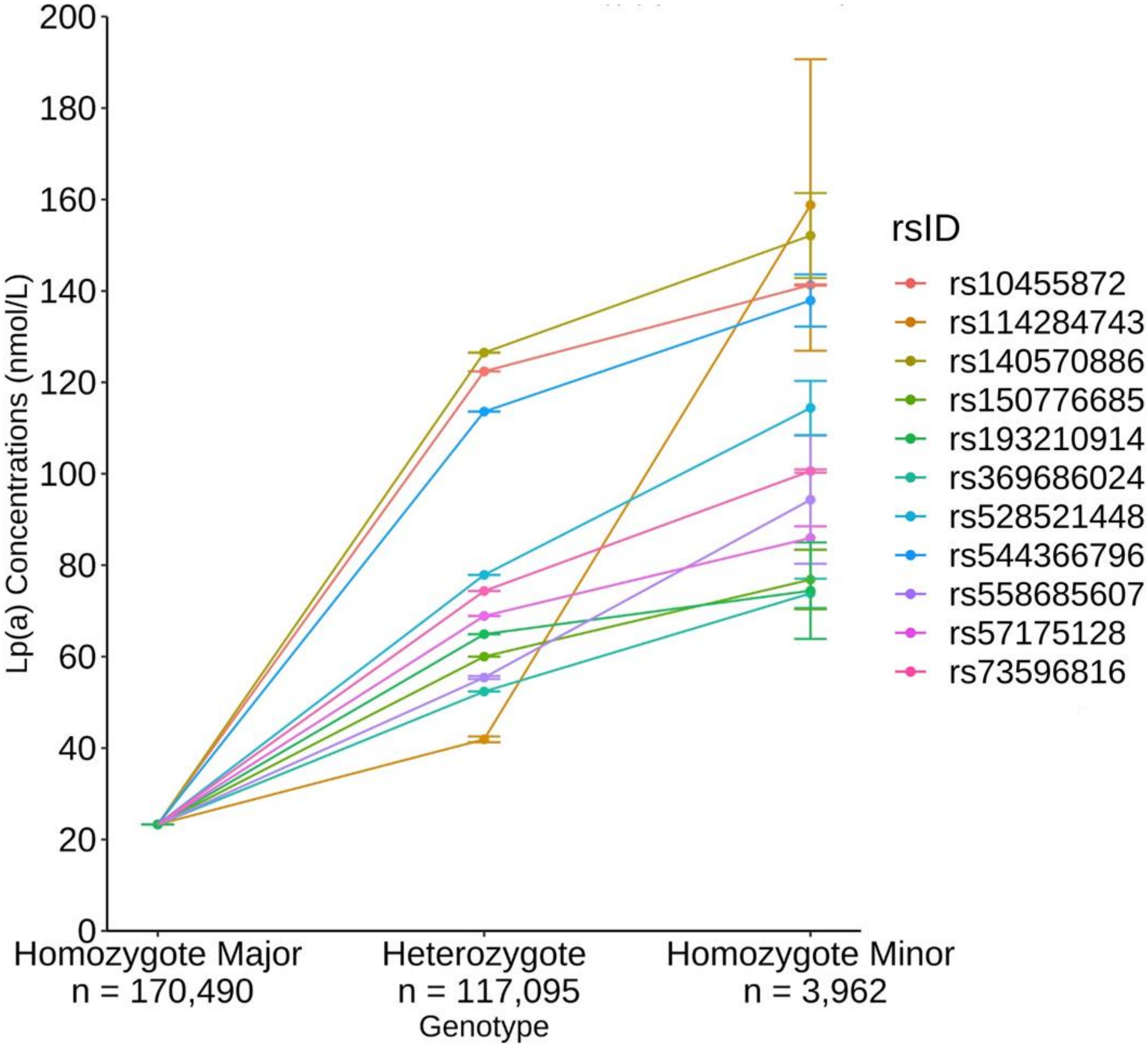


Figure 1. Average Lp(a) Concentrations (nmol/L) by Genotype in UKB White British Individuals. Lp(a) GWAS variants included have homozygote minor individuals with an average Lp(a) concentration of at least 70 nmol/L. The point for homozygote major individuals is an average of all plotted variants. Heterozygote and homozygote minor individuals were homozygous major for all other variants except the one plotted for each point. Error bars represent confidence intervals. Calculations included 291,547 UKB White British individuals. GWAS: Genome-Wide Association Study; Lp(a): Lipoprotein(a); UKB: UK Biobank.

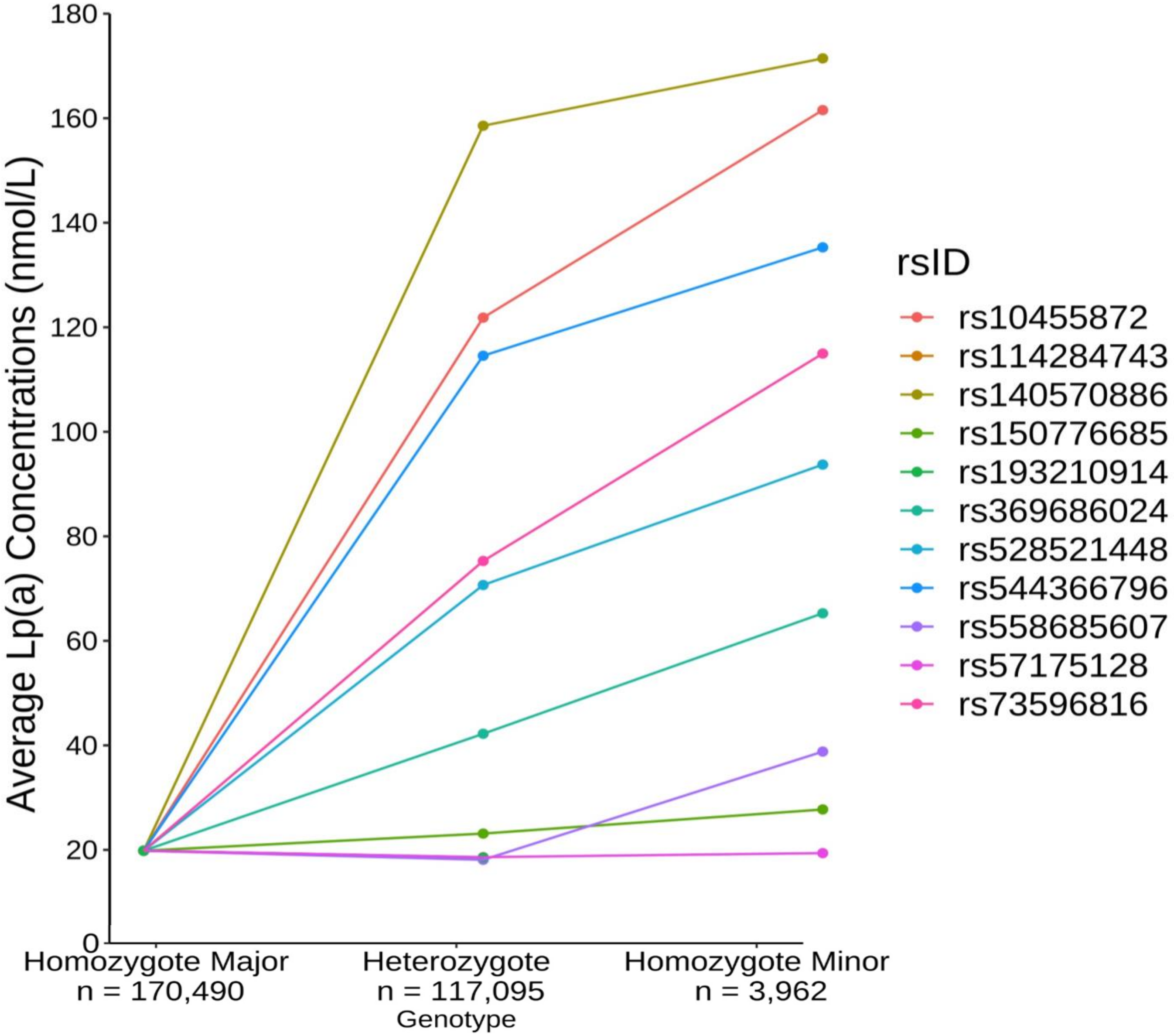


Figure 2. Average Haplotypic Lp(a) Concentrations (nmol/L) in UKB White British Individuals. Lp(a) GWAS variants included have homozygote minor individuals with an average Lp(a) concentration of at least 70 nmol/L. The point for homozygote major individuals is an average of all plotted variants. Heterozygote and homozygote minor individuals were homozygous major for all other variants except the one plotted for each point. Calculations included 291,547 UKB White British individuals. GWAS: Genome-Wide Association Study; Lp(a): Lipoprotein(a); UKB: UK Biobank.

Reference List

1. Boffa MB, Koschinsky ML. Oxidized phospholipids as a unifying theory for lipoprotein(a) and cardiovascular disease. *Nat Rev Cardiol.* 2019;16(5):305-318.
2. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med.* 2009;361(26):2518-2528.
3. Rader DJ, Cain W, Zech LA, Usher D, Brewer HB, Jr. Variation in lipoprotein(a) concentrations among individuals with the same apolipoprotein (a) isoform is determined by the rate of lipoprotein(a) production. *J Clin Invest.* 1993;91(2):443-447.
4. 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.
5. Saleheen D, Haycock PC, Zhao W, et al. Apolipoprotein(a) isoform size, lipoprotein(a) concentration, and coronary artery disease: a mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* 2017;5(7):524-533.
6. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA.* 2009;301(22):2331-2339.
7. Kamstrup PR, Tybjaerg-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.
8. Thanassoulis G, Campbell CY, Owens DS, et al. Genetic associations with valvular calcification and aortic stenosis. *N Engl J Med.* 2013;368(6):503-512.
9. Helgadottir A, Thorleifsson G, Gretarsdottir S, et al. Genome-wide analysis yields new loci associating with aortic valve stenosis. *Nat Commun.* 2018;9(1):987.

10. Hoekstra M, Chen HY, Rong J, et al. Genome-wide association study highlights *APOH* as a novel locus for lipoprotein(a) levels-brief report. *Arterioscler Thromb Vasc Biol.* 2021;41(1):458-464.

11. 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;3(1):39-46.

12. Paré G, Çaku A, McQueen M, et al. Lipoprotein(a) levels and the risk of myocardial infarction among 7 ethnic groups. *Circulation.* 2019;139(12):1472-1482.

13. Tsimikas S, Clopton P, Brilakis ES, et al. Relationship of oxidized phospholipids on apolipoprotein B-100 particles to race/ethnicity, apolipoprotein(a) isoform size, and cardiovascular risk factors: results from the Dallas Heart Study. *Circulation.* 2009;119(13):1711-1719.

14. Marcovina SM, Albers JJ, Wijsman E, Zhang Z, Chapman NH, Kennedy H. Differences in Lp[a] concentrations and apo[a] polymorphs between black and white Americans. *J Lipid Res.* 1996;37(12):2569-2585.

15. Paultre F, Pearson TA, Weil HF, et al. High levels of Lp(a) with a small apo(a) isoform are associated with coronary artery disease in African American and white men. *Arterioscler Thromb Vasc Biol.* 2000;20(12):2619-2624.

16. 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;9(1):2606.

17. Schmidt K, Noureen A, Kronenberg F, Utermann G. Structure, function, and genetics of lipoprotein (a). *J Lipid Res.* 2016;57(8):1339-1359.

18. Steffen BT, Thanassoulis G, Duprez D, et al. Race-based differences in lipoprotein(a)-associated risk of carotid atherosclerosis. *Arterioscler Thromb Vasc Biol.* 2019;39(3):523-529.

19. Patel AP, Wang M, Pirruccello JP, et al. Lp(a) (Lipoprotein[a]) concentrations and incident atherosclerotic cardiovascular disease: new insights from a large national biobank. *Arterioscler Thromb Vasc Biol.* 2021;41(1):465-474.

20. Lee SR, Prasad A, Choi YS, et al. *LPA* gene, ethnicity, and cardiovascular events. *Circulation.* 2017;135(3):251-263.

21. Sulovari A, Li R, Audano PA, et al. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc Natl Acad Sci U S A.* 2019;116(46):23243-23253.

22. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.

23. Olson JL, Bild DE, Kronmal RA, Burke GL. Legacy of MESA. *Glob Heart.* 2016;11(3):269-274.

24. Banda Y, Kvale MN, Hoffmann TJ, et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics.* 2015;200(4):1285-1295.

25. Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science.* 2002;296(5566):261-262.

26. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81(3):559-575.

27. Mukamel RE, Handsaker RE, Sherman MA, et al. Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. *bioRxiv.* 2021;2021.01.19.427332.

28. Morgan BM, Brown AN, Deo N, et al. Nonsynonymous SNPs in *LPA* homologous to plasminogen deficiency mutants represent novel null apo(a) alleles. *J Lipid Res.* 2020;61(3):432-444.

29. Wang Z, Chen H, Bartz TM, et al. Role of rare and low-frequency variants in gene-alcohol interactions on plasma lipid levels. *Circ Genom Precis Med.* 2020;13(4):e002772.

30. Lim ET, Würtz P, Havulinna AS, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* 2014;10(7):e1004494.

31. Kyriakou T, Seedorf U, Goel A, et al. A common *LPA* null allele associates with lower lipoprotein(a) levels and coronary artery disease risk. *Arterioscler Thromb Vasc Biol.* 2014;34(9):2095-2099.

32. Templeton AR. Out of Africa? What do genes tell us? *Curr Opin Genet Dev.* 1997;7(6):841-847.

33. Stern MC, Fejerman L, Das R, et al. Variability in cancer risk and outcomes within US Latinos by national origin and genetic ancestry. *Curr Epidemiol Rep.* 2016;3:181-190.

34. Langsted A, Varbo A, Kamstrup PR, Nordestgaard BG. Elevated lipoprotein(a) does not cause low-grade inflammation despite causal association with aortic valve stenosis and

myocardial infarction: a study of 100,578 individuals from the general population. *J Clin Endocrinol Metab.* 2015;100(7):2690-2699.

35. 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.
36. Lu W, Cheng YC, Chen K, et al. Evidence for several independent genetic variants affecting lipoprotein (a) cholesterol levels. *Hum Mol Genet.* 2015;24(8):2390-2400.
37. Luke MM, Kane JP, Liu DM, et al. A polymorphism in the protease-like domain of apolipoprotein(a) is associated with severe coronary artery disease. *Arterioscler Thromb Vasc Biol.* 2007;27(9):2030-2036.
38. Ogorelkova M, Gruber A, Utermann G. Molecular basis of congenital lp(a) deficiency: a frequent apo(a) 'null' mutation in caucasians. *Hum Mol Genet.* 1999;8(11):2087-2096.
39. Stefanutti C, Pisciotta L, Favari E, et al. Lipoprotein(a) concentration, genetic variants, apo(a) isoform size, and cellular cholesterol efflux in patients with elevated Lp(a) and coronary heart disease submitted or not to lipoprotein apheresis: An Italian case-control multicenter study on Lp(a). *J Clin Lipidol.* 2020;14(4):487-497.e481.
40. Zeng L, Moser S, Mirza-Schreiber N, et al. Cis-epistasis at the *LPA* locus and risk of cardiovascular diseases. *Cardiovasc Res.* 2021;cvab136.
41. 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;70:214-223.
42. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet.* 2018;50(11):1514-1523.
43. Nothnagel M, Ellinghaus D, Schreiber S, Krawczak M, Franke A. A comprehensive evaluation of SNP genotype imputation. *Hum Genet.* 2009;125(2):163-171.

44. 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(21):3555-3557.

45. Abecasis GR, Auton A, Brooks LD, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature.* 2012;491(7422):56-65.

Supplemental Tables:

Table S1: Sample Sizes for Analyzed HGDP Populations

Composite Ethnicity	Population	N
European		**160**
	Adygei	17
	Basque	24
	French	29
	Italian	13
	Orcadin	16
	Russian	25
	Sardinian	28
	Tuscan	8
South Asians		**125**
	Brahui	25
	Burusho	25
	Kalash	25
	Makrani	25
	Sindhi	25
West Asians		**146**
	Bedouin	48
	Druze	47
	Palestinian	51
East Asians		**142**
	Daur	9
	Han	44
	Japanese	29
	Miao	10
	Oroqen	10
	She	10
	Tu	10
	Tujia	10
	Yi	10
Native American		**108**
	Colombian	13
	Karitiana	24
	Maya	25
	Pima	25
	Surui	21
Sub-Saharan Africans		**121**
	Bantu	20
	Biaka	32
	Mandenka	24
	Mbuti	15
	San	6
	Yoruba	24

(*Table S1* Cont'd) This table contains the name and sample size of each population assessed in HGDP analyses. The analyses included 802 individuals. HGDP: Human Genome Diversity Project.

Table S2: *LPA* Region Haplotype SNPs

SNP rsID	MAF	R^2 with rs6902102
rs9456552	0.36	1.00
rs6913833	0.36	1.00
rs7771129	0.36	1.00
rs7752408	0.36	1.00
rs13192132	0.36	1.00
rs6455691	0.36	1.00
rs6455692	0.36	1.00
rs6415085	0.37	0.99
rs9365179	0.36	1.00
rs6902102	0.36	Reference
rs6455693	0.36	1.00
rs6932014	0.36	1.00
rs6455695	0.36	1.00
rs6937879	0.36	1.00
rs9355814	0.36	1.00
rs9355815	0.36	1.00
rs6929299	0.36	1.00
rs7761377	0.37	0.98
rs3798224	0.36	1.00
rs62441718	0.37	0.99
rs36115561	0.36	1.00
rs6905422	0.36	1.00
rs6933576	0.36	1.00
rs9295130	0.36	1.00
rs10945682	0.36	1.00
rs10945683	0.36	1.00
rs7760585	0.36	1.00
rs1569933	0.36	1.00
rs2983236	0.36	1.00
rs2144726	0.37	0.98
rs1950564	0.36	1.00
rs1950563	0.37	0.99
rs2092931	0.36	1.00
rs2872764	0.36	1.00
rs1740428	0.36	1.00
rs1321196	0.36	1.00

The variants listed define the haplotypes in the *LPA* gene that are associated with lipoprotein(a) concentrations. The included SNPs were identified in 291,547 UKB White British individuals. Rs6902102 is used as a reference for the linkage disequilibrium between the *LPA* region SNPs. MAF: Minor Allele Frequency; SNPs: Single Nucleotide Polymorphisms; UKB: UK Biobank.

Table S3. UKB White British Lp(a) GWAS *LPA* Haplotypes

Haplotype	Haplotype Frequency	Haplotype Count	Beta	P-Value	Variant(s) Present
TACCATGGGAAA	0.48	277664	-0.41	$< 1 \times 10^{-315}$	Major haplotypes
TACCATGGAAAA	0.34	197424	-0.29	$< 1 \times 10^{-315}$	Minor haplotype
TACCGTGGGAAA	0.07	40235	1.72	$< 1 \times 10^{-315}$	rs10455872
TACCATAGGAAA	0.03	17376	0.94	$< 1 \times 10^{-315}$	rs73596816
TACCACGGGAAA	0.01	7678	1.47	$< 1 \times 10^{-315}$	rs140570886
TACGATGGAAAA	8.46×10^{-3}	4884	1.40	$< 1 \times 10^{-315}$	rs544366796 and minor haplotype
TACCATGAGAAA	8.35×10^{-3}	4820	0.62	$< 1 \times 10^{-315}$	rs369686024
CACCATGGGAAA	4.13×10^{-3}	2384	-0.33	2.13×10^{-32}	rs57175128
TACCATGGGTAA	4.01×10^{-3}	2315	1.12	$< 1 \times 10^{-315}$	rs528521448
CACCGTGGGAAA	4.01×10^{-3}	2315	1.91	$< 1 \times 10^{-315}$	rs57175128 and rs10455872
CACCATGGAAAA	3.35×10^{-3}	1934	-0.17	4.31×10^{-8}	rs57175128 and minor haplotype
TACCATGGGAGA	3.33×10^{-3}	1922	-0.06	0.03	rs150776685
TACCATGGAAGA	3.23×10^{-3}	1865	0.00	0.95	rs150776685 and minor haplotype
TACCATGGAAAT	2.80×10^{-3}	1616	-0.16	6.70×10^{-7}	rs193210914 and minor haplotype
TAACATGGGAAA	2.58×10^{-3}	1489	-0.26	8.77×10^{-15}	rs558685607
TACCGTGGGAGA	2.13×10^{-3}	1230	2.09	$< 1 \times 10^{-315}$	rs10455872 and rs150776685
TACCGTGGGAAT	1.89×10^{-3}	1091	2.05	$< 1 \times 10^{-315}$	rs10455872 and rs193210914
TACCATGGGAAT	1.87×10^{-3}	1080	-0.28	7.07×10^{-12}	rs193210914
TAACATGGAAAA	1.67×10^{-3}	964	-0.29	3.46×10^{-11}	rs558685607 and minor haplotype
TACCATAGGTAA	1.59×10^{-3}	918	1.14	5.98×10^{-194}	rs73596816 and rs528521448
TAACGTGGGAAA	1.50×10^{-3}	866	2.17	$< 1 \times 10^{-315}$	rs558685607 and rs10455872
TTCCATGGGAAA	1.23×10^{-3}	710	-0.30	5.48×10^{-9}	rs114284743
TTCCATGGAAAA	1.11×10^{-3}	641	-0.29	3.00×10^{-8}	rs114284743 and minor haplotype
TACCATGGATAA	9.06×10^{-4}	523	1.73	7.75×10^{-140}	rs528521448 and minor haplotype
TACGATGGGAAA	5.32×10^{-4}	307	1.30	5.96×10^{-46}	rs544366796
TACCACAGGAAA	4.34×10^{-4}	250	1.41	1.03×10^{-23}	rs140570886 and rs73596816
TACGGTGGGAAA	3.20×10^{-4}	185	2.55	2.65×10^{-81}	rs544366796 and rs10455872
CACCATAGGAAA	2.99×10^{-4}	173	2.04	9.25×10^{-72}	rs57175158 and rs73596816
TTCCGTGGGAAA	8.34×10^{-5}	48	4.64	1.80×10^{-58}	rs114284743 and rs10455872
TACCATAGAAAA	4.18×10^{-5}	24	-0.16	0.48	rs73596816 and minor haplotype

Major and minor are according to UKB non-Black haplotypes. The variants in this table were prioritized in an Lp(a) GWAS, and are located in the *LPA* region, except for one which is a marker for the *LPA* region yin-yang haplotypes that spans the KIV_2 repeats (rs10945682). The variants in the haplotype are in the following order: rs57175128, rs114284743, rs558685607, rs544366796, rs10455872, rs140570886, rs73596816, rs369686024, rs10945682, rs528521448, rs150776685, and rs193210914. Bases highlighted in red indicate the minor allele of a variant.

(*Table S3* Cont'd) Analyses included 288,632 UKB White British individuals. GWAS: Genome-Wide Association Study; Lp(a): Lipoprotein(a); UKB: UK Biobank.

Table S4. Haplotype Frequencies in Individuals of African Descent

Haplotypes	**UKB Haplotype Frequency**	**MESA Haplotype Frequency**	**GERA Haplotype Frequency**
Major	0.36	0.39	0.37
UKB Haplotype 1	0.19	0.15	0.31*
UKB Haplotype 2	0.18	0.16	
Minor	0.06	0.09	0.12
UKB Haplotype 3	0.06	0.04	0.05
UKB Haplotype 4	0.04	0.04	0.04
UKB Haplotype 5	0.03	0.04	0.03
UKB Haplotype 6	0.02	0.02	0.03
UKB Haplotype 7	0.02	0.02	0.03
UKB Haplotype 8	0.01	NA	0.05

Major and minor are according to UKB non-Black haplotypes. This table shows the frequency of haplotypes found in UKB Blacks at a frequency of at least 0.01. The haplotypes span the *LPA* gene from rs9456552 to rs1321196. One GERA haplotype represents two UKB haplotypes because only 33 of the original 36 haplotype variants were found in GERA. GERA: Genetic Epidemiology Research on Adult Health and Aging; MESA: Multi-Ethnic Study of Atherosclerosis; UKB: UK Biobank.

Chapter 3: General Discussion:

This study identified two common *LPA* region haplotypes that are associated with Lp(a) concentrations and are highly conserved across many populations. The effects of these haplotypes are modulated by other variants including rs10455872 and rs140570886, and more complete haplotypes may be proxies for another type of variation such as the KIV_2 repeats. The two haplotypes also showed an association with CVD, indicating that they may be useful clinical markers for such diseases. Many of the variants associated with Lp(a) are likely markers of KIV_2 repeats as Lp(a) increases linearly with each additional copy of risk variants. These results indicate that haplotypes may serve as useful markers for diagnostic assessment or treatment criteria in the future. Despite these results, it is important to note that the conservation of the *LPA* gene may not be as high as indicated by the present results. A previous study indicated that hundreds of variants in the KIV_2 region of *LPA* have been missed because of a poverty of studies assessing the KIV_2 region directly.[31] Also, the KIV_2 region can compose almost 70% of the amino acids in apo(a),[31] meaning a large portion of the gene is not being assessed for genetic associations. These missed variants, if causal, may contribute to the two *LPA* haplotypes that I identified as having opposing effects on Lp(a).

I also found that the two *LPA* haplotypes are highly conserved across multiple ethnicities except those of African descent. In HGDP, all assessed ethnicities have at least 99% of their haplotypes as one of the two *LPA* haplotypes except Sub-Saharan Africans, in which these two haplotypes account for only <92% of haplotypes. These haplotypes are at least 98% of haplotypes in UKB White British, South Asians, and Chinese, as well as in GERA Europeans, Latinos, and Chinese. More heterogeneity is present in populations of African descent, with only 42% of UKB Black and 48% of GERA African American haplotypes being one of the two *LPA* haplotypes.

However, individuals of African descent show similar haplotype structures and frequencies to each other. This indicates that clinically relevant markers for Lp(a) size and diagnosis of disease likely differ between those of African descent and other ethnicities.

A haplotype similar to the two identified in the present work has been found in previous research and this haplotype was also highly conserved and associated with Lp(a) concentrations across multiple ethnicities.[79] The haplotype identified in that study, similar to the ones in my study, spanned the KIV_2 repeats and did not span the entire *LPA* gene. This haplotype also contains one of the same variants identified in this study: rs10945682. However, their haplotype included an area outside of the *LPA* gene, unlike the present haplotypes, which consistently had boundaries within the *LPA* gene itself. That same study also found other haplotypes in the *LPA* region that are associated with Lp(a) concentrations, but that were located on different parts of the *LPA* gene.[79] Although work similar to a portion of the present work was completed a decade ago and showed important results, the genetics community has yet to implement haplotype analyses in their workflows. It is hoped that the results of the present work will encourage other researchers to implement such analyses in their work.

The major and minor haplotypes identified in the present study are “yin yang” haplotypes. These are haplotypes in which the allele differs at every assessed position.[110] Yin yang haplotypes are found across the majority of the genome and are evolutionarily ancient,[111] likely due to evolutionary pressures.[110] Yin yang haplotypes show much greater diversity in African populations than both Asian and European populations.[110] The results from my study confirms the increased haplotypic diversity found in African derived populations specifically for the *LPA* gene region. Further, both yin yang haplotypes are the predominant haplotypes in all assessed ethnicities in the present study, except in UKB and GERA Black populations. The yin yang haplotype in this study spans 64.722 kb. This is relatively short for a yin yang haplotype, as such haplotypes have been found to be as long as 14.3 Mb.[112] The MAFs of the variants that compose the yin yang haplotypes are quite high (MAF = 0.38–0.39), but such yin yang structures have also been found with rare variants as well.[113] Future research should include a greater number of rarer variants in the *LPA* region to assess if other, rarer haplotypes in this region may be in LD with the kringle repeats.

An important finding of my work is that rs3798220 is likely not a causative variant, despite being a missense variant, and that a variant in high LD, rs140570886, is associated with increased Lp(a) concentrations regardless of the presence of rs3798220. This is an especially notable finding, because the fact that rs3798220 is a missense variant led some researchers to believe that the association of rs3798220 with Lp(a) concentrations was driven by something other than the KIV_2 repeats.[114] This is not the first time rs140570886 has been found to be strongly associated with Lp(a) as this variant was identified at the genome-wide significance level for the first time in a study by Zekavat et al. (2018).[77] However, their study found an association with Lp(a) cholesterol levels, which are not as reliably measured as Lp(a) concentrations.[71] Another study found that rs3798220, rs10455872, and rs9457951 (an African American-specific variant) are all likely markers for other causative genetic variation.[83] That study also found that rs3798220 is associated with longer *LPA* genes in Latinos, in whom the variant is very common, while this variant is associated with shorter *LPA* gene lengths in Europeans, in whom the variant is relatively uncommon.[83] These results are compatible with the results of the present study, as they indicate that many key *LPA* variants are likely in LD with KIV_2 repeats. These variants may be markers for other genetic variation, besides KIV_2 repeats, as previous research has identified several variants associated with Lp(a) concentrations that are not correlated with KIV_2 repeat count.[25,77] For example, the variant rs4063600 identified in Estonian populations is associated with methylation and acetylation patterns in *LPA*.[77] There are also splice donor and acceptor variants found in the KIV_2 repeats that affect apo(a) length independent of KIV_2 repeat count.[25] That same study found variants in the 5’-UTR that affect transcription rate.

The lack of association of rs3798220 with Lp(a) concentrations indicates that including haplotype analysis in research can be revealing. When one analyzes variants in isolation, the observed effects could be the result of LD between a prioritized variant and the actual causative variant. Further, this result also indicates that extensive functional analyses should follow association studies. Rs3798220 is a missense variant with a strong association with Lp(a) concentrations, while rs140570886, the variant that is consistently associated with increased Lp(a) concentrations, is an intronic variant; thus, a variant is not necessarily important for a phenotype simply because it causes an amino acid change. It has also been proposed that assessing Lp(a) concentrations directly remains a better clinical tool than using genetic variants, but the present research indicates that genetic research can be important by using haplotypes as proxies for KIV_2 count, which requires a more laborious and complex process to accurately quantify.

Rs10455872, another key SNP, is associated with increased Lp(a) concentrations and increased odds of CAD and AS in all assessed ethnicities in UKB, GERA, and MESA. This is not surprising as rs10455872 has been shown previously to be strongly associated with Lp(a), as well as CVD.[15] Rs10455872 has been previously found to be associated with fewer KIV_2 repeats[18,57] and the KIV_2 repeats explain about 61% of variation in Lp(a) concentrations in European individuals and thus are the main determinant of Lp(a) concentrations.[25] That same study found that the effect of rs10455872 was entirely due to its strong LD with the KIV_2 repeats.

As for the genome-wide significant variants, several showed a trend of increasing Lp(a) concentrations for each additional copy of their respective minor allele in UKB White British. The trend appears to be additive for some of the assessed variants, such as rs73596816 and rs150776685, perhaps indicating that these variants are associated with very specific KIV_2 counts. Other variants that showed a less linear trend, such as rs10455782 and rs140570886, are likely the result of using truncated UKB data that resulted in the removal of high (out of range) Lp(a) values. These variants would likely show a linear trend if these missing Lp(a) measurements were included because individuals homozygous for the minor allele would likely have higher mean Lp(a) concentrations. This also means that these other variants that deviate moderately from a linear trend are also likely markers for KIV_2 repeat count. In fact, all eleven of these variants are likely proxies for KIV_2 repeat count as a previous study identified their association with KIV_2 counts, estimated using whole exome sequencing data.[25] My analysis was only completed in European individuals because of the small sample sizes for these Lp(a) GWAS variants in non-European ethnicities; future studies should use larger non-European cohorts to make a similar assessment of genotypic Lp(a) concentrations.

As for the null alleles assessed, three of them were homozygous in some UKB White British individuals: rs143431368, rs14272114, and rs4125944; rs4125944 was also homozygous in South Asians. None of the Black individuals were homozygous for the null variants. Homozygous null alleles usually resulted in Lp(a) concentrations below the detection threshold (3.8 nmol/L). UKB South Asians show similar associations to White British for the heterozygosity of the three null alleles and Lp(a) level value missingness rates; in South Asians, rs143431368 had a lower missingness rate in heterozygotes than homozygote major individuals, and the other two variants had two to three times the missingness rate in heterozygotes compared to homozygote major individuals. Rs143431368 is a splice acceptor variant and rs41272114 is a splice donor variant, and these variants likely interfere with proper splicing of the *LPA* mRNA, which results in nonsense-mediated decay. Rs41259144 is a missense variant that may result in proteasomal degradation of malformed apo(a) produced from *LPA* mRNA containing this variant. The downstream effects of these variants demonstrate the fact that KIV_2 repeat count is not the only crucial variation in the *LPA* region. Future research will need to assess the mechanism by which these variants result in undetectable Lp(a) levels.

Adding rs10455872 to the major haplotype resulted in a haplotype with an effect opposite the direction of the original major haplotype effect on Lp(a). Adding rs10455872, rs3798220, and rs140570886 to the minor *LPA* haplotype resulted in small odds ratios in the CVD analyses. These results indicate that the major and minor haplotypes are simply markers for other genetic variation, as accounting for more variants appears to diminish the effects of the two haplotypes. It is likely that an accumulation of rarer variants occurred on one haplotypic background which led to the observed opposing effects of the two haplotypes on Lp(a). Of 127 independent variants of varying frequency identified in the *LPA* region,[93] I investigated 11 of these variants and found each one to occur on a specific haplotypic background the majority of the time. Some of these variants demonstrate approximately linear trends in Lp(a) increase by genotype. If the major and minor haplotypes are proxies for KIV_2 repeats, they are most likely markers for a collection of KIV_2 repeats that are associated with the important variants in the *LPA* region. The major haplotype is likely a marker for several shorter *LPA* gene lengths and thus shorter average length, while the minor haplotype is likely a marker for longer *LPA* gene lengths.

It is also possible that the *LPA* haplotypes identified here are associated with other types of variation. A previous study found that the KIV_2 repeats and 23 independent variants together explain 90% of the heritability of Lp(a) concentrations in individuals of European descent.[25]

Chapter 4: Conclusions and Future Directions:

The present study has identified two haplotypes that are highly conserved across multiple populations and that are associated with Lp(a) concentrations. Further haplotype analysis revealed that rs3798220 was a variant in LD with rs140570886 which had a much more consistent association with Lp(a) levels. While non-European ethnicities were found to have insufficient allele counts for some of the assessed variation, important information was discovered in these ethnicities, nonetheless. Black individuals were shown to have much more haplotypic diversity than several ethnicities from five different continents. They also demonstrated high similarity in haplotype structure and frequency between different populations of African descent. Individuals of African descent also show some differing associations between the *LPA* haplotypes and Lp(a), indicating that the haplotypes are likely proxies for other genetic variation such as the KIV_2 repeats.

Future research should assess if the *LPA* haplotypes identified in this study are associated with specific KIV_2 repeats counts. This could be done using third-generation sequencing technologies such as Oxford Nanopore and Pacific Biosciences. An issue with second-generation sequencing is that each sequencing read is short, which means that genetic loci with variable numbers of repeats are difficult to sequence with such a technology.[115] Third generation sequencing has much longer reads that allow for the entire genetic locus coding for the KIV_2 repeats to be sequenced and phased accurately.[115] Beyond the KIV_2 repeats, third generation sequencing will allow the sequencing of long structural variants that second-generation sequencing could not capture.[115] However, third generation sequencing has a high error rate (~10–13%) that must be accounted for with multiple rounds of sequencing.[115] Further, a previous study attempted to use PacBio SMRT sequencing to sequence the *LPA* KIV_2 repeats, but had a low success rate because of insufficient read lengths;[116] this means that very long read third-generation technologies are required for the accurate sequencing of the KIV_2 repeats, which is one of the longest repeat polymorphisms found in the human genome.[116] In addition to such sequencing technologies, a new version of the human genome was recently sequenced that has no gaps and that accurately assayed repeat polymorphisms across the genome.[117] This genome, known as T2T-CH13,[117] can be used as a reference genome for future research studying copy number variations such as the KIV_2 repeats.

Future studies should also assess inter-haplotype interactions. One way such an analysis can be performed is through a gene-set-like analysis, with haplotypes as opposed to genes.[118] These analyses would assess if haplotypes affect Lp(a) concentrations, or variability in another phenotype, based on the sequence of other haplotypes on a chromosome. An alternative would be to select haplotype-tagging SNPs and perform SNP-level analyses.[110] Since the two haplotypes produced in the present work were highly conserved and very common, one could use a single variant to tag both haplotypes.

Since haplotypes have been informative in assessing genetic associations with Lp(a) concentrations in my study, it is likely that haplotypes in other regions of the genome will contribute to an improved understanding of genetic associations. Rs3798220 has been published as a key variant across many studies, and using a single haplotype analysis essentially disproved the causal role of this variant. Future research should incorporate haplotype analyses to better identify causative loci and understand the precise contribution of all variants at a locus.

Chapter 5: References: Master Reference List:

1. Berg K. A new serum type system in man--the LP system. *Acta Pathol Microbiol Scand.* 1963;59:369-382.
2. Witztum Joseph LJ, Ginsberg HN. Lipoprotein (a): coming of age at last. *J Lipid Res.* 2016;57(3):336-339.
3. Berg K, Dahlén G, Frick MH. Lp(a) lipoprotein and pre-beta1-lipoprotein in patients with coronary heart disease. *Clin Genet.* 1974;6(3):230-235.
4. Albers JJ, Adolphson JL, Hazzard WR. Radioimmunoassay of human plasma Lp(a) lipoprotein. *J Lipid Res.* 1977;18(3):331-338.
5. Rhoads GG, Dahlen G, Berg K, Morton NE, Dannenberg AL. Lp(a) lipoprotein as a risk factor for myocardial infarction. *JAMA.* 1986;256(18):2540-2544.
6. Zenker G, Költringer P, Boné G, Niederkorn K, Pfeiffer K, Jürgens G. Lipoprotein(a) as a strong indicator for cerebrovascular disease. *Stroke.* 1986;17(5):942-945.
7. McLean JW, Tomlinson JE, Kuang WJ, et al. cDNA sequence of human apolipoprotein(a) is homologous to plasminogen. *Nature.* 1987;330(6144):132-137.
8. Cobbaert C, Kesteloot H. Serum lipoprotein(a) levels in racially different populations. *Am J Epidemiol.* 1992;136(4):441-449.
9. Seed M, Hoppichler F, Reaveley D, et al. Relation of serum lipoprotein(a) concentration and apolipoprotein(a) phenotype to coronary heart disease in patients with familial hypercholesterolemia. *N Engl J Med.* 1990;322(21):1494-1499.
10. Hoefler G, Harnoncourt F, Paschke E, Mirtl W, Pfeiffer KH, Kostner GM. Lipoprotein Lp(a). A risk factor for myocardial infarction. *Arteriosclerosis.* 1988;8(4):398-401.

11. Gries A, Malle E, Wurm H, Kostner GM. Influence of dietary fish oils on plasma Lp(a) levels. *Thromb Res.* 1990;58(6):667-668.
12. Kostner GM. Interaction of Lp(a) and of apo(a) with liver cells. *Arterioscler Thromb.* 1993;13(7):1101-1109.
13. Bostom AG, Gagnon DR, Cupples LA, et al. A prospective investigation of elevated lipoprotein (a) detected by electrophoresis and cardiovascular disease in women. The Framingham Heart Study. *Circulation.* 1994;90(4):1688-1695.
14. Ridker PM, Hennekens CH, Stampfer MJ. A prospective study of lipoprotein(a) and the risk of myocardial infarction. *JAMA.* 1993;270(18):2195-2199.
15. Boffa MB, Koschinsky ML. Oxidized phospholipids as a unifying theory for lipoprotein(a) and cardiovascular disease. *Nat Rev Cardiol.* 2019;16(5):305-318.
16. Rifai N, Ma J, Sacks FM, et al. Apolipoprotein(a) size and lipoprotein(a) concentration and future risk of angina pectoris with evidence of severe coronary atherosclerosis in men: The Physicians' Health Study. *Clin Chem.* 2004;50(8):1364-1371.
17. Kostner KM, Kostner GM. Lipoprotein (a): a historical appraisal. *J Lipid Res.* 2017;58(1):1-14.
18. Clarke R, Peden JF, Hopewell JC, et al. Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N Engl J Med.* 2009;361(26):2518-2528.
19. Erqou S, Kaptoge S, Perry PL, et al. Lipoprotein(a) concentration and the risk of coronary heart disease, stroke, and nonvascular mortality. *JAMA.* 2009;302(4):412-423.
20. Kamstrup PR, Tybjaerg-Hansen A, Steffensen R, Nordestgaard BG. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA.* 2009;301(22):2331-2339.

21. Thanassoulis G, Campbell CY, Owens DS, et al. Genetic associations with valvular calcification and aortic stenosis. *N Engl J Med.* 2013;368(6):503-512.

22. Kamstrup PR, Tybjaerg-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.

23. van der Hoek YY, Wittekoek ME, Beisiegel U, Kastelein JJ, Koschinsky ML. The apolipoprotein(a) kringle IV repeats which differ from the major repeat kringle are present in variably-sized isoforms. *Hum Mol Genet.* 1993;2(4):361-366.

24. Lackner C, Cohen JC, Hobbs HH. Molecular definition of the extreme size polymorphism in apolipoprotein(a). *Hum Mol Genet.* 1993;2(7):933-940.

25. Mukamel RE, Handsaker RE, Sherman MA, et al. Protein-coding repeat polymorphisms strongly shape diverse human phenotypes. *bioRxiv.* 2021.

26. Kraft HG, Köchl S, Menzel HJ, Sandholzer C, Utermann G. The apolipoprotein (a) gene: a transcribed hypervariable locus controlling plasma lipoprotein (a) concentration. *Hum Genet.* 1992;90(3):220-230.

27. Lackner C, Boerwinkle E, Leffert CC, Rahmig T, Hobbs HH. Molecular basis of apolipoprotein (a) isoform size heterogeneity as revealed by pulsed-field gel electrophoresis. *J Clin Invest.* 1991;87(6):2153-2161.

28. Li Y, Luke MM, Shiffman D, Devlin JJ. Genetic variants in the apolipoprotein(a) gene and coronary heart disease. *Circ Cardiovasc Genet.* 2011;4(5):565-573.

29. Brunner C, Kraft HG, Utermann G, Müller HJ. Cys4057 of apolipoprotein(a) is essential for lipoprotein(a) assembly. *Proc Natl Acad Sci U S A.* 1993;90(24):11643-11647.

30. Schmidt K, Noureen A, Kronenberg F, Utermann G. Structure, function, and genetics of lipoprotein (a). *J Lipid Res.* 2016;57(8):1339-1359.

31. Coassin S, Schönherr S, Weissensteiner H, et al. A comprehensive map of single-base polymorphisms in the hypervariable *LPA* kringle IV type 2 copy number variation region. *J Lipid Res.* 2019;60(1):186-199.

32. Ichinose A. Multiple members of the plasminogen-apolipoprotein(a) gene family associated with thrombosis. *Biochemistry.* 1992;31(12):3113-3118.

33. Mooser V, Marcovina SM, White AL, Hobbs HH. Kringle-containing fragments of apolipoprotein(a) circulate in human plasma and are excreted into the urine. *J Clin Invest.* 1996;98(10):2414-2424.

34. Boffa MB, Koschinsky ML. Lipoprotein (a): truly a direct prothrombotic factor in cardiovascular disease? *J Lipid Res.* 2016;57(5):745-757.

35. Guevara J, Jr., Jan AY, Knapp R, Tulinsky A, Morrisett JD. Comparison of ligand-binding sites of modeled apo[a] kringle-like sequences in human lipoprotein[a]. *Arterioscler Thromb.* 1993;13(5):758-770.

36. Gabel BR, May LF, Marcovina SM, Koschinsky ML. Lipoprotein(a) assembly. Quantitative assessment of the role of apo(a) kringle IV types 2-10 in particle formation. *Arterioscler Thromb Vasc Biol.* 1996;16(12):1559-1567.

37. Koschinsky ML, Côté GP, Gabel B, van der Hoek YY. Identification of the cysteine residue in apolipoprotein(a) that mediates extracellular coupling with apolipoprotein B-100. *J Biol Chem.* 1993;268(26):19819-19825.

38. Rader DJ, Cain W, Ikewaki K, et al. The inverse association of plasma lipoprotein(a) concentrations with apolipoprotein(a) isoform size is not due to differences in Lp(a) catabolism but to differences in production rate. *J Clin Invest.* 1994;93(6):2758-2763.

39. Frischmann ME, Ikewaki K, Trenkwalder E, et al. *In vivo* stable-isotope kinetic study suggests intracellular assembly of lipoprotein(a). *Atherosclerosis.* 2012;225(2):322-327.

40. Demant T, Seeberg K, Bedynek A, Seidel D. The metabolism of lipoprotein(a) and other apolipoprotein B-containing lipoproteins: a kinetic study in humans. *Atherosclerosis.* 2001;157(2):325-339.

41. White AL, Lanford RE. Cell surface assembly of lipoprotein(a) in primary cultures of baboon hepatocytes. *J Biol Chem.* 1994;269(46):28716-28723.

42. Cain WJ, Millar JS, Himebauch AS, et al. Lipoprotein [a] is cleared from the plasma primarily by the liver in a process mediated by apolipoprotein [a]. *J Lipid Res.* 2005;46(12):2681-2691.

43. Kronenberg F, Utermann G, Dieplinger H. Lipoprotein(a) in renal disease. *Am J Kidney Dis.* 1996;27(1):1-25.

44. Kronenberg F, Trenkwalder E, Lingenhel A, et al. Renovascular arteriovenous differences in Lp[a] plasma concentrations suggest removal of Lp[a] from the renal circulation. *J Lipid Res.* 1997;38(9):1755-1763.

45. Romagnuolo R, Scipione CA, Boffa MB, Marcovina SM, Seidah NG, Koschinsky ML. Lipoprotein(a) catabolism is regulated by proprotein convertase subtilisin/kexin type 9 through the low density lipoprotein receptor. *J Biol Chem.* 2015;290(18):11649-11662.

46. Boffa MB, Koschinsky ML. Update on lipoprotein(a) as a cardiovascular risk factor and mediator. *Curr Atheroscler Rep.* 2013;15(10):360.

47. 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;121(2):306-314.

48. 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;352(23):2389-2397.
49. Rossebø AB, Pedersen TR, Boman K, et al. Intensive lipid lowering with simvastatin and ezetimibe in aortic stenosis. *N Engl J Med.* 2008;359(13):1343-1356.
50. Canuel M, Sun X, Asselin MC, Paramithiotis E, Prat A, Seidah NG. Proprotein convertase subtilisin/kexin type 9 (PCSK9) can mediate degradation of the low density lipoprotein receptor-related protein 1 (LRP-1). *PloS One.* 2013;8(5):e64145.
51. Poirier S, Mayer G, Benjannet S, et al. The proprotein convertase PCSK9 induces the degradation of low density lipoprotein receptor (LDLR) and its closest family members VLDLR and ApoER2. *J Biol Chem.* 2008;283(4):2363-2372.
52. Tam SP, Zhang X, Koschinsky ML. Interaction of a recombinant form of apolipoprotein[a] with human fibroblasts and with the human hepatoma cell line HepG2. *J Lipid Res.* 1996;37(3):518-533.
53. Trieu VN, McConathy WJ. Lipoprotein(a) binding to other apolipoprotein B containing lipoproteins. *Biochemistry.* 1990;29(25):5919-5924.
54. Rader DJ, Cain W, Zech LA, Usher D, Brewer HB, Jr. Variation in lipoprotein(a) concentrations among individuals with the same apolipoprotein (a) isoform is determined by the rate of lipoprotein(a) production. *J Clin Invest.* 1993;91(2):443-447.
55. White AL, Hixson JE, Rainwater DL, Lanford RE. Molecular basis for "null" lipoprotein(a) phenotypes and the influence of apolipoprotein(a) size on plasma lipoprotein(a) level in the baboon. *J Biol Chem.* 1994;269(12):9060-9066.

56. Puckey LH, Lawn RM, Knight BL. Polymorphisms in the apolipoprotein(a) gene and their relationship to allele size and plasma lipoprotein(a) concentration. *Hum Mol Genet.* 1997;6(7):1099-1107.

57. Saleheen D, Haycock PC, Zhao W, et al. Apolipoprotein(a) isoform size, lipoprotein(a) concentration, and coronary artery disease: a mendelian randomisation analysis. *Lancet Diabetes Endocrinol.* 2017;5(7):524-533.

58. Kratzin H, Armstrong VW, Niehaus M, Hilschmann N, Seidel D. Structural relationship of an apolipoprotein (a) phenotype (570 kDa) to plasminogen: homologous kringle domains are linked by carbohydrate-rich regions. *Biol Chem Hoppe Seyler*. 1987;368(12):1533-1544.

59. Hixson JE, Jett C, Birnbaum S. Identification of promoter sequences in the 5' untranslated region of the baboon apolipoprotein[a] gene. *J Lipid Res.* 1996;37(11):2324-2331.

60. Negi S, Singh SK, Pati N, Handa V, Chauhan R, Pati U. A proximal tissue-specific module and a distal negative regulatory module control apolipoprotein(a) gene transcription. *Biochem J.* 2004;379(Pt 1):151-159.

61. Coassin S, Erhart G, Weissensteiner H, et al. A novel but frequent variant in *LPA* KIV-2 is associated with a pronounced Lp(a) and cardiovascular risk reduction. *Eur Heart J.* 2017;38(23):1823-1831.

62. Boffelli D, Zajchowski DA, Yang Z, Lawn RM. Estrogen modulation of apolipoprotein(a) expression. Identification of a regulatory element. *J Biol Chem.* 1999;274(22):15569-15574.

63. Müller N, Schulte DM, Türk K, et al. IL-6 blockade by monoclonal antibodies inhibits apolipoprotein (a) expression and lipoprotein (a) synthesis in humans. *J Lipid Res.* 2015;56(5):1034-1042.

64. Lim ET, Würtz P, Havulinna AS, et al. Distribution and medical impact of loss-of-function variants in the Finnish founder population. *PLoS Genet.* 2014;10(7):e1004494.

65. Bergmark C, Dewan A, Orsoni A, et al. A novel function of lipoprotein [a] as a preferential carrier of oxidized phospholipids in human plasma. *J Lipid Res.* 2008;49(10):2230-2239.

66. Leibundgut G, Scipione C, Yin H, et al. Determinants of binding of oxidized phospholipids on apolipoprotein (a) and lipoprotein (a). *J Lipid Res.* 2013;54(10):2815-2830.

67. Yano Y, Shimokawa K, Okada Y, Noma A. Immunolocalization of lipoprotein(a) in wounded tissues. *J Histochem Cytochem.* 1997;45(4):559-568.

68. Brown MS, Goldstein JL. Plasma lipoproteins: teaching old dogmas new tricks. *Nature.* 1987;330(6144):113-114.

69. von Zychlinski A, Kleffmann T, Williams MJ, McCormick SP. Proteomics of Lipoprotein(a) identifies a protein complement associated with response to wounding. *J Proteomics.* 2011;74(12):2881-2891.

70. Feric NT, Boffa MB, Johnston SM, Koschinsky ML. Apolipoprotein(a) inhibits the conversion of Glu-plasminogen to Lys-plasminogen: a novel mechanism for lipoprotein(a)-mediated inhibition of plasminogen activation. *J Thromb Haemost.* 2008;6(12):2113-2120.

71. Kronenberg F, Tsimikas S. The challenges of measuring Lp(a): A fight against Hydra? *Atherosclerosis.* 2019;289:181-183.

72. Lawn RM, Boonmark NW, Schwartz K, et al. The recurring evolution of lipoprotein(a). Insights from cloning of hedgehog apolipoprotein(a). *J Biol Chem.* 1995;270(41):24004-24009.

73. Lawn RM, Schwartz K, Patthy L. Convergent evolution of apolipoprotein(a) in primates and hedgehog. *Proc Natl Acad Sci U S A.* 1997;94(22):11992-11997.

74. Tomlinson JE, McLean JW, Lawn RM. Rhesus monkey apolipoprotein(a). Sequence, evolution, and sites of synthesis. *J Biol Chem.* 1989;264(10):5957-5965.

75. Reyes-Soffer G. The impact of race and ethnicity on lipoprotein(a) levels and cardiovascular risk. *Curr Opin Lipidol.* 2021;32(3):163-166.

76. Patel AP, Wang M, Pirruccello JP, et al. Lp(a) (Lipoprotein[a]) concentrations and incident atherosclerotic cardiovascular disease: new insights from a large national biobank. *Arterioscler Thromb Vasc Biol.* 2021;41(1):465-474.

77. 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;9(1):2606.

78. Tsimikas S, Clopton P, Brilakis ES, et al. Relationship of oxidized phospholipids on apolipoprotein B-100 particles to race/ethnicity, apolipoprotein(a) isoform size, and cardiovascular risk factors: results from the Dallas Heart Study. *Circulation.* 2009;119(13):1711-1719.

79. 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;3(1):39-46.

80. Marcovina SM, Albers JJ, Wijsman E, Zhang Z, Chapman NH, Kennedy H. Differences in Lp[a] concentrations and apo[a] polymorphs between black and white Americans. *J Lipid Res.* 1996;37(12):2569-2585.

81. 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;6(1):e14581.

82. Paré G, Çaku A, McQueen M, et al. Lipoprotein(a) levels and the risk of myocardial infarction among 7 ethnic groups. *Circulation.* 2019;139(12):1472-1482.

83. Lee SR, Prasad A, Choi YS, et al. *LPA* gene, ethnicity, and cardiovascular events. *Circulation.* 2017;135(3):251-263.

84. 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;42(1):2-9.

85. Chretien JP, Coresh J, Berthier-Schaad Y, et al. Three single-nucleotide polymorphisms in *LPA* account for most of the increase in lipoprotein(a) level elevation in African Americans compared with European Americans. *J Med Genet.* 2006;43(12):917-923.

86. Cann HM, de Toma C, Cazes L, et al. A human genome diversity cell line panel. *Science.* 2002;296(5566):261-262.

87. Bergström A, McCarthy SA, Hui R, et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science.* 2020;367(6484).

88. Cavalli-Sforza LL. The Human Genome Diversity Project: past, present and future. *Nat Rev Genet.* 2005;6(4):333-340.

89. Crawford DC, Nickerson DA. Definition and clinical importance of haplotypes. *Annu Rev Med.* 2005;56:303-320.

90. Adekile A. The genetic and clinical significance of fetal hemoglobin expression in sickle cell disease. *Med Princ Pract.* 2020;30(3):201-211.

91. Drysdale CM, McGraw DW, Stack CB, et al. Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict *in vivo* responsiveness. *Proc Natl Acad Sci U S A.* 2000;97(19):10483-10488.

92. Tregouet DA, Konig IR, Erdmann J, et al. Genome-wide haplotype association study identifies the *SLC22A3-LPAL2-LPA* gene cluster as a risk locus for coronary artery disease. *Nat Genet.* 2009;41(3):283-285.

93. Hoekstra M, Chen HY, Rong J, et al. Genome-wide association study highlights *APOH* as a novel locus for lipoprotein(a) levels-brief report. *Arterioscler Thromb Vasc Biol.* 2021;41(1):458-464.

94. 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;58(9):1834-1844.

95. Lindman BR, Clavel MA, Mathieu P, et al. Calcific aortic stenosis. *Nat Rev Dis Primers.* 2016;2:16006.

96. Boffa MB, Koschinsky ML. The journey towards understanding lipoprotein(a) and cardiovascular disease risk: are we there yet? *Curr Opin Lipidol.* 2018;29(3):259-267.

97. Nielsen LB, Stender S, Kjeldsen K, Nordestgaard BG. Specific accumulation of lipoprotein(a) in balloon-injured rabbit aorta *in vivo*. *Circ Res.* 1996;78(4):615-626.

98. Dweck MR, Boon NA, Newby DE. Calcific aortic stenosis: a disease of the valve and the myocardium. *J Am Coll Cardiol.* 2012;60(19):1854-1863.

99. Bouchareb R, Mahmut A, Nsaibia MJ, et al. Autotaxin derived from lipoprotein(a) and valve interstitial cells promotes inflammation and mineralization of the aortic valve. *Circulation.* 2015;132(8):677-690.

100. Malakar AK, Choudhury D, Halder B, Paul P, Uddin A, Chakraborty S. A review on coronary artery disease, its risk factors, and therapeutics. *J Cell Physiol.* 2019;234(10):16812-16823.

101. Maranhão RC, Carvalho PO, Strunz CC, Pileggi F. Lipoprotein (a): structure, pathophysiology and clinical implications. *Arq Bras Cardiol.* 2014;103(1):76-84.

102. Thanassoulis G. Lipoprotein (a) in calcific aortic valve disease: from genomics to novel drug target for aortic stenosis. *J Lipid Res.* 2016;57(6):917-924.

103. Seimon TA, Nadolski MJ, Liao X, et al. Atherogenic lipids and lipoproteins trigger CD36-TLR2-dependent apoptosis in macrophages undergoing endoplasmic reticulum stress. *Cell Metab.* 2010;12(5):467-482.

104. Tsimikas S, Viney NJ, Hughes SG, et al. Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study. *Lancet.* 2015;386(10002):1472-1483.

105. Katzmann JL, Packard CJ, Chapman MJ, Katzmann I, Laufs U. Targeting RNA with antisense oligonucleotides and small interfering RNA: JACC state-of-the-art review. *J Am Coll Cardiol.* 2020;76(5):563-579.

106. Afshar M, Yazdan-Ashoori S, Engert JC, Thanassoulis G. Drugs for prevention and treatment of aortic stenosis: how close are we? *Can J Cardiol.* 2021;37(7):1016-1026.

107. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *N Engl J Med.* 2017;376(18):1713-1722.

108. Gaudet D, Kereiakes DJ, McKenney JM, et al. Effect of alirocumab, a monoclonal proprotein convertase subtilisin/kexin 9 antibody, on lipoprotein(a) concentrations (a pooled analysis of 150 mg every two weeks dosing from phase 2 trials). *Am J Cardiol.* 2014;114(5):711-715.

109. Vavuranakis MA, Jones SR, Cardoso R, Gerstenblith G, Leucker TM. The role of lipoprotein(a) in cardiovascular disease: current concepts and future perspectives. *Hellenic J Cardiol.* 2020;61(6):398-403.

110. Zhang J, Rowe WL, Clark AG, Buetow KH. Genomewide distribution of high-frequency, completely mismatching SNP haplotype pairs observed to be common across human populations. *Am J Hum Genet.* 2003;73(5):1073-1081.

111. Elliott P, Chambers JC, Zhang W, et al. Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA.* 2009;302(1):37-48.

112. Curtis D, Vine AE. Yin yang haplotypes revisited - long, disparate haplotypes observed in European populations in regions of increased homozygosity. *Hum Hered.* 2010;69(3):184-192.

113. Curtis D. Rare missense variants within a single gene form yin yang haplotypes. *EJHG.* 2016;24(1):139-141.

114. Luke MM, Kane JP, Liu DM, et al. A polymorphism in the protease-like domain of apolipoprotein(a) is associated with severe coronary artery disease. *Arterioscler Thromb Vasc Biol.* 2007;27(9):2030-2036.

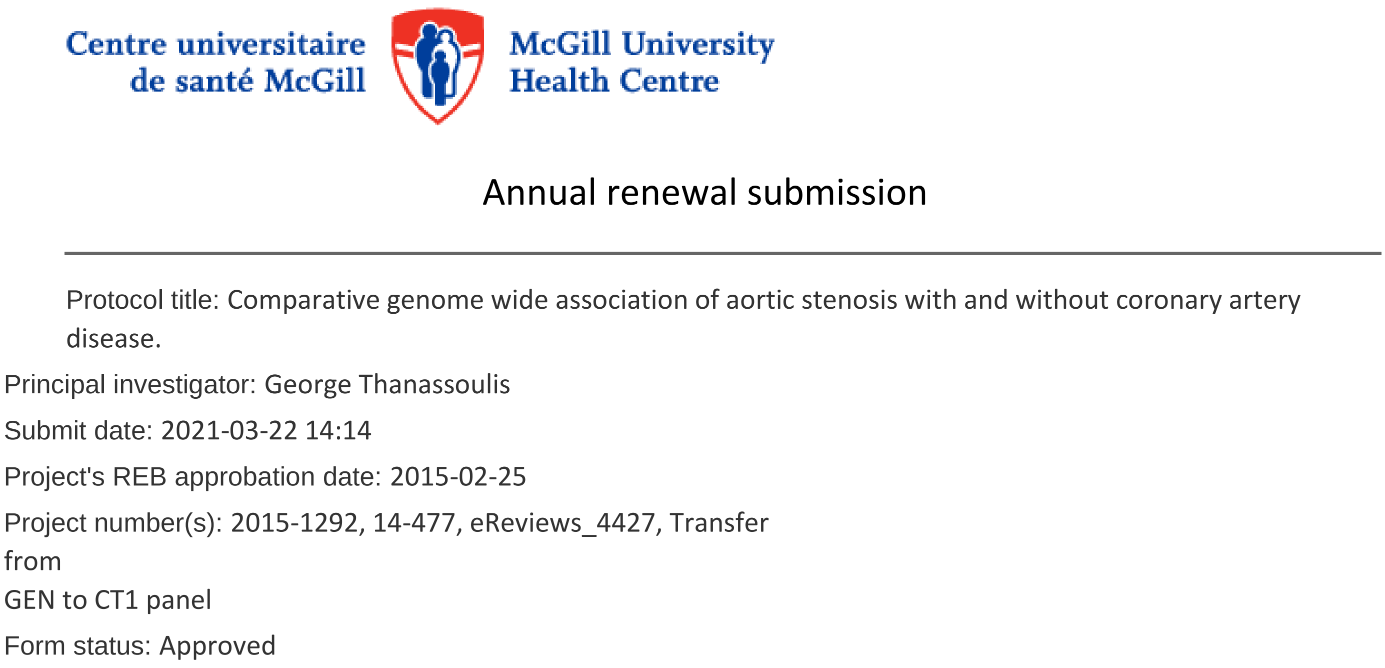
115. van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The third revolution in sequencing technology. *Trends Genet.* 2018;34(9):666-681.

116. Sulovari A, Li R, Audano PA, et al. Human-specific tandem repeat expansion and differential gene expression during primate evolution. *Proc Natl Acad Sci U S A.* 2019;116(46):23243-23253.

117. Vollger MR, Guitart X, Dishuck PC, et al. Segmental duplications and their variation in a complete human genome. *bioRxiv.* 2021:2021.2005.2026.445678.

118. Dutta R, Mainsah J, Yatskiv Y, et al. Intricacies in arrangement of SNP haplotypes suggest "Great Admixture" that created modern humans. *BMC Genomics.* 2017;18(1):433.

Appendix: Ethics Forms:



Centre universitaire de santé McGill

McGill University Health Centre

Annual renewal submission

Protocol title: Comparative genome wide association of aortic stenosis with and without coronary artery disease.

Principal investigator: George Thanassoulis

Submit date: 2021-03-22 14:14

Project's REB approbation date: 2015-02-25

Project number(s): 2015-1292, 14-477, eReviews_4427, Transfer from GEN to CT1 panel

Form status: Approved

Administration

Nagano identifier: 14-477 GEN

1.

MUHC REB Panel & Co-chair(s):

Clinical Trials 1 (CT1)

Co-chair: Vincent Lajoie

2.

REB Decision:

Approved - REB delegated review

3.

Renewal Period Granted:

From 2021-04-20 Until 2022-04-19

4.

Date of the REB final decision & signature

2021-03-24

Signature

Ms. Elizabeth Craven, Coordinator, MUHC REB

MUHC REB Coordinator
for MUHC Co-chair mentioned above

5.

FWA 00000840 - FWA 00004545

6.

Local REB number

IRB00010120

A. General information

1.

Indicate the full title of the research study

Comparative genomewide association of aortic stenosis with and without coronary artery disease.

---

2. If relevant, indicate the full study title in French

3.

Indicate the name of the Principal Investigator in our institution (MUHC)

Thanassoulis, George

4.

Are there local co-investigators & collaborators involved in this project?

Yes

List all the local co-investigators & collaborators involved in the research study

Engert, James

Huynh, Thao

Department of co-investigators & collaborators

Oncology

---

5. Identify the study coordinator(s)

Mohammadnetaj, Maedeh

Indicate the role of the collaborator(s)

Coordinator

B. Project development

Study start date:

1. 2015-02-02

---

2. Expected ending date of the study:

☐ Determined date
☑ Undetermined date

---

3. Date of recruitment of the first participant?

☐ 1st enrollment date
☑

is... No participant
enrolled Please justify:

This is a database project. All participants have previously been recruited.

4. Add a brief statement on the study status

Study is progressing as planned. More data analysis is in progress

---

5. Information about the participants at this institution, since the beginning of the project

Number of participants to be recruited according to
protocol 0

Number of participants who have been recruited

0

Number of minors

0

Number of incompetent adults

0

Number of participants who have not yet completed the study (still in
progress) 0

Number of participants who've completed the
study 0

Number of participants who were recruited to the study, but who were then excluded or withdrawn:

0

Indicate the reasons the participants were
withdrawn 0

Number of participants who dropped out (voluntary withdrawal):

0

Please indicate the reasons for the participant(s) withdrawing from the
study 0

Number of participants who died during the study

0

6. Information about the participants at this institution (MUHC) since the previous REB approval

Number of participants who have been recruited

0

Number of minors

0

Number of incompetent adults

0

Number of participants who have not yet completed the study (still in progress) 0

Number of participants who've completed the study 0

Number of participants who were recruited to the study, but who were then excluded or withdrawn:

0

Indicate the reasons participants were excluded/withdrawn 0

Number of participants who dropped out (voluntary withdrawal):

0

Please indicate the reasons the participant(s) stopped 0

Number of participants who died during the study

0

7. Since the previous REB approval (annual renewal or initial approval):

Were there any changes to the protocol (or to the databank management framework) ?

No

Specify the current version/date:

2019-12-19

Date approved by the REB:

2020-01-06

Were there any changes to the information and consent form? No

Were there any reportable adverse events at this site (or, for multi-center projects, at an institution under the jurisdiction of our REB) that should be reported to the REB under section 5.2.1 of " SOP- REB404001 " ? https://muhc.ca/cae/page/standard-operating-procedures-sops

No

Has there has been any new information likely to affect the ethics of the project or influence the decision of a participant as to their continued participation in the project ?

No

Were there any deviations / major violations protocol (life -threatening or not meeting the inclusion / exclusion criteria)?

No

Was there a temporary interruption of the project?

No

Have the project results been submitted for publication, presented or published?

Yes

Please specify:

1) Race-Based Differences in Lipoprotein(a)-Associated Risk of Carotid Atherosclerosis.Arterioscler Thromb Vasc Biol. 2019 Mar;39(3):523-529. doi:10.1161/ATVBAHA.118.312267. PMID: 30727753

2)Observational and Genetic Associations of Resting Heart Rate With Aortic Valve Calcium.
Am J Cardiol. 2018 May 15;121(10):1246-1252. doi: 10.1016/j.amjcard.2018.01.048. Epub 2018 Feb 13.
PMID: 29656781

3)Association of Triglyceride-Related Genetic Variants With Mitral Annular Calcification. J Am Coll Cardiol.
2017 Jun 20;69(24):2941-2948. doi: 10.1016/j.jacc.2017.04.051. PMID: 28619195

4) Genome-Wide Association Study Highlights APOH as a Novel Locus for Lipoprotein(a) Levels. PMID: 33115273 DOI: 10.1161/ATVBAHA.120.314965

5) Genome-wide association meta-analysis in 652,134 participants identifies 9 novel susceptibility loci for aortic stenosis. Oral presentation at the European Society of Cardiology. August 2020.

Has the REB been notified of a conflict of interest - (apparent , potential or actual), of one or more members of the research team - that was not known when it was last approved project? No

Do you want to bring any other info to the REB's attention?

No

8. Is there a data safety monitoring committee analyzing data on the safety and efficacy of the treatment? No

C. Signature

1. I confirm that all information is complete & accurate.

First & last name of person who completed the submission

Maedeh Mohammadnetaj
2021-03-16 15:03