SUPPLEMENTAL MATERIAL

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

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

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

*co-senior authors.

Supplemental Methods

Phenotype Definitions in the UK Biobank

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

on a Beckman Coulter AU5800. Corrected LDL-C was calculated by converting both LDL-C and lipoprotein(a) (Lp[a]) into mg/dL and subtracting 30% of the Lp(a) value from the LDL-C value, as previously described[1]. High-density lipoprotein cholesterol (nmol/L) was measured by enzyme immunoinhibition analysis on a Beckman Coulter AU5800. Systolic and diastolic blood pressure (mmHg) were each measured automatically using an Omron device. Body mass index (kg/m^2) was calculated using weight and height measurements from the initial assessment visit.

Replication Cohorts

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

The Framingham Offspring Study (FOS) is a longitudinal population-based study that recruited 5,124 of the original Framingham Heart Study participants' offspring and their spouses, as previously described[4]. During the 5^{th} examination cycle (1991-1995), participants underwent a medical history, physical examination, and had blood drawn for plasma lipid and lipoprotein measurements. Lp(a) was measured in mg/dL using an immunoturbidimetric assay from Wako Chemicals USA (Richmond, VA). Genotyping was performed on the Affymetrix GeneChip Human Mapping 500K Array and 50K Human Gene Focused Panel and genotypes were imputed using the Haplotype Reference Consortium panel[5]. Our analysis included 3,009 unrelated individuals of European ancestry. Associations between genetic variants and natural log-transformed Lp(a) were tested using linear regression models adjusted for age, sex, and 10 principal components of ancestry.

Supplemental Figures

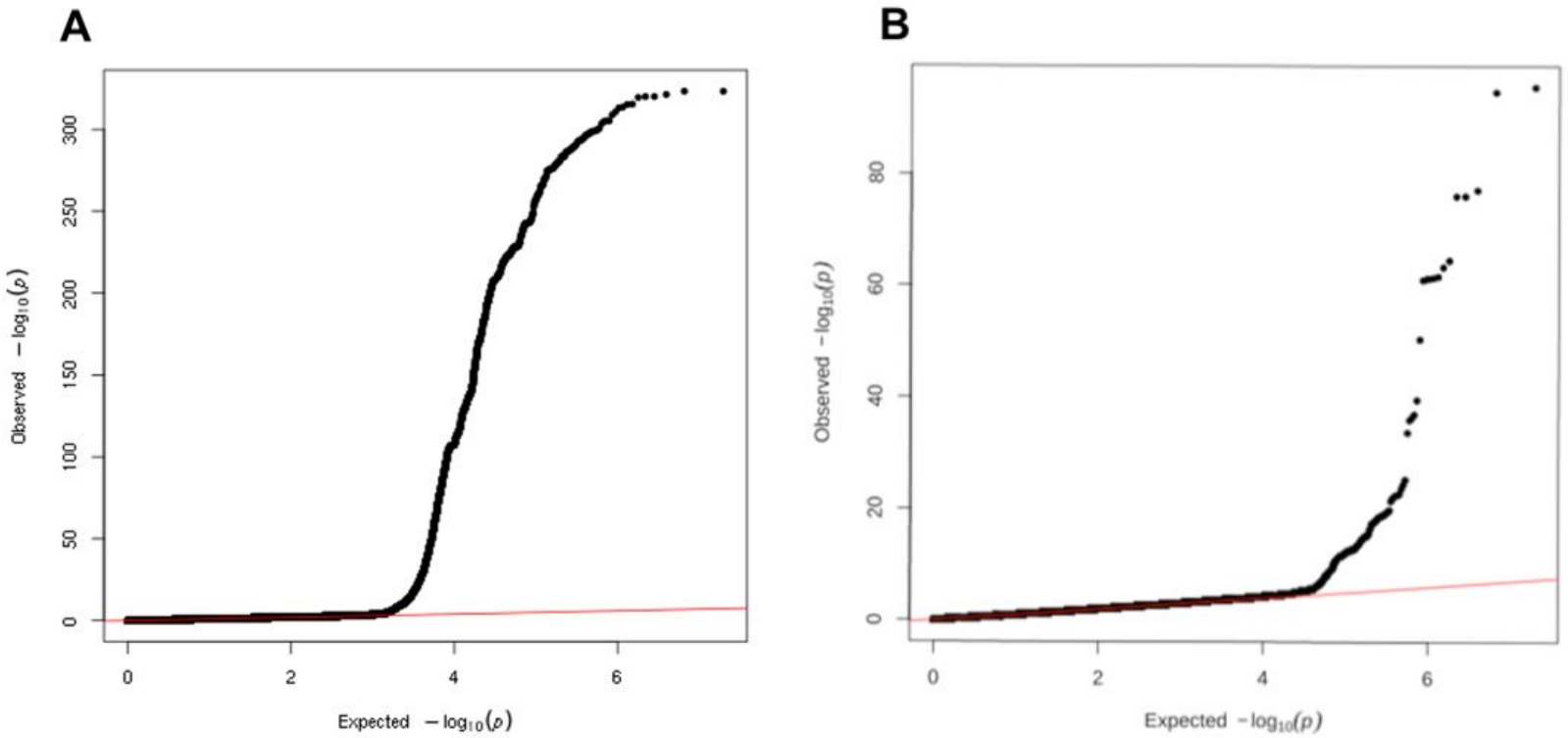

Figure I: QQ-plots for the GWAS of natural log-transformed Lp(a) adjusted for age, sex, genotype batch, and 20 principal components with A) all variants, and B) exclusion of variants in the *LPA* region on chromosome 6. Associations with $P < 4.9 \times 10^{-324}$ are not shown due to limitations in the plotting software.

Supplemental Tables

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

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

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

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

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

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

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

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

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

Variant	Locus	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*
rs10455872	*LPA*	G (0.076)	1.7 [1.7, 1.7]	$8.9 \times 10^{-22,140}$
rs1065853	*APOE*	T (0.080)	-0.11 [-0.12, -0.10]	1.9×10^{-96}
rs247617	*CETP*	A (0.32)	-0.023 [-0.030, -0.017]	1.1×10^{-13}
rs8178824	*APOH*	T (0.030)	0.064 [0.047, 0.081]	2.9×10^{-13}
rs826128	*AC093639.1*	A (0.054)	-0.040 [-0.053, -0.0026]	5.4×10^{-9}

Abbreviations: CI, confidence interval.

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

Variant	Locus	Group	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*
rs10455872	*LPA*	SA	G (0.011)	1.01 [0.82, 1.19]	8.5×10^{-26}
		BA	G (0.00063)	0.68 [-0.24, 1.6]	0.15
		BC	G (0.010)	0.81 [0.59, 1.03]	1.4×10^{-12}
		WB	G (0.076)	1.7 [1.7, 1.7]	$6.2 \times 10^{-22,136}$
rs1065853	*APOE*	SA	T (0.045)	-0.17 [-0.25, -0.081]	1.3×10^{-4}
		BA	T (0.12)	-0.28 [-0.34, -0.21]	9.4×10^{-16}
		BC	T (0.12)	-0.21 [-0.27, -0.15]	1.2×10^{-11}
		WB	T (0.080)	-0.11 [-0.12, -0.10]	2.8×10^{-96}
rs247617	*CETP*	SA	A (0.34)	-0.0046 [-0.041, 0.032]	0.81
		BA	A (0.26)	-0.020 [-0.70, 0.031]	0.45
		BC	A (0.25)	0.0086 [-0.038, 0.055]	0.72
		WB	A (0.32)	-0.023 [-0.030, -0.017]	1.0×10^{-13}

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

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

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

Variant	**Locus**	**Minor Allele (Freq*)**	**FOS**		**MESA**		**Meta-analysis**	
			β [95% CI] (ln mg/dL)	***P***	**β [95% CI] (ln mg/dL)**	***P***	**β [95% CI] (ln mg/dL)**	***P***
rs10455872	*LPA*	G (0.066)	2.3 [2.2, 2.4]	2.0×10^{-216}	1.6 [1.5, 1.8]	2.5×10^{-85}	2.1 [2.0, 2.2]	6.0×10^{-534}
rs7412†	*APOE*	T (0.077)	0.0014 [-0.12, 0.13]	0.99	-0.15 [-0.29, -0.0084]	0.038	-0.065 [-0.16, 0.028]	0.17
rs247617	*CETP*	A (0.32)	-0.026 [-0.082, 0.030]	0.54	0.018 [-0.043, 0.078]	0.57	-0.0057 [-0.047, 0.035]	0.79
rs8178824	*APOH*	T (0.035)	0.19 [0.049, 0.34]	0.082	0.11 [-0.089, 0.31]	0.28	0.16 [0.044, 0.28]	0.0071
rs826128	*AC093639.1*	A (0.049)	0.050 [-0.072, 0.17]	0.58	-0.047 [-0.17, 0.074]	0.45	0.0013 [-0.084, 0.087]	0.98

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

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

Variant	Locus	Minor Allele (Freq)	β [95% CI] (ln nmol/L)	*P*
rs1065853	*APOE*	T (0.080)	-0.17 [-0.17, -0.16]	3.6×10^{-363}
rs247617	*CETP*	A (0.32)	-0.029 [-0.034, -0.024]	1.2×10^{-34}
rs8178824	*APOH*	T (0.030)	0.089 [0.076, 0.10]	3.8×10^{-42}
rs826128	*AC093639.1*	A (0.054)	-0.016 [-0.026, -0.0058]	0.0019

Abbreviations: CI, confidence interval.

Major Resources Table

Data & Code Availability

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

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

References

1. Viney NJ, van Capelleveen JC, Geary RS, et al. Antisense oligonucleotides targeting apolipoprotein(a) in people with raised lipoprotein(a): two randomised, double-blind, placebo-controlled, dose-ranging trials. *Lancet*. 2016;388:2239-2253.
2. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol*. 2002; 156:871-881.
3. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. *Nature*. 2015;26:68-74.
4. Kannel WB, Feinleib M, MeNAMARA PM, Garrison RJ, Castelli WP. An Investigation of Coronary Heart Disease in Families: The Framingham Offspring Study. *Am J Epidemiol*. 2017;185:1103-1108.
5. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279-1283.
6. Machiela MJ, Chanock SJ. LDlink: A web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics*. 2015;31:3555-3557.