



# Global genetic diversity of human apolipoproteins and effects on cardiovascular disease risk<sup>S</sup>

Yitian Zhou,\* Reedik Mägi,<sup>†</sup> Lili Milani,<sup>†,§</sup> and Volker M. Lauschke<sup>1,\*</sup>

Department of Physiology and Pharmacology, Section of Pharmacogenetics,\* Karolinska Institutet, Stockholm, Sweden; Estonian Genome Center,<sup>†</sup> University of Tartu, Tartu, Estonia; and Science for Life Laboratory, Department of Medical Sciences,<sup>§</sup> Uppsala University, Uppsala, Sweden

ORCID IDs: 0000-0002-1140-6204 (V.M.L.)

**Abstract** Abnormal plasma apolipoprotein levels are consistently implicated in CVD risk. Although 30% to 60% of their interindividual variability is genetic, common genetic variants explain only 10% to 20% of these differences. Rare genetic variants may be major sources of the missing heritability, yet quantitative evaluations of their contribution to phenotypic variability are lacking. Here, we analyzed whole-genome and whole-exome sequencing data from 138,632 individuals across seven major human populations to present a systematic overview of genetic apolipoprotein variability. We provide population-specific frequencies of 38 clinically important apolipoprotein alleles and identify further 6,875 genetic variants, 33% of which are novel and 98.7% of which are rare with minor allele frequencies <1%. We predicted the functional impact of rare variants and found that their relative importance differed drastically between genes and among ethnicities. Importantly, we validated the clinical relevance of multiple variants with predicted effects by leveraging association data from the CARDIoGRAM (Coronary Artery Disease Genomewide Replication and Meta-analysis) and Global Lipids Genetics consortia.<sup>1,2</sup> Overall, we provide a consolidated overview of population-specific apolipoprotein genetics as a valuable data resource for scientists and clinicians, estimate the importance of rare genetic variants for the missing heritability of apolipoprotein-associated disease traits, and pinpoint multiple novel apolipoprotein variants with putative population-specific impacts on serum lipid levels.—Zhou, Y., R. Mägi, L. Milani, and V. M. Lauschke. Global genetic diversity of human apolipoproteins and effects on cardiovascular disease risk. *J. Lipid Res.* 2018. 59: 1987–2000.

**Supplementary key words** cholesterol • lipid traits • Alzheimer's disease • population genetics

L.M. is supported by Eesti Teadusagentuur (Estonian Research Council) Grant PRG184 and by an Uppsala Universitet Strategic Research Grant as part of the Science for Life Laboratory fellowship program. V.M.L. is supported by Swedish Research Council Grants 2016-01153 and 2016-01154, the Strategic Research Programme in Diabetes at Karolinska Institutet, the Malin and Lennart Philipson Foundation, and the Harald och Greta Jeansson's Stiftelse. V.M.L. is a co-founder and owner of HepaPredict AB.

\***Author's Choice**—Final version open access under the terms of the Creative Commons CC-BY license.

Manuscript received 8 May 2018 and in revised form 16 July 2018.

Published, *JLR Papers in Press*, August 3, 2018

DOI <https://doi.org/10.1194/jlr.P086710>

Copyright © 2018 Zhou et al. Published by The American Society for Biochemistry and Molecular Biology, Inc.

This article is available online at <http://www.jlr.org>

CVD is a major cause of morbidity and mortality with an estimated worldwide aggregated lifetime risk of over 60% (1). Globally, 422 million patients suffer from CVD, and CVD causes more than 4 million deaths annually in Europe alone, accounting for 45% of all deaths (2, 3). Moreover, 10–18% of the number of years lost due to poor health, disability, or early death are allotted to CVD, resulting in a projected global cost of USD 47 trillion worldwide in the next 25 years (4).

Abnormal plasma lipid profiles constitute the single most important risk factor for CVD (5–7). Particularly, elevated non-HDL cholesterol levels and hypertriglyceridemia are considered as the main risk indicators (8, 9). Mechanistically, LDLs and VLDLs, as well as their remnants, can penetrate the endothelial lining of the arterial walls and be retained in the underlying intima, where they promote inflammation, as well as atherosclerotic plaque formation and progression (10).

Lipoprotein cargo, processing, and transport are tightly linked to apolipoprotein composition, and measurement of apolipoprotein plasma concentrations instead of lipid levels can serve as a simplified method for risk assessment in vascular disease (11, 12). In total, the human genome encodes 21 apolipoproteins, of which APOAs (APOA1, APOA2, APOA4, and APOA5), APOCs (APOC1, APOC2, and APOC3), APOB, and APOE have been demonstrated to play essential roles in triglyceride and cholesterol transport and metabolism (13). Furthermore, the atypical apolipoproteins APOH and APOM have been convincingly

Abbreviations: CAD, coronary artery disease; CNV, copy number variant; GWAS, genome-wide association study; LDLR, LDL receptor; MAF, minor allele frequency; NAFLD, nonalcoholic fatty liver disease; NGS, next-generation sequencing; OMIM, Online Mendelian Inheritance in Man; OR, odds ratio; SNV, single nucleotide variant; UTR, untranslated region.

<sup>1</sup>To whom correspondence should be addressed.

e-mail: [volker.lauschke@ki.se](mailto:volker.lauschke@ki.se)

<sup>S</sup>The online version of this article (available at <http://www.jlr.org>) contains a supplement.

implicated in the regulation of postprandial triglyceride clearance, as well as lipid and HDL metabolism (14–16).

Seminal studies have firmly established the importance of selected SNPs in *APO* genes as important genetic risk factors for dyslipidemias as well as its comorbidities and sequelae. Prominent examples include the association between familial hypercholesterolemia [Online Mendelian Inheritance in Man (OMIM) identifier 144010]; increased risk of ischemic heart disease and genetic variants in *APOB* (17); and the association of the *APOA5* variants with hypertriglyceridemia (OMIM identifier 145750) in African-Americans, Spanish, and Caucasians (18); as well as *APOE* genotypes with circulating lipid levels and with coronary risk (19, 20). Besides constituting a risk factor for CVD, genetic variants in *APOE* are strongly linked to the risk of developing Alzheimer's disease (OMIM identifier 104310) with the  $\epsilon 4$  haplotype increasing risk about 3.7-fold per copy, whereas the  $\epsilon 2$  allele is neuroprotective with an odds ratio (OR) of 0.5 per copy (21, 22).

Although these and other studies have provided important data on the prevalence of selected *APO* variants with clinical importance and their interethnic differences, frequencies of the majority of variants were mostly assessed in few heterogeneous populations with relatively small sample sizes. Furthermore, the genetic variability in *APO* genes beyond the interrogated selected subset of SNPs has not been systematically addressed. Importantly, the rapidly increasing extent of available next-generation sequencing (NGS) data provided by a multitude of population-scale sequencing projects allows us for the first time to comprehensively analyze and portray the landscape of genetic diversity and interethnic variability in *APO* genes across major worldwide populations.

In this study, we integrated whole-exome sequencing and whole-genome sequencing data from 138,632 individuals across seven major human populations to comprehensively profile the genetic diversity of 11 *APO* genes with the highest relevance for human lipid and cholesterol metabolism and transport. Based on these data, we provide a consolidated overview of population-specific frequencies of clinically important *APO* variants on an unprecedented scale. In addition, we analyzed the overall pattern of *APO* genetic diversity and identified 6,875 genetic variants, 2,270 of which were novel. We predict the functional impact of this genetic variability using computational predictions and by mapping variants to the domain structures of *APOB* and *APOE* and provide estimates for the functional importance of rare genetic variants and their contribution to the unexplained heritability in lipidemic phenotypes. By leveraging genome-wide association study (GWAS) data provided by the CARDIoGRAM (Coronary Artery Disease Genomewide Replication and Meta-analysis) and Global Lipids Genetics consortia, we confirm the overall accuracy of our computational variant assessments for predicting genetic associations with blood lipid traits and coronary artery disease (CAD) risk. The presented data constitute, to our knowledge, the most comprehensive overview of genetic variability in apolipoproteins published to date and provide important information to refine population-specific

genotyping strategies for dyslipidemias, as well as CVD and neurological disease risk.

## METHODS

### Data sources

*APO* variants and their frequencies were derived from human sequencing data (123,136 whole-exome sequences and 15,496 whole-genome sequences) of 138,632 individuals (63,369 non-Finnish Europeans, 12,897 Finnish, 12,020 Africans, 9,435 East Asians, 15,391 South Asians, 17,210 Latinos, 5,076 Ashkenazi Jews, and 3,234 from other populations) provided by the Genome Aggregation Database (GnomAD) (23). Apolipoprotein copy number variant (CNV) data from 56,945 individuals were obtained from the Exome Aggregation Consortium repository. *APOE* and *APOB* protein domain structures were derived from Uniprot (<http://www.uniprot.org>) and the published literature (24, 25). The provided numbering of amino acid positions includes the signal peptide sequence. The suitability of short-read sequencing technologies for genomic profiling of human *APO* loci was determined on the basis of GC content and paralogue similarities, which were identified using Ensembl BioMart, and the fraction of inaccessible genome for each gene using the data provided by the 1000 Genomes Project ("strict mask") in Python (supplemental Table S1) (26).

### In silico predictions

We assessed the functionality of all missense variants using five current computational functionality prediction algorithms (SIFT, Polyphen2, MutationAssessor, PROVEAN, and DANN) through ANNOVAR (27). In addition, variants were considered deleterious when they resulted in frameshifts, premature stop codons, loss of the start codons, or disruption of splice donor or acceptor sites. Predictive performance, utilized thresholds, and descriptions of the underlying assessment parameters, as well as the associated references, are provided in supplemental Table S2. We classified a variant as putatively functional if at least two methods predicted a deleterious effect. Signal peptides were analyzed using SignalP (Version 4.1) (28) and Signal-3L (Version 2.0) (29). Variants whose functionality could not be predicted by any algorithm were excluded.

### Variant and haplotype frequency analyses

Novel variants were defined relative to SNP database (dbSNP) release 135. Total numbers and aggregated frequencies of functional variants were calculated by averaging the values generated by five predictive algorithms. Representative haplotype frequencies were calculated by integrating variant frequencies with population-specific linkage information from the 1000 Genomes Project using LDLink (30). Rare and common variants were defined as variants with minor allele frequency (MAF)  $\leq 0.01$  and  $MAF > 0.01$ , respectively. The fraction of functional variability allotted to rare variants was computed as the aggregated frequency of rare functional variants divided by the total frequency of all functional variants for each gene.

### Associations with GWAS data

Genetic association data for lipid traits were obtained from the Global Lipids Genetics Consortium for 114,229 individuals; fasting lipid profiles were available for 97,248 (85.1%) of these (31). GWAS data for CVD risk was obtained from the CARDIoGRAM consortium [22,233 individuals with CAD and 64,762 controls,

respectively (32). The functionality of genetic variants identified in GnomAD for which GWAS data were available was predicted as above, and effect sizes were compared between putatively deleterious and neutral variants.

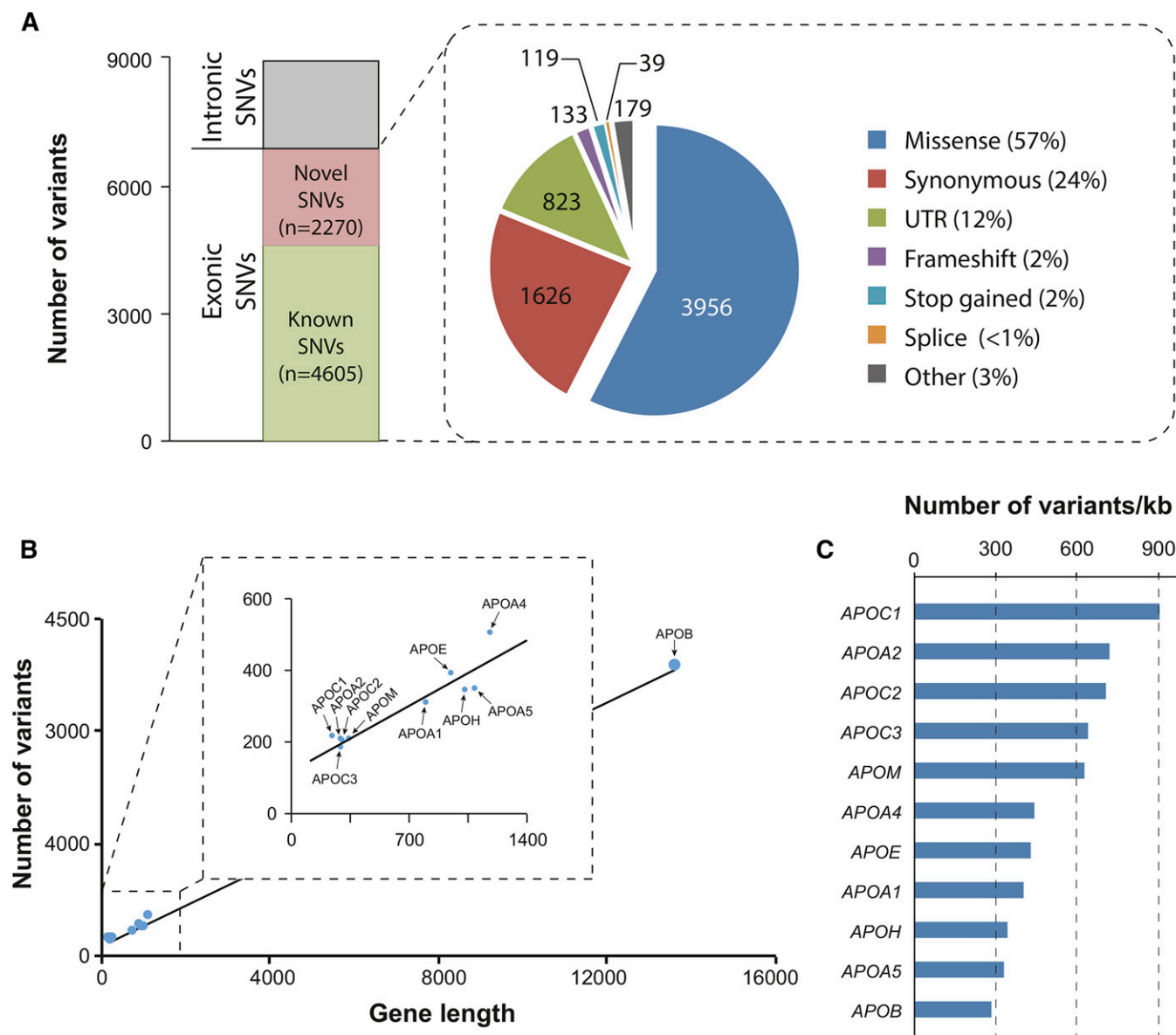
## RESULTS

### Overview of the genetic variability profile in human *APO* genes

In this study, we analyzed the genetic variability in 11 human *APO* loci with documented clinical relevance using exome and whole-genome sequencing data from 138,632 unrelated individuals from seven major human populations. In total, we identified 8,886 variants, of which 6,875

variants are located in exons (Fig. 1A). The majority of exonic variants are missense variants ( $n = 3,956$ ; corresponding to 57% of all exonic variants), synonymous variants ( $n = 1,626$ ; 24%), and variants in the untranslated region (UTR) of the mRNA ( $n = 823$ ; 12%). Notably, 2,270 of these 6,875 variants are identified as novel as compared with dbSNP release 135 (Fig. 1A).

Most variants were found in the *APOB* gene ( $n = 3,839$ ), followed by *APOA4* ( $n = 524$ ), *APOE* ( $n = 408$ ), and *APOA5* ( $n = 363$ ; Fig. 1B). Yet, when analyzing the overall mutational constraints in each gene by normalizing the number of single nucleotide variants (SNVs) to the length of the corresponding transcript, we found that the gene encoding the essential apolipoprotein *APOB* was overall most



**Fig. 1.** Landscape of genetic variability in human *APO* genes. **A:** Stacked column and pie chart showing the variant composition of the 11 analyzed *APO* genes. In total, 8,886 variants were identified in 138,632 individuals, of which 6,875 are located in exons. The majority of exonic variants are missense mutations (57%), followed by synonymous variants (24%). **B:** Scatter plot in which the number of variants identified in each gene is plotted against the respective gene length. Linear regression line is shown. *APOB* is by far the largest *APO* gene and harbors the most variants. **C:** However, when the number of variants is normalized by gene length, *APOB* was the most conserved, harboring approximately three times less variants per kilobase compared with the least conserved gene, *APOC1*.



highly conserved (280 SNVs per kilobase of transcript), whereas the exchangeable apolipoproteins *APOC1* (908 SNVs per kilobase), *APOA2* (717 SNVs per kilobase), and *APOC2* (706 SNVs per kilobase) showed the highest mutational burdens (Fig. 1C). Of note, however, compared with all genes in the human genome, evolutionary constraints on apolipoproteins was overall rather low (supplemental Table S3).

### Worldwide frequencies of clinically important variants in the human *APO* gene family

Genetic variants in apolipoprotein have been reproducibly linked to alterations in serum lipid profiles, progression of atherosclerosis, risk of coronary heart disease, and nonalcoholic fatty liver disease (NAFLD) (33–37). Here, we analyzed the prevalence of 38 human *APO* alleles with the highest clinical relevance in major human populations. In *APOB*, we analyzed nine SNVs (Table 1). The missense variant rs1042031 that was strongly associated with reduced risks of ischemic cerebrovascular disease (hazard ratio = 0.5) and stroke (hazard ratio = 0.2) in the 23-year prospective Copenhagen City Heart Study (38) differed across populations with MAFs between 4.8% in East Asians and 18.3% in Europeans. Similarly, the allele with the lead SNP rs1042034, which correlated with reduced triglyceride levels in a metaanalysis of >100,000 individuals of European ancestry ( $P < 10^{-45}$ ) (39), differed drastically in population frequencies between 27.3% in East Asians and 85.2% in African populations. In contrast, SNV rs1367117, which reproducibly correlated with increased LDL cholesterol in large metaanalyses of >180,000 individuals ( $P < 10^{-196}$ ) (40), was most and least prevalent in Europeans (MAF = 31.9%) and Africans (MAF = 11.2%), respectively. One additional *APOB* variant with strong associations with ischemic heart disease (rs5742904; OR = 7) was rare in every population analyzed (17).

Genetic variability in the *APOE* gene has been reproducibly associated with differences in LDL cholesterol (19). Furthermore, *APOE* polymorphisms are the strongest risk factor for Alzheimer's disease across genders and ethnicities with ORs for homozygous carriers of the  $\epsilon 4$  allele between 2.2 in Latinos and 33.1 in East Asian populations (41, 42). In agreement with previous reports (43), we found the highest frequencies of the  $\epsilon 2$  and  $\epsilon 4$  alleles in African populations, with a frequency of 10.8% and 21.7%, respectively (Table 1). In contrast, the  $\epsilon 2$  allele was least prevalent in Latino populations (frequency = 3.2%), whereas the frequency of the  $\epsilon 4$  allele was lowest in East Asians (frequency = 8.9%).

In addition, we analyzed the frequencies of 10 variant alleles in *APOA1*, *APOA2*, *APOA4*, and *APOA5* that have been consistently linked to hypertriglyceridemia and elevated risks of CAD (Table 1). *APOA5\*2* and *APOA5\*3* are both independently associated with high plasma triglyceride levels (18, 44). Importantly, whereas *APOA5\*2* was highly prevalent in East and South Asians (allele frequency = 23.8% and 17.7%, respectively), the allele was absent in Africans. In contrast, *APOA5\*3* was common in Europeans, Africans, Latinos, Ashkenazim, and South Asians with allele

frequencies between 3.8% and 15.3%, but the variant was very rare in East Asian populations. The missense variant rs2075291 that strongly increases the likelihood of developing hypertriglyceridemia (OR = 11.7) (45) was only common in East Asians (MAF = 6.9%), whereas it was rare or absent in all other populations studied.

The loss-of-function mutations rs147210663, rs76353203, and rs138326449 in *APOC3* have been robustly linked to favorable lipid profiles and reduced risks of CVD (46, 47). rs147210663 has been reported to have particularly strong effects on serum triglyceride levels in Pima Indians, lowering triglyceride levels by 42% (48). In our data set, the frequency of this variant was highest in Ashkenazi Jews (MAF = 1.1%) and rare in all other populations, including Latinos (MAF < 0.1%). Similarly, the stop-gain variant rs76353203 and the splice mutation rs138326449 were rare or very rare across all populations (Table 1).

Missense and promoter SNVs in the genes encoding the atypical apolipoproteins *APOH* and *APOM* showed drastic interpopulation differences (Table 1). Prevalence of the *APOH* variant rs1801689, which was strongly associated with increased serum LDL (31), was highest in Ashkenazi Jews (MAF = 5.8%) but rare in both African (MAF = 0.5%) and East Asian (MAF < 0.1%) populations. Conversely, rs805296 located in the *APOM* promoter was most abundant in Africans (MAF = 11.5%) and East Asians (MAF = 11.6%), but less prevalent in Europeans (MAF = 1.3%) and Ashkenazim (MAF = 1%). Combined, the presented data reveal the extent of interpopulation differences in apolipoprotein alleles with demonstrated clinical relevance and provide a powerful resource for researchers and clinicians to design population-specific genotyping strategies for biomarker identification and disease-risk analyses.

### Rare genetic variants are predicted to contribute substantially to the functional variability in human apolipoproteins

Importantly, the vast majority of identified exonic variants were rare (98.7%) or very rare (96.9%) with MAFs  $\leq 1\%$  or  $\leq 0.1\%$ , respectively, highlighting the genetic complexity of human *APO* genes (Fig. 2A). To estimate the contribution of rare genetic variants to the lipid trait variability, we predicted the overall genetically encoded functional variation in human apolipoproteins using population-scale NGS data and compared the impacts of common and rare *APO* variants. To this end, we used five functionality prediction algorithms and ensemble scores (SIFT, Polyphen2, MutationAssessor, PROVEAN, and DANN) that predict the functional impact of a variant based on a diverse set of features, including amino acid properties, secondary structure, and evolutionary conservation (supplemental Table 2).

Based on these algorithms, we detected overall 1,829 different putatively functional SNVs across human *APO* genes (27% of all exonic variants), and rare variants were significantly enriched in mutations with deleterious effects (30% of all exonic rare variants were predicted to be deleterious compared with 12.2% of common variants; Fig. 2B and supplemental Table S4). Additionally, we found novel rare CNVs in

TABLE 1. Genetic diversity of a selection of clinically important APO variants across major human populations

Defining variants as RSID (HGVS)	Population frequencies (in %)							Variant type	Clinical parameters	Effect size or strength of association	Reference
	EUR	AFR	EAS	SAS	AMR	AJ	AJ				
<i>APOB</i>											
rs5742904 (NC_000002.11:g.21229160C>T)	<0.1	<0.1	0	0	<0.1	0	0	Missense (R3527Q)	Ischemic heart disease	OR = 7	(17)
rs1042031 (NC_000002.11:g.21225753C>T)	18.3	15.4	4.8	10.3	12.3	14.8	14.8	Missense (E4181K)	Ischemic cerebrovascular disease	HR = 0.5	(38)
rs1367117 (NC_000002.11:g.21263900G>A)	31.9	11.2	12.7	16	28.9	18.7	18.7	Missense (T98I)	Ischemic stroke	HR = 0.2	(40)
rs1042034 (NC_000002.11:g.21225281C>T)	78.4	85.2	27.3	48.6	74.2	80.5	80.5	Missense (S4338N)	CAD	$\beta=0.035$	(39)
rs693 (NC_000002.11:g.21232195G>A)	50	22.1	5.6	26.7	38	34.9	34.9	Synonymous (T2515T)	LDL-C	$\beta=0.12$	(40)
rs562338 (NC_000002.11:g.21288321A>G)	18	59.6	0.19	N.A.	16.5	29.8	29.8	UTR	LDL	$\beta'=4.05$	(39)
rs754523 (NC_000002.11:g.21311691A>G)	31.1	21.9	29.5	N.A.	29.1	29.1	29.1	UTR	TG	$\beta'=5.99$	(39)
rs515135 (NC_000002.11:g.21286057T>C)	18.2	47.7	9.5	N.A.	18.8	30.1	30.1	UTR	LDL-C	$P=7.1*10^{-7}$	(71)
rs673548 (NC_000002.11:g.21237544G>A)	22.7	21.1	73	N.A.	22	14.2	14.2	Intron	LDL	$\beta=0.123$	(72)
<i>APOE</i>											
$\epsilon 3$	77.4	67.5	83.6	85.7	86.4	80.6	80.6	Wild-type	CAD	OR = 0.6-2.6	(41)
$\epsilon 2$ :rs7412 (NC_000019.9:g.45412079C>T)	7.7	10.8	7.5	4.2	3.2	7.8	7.8	Missense (R176C)	AD	OR = 2.2-33.1	(41)
$\epsilon 4$ :rs429358 (NC_000019.9:g.45411941T>C)	14.9	21.7	8.9	10.1	10.4	11.6	11.6	Missense (C130R)	AD	OR = 1.06	(19)
rs4420638 (NC_000019.9:g.45422946A>G)	18.4	20.2	11.4	N.A.	10.8	14.6	14.6	UTR	CAD	OR = 0.51	(74)
rs439401 (NC_000019.9:g.45414451T>C)	36.8	15.3	58.6	N.A.	54.9	43	43	UTR	NAFLD	$\beta'=7.14$	(39)
<i>APOA1</i>											
rs670 (NC_000011.9:g.116708413C>T)	17.3	14.8	27.8	N.A.	27.7	15.2	15.2	Promoter	LDL	$P=3.4*10^{-13}$	(71)
<i>APOA2</i>											
rs5082 (NC_000001.10:g.161193683G>A)	40.6	22.5	8.2	N.A.	23.4	35.8	35.8	Promoter	TG	$\beta=0.19$	(75)
<i>APOA4</i>											
rs675 (NC_000011.9:g.116691675T>A)	19.7	11.5	<0.1	12.9	9.4	21.5	21.5	Missense (T367S)	TC	$\beta'=5.5$	(39)
rs5110 (NC_000011.9:g.116691634C>A)	7.8	1.5	<0.1	1.9	3.6	6.7	6.7	Missense (Q380H)	LDL	OR = 1.66	(76)
rs1729407 (NC_000011.9:g.116677370C>G)	50.8	12.7	30.4	N.A.	50.7	35.4	35.4	Intergenic	LDL	OR = 1.77	(76)
<i>APOA5</i>											
*2:rs662799, rs651821, rs2072560, rs2266788 (NC_000011.9:g.116663707G>A, 116662579C>T, 116661826T>C, 116660686G>A)	8.1	0	23.8	17.7	13	N.A.	N.A.	Promoter, Kozak, Intron and UTR	Obesity	OR = 1.84	(77)
*3:rs3135506 (NC_000011.9:g.116662407G>C)	6.4	6.2	<0.1	3.8	15.3	6.8	6.8	Missense (S19W)	CAD	HR = 2.07	(78)
rs662799 (NC_000011.9:g.116663707G>A)	6.9	12.1	29.2	N.A.	15	10.6	10.6	Promoter	TG	$P=0.035 (+)$	(79)
rs2266788 (NC_000011.9:g.116660686G>A)	7.4	1.6	21	N.A.	13.6	10.9	10.9	UTR	VLDL	$P=0.035 (+)$	(80)
rs2075291 (NC_000011.9:g.116661392C>A)	<0.1	0.3	6.9	0.8	<0.1	0	0	Missense (G185C)	HDL	$P=0.0005 (-)$	(80)
20-30% elevation											
<i>TG</i>											
<i>TG</i>											
<i>TG</i>											
<i>TG</i>											
<i>HDL</i>											
<i>TG</i>											
<i>CAD</i>											
<i>TG</i>											
<i>CAD</i>											

TABLE 1. Continued.

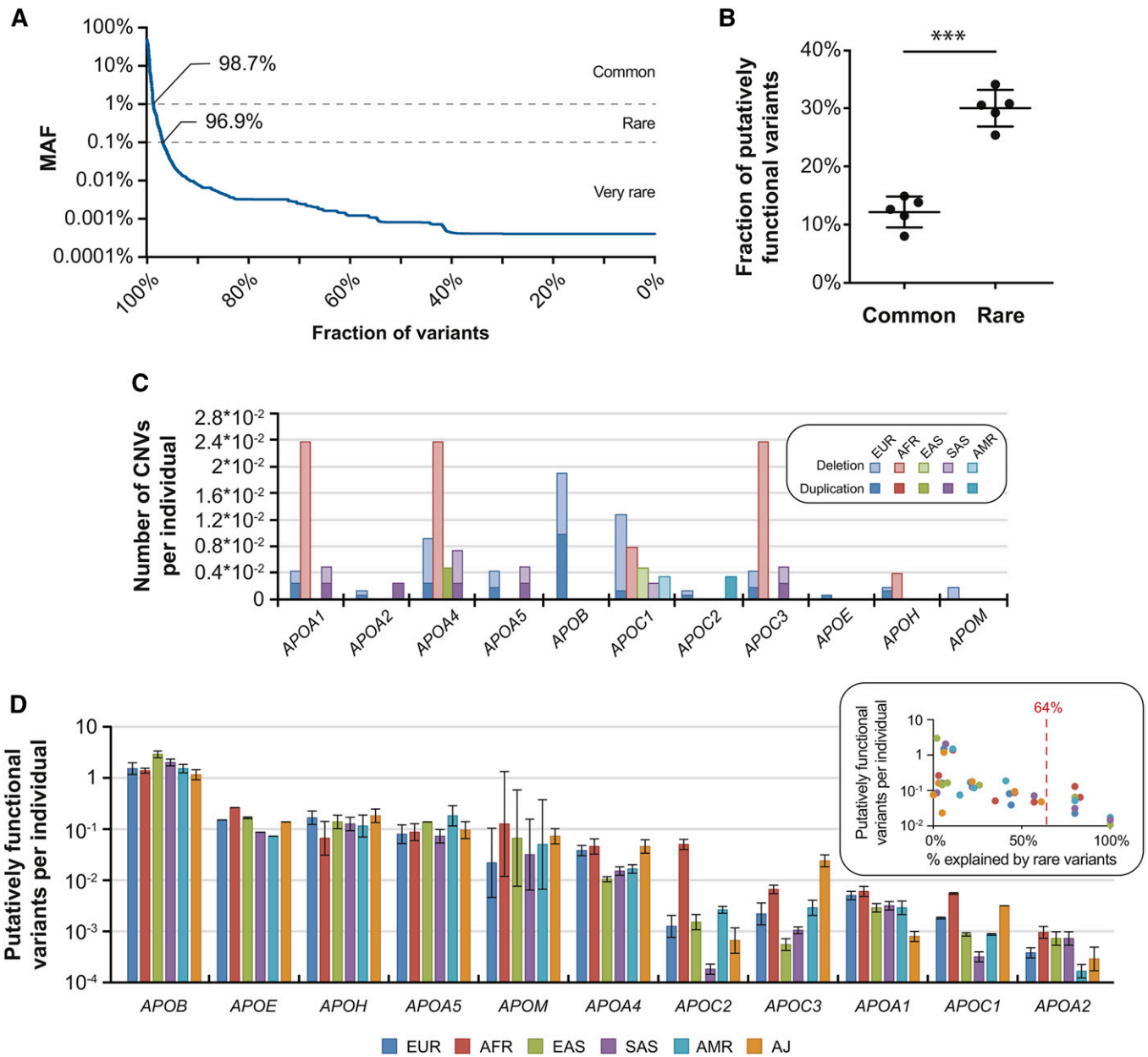
Defining variants as RSID (HGVS)	Population frequencies (in %)							Clinical parameters	Effect size or strength of association	Reference
	EUR	AFR	EAS	SAS	AMR	AJ				
<i>APOC1</i>										
rs11568822 NC_000019.9:g:45417640_45417641insCGTT	23 <sup>a</sup>	35 <sup>b</sup>	N.A.	N.A.	16.5 <sup>c</sup>	N.A.		AD	OR = 1.84	(86)
<i>APOC3</i>										
rs5128 NC_000011.9:g:116703640G>C	9.5	15.5	30.3	30.3	21.2	14		CAD	OR = 1.3	(87)
rs2854116 NC_000011.9:g:116700169C>T	39.1	70.7	41.6	N.A.	34.8	47		Metabolic syndrome CAD	OR = 1.73 OR = 1.28	(88) (87)
rs2854117 NC_000011.9:g:116700142T>C	28.9	68.3	42.4	N.A.	30.7	35.4		NAFLD TG HDL	$P < 0.001 (+)$ $P = 0.041 (+)$ $P = 0.005 (-)$	(89) (83)
rs147210663 NC_000011.9:g:116701560G>A	<0.1	0.2	<0.1	<0.1	<0.1	1.1		NAFLD TG HDL	$P < 0.001 (+)$ $P = 0.01 (-)$ $P = 0.004 (+)$	(89) (47)
rs76353203 NC_000011.9:g:116701353C>G	<0.1	<0.1	<0.1	0.2	<0.1	0		CAC CHD	OR = 0.35 HR = 0.68	(46)
rs138326449 NC_000011.9:g:116701354G>A	0.2	<0.1	0	<0.1	0.1	0.2		TG	$P = 8^* 10^{-8} (-)$	(90)
<i>APOH</i>										
rs8178847 NC_000017.10:g:64216815C>T	6.7	9.8	6.3	5.1	2.9	6.7		VT	OR = 1.55	(91)
rs1801689 NC_000017.10:g:64210580A>C	3.3	0.5	<0.1	1.6	2.4	5.8		LDL	$P = 1^* 10^{-11} (+)$	(31)
rs1801690 NC_000017.10:g:64208285C>G	5.6	0.8	6.8	4.7	1.8	3.3		TG	$P = 0.018 (+)$	(92)
rs3760291 NC_000017.10:g:64226197G>T	26.1	7.2	6.3	N.A.	10.7	11.9		APOA-I levels TC LDL	$P = 0.026 (-)$ $P = 0.006 (-)$ $P = 0.03 (-)$	(14)
<i>APOM</i>										
rs707922 NC_000006.11:g:31625507G>T	5.2	31.9	15.9	7.2	12.6	18.9		TC LDL	$P = 0.006 (+)$ $P = 0.009 (+)$	(93)
rs805296 NC_000006.11:g:31622893T>C	1.3	11.5	11.6	N.A.	4.5	1		CAD	OR = 1.9	(94)
rs940494 NC_000007.13:g:56348924A>G	22.4	14.1	9.4	N.A.	27.7	17.9		T2DM CAD	OR = 2.29 OR = 1.82	(95) (96)

AD, Alzheimer's disease; AFR, Africans; AJ, Ashkenazi Jews; AMR, Latinos; CAC, coronary artery calcification; CHD, coronary heart disease; EAS, East Asians; EUR, Europeans; HDL, HDL levels; HGVS, Human Genome Variation Society; HR, hazard ratio; LDL, LDL levels; LDL-C, LDL cholesterol; N.A., not available; RA, rheumatoid arthritis; RSID, reference SNP cluster ID; SAS, South Asians; T2DM, T2D mellitus; TC, total cholesterol levels; TG, triglyceride levels; VLDL, VLDL levels; VT, venous thrombosis;  $\beta$ , standardized effect size;  $\beta^*$ , nonstandardized effect size.

<sup>a</sup>Obtained from Gayà-Vidal et al. (97).

<sup>b</sup>Obtained from Gao et al. (98).

<sup>c</sup>Obtained from Lucatelli et al. (99).



**Fig. 2.** Rare genetic variants are important contributors to the functional variability of apolipoproteins. **A:** The vast majority of the identified apolipoprotein variants were rare (98.7%) or very rare (96.7%) with MAF < 1% or < 0.1%, respectively. **B:** The fraction of common and rare *APO* variants with putatively functional effects, as predicted by five computational algorithms. Note that all algorithms indicate that rare variants are enriched in mutations with functional effects. \*\*\*  $P < 0.0001$  in paired heteroscedastic *t*-test. Error bars indicate SD. **C:** The number of CNVs per individual in *APO* genes is shown for six major human populations. AFR, Africans; AMR, Latinos; EAS, East Asians; EUR, Europeans; SAS, South Asians. Deletions and duplications are indicated in light and dark shades, respectively. Note that *APOA1*, *APOA4*, and *APOC3* are in the same locus. **D:** The number of variants with putatively functional consequences per individual is shown for each analyzed *APO* gene across six populations. The inset indicates the fractions of putatively functional variants that are explained by rare variants (see also supplemental Fig. S1). Average value is indicated by red dashed line. Genes with an aggregated functional variant frequency is < 1% are by definition explained exclusively by rare variants and are not shown.

all *APO* genes studied, many of which were population-specific (Fig. 2C and supplemental Table S4). Deletions of the *APOA1/APOA4/APOC3* locus in Africans (frequency = 0.11%) and CNVs of the *APOB* gene in Europeans (0.09%) were overall most common. In contrast, very few CNVs were observed in *APOA2*, *APOM*, *APOE*, and *APOC2*.

When combining information of putatively functional SNVs and CNVs, we observed that the frequency of genetically

encoded functional variability differed by more than 4,000-fold between *APO* genes. Of all analyzed *APO* genes, *APOB* harbored the most variants, with predicted functional impacts (1.8 functional variants per individual), whereas *APOA2* was highly invariant ( $5.4 \times 10^{-4}$  functional variants per individual; Fig. 2D). In addition to these drastic differences in functional variability between genes, we observed large variability across populations. For *APOC2*, the genetically



encoded functional variability differed 280-fold between  $1.8 \times 10^{-4}$  variants per individual in South Asians and 0.051 in Africans. In contrast, functional variability in *APOA5* (0.071–0.19 per individual), *APOB* (1.2–3 per individual), and *APOH* (0.065–0.18 per individual) differed less than 3-fold across populations (Fig. 2D).

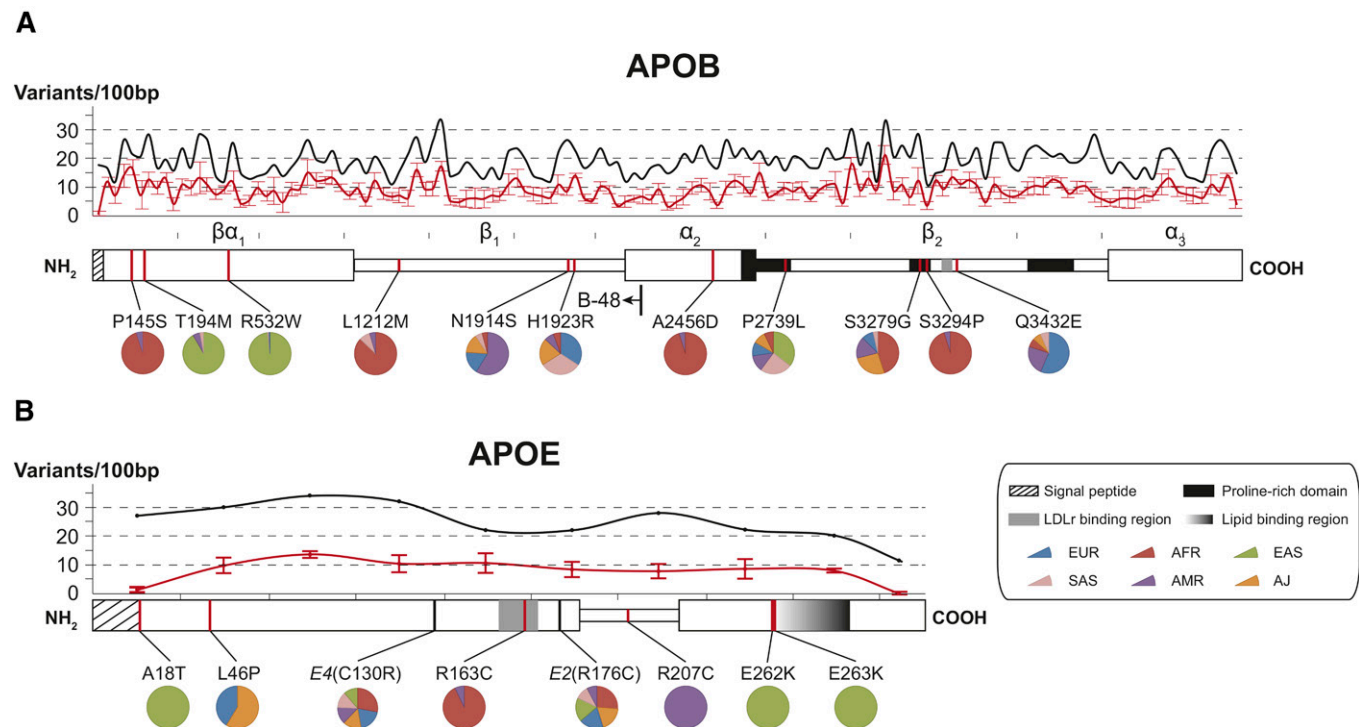
Based on our predictions, we estimate that rare genetic variability accounts for 8% and 10% of the genetically encoded functional variability in *APOE* and *APOB*, respectively (supplemental Fig. S1). In contrast, the functional variability of *APOA4*, *APOM*, *APOA1*, *APOA2*, and members of the *APOC* gene family was fully allotted to rare variants. On average, across *APO* genes and populations, we predict that an average of 64% of the functional variability in coding sequences was due to rare genetic variants (Fig. 2D, inset). Combined, these results suggest that a substantial fraction of the genetically encoded functional variability in apolipoproteins is missed when only considering common genetic variants, thus incentivizing the consideration of rare SNVs in apolipoprotein-encoding genes for the guidance of personalized disease-risk predictions.

### Structural variability profiles of *APOB* and *APOE*

Next, we mapped the genetic variability in *APOB* and *APOE* onto the respective protein domain structures (Fig. 3). With the exception of the N-terminal sequence encoding the signal peptide, genetic variability in *APOB* fluctuates

between 11 and 34 variants per 100 bp, whereas the numbers of putatively deleterious variants range from 3.4 to 21.4 variants per 100 bp (Fig. 3A). The functionally most conserved regions are located in the amphipathic lipid-associating  $\alpha$ -helix  $\alpha_2$  and in the proline-rich domains of the  $\beta_2$   $\beta$ -sheet that assumes a confirmation parallel to the phospholipid monolayer of LDL (49), with 3.4 and 3.6 variants per 100 bp, respectively. Notably, whereas the proline-rich domains in  $\beta_2$  are highly conserved (3.6 variants per 100 bp), substantially more variants were found in the adjacent LDL receptor (LDLR) binding site (9.4–13.6 variants per 100 bp). Regions of overall highest putatively functional variability are located in  $\beta$ -sheets  $\beta_1$  and  $\beta_2$  (17.4 and 21.4 variants per 100 bp, respectively).

To pinpoint novel variants with putative functional consequences and clinical relevance, we filtered variants that were classified as deleterious with high confidence by all algorithms employed and that were common with MAF > 1% in at least one population studied (Table 2). The population-specific variants rs6752026, rs13306198, and rs13306194 are located in the N-terminal  $\alpha\beta_1$  domain that forms a lipid pocket that is necessary for VLDL as well as chylomicron particle assembly (50). Variant rs676210, which has been previously implicated in differential lipid-lowering response to fenofibrate in Europeans (51), mapped to the proline-rich domain at the interface between the  $\alpha_2$  and  $\beta_2$  regions and was the most prevalent of the putatively functional



**Fig. 3.** Structural variability of *APOB* and *APOE*. A: Domain map of *APOB* containing signal peptide (amino acids 1–27), proline-rich domains (amino acids 2,578–2,767; 3,243–3,318; and 3,714–3,892) (see Ref. 99), and the LDLR binding region (amino acids 3,386–3,396) (see Ref. 25). Additionally, the C terminus of the truncated *APOB48* isoform is indicated. B: The domain map of *APOE* is shown with signal peptide (amino acids 1–18), LDLR binding region (amino acids 152–168), and lipid-binding region (amino acids 262–290) highlighted (see Ref. 100). Aligned line plots indicate the total number of variants (black line) and number of functional variants (red line) per 100 bp. Functionality is predicted by five orthogonal computational algorithms, and the average  $\pm$  SD is shown. Variants that were common (MAF > 1%) in at least one population and deemed functional by all employed algorithms are highlighted. Pie charts indicate their relative abundance in the six major populations analyzed. AFR, Africans; AJ, Ashkenazi Jews; AMR, Latinos; EAS, East Asians; EUR, Europeans; SAS, South Asians.



TABLE 2. Selection of putatively functional variants in *APOB* and *APOE*

Defining Variants as RSID (HGVS)	Variant Type	Population Frequencies (%)						Global Lipid Genetics Consortium			CARDIoGRAM	
		EUR	AFR	EAS	SAS	AMR	AJ	Effect Size	P	Log <sub>e</sub> odds	P	
<i>APOB</i>												
rs6752026 (NC_000002.11:g:21260934G>A)	Missense (P145S)	<0.1	12.8	0	<0.1	0.7	<0.1	N.A.	N.A.	N.A.	N.A.	
rs13306198 (NC_000002.11:g:21260084C>A)	Missense (T194M)	<0.1	<0.1	5.5	0.2	0.3	0	N.A.	N.A.	N.A.	N.A.	
rs13306194 (NC_000002.11:g:21252534G>A)	Missense (R532W)	0.1	<0.1	13.4	0.2	<0.1	<0.1	N.A.	N.A.	N.A.	N.A.	
rs61736761 (NC_000002.11:g:21238007G>T)	Missense (L1212M)	<0.1	8.9	0	0.9	0.5	<0.1	N.A.	N.A.	N.A.	N.A.	
rs1801699 (NC_000002.11:g:21233999T>C)	Missense (N1914S)	1.9	0.5	<0.1	0.6	6.5	1.6	2.6 × 10 <sup>-6</sup>	N.A.	N.A.	N.A.	
rs533617 (NC_000002.11:g:21233972T>C)	Missense (H1923R)	4	0.7	<0.1	3.7	0.9	2.4	9.6 × 10 <sup>-45</sup>	0.091	0.14	0.034	
rs12713675 (NC_000002.11:g:21232373G>T)	Missense (A2456D)	<0.1	5.3	0	<0.1	0.3	0	N.A.	N.A.	N.A.	N.A.	
rs676210 (NC_000002.11:g:21231524G>A)	Missense (P2739L)	21.6	14.7	72.5	50.2	25.7	19.4	4.1 × 10 <sup>-39</sup>	0.059	0.03	0.077	
rs12720854 (NC_000002.11:g:21229905T>C)	Missense (S3279G)	0.3	1.7	<0.1	0.1	0.6	1	N.A.	N.A.	N.A.	N.A.	
rs12720855 (NC_000002.11:g:21229860A>G)	Missense (S3294P)	<0.1	5.3	0	<0.1	0.3	0	N.A.	N.A.	N.A.	N.A.	
rs1042023 (NC_000002.11:g:21229446G>C)	Missense (Q3432E)	1.1	0.1	0	0.1	0.5	0.1	N.A.	N.A.	N.A.	N.A.	
<i>APOE</i>												
rs533904656 (NC_000019.9:g:45411025G>A)	Missense (A18T)	0	0	0.2	0	0	0	N.A.	N.A.	N.A.	N.A.	
rs769452 (NC_000019.9:g:45411110T>C)	Missense (L46P)	0.3	<0.1	0	<0.1	<0.1	0.5	N.A.	N.A.	N.A.	N.A.	
rs769455 (NC_000019.9:g:45412040G>T)	Missense (R163C)	<0.1	2	<0.1	<0.1	0.2	0	N.A.	N.A.	N.A.	N.A.	
rs749750245 (NC_000019.9:g:45412172C>T)	Missense (R207C)	0	0	0	0	0.2	0	N.A.	N.A.	N.A.	N.A.	
rs140808909 (NC_000019.9:g:45412337G>A)	Missense (E262K)	0	0	0.3	0	0	0	N.A.	N.A.	N.A.	N.A.	
rs190853081 (NC_000019.9:g:45412340G>A)	Missense (E263K)	0	0	0.3	0	0	0	N.A.	N.A.	N.A.	N.A.	

AFR, Africans; AJ, Ashkenazi Jews; AMR, Latinos; EAS, East Asians; EUR, Europeans; HGVS, Human Genome Variation Society; N.A., not available; RSID, reference SNP cluster ID; SAS, South Asians.

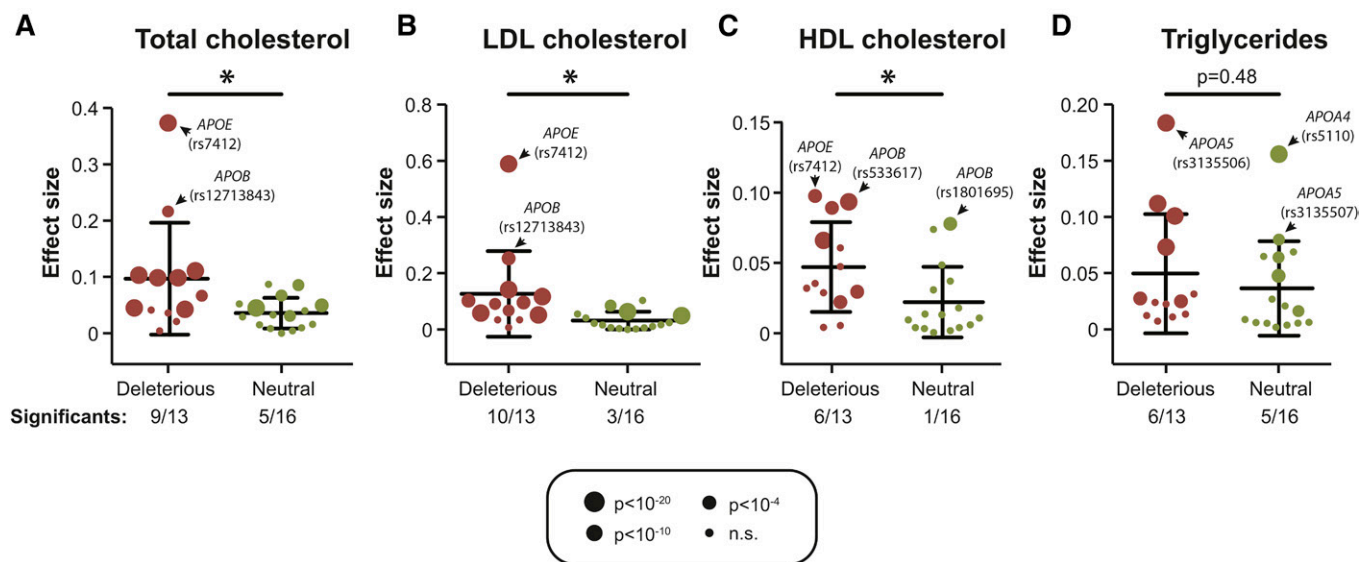
*APOB* variants with population frequencies ranging between 14.7% in Africans and 72.5% in East Asians.

Familial hypobetalipoproteinemia type 1 (FHBL1; OMIM identifier 615558) is caused by genetic variants that result in truncated forms of *APOB* protein. Heterozygous carriers of such variants (1:500 to 1:1,000 in Western populations) are often clinically asymptomatic, whereas individuals homozygous for *APOB* truncating mutations often exhibit very low LDL levels, fat malabsorption in the intestine, hepatic steatosis due to impaired VLDL secretion by the liver, and high prevalence of severe fibrosis (52–54). Additionally, if the truncating variant occurs within the *APOB*48 isoform, chylomicron secretion from the enterocytes is affected. In the 138,632 individuals analyzed here, we found only very few carriers of *APOB* truncating variants (supplemental Table S5). However, four *APOB* missense variants recently associated with FHBL1 were identified with frequencies between 0.8% and 2.5% in Africans (55) (supplemental Table S5).

In contrast to *APOB*, variability in *APOE* was distributed uniformly across the open-reading frame sequence with a local minimum at the LDLR binding region (Fig. 3B). In-depth computational analysis of the *APOE* variant inventory revealed multiple rare SNVs with high-confidence functional consequences. Variants rs533904656 (A18T), rs140808909 (E262K), and rs190853081 (E263K) were specific to East Asian populations with MAFs between 0.2% and 0.3%. The A18T alters the signal peptide sequence and obscures cleavage-site recognition (28, 29), potentially modulating *APOE* secretory efficiency. E262 and E263 are located in the CT domain, and the latter forms salt bridges with R121 and R165 in the LDLR-binding region, which shields the LDLR-binding domain in the absence of bound lipids (24). Thus, we hypothesize that SNVs altering the charge of this residue entail destabilization and favor premature binding of the lipid-free *APOE* to its receptor. A similar weakening of domain interactions can be expected for the R163C mutation that is found in 2% of African alleles and which abolishes interactions with Q59 in the NT domain (24).

### Validation of predicted associations using GWAS data

We aimed to estimate the accuracy of these predictions by leveraging preexisting genotyping data from 114,229 individuals provided by the Global Lipids Genetics consortium (31). Of the 6,875 exonic apolipoprotein variants identified in this study, only 51 overlapped with GWAS data (0.7%; supplemental Table S6). This demonstrates the vast extent of genetic complexity not interrogated by genome-wide and custom genotyping arrays and emphasizes the added value of sequencing-based profiling techniques. Importantly, we found that apolipoprotein variants that were predicted to affect the functionality of the corresponding gene product showed significantly higher effect sizes for cholesterol traits, including total, LDL, and HDL cholesterol ( $P < 0.05$  for each correlation; Fig. 4A–C). Of variants with putatively deleterious effects, 69% (9/13), 77% (10/13), and 46% (6/13) significantly correlated with changes in total, LDL, and HDL cholesterol levels, respectively, whereas only 31%, 19%, and 6% of neutral variants correlated with the respective lipid traits. Although putatively deleterious *APO* variants



**Fig. 4.** Putatively deleterious variants are enriched in mutations with effects on blood lipid traits. Identified *APO* variants were overlaid with GWAS data provided by the Global Lipids Genetics Consortium for total cholesterol (A), LDL cholesterol (B), HDL cholesterol (C), and serum triglycerides (D). Sizes of dots indicate *P* values of the associations between variant and the respective clinical parameter.  $P < 10^{-4}$  indicates significance of association after Bonferroni correction. Importantly, variants predicted to be deleterious (indicated in red) were significantly enriched in mutations affecting lipid traits ( $P < 0.001$ ; chi-squared test) compared with variants predicted to be functionally neutral (indicated in green). When individual lipid parameters were compared, variant associations were significant for cholesterol traits (total, LDL, and HDL cholesterol;  $P < 0.05$ ) but not serum triglyceride levels ( $P = 0.48$ ; heteroscedastic two-tailed *t*-test). \*  $P < 0.05$ ; n.s., not significant.

significantly correlated with cholesterol levels, no significant correlations were observed for serum triglyceride levels ( $P = 0.48$ ; Fig. 4D). Notably, variants with the largest effect sizes on cholesterol traits were observed in *APOE* and *APOB*, whereas variants in *APOA4* and *APOA5* had the highest effects on triglyceride levels.

Because of the intimate link between lipid traits and CVD, we utilized the CARDIoGRAM GWAS data of 22,233 individuals with CAD and 64,762 controls (32) to further assess whether identified *APO* variants directly associated with disease risk. Sixteen missense apolipoprotein variants were covered by the CARDIoGRAM data, of which eight each were predicted to be deleterious or neutral (supplemental Table S6). Notably, whereas three of the putatively deleterious variants nominally correlated with CAD risk ( $P < 0.05$ ; 38%), none of the correlations were found to be significant for the putatively neutral variants (0%). Combined, these data provide proof of concept that computational functionality predictions of apolipoprotein variants can highlight variants with putative effects on blood lipid traits and associated disease risk. We therefore conclude that the integration of NGS-based sequencing methods and in silico variant assessment provides a useful approach for the interpretation of the vast extent of rare or novel variants identified by large-scale sequencing projects for which clinical validations or GWAS data are not available.

## DISCUSSION

Serum lipid levels and risk of CVD are highly heritable with estimates ranging from 30% to 60% (56–58), and genetic

polymorphisms in apolipoprotein-encoding genes constitute important modulators of serum lipid profiles and CVD susceptibility. Here, we analyzed the worldwide frequencies of 38 human *APO* alleles that have been consistently implicated in lipid traits and disease risk (Table 1). Most alleles exhibited large interethnic differences in frequencies, indicating that accurate genetic prediction of dyslipidemia and CVD risk requires population-specific genotyping strategies.

Importantly, common variants identified in large-scale GWASs only explain around 10–20% of the heritability of lipid traits (31, 39, 59). Rare SNVs are enriched in variants with functional effects and large effect sizes (60, 61) and have been suggested as an important source of this unexplained heritability (62, 63). However, the extent to which rare variants contribute to the overall functional variability in apolipoproteins had not been assessed. Thus, we leveraged population-scale NGS data to directly estimate the relative importance of rare genetic variability for the missing heritability of apolipoprotein-associated disease traits. By integrating the results of five partly orthogonal methods, we predict that for *APOE*, *APOB*, *APOH*, and *APOA5*, common variants explain 50–90% of the genetically encoded functional variability in coding sequences. In contrast, no common deleterious variant was detected in *APOC2* and *APOC3*. Thus, rare variants and CNVs are expected to explain the entire functional variability of *APOC3*, a gene for which loss-of-function mutations have been strongly linked with favorable lipid profiles (64), and of *APOC2*, which has been consistently associated with hypertriglyceridemia (65). These findings align with the causal implication of a multitude of rare variants in these genes

with lipid traits, whereas no common deleterious variants in the coding sequences have been described.

Although the presented analyses provide the most comprehensive overview of genetic variability in *APO* genes described to date, it is important to note that variant and allele frequencies can differ drastically between ethnic groups within these aggregated superpopulations (66). Furthermore, genetic profiles of populations not represented in this data set might yield exciting additional information about apolipoprotein diversity. Examples for such insights based on population isolates or founder populations are effects of the triglyceride-lowering *APOC3* variants rs147210663 in Pima Indians (MAF = 2.6%) (48) and rs138326449 in Hutterites (MAF = 2.2%) (67), as well as the LDL risk variant rs5742904 in *APOB* in Old Order Amish (MAF = 12%) (68). With decreasing sequencing costs, we anticipate that the sequencing of founder populations will continue to represent a powerful tool for genetic research of apolipoproteins.

We used in silico prediction algorithms that can distinguish deleterious from functionally neutral missense variants with relatively high confidence as judged by areas under the receiver operating characteristic curve between 0.8 and 0.95 for genome-wide analyses (supplemental Table S2). However, for families of genes with low evolutionary pressure, the quality of predictions can be substantially lower (69). Importantly, we show that functionality scores for apolipoprotein variants were overall predictive for their effects on serum cholesterol levels despite complex linkage disequilibria, which might obscure functional effects (supplemental Fig. S2). For instance, rs1367177 and rs679899 were both predicted to affect APOB functionality and associate clearly with LDL ( $p_{LD} = 9 \times 10^{-183}$  and  $4 \times 10^{-39}$ , respectively) and total cholesterol levels ( $p_{TC} = 2 \times 10^{-139}$  and  $4 \times 10^{-22}$ , respectively). However, these variants are in linkage with the putatively neutral variant rs1801701 ( $R^2$  in Europeans = 0.18), which also correlated with LDL and cholesterol levels, albeit less strongly ( $p_{LDL} = 8 \times 10^{-15}$  and  $p_{TC} = 2 \times 10^{-21}$ ).

Although this predictive power is sufficient to provide faithful estimates of the overall functional mutational burden in the coding sequence of loci of interest on a population scale, these functional analyses are currently limited to individuals of European descent, as functional GWAS data for other populations of similar extent are currently lacking. Moreover, extensions of computational algorithms are needed to further improve the prediction of personalized dyslipidemia risks. Potential refinements include the consideration of population-specific linkage information, as well as adjustments that allow the functional interpretation of regulatory variants in promoters, enhancers, or UTRs that cannot currently be accurately evaluated using most computational methods. Finally, the evaluation of expression quantitative trait loci in relevant tissues, such as liver and small intestine, as provided by the Genotype-Tissue Expression Consortium, might allow further mechanistic interpretations of genetic apolipoprotein variation (70).

In summary, our analyses reveal that the genetic landscape in human apolipoproteins is highly complex, and

every individual was found to harbor on average 19 *APO* variants, of which 2 had putative functional effects. The vast majority of variants were rare, and these rare variants contributed substantially to the genetically encoded apolipoprotein variability. Furthermore, by leveraging GWAS data from the CARDIoGRAM and Global Lipids Genetics consortia, we found that computational methods provide overall useful predictions for the functional effects of apolipoprotein variants on lipid traits and apolipoprotein-associated disease risk. **FIG**

The authors thank the CARDIoGRAM, Global Lipids Genetics, and Exome Aggregation consortia and all contributing groups for sharing their data, which were instrumental for this work.

## REFERENCES

1. Benjamin, E. J., M. J. Blaha, S. E. Chiuve, M. Cushman, S. R. Das, R. Deo, S. D. de Ferranti, J. Floyd, M. Fornage, C. Gillespie, et al. 2017. Heart disease and stroke statistics—2017 update: a report from the American Heart Association. *Circulation*. **135**: e146–e603.
2. Townsend, N., L. Wilson, P. Bhatnagar, K. Wickramasinghe, M. Rayner, and M. Nichols. 2016. Cardiovascular disease in Europe: epidemiological update 2016. *Eur. Heart J.* **37**: 3232–3245.
3. Roth, G. A., C. Johnson, A. Abajobir, F. Abd-Allah, S. F. Abera, G. Abyu, M. Ahmed, B. Aksut, T. Alam, K. Alam, et al. 2017. Global, regional, and national burden of cardiovascular diseases for 10 causes, 1990 to 2015. *J. Am. Coll. Cardiol.* **70**: 1–25.
4. Laslett, L. J., P. Alagona, B. A. Clark, J. P. Drozda, F. Saldivar, S. R. Wilson, C. Poe, and M. Hart. 2012. The worldwide environment of cardiovascular disease: prevalence, diagnosis, therapy, and policy issues: a report from the American College of Cardiology. *J. Am. Coll. Cardiol.* **60**: S1–S49.
5. Lewington, S., G. Whitlock, R. Clarke, P. Sherliker, J. Emberson, J. Halsey, N. Qizilbash, R. Peto, and R. Collins. 2007. Blood cholesterol and vascular mortality by age, sex, and blood pressure: a meta-analysis of individual data from 61 prospective studies with 55,000 vascular deaths. *Lancet*. **370**: 1829–1839.
6. van Holten, T. C., L. F. Waanders, P. G. de Groot, J. Vissers, I. E. Hoefer, G. Pasterkamp, M. W. J. Prins, and M. Roest. 2013. Circulating biomarkers for predicting cardiovascular disease risk; a systematic review and comprehensive overview of meta-analyses. *PLoS One*. **8**: e62080.
7. Graham, I., M-T. Cooney, D. Bradley, A. Dudina, and Ž. Reiner. 2012. Dyslipidemias in the prevention of cardiovascular disease: risks and causality. *Curr. Cardiol. Rep.* **14**: 709–720.
8. Catapano, A. L., I. Graham, G. De Backer, O. Wiklund, M. J. Chapman, H. Drexel, A. W. Hoes, C. S. Jennings, U. Landmesser, T. R. Pedersen, et al. 2016. 2016 ESC/EAS guidelines for the management of dyslipidaemias. *Eur. Heart J.* **37**: 2999–3058.
9. Miller, M., N. J. Stone, C. Ballantyne, V. Bittner, M. H. Criqui, H. N. Ginsberg, A. C. Goldberg, W. J. Howard, M. S. Jacobson, P. M. Kris-Etherton, et al. 2011. Triglycerides and cardiovascular disease: a scientific statement from the American Heart Association. *Circulation*. **123**: 2292–2333.
10. Bentzon, J. F., F. Otsuka, R. Virmani, and E. Falk. 2014. Mechanisms of plaque formation and rupture. *Circ. Res.* **114**: 1852–1866.
11. Brunzell, J. D., M. Davidson, C. D. Furberg, R. B. Goldberg, B. V. Howard, J. H. Stein, and J. L. Witztum. 2008. Lipoprotein management in patients with cardiometabolic risk: consensus conference report from the American Diabetes Association and the American College of Cardiology Foundation. *J. Am. Coll. Cardiol.* **51**: 1512–1524.
12. Di Angelantonio, E., N. Sarwar, P. Perry, S. Kaptoge, K. K. Ray, A. Thompson, A. M. Wood, S. Lewington, N. Sattar, et al. 2009. Major lipids, apolipoproteins, and risk of vascular disease. *JAMA*. **302**: 1993–2000.
13. Dominiczak, M. H., and M. J. Caslake. 2011. Apolipoproteins: metabolic role and clinical biochemistry applications. *Ann. Clin. Biochem.* **48**: 498–515.



14. Leduc, M. S., L. C. Shimmin, K. L. E. Klos, C. Hanis, E. Boerwinkle, and J. E. Hixson. 2008. Comprehensive evaluation of apolipoprotein H gene (APOH) variation identifies novel associations with measures of lipid metabolism in GENOA. *J. Lipid Res.* **49**: 2648–2656.
15. Christoffersen, C., and L. B. Nielsen. 2013. Apolipoprotein M: bridging HDL and endothelial function. *Curr. Opin. Lipidol.* **24**: 295–300.
16. Christoffersen, C., C. K. Federspiel, A. Borup, P. M. Christensen, A. N. Madsen, M. Heine, C. H. Nielsen, A. Kjaer, B. Holst, J. Heeren, et al. 2018. The apolipoprotein M/SIP axis controls triglyceride metabolism and brown fat activity. *Cell Reports.* **22**: 175–188.
17. Tybjaerg-Hansen, A., R. Steffensen, H. Meinertz, P. Schnohr, and B. G. Nordestgaard. 1998. Association of M mutations in the apolipoprotein B gene with hypercholesterolemia and the risk of ischemic heart disease. *N. Engl. J. Med.* **338**: 1577–1584.
18. Pennacchio, L. A., M. Olivier, J. A. Hubacek, R. M. Krauss, E. M. Rubin, and J. C. Cohen. 2002. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. *Hum. Mol. Genet.* **11**: 3031–3038.
19. Bennet, A. M., E. Di Angelantonio, Z. Ye, F. Wensley, A. Dahlin, A. Ahlbom, B. Keavney, R. Collins, B. Wiman, U. de Faire, et al. 2007. Association of apolipoprotein E genotypes with lipid levels and coronary risk. *JAMA.* **298**: 1300–1311.
20. IBC 50K CAD Consortium. 2011. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**: e1002260. [Erratum. 2011. *PLoS Genet.* **7**: e1002260.]
21. Yu, J.-T., L. Tan, and J. Hardy. 2014. Apolipoprotein E in Alzheimer's disease: an update. *Annu. Rev. Neurosci.* **37**: 79–100.
22. Huynh, T.-P. V., A. A. Davis, J. D. Ulrich, and D. M. Holtzman. 2017. Apolipoprotein E and Alzheimer's disease: the influence of apolipoprotein E on amyloid- $\beta$  and other amyloidogenic proteins. *J. Lipid Res.* **58**: 824–836.
23. Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* **536**: 285–291.
24. Chen, J., Q. Li, and J. Wang. 2011. Topology of human apolipoprotein E3 uniquely regulates its diverse biological functions. *Proc. Natl. Acad. Sci. USA.* **108**: 14813–14818.
25. Segrest, J. P., M. K. Jones, H. De Loof, and N. Dashti. 2001. Structure of apolipoprotein B-100 in low density lipoproteins. *J. Lipid Res.* **42**: 1346–1367.
26. 1000 Genomes Project. FTP webserver of the 1000 Genomes Project. Accessed January 3, 2018, at [ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis\\_results/supporting/accessible\\_genome\\_masks/](ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/accessible_genome_masks/)
27. Wang, K., M. Li, and H. Hakonarson. 2010. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**: e164.
28. Petersen, T. N., S. Brunak, G. von Heijne, and H. Nielsen. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Methods.* **8**: 785–786.
29. Zhang, Y.-Z., and H.-B. Shen. 2017. Signal-3L 2.0: a hierarchical mixture model for enhancing protein signal peptide prediction by incorporating residue-domain cross-level features. *J. Chem. Inf. Model.* **57**: 988–999.
30. Machiela, M. J., and S. J. Chanock. 2015. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. *Bioinformatics.* **31**: 3555–3557.
31. Willer, C. J., E. M. Schmidt, S. Sengupta, G. M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, M. L. Buchkovich, S. Mora, et al. 2013. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**: 1274–1283.
32. Schunkert, H., I. R. König, S. Kathiresan, M. P. Reilly, T. L. Assimes, H. Holm, M. Preuss, A. F. R. Stewart, M. Barbalic, C. Gieger, D. Absher, et al. 2011. Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43**: 333–338.
33. Vrablik, M., and J. A. Hubacek. 2010. Genetic determination of triglyceridemia with special focus on apolipoprotein gene variants. *Clin. Lipidol.* **5**: 543–554.
34. Wood, K. L., M. H. Miller, and J. F. Dillon. 2015. Systematic review of genetic association studies involving histologically confirmed non-alcoholic fatty liver disease. *BMJ Open Gastroenterol.* **2**: e000019.
35. Rasmussen, K. L. 2016. Plasma levels of apolipoprotein E, APOE genotype and risk of dementia and ischemic heart disease. *Atherosclerosis.* **255**: 145–155.
36. Greenow, K., N. J. Pearce, and D. P. Ramji. 2005. The key role of apolipoprotein E in atherosclerosis. *J. Mol. Med. (Berl.).* **83**: 329–342.
37. Zhang, J.-Z., Y.-Y. Zheng, Y.-N. Yang, X.-M. Li, Z.-Y. Fu, C.-F. Dai, F. Liu, B.-D. Chen, M.-T. Gai, Y.-T. Ma, et al. 2015. Association between apolipoprotein B gene polymorphisms and the risk of coronary heart disease (CHD): an update meta-analysis. *J. Renin Angiotensin Aldosterone Syst.* **16**: 827–837.
38. Benn, M., B. G. Nordestgaard, J. S. Jensen, and A. Tybjaerg-Hansen. 2007. Polymorphisms in apolipoprotein B and risk of ischemic stroke. *J. Clin. Endocrinol. Metab.* **92**: 3611–3617.
39. Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Koseki, J. P. Pirruccello, S. Ripatti, D. I. Chasman, C. J. Willer, et al. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature.* **466**: 707–713.
40. Do, R., C. J. Willer, E. M. Schmidt, S. Sengupta, C. Gao, G. M. Peloso, S. Gustafsson, S. Kanoni, A. Ganna, J. Chen, et al. 2013. Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**: 1345–1352.
41. Farrer, L. A., L. A. Cupples, J. L. Haines, B. Hyman, W. A. Kukull, R. Mayeux, R. H. Myers, M. A. Pericak-Vance, N. Risch, and C. M. van Duijn. 1997. Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA.* **278**: 1349–1356.
42. Neu, S. C., J. Pa, W. Kukull, D. Beeky, A. Kuzma, P. Gangadharan, L.-S. Wang, K. Romero, S. P. Arneric, A. Redolfi, et al. 2017. Apolipoprotein E genotype and sex risk factors for Alzheimer disease. *JAMA Neurol.* **74**: 1178–1189.
43. Singh, P. P., M. Singh, and S. S. Mastana. 2006. APOE distribution in world populations with new data from India and the UK. *Ann. Hum. Biol.* **33**: 279–308.
44. Pennacchio, L. A., M. Olivier, J. A. Hubacek, J. C. Cohen, D. R. Cox, J. C. Fruchart, R. M. Krauss, and E. M. Rubin. 2001. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. *Science.* **294**: 169–173.
45. Kao, J.-T., H.-C. Wen, K.-L. Chien, H.-C. Hsu, and S.-W. Lin. 2003. A novel genetic variant in the apolipoprotein A5 gene is associated with hypertriglyceridemia. *Hum. Mol. Genet.* **12**: 2533–2539.
46. Pollin, T. I., C. M. Damcott, H. Shen, S. H. Ott, J. Shelton, R. B. Horenstein, W. Post, J. C. McLenithan, L. F. Bielak, P. A. Peyser, et al. 2008. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science.* **322**: 1702–1705.
47. Jørgensen, A. B., R. Frikke-Schmidt, B. G. Nordestgaard, and A. Tybjaerg-Hansen. 2014. Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N. Engl. J. Med.* **371**: 32–41.
48. Hsueh, W.-C., A. K. Nair, S. Kobes, P. Chen, H. H. H. Göring, T. I. Pollin, A. Malhotra, W. C. Knowler, L. J. Baier, and R. L. Hanson. 2017. Identity-by-descent mapping identifies major locus for serum triglycerides in Amerindians largely explained by an APOC3 founder mutation. *Circ Cardiovasc Genet.* **10**: e001809.
49. Johs, A., M. Hammel, I. Waldner, R. P. May, P. Laggner, and R. Prassl. 2006. Modular structure of solubilized human apolipoprotein B-100. Low resolution model revealed by small angle neutron scattering. *J. Biol. Chem.* **281**: 19732–19739.
50. Segrest, J. P., M. K. Jones, and N. Dashti. 1999. N-terminal domain of apolipoprotein B has structural homology to lipovitellin and microsomal triglyceride transfer protein: a “lipid pocket” model for self-assembly of apob-containing lipoprotein particles. *J. Lipid Res.* **40**: 1401–1416.
51. Wojczynski, M. K., G. Gao, I. Borecki, P. N. Hopkins, L. Parnell, C.-Q. Lai, J. M. Ordovas, B. H. Chung, and D. K. Arnett. 2010. Apolipoprotein B genetic variants modify the response to fenofibrate: a GOLDN study. *J. Lipid Res.* **51**: 3316–3323.
52. Linton, M. F., R. V. Farese, and S. G. Young. 1993. Familial hypobetalipoproteinemia. *J. Lipid Res.* **34**: 521–541.
53. Tarugi, P., M. Aversa, E. Di Leo, A. B. Cefalù, D. Noto, L. Magnolo, L. Cattin, B. Bertolini, and S. Calandra. 2007. Molecular diagnosis of hypobetalipoproteinemia: an ENID review. *Atherosclerosis.* **195**: e19–e27.
54. Di Filippo, M., P. Moulin, P. Roy, M. E. Samson-Bouma, S. Collardeau-Frachon, S. Chebel-Dumont, N. Peretti, J. Dumortier, F. Zoulim, T. Fontanges, et al. 2014. Homozygous MTP and APOB mutations may lead to hepatic steatosis and fibrosis despite metabolic differences in congenital hypocholesterolemia. *J. Hepatol.* **61**: 891–902.



55. Wang, L. R., A. D. McIntyre, and R. A. Hegele. 2018. Complex genetic architecture in severe hypobetalipoproteinemia. *Lipids Health Dis.* **17**: 48.
56. Tada, H., H-H. Won, O. Melander, J. Yang, G. M. Peloso, and S. Kathiresan. 2014. Multiple associated variants increase the heritability explained for plasma lipids and coronary artery disease. *Circ Cardiovasc Genet.* **7**: 583–587.
57. Schmidt, E. M., and C. J. Willer. 2015. Insights into blood lipids from rare variant discovery. *Curr. Opin. Genet. Dev.* **33**: 25–31.
58. Cole, C. B., M. Nikpay, and R. McPherson. 2015. Gene-environment interaction in dyslipidemia. *Curr. Opin. Lipidol.* **26**: 133–138.
59. Deloukas, P., S. Kanoni, C. Willenborg, M. Farrall, T. L. Assimes, J. R. Thompson, E. Ingelsson, D. Saleheen, J. Erdmann, B. A. Goldstein, et al. 2013. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**: 25–33.
60. Tennessen, J. A., A. W. Bigham, T. D. O'Connor, W. Fu, E. E. Kenny, S. Gravel, S. McGee, R. Do, X. Liu, G. Jun, et al.; and NHLBI Exome Sequencing Project. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science.* **337**: 64–69.
61. Nelson, M. R., D. Wegmann, M. G. Ehm, D. Kessner, P. St Jean, C. Verzilli, J. Shen, Z. Tang, S. A. Bacanu, D. Fraser, et al. 2012. An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science.* **337**: 100–104.
62. Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hunter, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, et al. 2009. Finding the missing heritability of complex diseases. *Nature.* **461**: 747–753.
63. Hegele, R. A. 2009. Plasma lipoproteins: genetic influences and clinical implications. *Nat. Rev. Genet.* **10**: 109–121.
64. Norata, G. D., S. Tsimikas, A. Pirillo, and A. L. Catapano. 2015. Apolipoprotein C-III: from pathophysiology to pharmacology. *Trends Pharmacol. Sci.* **36**: 675–687.
65. Wolska, A., R. L. Dunbar, L. A. Freeman, M. Ueda, M. J. Amar, D. O. Sviridov, and A. T. Remaley. 2017. Apolipoprotein C-II: new findings related to genetics, biochemistry, and role in triglyceride metabolism. *Atherosclerosis.* **267**: 49–60.
66. Zaman, M. M., S. Ikemoto, N. Yoshiike, C. Date, T. Yokoyama, and H. Tanaka. 1997. Association of apolipoprotein genetic polymorphisms with plasma cholesterol in a Japanese rural population: the Shibata study. *Arterioscler. Thromb. Vasc. Biol.* **17**: 3495–3504.
67. Igartua, C., S. V. Mozaffari, D. L. Nicolae, and C. Ober. 2017. Rare non-coding variants are associated with plasma lipid traits in a founder population. *Sci. Rep.* **7**: 1645.
68. Shen, H., C. M. Damcott, E. Rampersaud, T. I. Pollin, R. B. Horenstein, P. F. McArdle, P. A. Peyser, L. F. Bielak, W. S. Post, Y.-P. C. Chang, et al. 2010. Familial defective apolipoprotein B-100 and increased low-density lipoprotein cholesterol and coronary artery calcification in the old order Amish. *Arch. Intern. Med.* **170**: 1850–1855.
69. Zhou, Y., S. Mkrtychian, M. Kumondai, M. Hiratsuka, and V. M. Lauschke. 2018. An optimized prediction framework to assess the functional impact of pharmacogenetic variants. *Pharmacogenomics J.* In press.
70. Battle, A., C. D. Brown, B. E. Engelhardt, and S. B. Montgomery. 2017. Genetic effects on gene expression across human tissues. *Nature.* **550**: 204–213.
71. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, R. Saxena, B. F. Voight, V. Lyssenko, N. P. Burtt, P. I. W. de Bakker, H. Chen, J. J. Roix, S. Kathiresan, J. N. Hirschhorn, et al. 2007. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science.* **316**: 1331–1336.
72. Sabatti, C., S. K. Service, A.-L. Hartikainen, A. Pouta, S. Ripatti, J. Brodsky, C. G. Jones, N. A. Zaitlen, T. Varilo, M. Kaakinen, et al. 2009. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. *Nat. Genet.* **41**: 35–46.
73. Willer, C. J., S. Sanna, A. U. Jackson, A. Scuteri, L. L. Bonnycastle, R. Clarke, S. C. Heath, N. J. Timpson, S. S. Najjar, H. M. Stringham, et al. 2008. Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**: 161–169.
74. De Feo, E., C. Cefalo, D. Arzani, R. Amore, R. Landolfi, A. Grieco, W. Ricciardi, L. Miele, and S. Boccia. 2012. A case-control study on the effects of the apolipoprotein E genotypes in nonalcoholic fatty liver disease. *Mol. Biol. Rep.* **39**: 7381–7388.
75. Kathiresan, S., O. Melander, C. Guiducci, A. Surti, N. P. Burtt, M. J. Rieder, G. M. Cooper, C. Roos, B. F. Voight, A. S. Havulinna, et al. 2008. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**: 189–197.
76. Al-Bustan, S. A., A. E. Al-Serri, B. G. Annice, M. A. Alnaqeeb, and G. A. Ebrahim. 2013. Re-sequencing of the APOAI promoter region and the genetic association of the -75G>A polymorphism with increased cholesterol and low density lipoprotein levels among a sample of the Kuwaiti population. *BMC Med. Genet.* **14**: 90.
77. Corella, D., G. Peloso, D. K. Arnett, S. Demissie, L. A. Cupples, K. Tucker, C.-Q. Lai, L. D. Parnell, O. Coltell, Y.-C. Lee, et al. 2009. APOA2, dietary fat, and body mass index replication of a gene-diet interaction in 3 independent populations. *Arch. Intern. Med.* **169**: 1897–1906.
78. Wong, W. M., E. Hawe, L. K. Li, G. J. Miller, V. Nicaud, L. A. Pennacchio, S. E. Humphries, and P. J. Talmud. 2003. Apolipoprotein AIV gene variant S347 is associated with increased risk of coronary heart disease and lower plasma apolipoprotein AIV levels. *Circ. Res.* **92**: 969–975.
79. Ota, V. K., E. S. Chen, T. F. Ejchel, T. K. Furuya, D. R. Mazzotti, M. S. Cendoroglo, L. R. Ramos, L. Q. Araujo, R. R. Burbano, and M. d. A. C. Smith. 2011. APOA4 polymorphism as a risk factor for unfavorable lipid serum profile and depression: a cross-sectional study. *J. Invest. Med.* **59**: 966–970.
80. Lamina, C., S. Friedel, S. Coassin, R. Rueedi, N. A. Yousri, I. Seppälä, C. Gieger, S. Schönherr, L. Forer, G. Erhart, et al. 2016. A genome-wide association meta-analysis on apolipoprotein A-IV concentrations. *Hum. Mol. Genet.* **25**: 3635–3646.
81. Wang, J., M. R. Ban, G. Y. Zou, H. Cao, T. Lin, B. A. Kennedy, S. Anand, S. Yusuf, M. W. Huff, R. L. Pollex, et al. 2008. Polygenic determinants of severe hypertriglyceridemia. *Hum. Mol. Genet.* **17**: 2894–2899.
82. Martinelli, N., E. Trabetti, A. Bassi, D. Girelli, S. Friso, F. Pizzolo, M. Sandri, G. Malerba, P. F. Pignatti, R. Corrocher, et al. 2007. The -1131 T>C and S19W APOA5 gene polymorphisms are associated with high levels of triglycerides and apolipoprotein C-III, but not with coronary artery disease: an angiographic study. *Atherosclerosis.* **191**: 409–417.
83. Li, G.-P., J.-Y. Wang, S.-K. Yan, B.-S. Chen, H. Xue, and G. Wu. 2004. Genetic effect of two polymorphisms in the apolipoprotein A5 gene and apolipoprotein C3 gene on serum lipids and lipoproteins levels in a Chinese population. *Clin. Genet.* **65**: 470–476.
84. Shou, W., Y. Wang, F. Xie, B. Wang, L. Yang, H. Wu, Y. Wang, Z. Wang, J. Shi, and W. Huang. 2014. A functional polymorphism affecting the APOA5 gene expression is causally associated with plasma triglyceride levels conferring coronary atherosclerosis risk in Han Chinese Population. *Biochim. Biophys. Acta.* **1842**: 2147–2154.
85. Tang, Y., P. Sun, D. Guo, A. Ferro, Y. Ji, Q. Chen, and L. Fan. 2006. A genetic variant c.553G > T in the apolipoprotein A5 gene is associated with an increased risk of coronary artery disease and altered triglyceride levels in a Chinese population. *Atherosclerosis.* **185**: 433–437.
86. Zhou, Q., F. Zhao, Z.-p. Lv, C.-g. Zheng, W.-d. Zheng, L. Sun, N.-n. Wang, S. Pang, F. M. de Andrade, M. Fu, et al. 2014. Association between APOC1 polymorphism and Alzheimer's disease: a case-control study and meta-analysis. *PLoS One.* **9**: e87017.
87. Bacos, K., L. Gillberg, P. Volkov, A. H. Olsson, T. Hansen, O. Pedersen, A. P. Gjesing, H. Eiberg, T. Tuomi, P. Almgren, et al. 2016. Blood-based biomarkers of age-associated epigenetic changes in human islets associate with insulin secretion and diabetes. *Nat. Commun.* **7**: 11089.
88. Pollex, R. L., M. R. Ban, T. K. Young, P. Bjerregaard, S. S. Anand, S. Yusuf, B. Zinman, S. B. Harris, A. J. Hanley, P. W. Connelly, et al. 2007. Association between the -455T>C promoter polymorphism of the APOC3 gene and the metabolic syndrome in a multi-ethnic sample. *BMC Med. Genet.* **8**: 80.
89. Petersen, K. F., S. Dufour, A. Hariri, C. Nelson-Williams, J. N. Foo, X.-m. Zhang, J. Dziura, R. P. Lifton, and G. I. Shulman. 2010. Apolipoprotein C3 gene variants in nonalcoholic fatty liver disease. *N. Engl. J. Med.* **362**: 1082–1089.
90. Timpson, N. J., K. Walter, J. L. Min, I. Tachmazidou, G. Malerba, S.-Y. Shin, L. Chen, M. Futema, L. Southam, V. Iotchkova, et al. 2014. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat. Commun.* **5**: 4871.
91. Tang, L., W. Zeng, X. Lu, Q. Y. Wang, H. Liu, Z. P. Cheng, Y. Y. Wu, B. Hu, X. R. Jian, T. Guo, et al. 2014. Identification of APOH

- polymorphisms as common genetic risk factors for venous thrombosis in the Chinese population. *J. Thromb. Haemost.* **12**: 1616–1625.
92. Guo, T., R-X. Yin, H. Li, Y-M. Wang, J-Z. Wu, and D-Z. Yang. 2015. Association of the Trp316Ser variant (rs1801690) near the apolipoprotein H ( $\beta$ 2-glycoprotein-I) gene and serum lipid levels. *Int. J. Clin. Exp. Pathol.* **8**: 7291–7304.
  93. Zhou, J-W., S. K. W. Tsui, M. C. Y. Ng, H. Geng, S-K. Li, W. Y. So, R. C. Ma, Y. Wang, Q. Tao, Z-Y. Chen, et al. 2011. Apolipoprotein M gene (APOM) polymorphism modifies metabolic and disease traits in type 2 diabetes. *PLoS One.* **6**: e17324.
  94. Jiao, G. Q., Z. X. Yuan, Y. S. Xue, C. J. Yang, C. B. Lu, Z. Q. Lu, and M. D. Xiao. 2007. A prospective evaluation of apolipoprotein M gene T-778C polymorphism in relation to coronary artery disease in Han Chinese. *Clin. Biochem.* **40**: 1108–1112.
  95. Niu, N., X. Zhu, Y. Liu, T. Du, X. Wang, D. Chen, B. Sun, H. F. Gu, and Y. Liu. 2007. Single nucleotide polymorphisms in the proximal promoter region of apolipoprotein M gene (apoM) confer the susceptibility to development of type 2 diabetes in Han Chinese. *Diabetes Metab. Res. Rev.* **23**: 21–25.
  96. Xu, W-W., Y. Zhang, Y-B. Tang, Y-L. Xu, H-Z. Zhu, A. Ferro, Y. Ji, Q. Chen, and L-M. Fan. 2008. A genetic variant of apolipoprotein M increases susceptibility to coronary artery disease in a Chinese population. *Clin. Exp. Pharmacol. Physiol.* **35**: 546–551.
  97. Gayà-Vidal, M., G. Athanasiadis, R. Carreras-Torres, M. Via, E. Esteban, M. Villena, R. Vasquez, J-M. Dugoujon, and P. Moral. 2012. Apolipoprotein E/C1/C4/C2 gene cluster diversity in two native Andean populations: Aymaras and Quechuas. *Ann. Hum. Genet.* **76**: 283–295.
  98. Gao, L., C. Gabriel, T. Lavoie, and S. Q. Ye. 2002. An improved RSP method to detect HpaI polymorphism in the apolipoprotein C-1 gene promoter. *BMC Med. Genet.* **3**: 13.
  99. Lucatelli, J. F., A. C. Barros, V. K. d. Silva, F. S. Machado, P. C. Constantin, A. A. C. Dias, M. H. Hutz, and F. M. de Andrade. 2011. Genetic influences on Alzheimer's disease: evidence of interactions between the genes APOE, APOC1 and ACE in a sample population from the South of Brazil. *Neurochem. Res.* **36**: 1533–1539.
  100. Boren, J., I. Lee, W. Zhu, K. Arnold, S. Taylor, and T. L. Innerarity. 1998. Identification of the low density lipoprotein receptor-binding site in apolipoprotein B100 and the modulation of its binding activity by the carboxyl terminus in familial defective apo-B100. *J. Clin. Invest.* **101**: 1084–1093.