

ARTICLE

Resequencing of *LPL* in African Blacks and associations with lipoprotein–lipid levels

Dilek Pirim¹, Xingbin Wang^{1,2}, Zaheda H Radwan¹, Vipavee Niemsiri¹, Clareann H Bunker³, M Michael Barmada¹, M Ilyas Kamboh^{*,1} and F Yesim Demirci^{*,1}

Genome-wide association studies have identified several loci associated with plasma lipid levels but those common variants together account only for a small proportion of the genetic variance of lipid traits. It has been hypothesized that the remaining heritability may partly be explained by rare variants with strong effect sizes. Here, we have comprehensively investigated the associations of both common and uncommon/rare variants in the lipoprotein lipase (*LPL*) gene in relation to plasma lipoprotein–lipid levels in African Blacks (ABs). For variant discovery purposes, the entire *LPL* gene and flanking regions were resequenced in 95 ABs with extreme high-density lipoprotein cholesterol (HDL-C) levels. A total of 308 variants were identified, of which 64 were novel. Selected common tagSNPs and uncommon/rare variants were genotyped in the entire sample ($n=788$), and 126 QC-passed variants were evaluated for their associations with lipoprotein–lipid levels by using single-site, haplotype and rare variant (SKAT-O) association analyses. We found eight not highly correlated ($r^2 < 0.40$) signals (rs1801177:G > A, rs8176337: G > C, rs74304285:G > A, rs252:delA, rs316:C > A, rs329:A > G, rs12679834:T > C, and rs4921684:C > T) nominally ($P < 0.05$) associated with lipid traits (HDL-C, LDL-C, ApoA1 or ApoB levels) in our sample. The most significant SNP, rs252:delA, represented a novel association observed with LDL-C ($P=0.002$) and ApoB ($P=0.012$). For TG and LDL-C, the haplotype analysis was more informative than the single-site analysis. The SKAT-O analysis revealed that the bin (group) containing 22 rare variants with MAF ≤ 0.01 exhibited nominal association with TG ($P=0.039$) and LDL-C ($P=0.027$). Our study indicates that both common and uncommon/rare *LPL* variants/haplotypes may affect plasma lipoprotein–lipid levels in general African population.

European Journal of Human Genetics (2015) 23, 1244–1253; doi:10.1038/ejhg.2014.268; published online 28 January 2015

INTRODUCTION

Plasma lipoprotein–lipid levels modulate the risk for atherosclerosis and cardiovascular disease (CVD).¹ Plasma lipid–lipoprotein levels are under genetic control,² and genome-wide association studies (GWAS) have revealed several loci associated with plasma lipids.^{3–7} However, most of the identified loci have small effect sizes and explain only ~30% of the genetic variance of lipid phenotypes.⁶ It has been suggested that rare or low-frequency variants with moderate/strong effects that are not captured by GWAS could explain a part of the ‘missing heritability’.^{8,9} An effective way to identify these rare/low-frequency variants is to resequence the candidate genes in subjects with extreme phenotypes.¹⁰ This strategy has already been successfully employed for various candidate genes involved in lipid metabolism, where multiple rare variants were found to contribute to inter-individual variation in plasma lipid levels.^{11–16}

Lipoprotein lipase (encoded by *LPL*) is an enzyme that hydrolyzes triglyceride (TG) rich particles into free fatty acids and glycerol. High *LPL* activity is associated with lower TG and higher high-density lipoprotein cholesterol (HDL-C) levels.¹⁷ Several common *LPL* variants were reported to modulate lipid levels and CVD risk.¹⁸ Resequencing of *LPL* in patients with hypertriglyceridemia identified mutations associated with extremely high TG levels.^{12–14,19} However, the impact of rare *LPL* variants on plasma lipids in general population remains largely

unknown. In this study, we sequenced *LPL* in 95 African Blacks (ABs) with extreme HDL-C levels and genotyped-selected variants in 788 subjects to test their association with lipid levels.

MATERIALS AND METHODS

Subjects

The study sample was comprised of 788 ABs from Benin City, Nigeria (Table 1), recruited as part of a study on civil servants to investigate coronary heart disease (CHD)-related lifestyle factors in this generally lean population. Detailed information on the original study including the sample features and methods used for plasma lipoprotein–lipid measurements can be found elsewhere.^{20–24} Ninety-five subjects selected from the upper ($n=48$) and lower ($n=47$) 10th percentiles of HDL-C distribution in the study sample were used for *LPL* sequencing (Table 2).

DNA sequencing and genotyping

The entire *LPL* gene (27 993 bp) plus 1196 bp in 5′ flanking and 1 kb in 3′ flanking region were sequenced in both directions (see Supplementary Data for details). After quality filtering, the sequence chromatograms were individually inspected and variants were manually reviewed by at least two researchers. In addition to selecting all relevant common and rare variants for follow-up genotyping (see section Follow-up genotyping of selected *LPL* variants in the entire sample), any singleton novel variant surrounded by moderate-quality sequences (with some level of background noise) was treated as ‘suspicious’ and also included in follow-up genotyping for confirmation. Selected variants were

¹Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ²Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA; ³Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, USA

*Correspondence: Professor MI Kamboh or Assistant Professor FY Demirci, Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, 130 DeSoto Street, Pittsburgh, PA 15261, USA. Tel: +412 624 3066; Fax: +412 624 3020; E-mail: kamboh@pitt.edu or fyd1@pitt.edu

Received 31 January 2014; revised 24 October 2014; accepted 9 November 2014; published online 28 January 2015

Table 1 Biometric and quantitative data of the study sample

Variable	African Blacks (n = 788)	
	number or mean ± SD	
Male/female (n)	495/293	
Age (years)	40.95 ± 8.39	
BMI (kg/m ²)	22.87 ± 4.04	
LDL-C (mg/dl)	109.25 ± 34.40	
HDL-C (mg/dl)	47.88 ± 12.87	
TG (mg/dl)	72.96 ± 39.32	
TC (mg/dl)	172.01 ± 38.47	
ApoB (mg/dl)	66.98 ± 22.19	
ApoA1 (mg/dl)	137.03 ± 28.46	

Abbreviations: ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides.

Table 2 Demographics and characteristics of the sequencing sample (95 African Blacks)

Variable	High HDL-C (n = 48)	Low HDL-C (n = 47)	P-value ^a
	number or mean ± SD	number or mean ± SD	
Male/female	24/24	23/24	1
Age (years)	41.29 ± 8.72	40.87 ± 7.12	0.8
BMI (kg/m ²)	22.06 ± 4.70	23.91 ± 5.51	0.08
TC (mg/dl)	201 ± 39.68	141.68 ± 31.03	<0.001
LDL-C (mg/dl)	112.55 ± 39.75	95.04 ± 28.28	0.02
HDL-C (mg/dl)	76.05 ± 7.53	25.51 ± 5.66	<0.001
TG (mg/dl)	61.98 ± 19.85	95.79 ± 73.21	0.004
ApoB (mg/dl)	66.00 ± 20.22	69.64 ± 21.46	0.4
ApoA1 (mg/dl)	166.04 ± 28.19	103.84 ± 27.23	<0.001

Abbreviations: ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; TC, total cholesterol; TG, triglycerides.

^aP-values were calculated based on the original values by using *t*-test. No covariates were included.

genotyped in total sample using TaqMan (Applied Biosystems, Waltham, MA, USA) or iPLEX Gold (Sequenom, San Diego, CA, USA) methods (see Supplementary Data for details), except for *HindIII* (rs320:T>G) polymorphism genotyped by restriction fragment length polymorphism analysis in an earlier study.¹⁶

Statistical analyses

Haploview²⁵ was used to test the concordance with Hardy–Weinberg equilibrium (HWE) and to determine the allele frequencies, linkage disequilibrium (LD) patterns and pairwise correlations (*r*²). An additive linear regression model was used to test the effects of genotypes on the means of plasma lipid–lipoprotein traits. The nominal *P*-value of <0.05 was considered as suggestive evidence of association. Benjamini–Hochberg procedure was applied to control false discovery rate (FDR) in single-site analyses for each trait and FDR <0.20 was considered as statistically significant.²⁶ A *post hoc* power analysis²⁷ was performed to evaluate the power of detecting significant single-site associations for the observed proportions of the variance of lipid traits explained by tested SNPs. For haplotype analysis, the generalized linear model²⁸ was applied using the Haplo.Stats R package (Rochester, MN, USA). Cumulative effects of uncommon/rare variants were analyzed using the SKAT-O method²⁹ and three MAF thresholds (≤1%, ≤2% and <5%). Additional details on statistical analyses can be found in Supplementary Data.

RESULTS

DNA resequencing results

A total of 308 variants were identified, of which 130 were common (MAF ≥ 0.05), 118 were uncommon (0.01 ≤ MAF < 0.05) and 60 were

rare (MAF < 0.01; Figure 1a); 19 were indels and 2 triallelic SNPs (Supplementary Table 1); 246 were identified in introns, 30 in 3′-UTR, 14 in flanking regions, 12 in coding regions and 6 in 5′-UTR (Figures 1b and c). All identified coding variants were known SNPs; seven nonsynonymous and five synonymous. We successfully identified all but three common variants (rs1470187:G>T, rs59184895:T>C and rs328:C>G) reported in African-descent populations (dbSNP build 138); these variants were located at the beginning or end of resequencing amplicons where the sequence read quality is usually low in Sanger sequencing, which probably hampered their identification in our sequencing sample, however, they were successfully genotyped in our entire sample. Of 308 variants, 64 (2 common and 62 uncommon/rare) have not been previously reported in dbSNP and thus have been submitted to dbSNP database (http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?handle=KAMBOH). Four of 64 novel variants were located in flanking regions, 3 in 5′-UTR, 6 in 3′-UTR and 51 in introns; majority (69%) had MAF < 0.01. We identified nine novel indels (all located in introns) ranging in size from 1 to 15 bases. In addition to ‘novel variants’, we identified ‘novel uncommon/rare alleles’ at two nucleotide positions where common diallelic variations were previously reported (rs7002728:G>T and rs28599962:T>C), thus we detected triallelic variations at these positions where the least frequently observed allele was unique to our sample. Of 308 variants, 24 were found only in high HDL-C group vs 54 only in low HDL-C group.

Follow-up genotyping of selected *LPL* variants in the entire sample

LD and Tagger analyses were performed on 130 common variants (MAF ≥ 0.05) to identify and select the tagSNPs for follow-up genotyping in the entire sample; 92 common tagSNPs were selected using the *r*² cut-off of 0.9 (Supplementary Table 2). We also compared our Tagger results with those from the HapMap YRI population. All common tagSNP bins identified in HapMap YRI data for the same genomic region using the same parameters were also captured in our data. In addition to common tagSNPs, we selected the following variants for follow-up genotyping: (i) variants located in exons or intron–exon junctions, (ii) uncommon/rare variants (MAF < 0.05) present in ≥ 2 individuals included in sequencing and (iii) suspicious (borderline quality) novel variants identified in sequencing. Additionally, three known common *LPL* SNPs that were not detected in our sequencing (see section DNA resequencing results) were also included in follow-up genotyping. In total, 163 variants were selected for follow-up genotyping in addition to 1 variant (rs320:T>G) that was already genotyped as part of an earlier study.¹⁶ Of 164 variants (92 tagSNPs, 72 others) advanced into next stage, 30 failed the assay design for genotyping (23 tagSNPs, 7 others) and 2 suspicious variants turned out to be monomorphic after genotyping (indicating sequencing artifacts). Of 132 successfully genotyped variants (Supplementary Table 3), one was triallelic and not included in downstream analyses. As a part of QC, five additional SNPs (two tagSNPs, three others) were excluded from downstream analyses; two had low call rate and three did not meet the HWE. Thus, a total of 126 variants (67 common and 59 uncommon/rare variants based on their frequency in our entire sample) were included in downstream association analyses. The LD bins of these 126 variants are shown in Supplementary Table 4.

Association of common variants (MAF ≥ 0.05) with lipid/lipoprotein levels

Table 3 summarizes the single-site analyses results for 17 common variants that showed nominal associations (*P* < 0.05) with one or more of four lipid traits (HDL-C, LDL-C, ApoA1 or ApoB) in 788 ABs.

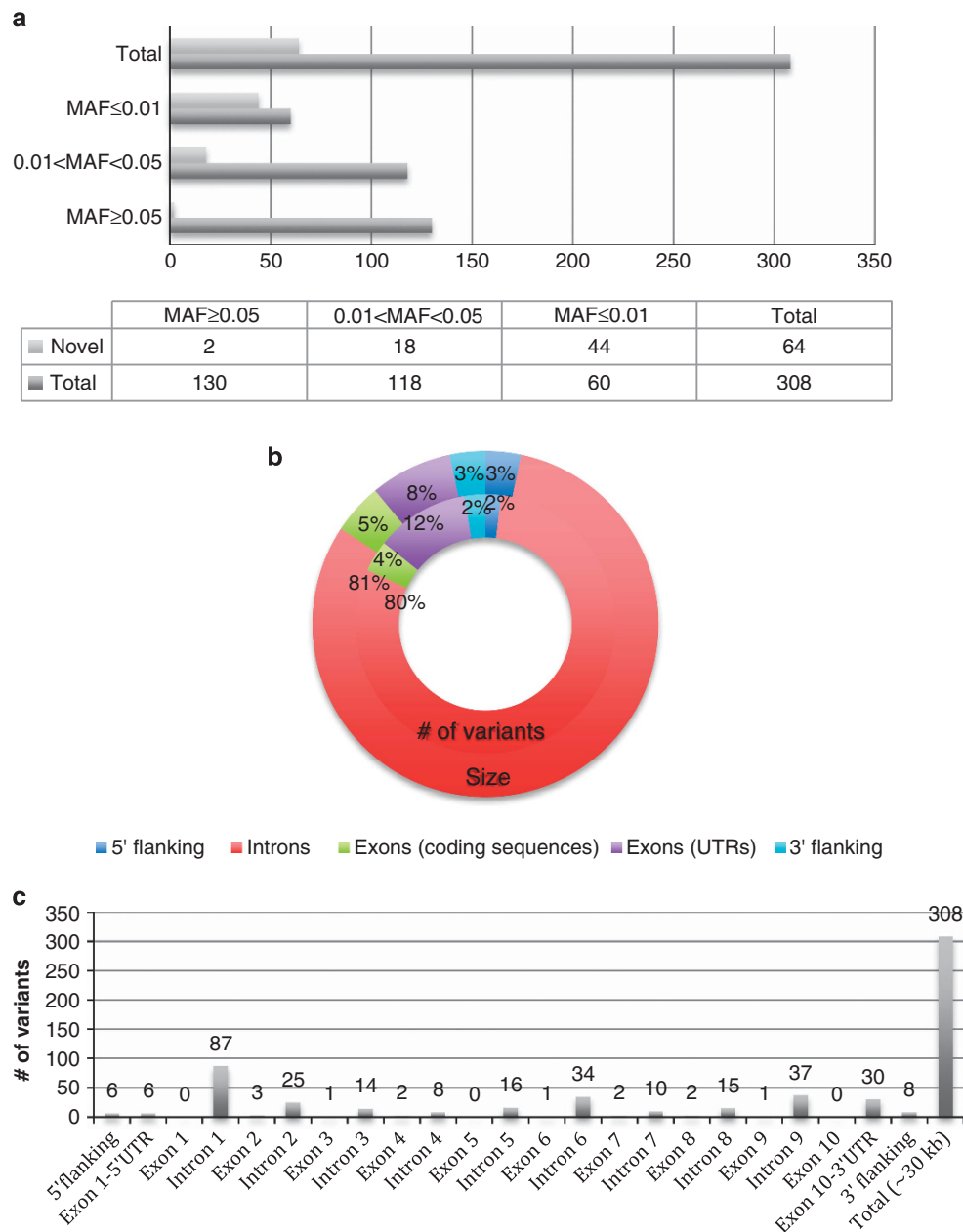


Figure 1 (a) Minor allele frequency distribution of *LPL* variants identified in African Blacks ($n=95$). (b) Distribution of *LPL* variants by location in the gene and comparison with size distribution of those locations. (c) Number and locations of *LPL* variants identified in African Blacks ($n=95$).

The results for all 67 common SNPs for all five tested lipid traits are summarized in Supplementary Tables 5. A *post hoc* power analysis showed that our power to detect a causal SNP would be 78–92%, given the proportion of the variance of the lipid traits explained by these SNPs (ranging from 1.0% to 1.4%) and a significance level of 0.05.

The most significant SNP, rs252:delA (intronic variant), was associated with LDL-C ($\beta = -1.037$; $P = 0.002$; $FDR = 0.134$) and ApoB ($\beta = -2.364$; $P = 0.012$; $FDR = 0.328$) levels. This is a novel association given that the rs252:delA SNP was not in high LD ($r^2 \leq 0.20$) with other nominally associated SNPs (Figure 2), including two functional *LPL* variants rs1801177:G>A (p.(Asp36Asn)) and rs13702:C>T (resides in 3'-UTR and disrupts a microRNA-410 recognition element seed site).³⁰ One additional SNP, rs74304285:

G>A, which again was not highly correlated ($r^2 \leq 0.11$) with other relevant SNPs, was also associated with both LDL-C ($\beta = 0.995$; $P = 0.019$; $FDR = 0.627$) and ApoB ($\beta = 3.014$; $P = 0.010$; $FDR = 0.328$). In addition, four other weakly correlated ($r^2 \leq 0.20$) SNPs (rs1801177:G>A, rs8176337:G>C, rs329:A>G and rs12679834:T>C) were nominally associated with ApoB levels, of which rs12679834:T>C was also associated with ApoA1 levels. Two of these SNPs (rs1801177:G>A and rs8176337:G>C) were previously reported to be associated with TG,^{31,32} but showed only $0.05 < P < 0.20$ for TG levels in our study.

The second most significant SNP, rs316:C>A (synonymous variant), was associated with HDL-C ($\beta = 0.68$; $P = 0.003$; $FDR = 0.178$) and ApoA1 ($\beta = 1.292$; $P = 0.022$; $FDR = 0.288$) levels. The rs316:C>A SNP was in LD with four intronic SNPs, including

Table 3 Common *LPL* variants nominally ($P < 0.05$) associated with plasma lipid/lipoprotein levels

Ref SNP ID	Position on chr8 ^a	Location in gene	Minor allele	MAF	TG			LDL-C			HDL-C			ApoB			ApoA1			RegulomeDB score
					Beta	P	FDR	Beta	P	FDR	Beta	P	FDR	Beta	P	FDR	Beta	P	FDR	
rs1801177	19805708	Exon 2	A	0.05	0.023	0.153	0.952	0.686	0.112	0.751	-0.066	0.877	0.958	2.715	0.027	0.359	0.657	0.536	0.869	5
rs8176337	19806671	Intron 2	C	0.313	-0.01	0.177	0.952	-0.332	0.103	0.751	-0.088	0.663	0.858	-1.298	0.021	0.344	-0.482	0.333	0.869	ND
rs74304285	19808030	Intron 2	A	0.056	-0.004	0.788	0.952	0.995	0.019	0.627	0.232	0.586	0.858	3.014	0.01	0.328	1.724	0.099	0.512	ND
rs252	19811232	Intron 4	InsA	0.093	-0.013	0.281	0.952	-1.037	0.002	0.134	0.03	0.929	0.958	-2.364	0.012	0.328	-0.832	0.313	0.869	ND
rs279	19814696	Intron 6	G	0.15	-0.002	0.825	0.952	0.078	0.776	0.776	0.595	0.029	0.233	-0.618	0.418	0.898	0.67	0.322	0.869	ND
rs295	19816238	Intron 6	C	0.39	-0.004	0.575	0.952	0.32	0.092	0.751	0.384	0.043	0.291	0.534	0.314	0.898	0.922	0.05	0.315	ND
rs301	19816934	Intron 7	C	0.288	-0.006	0.466	0.952	0.285	0.176	0.824	0.555	0.008	0.212	0.817	0.164	0.898	1.103	0.035	0.293	ND
rs316	19818436	Exon 8	A	0.225	-0.001	0.891	0.978	0.16	0.49	0.955	0.68	0.003	0.178	0.192	0.764	0.947	1.292	0.022	0.288	1f
rs320	19819077	Intron 8	G	0.306	-0.002	0.773	0.952	0.447	0.029	0.646	0.485	0.018	0.212	0.825	0.147	0.898	1.102	0.031	0.293	5
rs329	19820086	Intron 9	G	0.097	-0.003	0.813	0.952	-0.558	0.092	0.751	0.405	0.229	0.745	-1.806	0.048	0.533	-0.537	0.518	0.869	4
rs330	19820396	Intron 9	A	0.085	-0.005	0.708	0.952	0.354	0.317	0.891	0.669	0.055	0.305	0.886	0.364	0.898	2.031	0.019	0.288	5
rs12679834	19820433	Intron 9	C	0.058	0.004	0.812	0.952	0.424	0.289	0.891	0.278	0.487	0.858	2.711	0.015	0.328	2.207	0.024	0.288	1f
rs1059507	19823963	3'-UTR-exon 10	T	0.103	0.001	0.905	0.978	0.362	0.267	0.891	0.689	0.031	0.233	0.995	0.267	0.898	1.77	0.025	0.288	6
rs13702	19824492	3'-UTR-exon 10	T	0.459	0.004	0.623	0.952	-0.209	0.279	0.891	-0.49	0.011	0.212	-0.08	0.881	0.948	-0.386	0.416	0.869	6
rs3916027	19824868	3' flanking	A	0.413	-0.005	0.523	0.952	0.298	0.123	0.751	0.421	0.028	0.233	0.318	0.554	0.937	0.814	0.085	0.476	5
rs4921683	19825068	3' flanking	A	0.102	-1.17E-04	0.992	0.992	0.438	0.184	0.824	0.757	0.019	0.212	0.856	0.345	0.898	1.911	0.017	0.288	ND
rs4921684	19825128	3' flanking	T	0.079	-0.012	0.354	0.952	0.518	0.154	0.824	0.869	0.016	0.212	0.246	0.807	0.948	1.983	0.026	0.288	5

Abbreviations: ApoA1, apolipoprotein A1; ApoB, apolipoprotein B; FDR, false discovery rate; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; MAF, minor allele frequency; ND, no data; TG, total cholesterol; TG, triglycerides.

RegulomeDB scores were generated by using <http://regulome.stanford.edu/>. Scores represent; category 1 (likely to affect binding and linked to expression of a gene target); 1a - eQTL+TF binding+matched TF motif+matched DNase footprint+DNase peak; 1b - eQTL+TF binding+any motif+DNase peak; 1c - eQTL+TF binding+matched TF motif+DNase peak; 1d - eQTL+TF binding+any motif+DNase peak; 1e - eQTL+TF binding+matched TF motif; 1f - eQTL+TF binding+DNase peak; category 2 (likely to affect binding); 2a - TF binding+matched TF motif+matched DNase footprint+DNase peak; 2b - TF binding+any motif+DNase footprint+DNase peak; 2c - TF binding+matched TF motif+DNase peak; category 3 (less likely to affect binding); 3a - TF binding+any motif+DNase peak; 3b - TF binding+matched TF motif; category 4-6 (minimal binding evidence); 4 - TF binding+DNase peak; 5 - TF binding or DNase peak; 6 - other.

^aChromosomal position is according to the NCBI reference sequence: hg19; bold indicates nominally significant P -value (< 0.05); exons and introns are numbered based on reference sequence NG_008855.1.

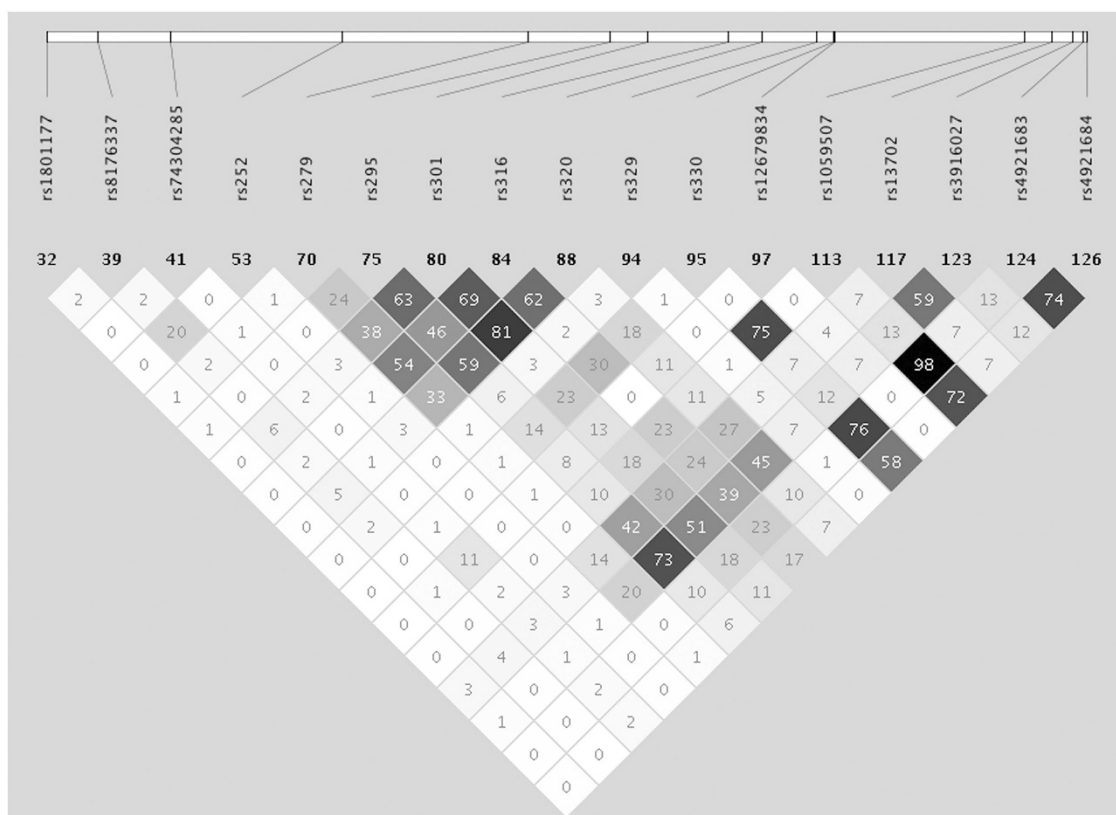


Figure 2 LD structure of 17 *LPL* SNPs associated with one or more lipid/lipoprotein traits (HDL-C, LDL-C, ApoA1 or ApoB levels). The values in the cells are the pairwise degree of LD indicated by $r^2 \times 100$. Shades of white indicate $r^2 = 0$, shades of gray indicate $0 < r^2 < 1$ and shades of black indicate $r^2 = 1$.

rs279:C>G ($P=0.029$; FDR=0.233; $r^2=0.54$) and rs295:A>C ($P=0.043$; FDR=0.291; $r^2=0.46$) associated with HDL-C, and rs301:T>C ($P=0.008$; FDR=0.212; $r^2=0.69$) and rs320:T>G ($P=0.018$; FDR=0.212; $r^2=0.62$) associated with HDL-C and ApoA1. The rs320:T>G SNP (*Hind*III polymorphism) has been well-known and consistently associated with HDL-C and TG levels in several studies.^{18,32,33} Five more SNPs located in 3'-UTR or 3' flanking region and in LD ($r^2 > 0.40$) with each other or with above five SNPs, also yielded nominal associations with HDL-C: rs1059507:C>T ($\beta=0.689$; $P=0.031$; FDR=0.233), rs13702:C>T ($\beta=-0.490$; $P=0.011$; FDR=0.212), rs3916027:G>A ($\beta=0.421$; $P=0.028$ FDR=0.233), rs4921683:T>A ($\beta=0.757$; $P=0.019$; FDR=0.212) and rs4921684:C>T ($\beta=0.869$; $P=0.016$; FDR=0.212). Of 10 SNPs associated with HDL-C in our sample, 7 (rs301:T>C, rs316:C>A, rs320:T>G, rs1059507:C>T, rs13702:C>T, rs3916027:G>A and rs4921683:T>A) exhibited suggestive evidence of association with HDL-C also in a recent meta-analysis.³⁴ Six of them (rs301:T>C, rs316:C>A, rs320:T>G, rs1059507:C>T, rs4921683:T>A, rs4921684:C>T) were also nominally associated with ApoA1 in our study. Another widely studied *LPL* variant, rs328:C>G/p.(Ser474Ter) (p.(Ser447Ter) in mature protein excluding the signal peptide), occurred at lower frequency (4.2%) in our African sample (>10% in Europeans) and its established association with TG and HDL-C in Europeans was not replicated here.

All together, we found eight nominally associated ($P < 0.05$) and not highly correlated ($r^2 < 0.40$) signals; rs1801177:G>A, rs8176337:G>C, rs74304285:G>A, rs252:delA, rs316:C>A, rs329:A>G, rs12679834:T>C and rs4921684:C>T. To our knowledge,

rs74304285:G>A and rs252:delA have not been tested before for association with lipids.

Association of uncommon/rare variants (MAF < 0.05) with lipid/lipoprotein levels

For the SKAT-O analysis of 59 uncommon/rare variants with lipid traits, 3 bins were generated using 3 MAF thresholds (bin 1 (59 variants with MAF < 0.05), bin 2 (32 variants with MAF ≤ 0.02) and bin 3 (22 variants with MAF ≤ 0.01)) and test statistics were calculated for each bin separately. Although bin 1 was associated with ApoB ($P=0.016$), bin 3 showed significant association with both TG ($P=0.039$) and LDL-C ($P=0.027$; Table 4). In *post hoc* single-site analysis of 22 rare variants in bin 3 (Supplementary Table 10), 7 (including 3 novel) showed nominal association with one or more lipid/lipoprotein traits (TG, HDL-C, LDL-C or ApoA1).

Haplotype-based association analysis results

A total of 123 overlapping windows, each containing four SNPs, were constructed using the sliding window approach (sliding one SNP at a time) for haplotype analysis of 126 variants with lipid traits (Figure 3). P -values were calculated based on the comparison of each haplotype with the most common reference haplotype. The strongest haplotype effects were observed on HDL-C, followed by ApoB, comprising 19 and 18 nominally significant ($P < 0.05$) global P -values, respectively (Supplementary Table 11). The most significant window ('window 81' containing rs313:A>G, rs314:A>G, rs77434393:G>A and rs316:C>A) was associated with both HDL-C (global $P=2.32E-04$) and ApoA1 (global $P=0.021$). The only SNP in this window that showed association in single-site analysis was rs316:C>A ($P=0.003$ with

Table 4 Results of rare variant association analyses (SKAT-O) of the *LPL* gene

MAF threshold	N.RV ^a	N.Sample_		TG		LDL-C		HDL-C		ApoB		ApoA1	
		RV ^b	NoRV ^c	Stat	P	Stat	P	Stat	P	Stat	P	Stat	P
Bin 1 (MAF < 0.05)													
	59	613	175	3.50E+05	0.170	2.26E+05	0.487	1.96E+05	0.265	2.08E+06	0.016	7.84E+05	0.128
Bin 2 (MAF ≤ 0.02)													
	32	223	565	1.92E+05	0.059	5.07E+05	0.103	1.20E+05	0.075	6.47E+04	0.706	1.08E+05	0.162
Bin 3 (MAF ≤ 0.01)													
	22	84	704	1.64E+05	0.039	4.57E+05	0.027	2.40E+05	0.086	4.81E+04	0.612	4.63E+04	0.332

Abbreviation: SKAT-O, optimal sequencing Kernel association test.

^aN.RV: number of rare variants below the defined MAF threshold.

^bN.Sample_RV: number of individuals carrying the rare allele with defined MAF threshold.

^cN.Sample_NoRV: number of individuals who do not carry any rare alleles with defined MAF threshold.

HDL-C; $P=0.022$ with ApoA1) and thus the observed haplotype association seemed to be primarily driven by this SNP. Although the haplotypes in intron 6–intron 8 region were associated with HDL-C and ApoA1, those in intron 8–intron 9 were associated with ApoB. Although no common *LPL* SNPs were associated with TG in single-site analysis, seven haplotype windows showed nominal associations with TG. Overall, haplotype analysis was more informative for TG and LDL-C than single-site analysis.

Functional annotation of significant SNPs using RegulomeDB

ENCODE annotations were retrieved for all variants using RegulomeDB³⁵ (Supplementary Table 1). Of 10 of 17 nominally associated SNPs scored in RegulomeDB (Table 3), 2 (rs1316:C>A and rs12679834:T>C) were cis-eQTL (expression quantitative trait loci) SNPs (score=1f) but the remaining showed minimal evidence for regulatory function (scores=4–6).

DISCUSSION

To our knowledge, this is the first study of its kind that reports a comprehensive catalog of sequence variation in the entire *LPL* gene in African individuals in relation to plasma lipid traits. Resequencing of 95 ABs with extreme HDL-C levels identified 308 variants, of which 64 were novel. Recently, our group also reported the resequencing of entire *LPL* gene in 95 non-Hispanic whites (NHWs) with extreme HDL-C levels;³⁶ 176 variants were identified in 95 NHWs, of which 113 were also found in 95 ABs. As expected, we observed more variation (especially more uncommon/rare variants) in ABs. Nickerson *et al*³⁷ sequenced a portion of *LPL* (9.7 kb) in 71 individuals randomly selected from three populations, including African Americans (AAs) from Jackson, Mississippi ($n=24$).³⁷ We observed more variants (97 vs 78, of which 71 shared) in this 9.7 kb region in ABs as compared with AAs in Nickerson *et al*.³⁷ Observed differences could be due to differences in genetic background (ABs vs AAs), sample size (95 vs 24), selection criteria (subjects with extreme lipid profiles vs randomly selected subjects) and/or software tools used for variant analyses (Variant Reporter vs PolyPhred).

Following the discovery stage, 164 variants were selected for follow-up in our entire sample ($n=788$), of which 126 successfully passed all genotyping steps (assay design, run and post-run QC) and were analyzed for association with lipid traits. Single-site analysis of common SNPs revealed 10 nominal associations ($P<0.05$) with HDL-C, 3 with LDL-C, 6 with ApoB and 8 with ApoA1 levels. Our observation of a higher number of associations with HDL-C is

consistent with the published data.³² In Europeans, several *LPL* SNPs have been reported to be associated with lipid levels, including six (rs268:A>G, rs326:G>A, rs320:T>G, rs328:C>G, rs1801177:G>A and rs13702:C>T) consistently associated with HDL-C and/or TG.^{32,38–41} Among these, rs268:A>G was absent in our African sample and in HapMap YRI data. The rs326:G>A SNP has also been shown to be associated with HDL-C in AAs.⁴² Although rs326:G>A showed no association in our sample, it was in LD with HDL-associated SNP rs13702:C>T ($P=0.011$; FDR=0.212; $r^2=0.64$). The association of rs13702:C>T with HDL-C has also been reported in AAs.^{43,44} Although rs13702:C>T was shown to be in LD with rs320:T>G in Europeans,³² the correlation was weak in our African sample ($r^2=0.27$). Replication of the association of rs13702:C>T with HDL-C in multiple African-derived samples and its potential to disrupt a microRNA recognition site³⁰ support its functional role in HDL metabolism in African-descent subjects. Common SNPs in/around *LPL* have also been reported in several lipid GWAS conducted mostly in European-descent subjects.^{45–47} A large GWAS conducted for CHD and risk factors in AAs⁴⁸ replicated 17 European loci, including two HDL-C associated SNPs (rs10503669:C>A and rs10096633:C>T) located downstream of *LPL* (outside of our target region).

The rs328:C>G (p.(Ser447Ter)) SNP has been shown to increase *LPL* activity and cause lower TG and higher HDL-C levels.^{16,18,33,38} Although its frequency is relatively high in Whites (MAF>10%), it is low in our African sample (4.2%) and in HapMAP YRI population (3.3%). Although the potential functional significance of rs328:C>G is appreciated in Europeans, recent studies in AAs have sparked debate on its role.^{43,44} An admixture mapping study of 3300 AAs determined that the effect size of rs328:C>G on TG was dependent on ancestral background and significantly diminished in subjects with African background,⁴³ suggesting that there might be other truly causal undiscovered variant(s) in LD (or acting synergistically) with rs328:C>G. A cross-sectional and longitudinal study found rs328:C>G to be significantly associated with both HDL-C and TG in European Americans but only with TG in AAs.⁴⁴ A recent large GWAS in AAs failed to show its association with HDL-C or TG.⁴⁸ Consistent with most reports in AAs, we did not observe association of this SNP with HDL-C or TG in our sample. The *HindIII* (rs320:T>G) polymorphism has been shown to have the same effects on TG and HDL-C as has rs328:C>G, and there has been a debate over whether their effects are independent.^{16,33,39,49–51} The rs320:T>G SNP is predicted to affect the binding of a transcription factor and so may be functional by itself.⁵² In our data, the association of rs320:T>G with HDL-C

($\beta=0.485$; $P=0.018$, $FDR=0.212$) was independent of rs328:C>G; however, it was in LD with four other HDL-associated SNPs: rs301:T>C ($P=0.008$; $FDR=0.212$; $r^2=0.81$), rs316:C>A ($P=0.003$; $FDR=0.178$; $r^2=0.62$), rs295:A>C ($P=0.043$; $FDR=0.291$; $r^2=0.59$) and rs3916027:G>A ($P=0.028$; $FDR=0.233$; $r^2=0.45$) that have also been reported to be associated with HDL-C.^{34,43,46} Further studies are needed to understand whether rs320:T>G explains other observed associations or yet to be discovered functional variant(s) exist in this region.

Associations of *LPL* SNPs with LDL-C have not been well documented in the literature, albeit some contrasting results exist for rs320:T>G.^{53,54} We have identified two independent ($r^2=0$) LDL-associated common SNPs, rs252:delA ($P=0.002$; $FDR=0.134$) and rs74304285:G>A ($P=0.019$; $FDR=0.627$), as novel observations that warrant further investigation in independent studies. We have also detected a group of 22 rare variants ($MAF\leq 0.01$; 11 in UTR, 5 coding and 6 intronic) that showed significant association with LDL-C ($P=0.027$) and TG ($P=0.039$), including 4 known

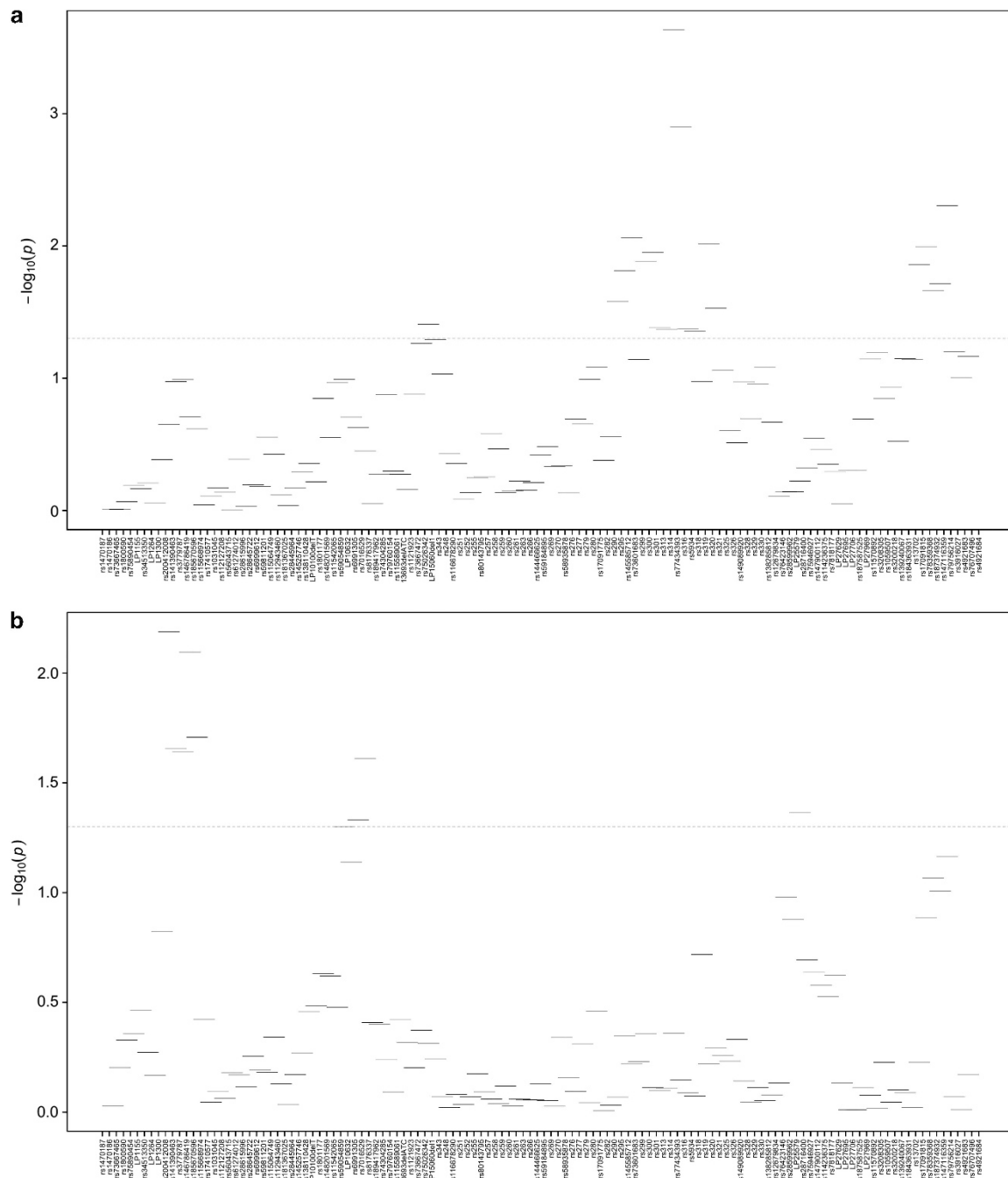


Figure 3 Haplotype association results for HDL (a), TG (b), LDL-C (c), ApoA1 (d) and ApoB (e) levels. The log of the global P -value is presented on the y axis and SNPs are presented across the x axis in chromosomal order. Horizontal lines are 4-SNP haplotype windows. The red horizontal line shows the significance threshold. A full color version of this figure is available at the *European Journal of Human Genetics* journal online.

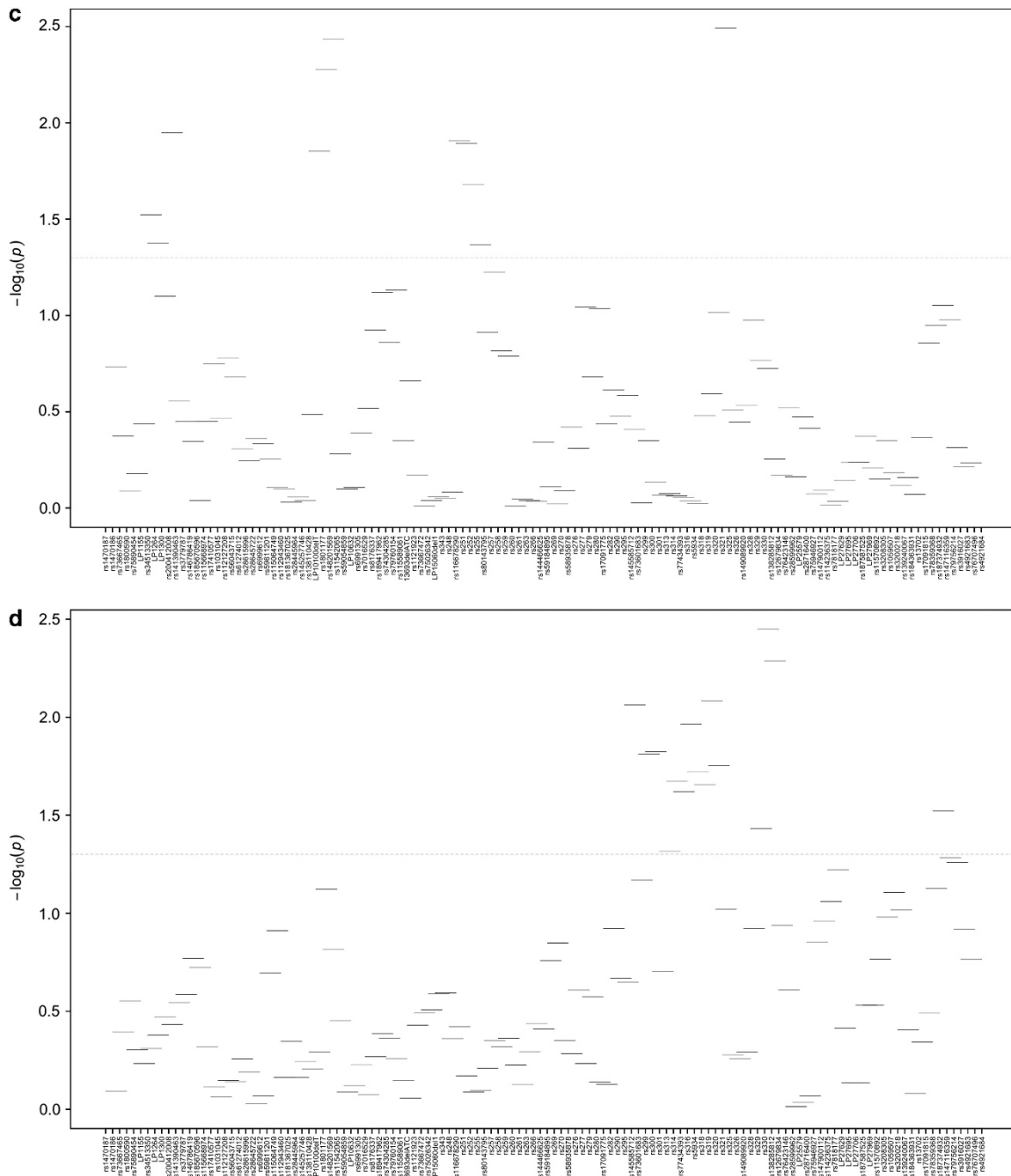


Figure 3 Continued.

nonsynonymous variants with no previous report of association with lipids and 10 novel variants. This association probably represents cumulative effects of functional rare variants; however, further studies are needed to understand the impact of these variants on plasma lipid profile. The results of haplotype and single-site analyses were largely comparable for HDL-C, ApoA1 and ApoB. However, for TG and LDL-C, several haplotype windows (some of which harboring only uncommon/rare variants) yielded significant associations despite the absence of (TG) or only few (LDL) nominal associations in single-site analyses.

To our knowledge, this is the first study that has comprehensively evaluated *LPL* genetic variation in relation to lipid traits in ABs. Our study provides new information in addition to supporting some previous observations, but has some limitations. Our sequencing sample was small (some relevant rare variants might have been missed) but our selective resequencing of subjects with ‘extreme’ phenotypes still enabled us to identify several rare variants. Given the known role of *LPL* in lipid metabolism and that several tested SNPs were not completely independent, the strict Bonferroni correction was not applied. After multiple testing correction using FDR, our top two

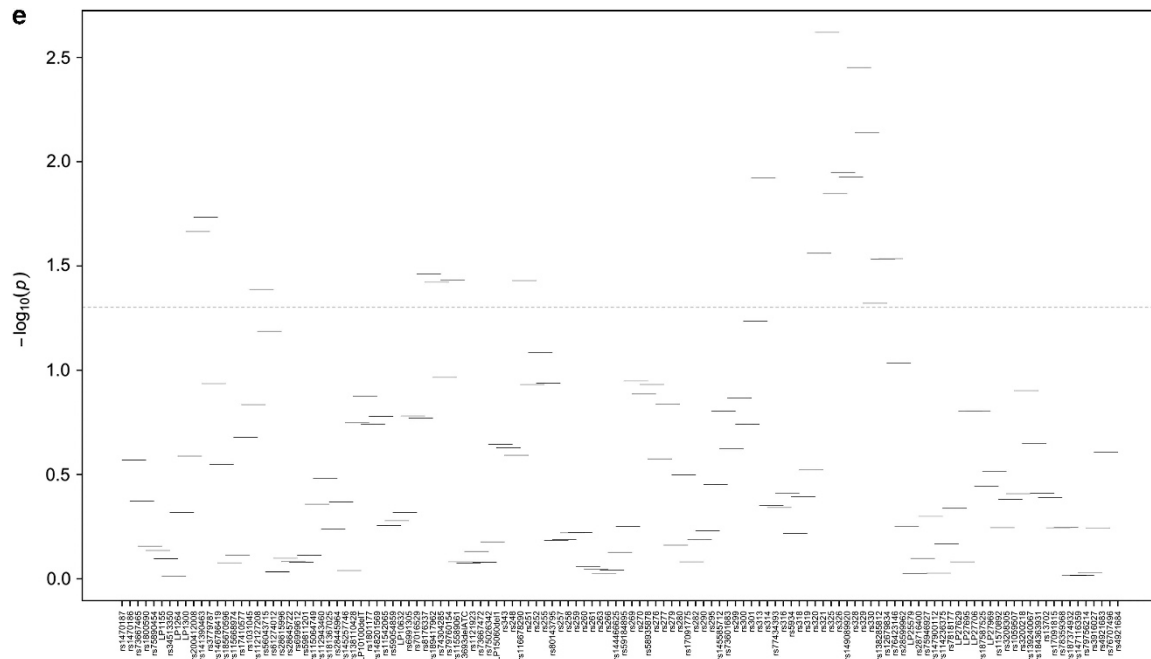


Figure 3 Continued.

associations (rs252:delA with LDL-C and rs316:C>A with HDL-C) still looked promising. However, given our modest sample size and generally small effect sizes of lipid-associated variants, our new observations warrant further evaluation in independent studies.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by US NIH grant HL084613.

- Castelli WP: Lipids, risk factors and ischaemic heart disease. *Atherosclerosis* 1996; **124**: S1–S9.
- Hunt SC, Hasstedt SJ, Kuida H *et al*: Genetic heritability and common environmental components of resting and stressed blood pressures, lipids, and body mass index in Utah pedigrees and twins. *Am J Epidemiol* 1989; **129**: 625–638.
- Kathiresan S, Melander O, Anevski D *et al*: Polymorphisms associated with cholesterol and risk of cardiovascular events. *N Engl J Med* 2008; **358**: 1240–1249.
- Willer CJ, Sanna S, Jackson AU *et al*: Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat Genet* 2008; **40**: 161–169.
- Aulchenko YS, Ripatti S, Lindqvist I *et al*: Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. *Nat Genet* 2009; **41**: 47–55.
- Teslovich TM, Musunuru K, Smith AV *et al*: Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 2010; **466**: 707–713.
- Willer CJ, Schmidt EM, Sengupta S *et al*: Discovery and refinement of loci associated with lipid levels. *Nat Genet* 2013; **45**: 1274–1283.
- Manolio TA, Collins FS, Cox NJ *et al*: Finding the missing heritability of complex diseases. *Nature* 2009; **461**: 747–753.
- Marian AJ, Belmont J: Strategic approaches to unraveling genetic causes of cardiovascular diseases. *Circ Res* 2011; **108**: 1252–1269.
- Cirulli ET, Goldstein DB: Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 2010; **11**: 415–425.
- Cohen JC, Kiss RS, Pertsemlidis A, Marcel YL, McPherson R, Hobbs HH: Multiple rare alleles contribute to low plasma levels of HDL cholesterol. *Science* 2004; **305**: 869–872.
- Evans D, Arzer J, Aberle J, Beil FU: Rare variants in the lipoprotein lipase (*LPL*) gene are common in hypertriglyceridemia but rare in type III hyperlipidemia. *Atherosclerosis* 2011; **214**: 386–390.

- Wang J, Cao H, Ban MR *et al*: Resequencing genomic DNA of patients with severe hypertriglyceridemia (MIM 144650). *Arterioscler Thromb Vasc Biol* 2007; **27**: 2450–2455.
- Johansen CT, Wang J, McIntyre AD *et al*: Excess of rare variants in non-genome-wide association study candidate genes in patients with hypertriglyceridemia. *Circ Cardiovasc Genet* 2012; **5**: 66–72.
- Khovidhunkit W, Chartyingcharoen P, Siriwong S *et al*: Resequencing *CETP*, *LIPC* and *LIPG* genes in Thai subjects with hyperalphalipoproteinemia. *Am J Cardiol* 2012; **110**: 62–66.
- Razzaghi H, Aston CE, Hamman RF, Kamboh MI: Genetic screening of the lipoprotein lipase gene for mutations associated with high triglyceride/low HDL-cholesterol levels. *Hum Genet* 2000; **107**: 257–267.
- Rip J, Nierman MC, Ross CJ *et al*: Lipoprotein lipase S447X: a naturally occurring gain-of-function mutation. *Arterioscler Thromb Vasc Biol* 2006; **26**: 1236–1245.
- Sagoo GS, Tatt I, Salanti G *et al*: Seven lipoprotein lipase gene polymorphisms, lipid fractions, and coronary disease: a HuGE association review and meta-analysis. *Am J Epidemiol* 2008; **168**: 1233–1246.
- Wright WT, Young IS, Nicholls DP, Graham CA: Genetic screening of the *LPL* gene in hypertriglyceridaemic patients. *Atherosclerosis* 2008; **199**: 187–192.
- Bunker CH, Ukoli FA, Okoro FI *et al*: Correlates of serum lipids in a lean black population. *Atherosclerosis* 1996; **123**: 215–225.
- Harris MR, Bunker CH, Hamman RF, Sanghera DK, Aston CE, Kamboh MI: Racial differences in the distribution of a low density lipoprotein receptor-related protein (LRP) polymorphism and its association with serum lipoprotein, lipid and apolipoprotein levels. *Atherosclerosis* 1998; **137**: 187–195.
- Richmond W: Preparation and properties of a cholesterol oxidase from *Nocardia* sp. and its application to the enzymatic assay of total cholesterol in serum. *Clin Chem* 1973; **19**: 1350–1356.
- Friedewald WT, Levy RI, Fredrickson DS: Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem* 1972; **18**: 499–502.
- Kamboh MI, Rewers M, Aston CE, Hamman RF: Plasma apolipoprotein A-I, apolipoprotein B, and lipoprotein(a) concentrations in normoglycemic Hispanics and non-Hispanic whites from the San Luis Valley, Colorado. *Am J Epidemiol* 1997; **146**: 1011–1018.
- Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.
- Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B* 1995; **57**: 289–300.
- Cohen J: *Statistical Power Analysis For The Behavioral Sciences*, second edition. Lawrence Erlbaum Associates: Hillsdale, NJ, USA, 1988.
- Lake SL, Lyon H, Tantisira K *et al*: Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 2003; **55**: 56–65.
- Lee S, Wu MC, Lin X: Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* 2012; **13**: 762–775.

- 30 Richardson K, Nettleton JA, Rotllan N *et al*: Gain-of-function lipoprotein lipase variant rs13702 modulates lipid traits through disruption of a microRNA-410 seed site. *Am J Hum Genet* 2013; **92**: 5–14.
- 31 Voruganti VS, Cole SA, Ebbesson SO *et al*: Genetic variation in APOJ, LPL, and TNFRSF10B affects plasma fatty acid distribution in Alaskan Eskimos. *Am J Clin Nutr* 2010; **91**: 1574–1583.
- 32 Boes E, Coassin S, Kollerits B, Heid IM, Kronenberg F: Genetic-epidemiological evidence on genes associated with HDL cholesterol levels: a systematic in-depth review. *Exp Gerontol* 2009; **44**: 136–160.
- 33 Fisher RM, Humphries SE, Talmud PJ: Common variation in the lipoprotein lipase gene: effects on plasma lipids and risk of atherosclerosis. *Atherosclerosis* 1997; **135**: 145–159.
- 34 Edmondson AC, Braund PS, Stylianou IM *et al*: Dense genotyping of candidate gene loci identifies variants associated with high-density lipoprotein cholesterol. *Circ Cardiovasc Genet* 2011; **4**: 145–155.
- 35 Boyle AP, Hong EL, Hariharan M *et al*: Annotation of functional variation in personal genomes using RegulomeDB. *Genome Res* 2012; **22**: 1790–1797.
- 36 Pirim D, Wang X, Radwan ZH *et al*: Lipoprotein lipase gene sequencing and plasma lipid profile. *J Lipid Res* 2014; **55**: 85–93.
- 37 Nickerson DA, Taylor SL, Weiss KM *et al*: DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. *Nat Genet* 1998; **19**: 233–240.
- 38 Wittrup HH, Tybjaerg-Hansen A, Nordestgaard BG: Lipoprotein lipase mutations, plasma lipids and lipoproteins, and risk of ischemic heart disease. A meta-analysis. *Circulation* 1999; **99**: 2901–2907.
- 39 Ahn YI, Kamboh MI, Hamman RF *et al*: Two DNA polymorphisms in the lipoprotein lipase gene and their associations with factors related to cardiovascular disease. *J Lipid Res* 1993; **34**: 421–428.
- 40 Komurcu-Bayrak E, Onat A, Poda M *et al*: The S447X variant of lipoprotein lipase gene is associated with metabolic syndrome and lipid levels among Turks. *Clin Chim Acta* 2007; **383**: 110–115.
- 41 Klos KL, Sing CF, Boerwinkle E *et al*: Consistent effects of genes involved in reverse cholesterol transport on plasma lipid and apolipoprotein levels in CARDIA participants. *Arterioscler Thromb Vasc Biol* 2006; **26**: 1828–1836.
- 42 Spirin V, Schmidt S, Pertsemlidis A, Cooper RS, Cohen JC, Sunyaev SR: Common single-nucleotide polymorphisms act in concert to affect plasma levels of high-density lipoprotein cholesterol. *Am J Hum Genet* 2007; **81**: 1298–1303.
- 43 Deo RC, Reich D, Tandon A *et al*: Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. *PLoS Genet* 2009; **5**: e1000342.
- 44 Tang W, Apostol G, Schreiner PJ *et al*: Associations of lipoprotein lipase gene polymorphisms with longitudinal plasma lipid trends in young adults: the Coronary Artery Risk Development in Young Adults (CARDIA) study. *Circ Cardiovasc Genet* 2010; **3**: 179–186.
- 45 Kathiresan S, Melander O, Guiducci C *et al*: Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* 2008; **40**: 189–197.
- 46 Heid IM, Boes E, Müller M *et al*: Genome-wide association analysis of high-density lipoprotein cholesterol in the population-based KORA study sheds new light on intergenic regions. *Circ Cardiovasc Genet* 2008; **1**: 10–20.
- 47 Kraja AT, Vaidya D, Pankow JS *et al*: A bivariate genome-wide approach to metabolic syndrome: STAMPEED consortium. *Diabetes* 2011; **60**: 1329–1339.
- 48 Lettre G, Palmer CD, Young T *et al*: Genome-wide association study of coronary heart disease and its risk factors in 8090 African Americans: the NHLBI CARE Project. *PLoS Genet* 2011; **7**: e1001300.
- 49 Holmer SR, Hengstenberg C, Mayer B *et al*: Lipoprotein lipase gene polymorphism, cholesterol subfractions and myocardial infarction in large samples of the general population. *Cardiovasc Res* 2000; **47**: 806–812.
- 50 Corella D, Guillén M, Sáiz C *et al*: Associations of *LPL* and *APOC3* gene polymorphisms on plasma lipids in a Mediterranean population: interaction with tobacco smoking and the APOE locus. *J Lipid Res* 2002; **43**: 416–427.
- 51 Ukkola O, Garenc C, Pérusse L *et al*: Genetic variation at the lipoprotein lipase locus and plasma lipoprotein and insulin levels in the Québec Family Study. *Atherosclerosis* 2001; **158**: 199–206.
- 52 Chen Q, Razzaghi H, Demirci FY, Kamboh MI: Functional significance of lipoprotein lipase HindIII polymorphism associated with the risk of coronary artery disease. *Atherosclerosis* 2008; **200**: 102–108.
- 53 Larson I, Hoffmann MM, Ordovas JM, Schaefer EJ, März W, Kreuzer J: The lipoprotein lipase HindIII polymorphism: association with total cholesterol and LDL-cholesterol, but not with HDL and triglycerides in 342 females. *Clin Chem* 1999; **45**: 963–968.
- 54 Munshi A, Babu MS, Kaul S, Rajeshwar K, Balakrishna N, Jyothy A: Association of *LPL* gene variant and LDL, HDL, VLDL cholesterol and triglyceride levels with ischemic stroke and its subtypes. *J Neurol Sci* 2012; **318**: 51–54.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)