Lipoprotein lipase gene sequencing and plasma lipid profile

Dilek Pirim,* Xingbin Wang,† Zaheda H. Radwan,* Vipavee Niemsiri,* John E. Hokanson,† Richard F. Hamman,‡ M. Michael Barmada,* F. Yesim Demirci,* and M. Ilyas Kamboh*§

Department of Human Genetics,* Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA; and Department of Epidemiology,§ Colorado School of Public Health, University of Colorado Denver, Aurora, CO

Abstract  Lipoprotein lipase (LPL) plays a crucial role in lipid metabolism by hydrolyzing triglyceride (TG)-rich particles and affecting HDL cholesterol (HDL-C) levels. In this study, the entire LPL gene plus flanking regions were resequenced in individuals with extreme HDL-C/TG levels (n = 95), selected from a population-based sample of 623 US non-Hispanic White (NHW) individuals. A total of 176 sequencing variants were identified, including 28 novel variants. A subset of 64 variants [common tag single nucleotide polymorphisms (tagSNP) and selected rare variants] were genotyped in the total sample, followed by association analyses with major lipid traits. A gene-based association test including all genotyped variants revealed significant association with HDL-C (P = 0.024) and TG (P = 0.006). Our single-site analysis revealed seven independent signals (P < 0.05; r² < 0.40) with either HDL-C or TG. The most significant association was for the SNP rs295 exerting opposite effects on TG and HDL-C levels with P values of 7.5 x 10⁻⁴ and 0.002, respectively. Our work highlights some common variants and haplotypes in LPL with significant associations with lipid traits; however, the analysis of rare variants using burden tests and SKAT-O method revealed negligible effects on lipid traits. Comprehensive resequencing of LPL in larger samples is warranted to further test the role of rare variants in affecting plasma lipid levels.—Pirim, D., X. Wang, Z. H. Radwan, V. Niemsiri, J. E. Hokanson, R. F. Hamman, M. M. Barmada, F. Y. Demirci, and M. Iyas Kamboh. Lipoprotein lipase gene sequencing and plasma lipid profile. J. Lipid Res. 2014. 55: 85–93.

Supplementary key words  triglycerides • lipid metabolism • rare variants • candidate gene • genotyping • genetic association

Coronary heart disease (CHD) is the most common cause of death in the Western world (1). Abnormal plasma lipid levels constitute major risk factors for CHD (2), and almost 50% of the variation in lipid levels is believed to be under genetic control (3, 4). During the past few years, great progress has been made in understanding the genetic architecture of inter-individual variation in lipid traits by genome-wide association studies (GWAS) (5–8). However, most of the identified loci have small effect sizes with odds ratios (OR) ranging from 1.05 to 1.20, and overall they explain only ~30% of the heritability estimated for major lipid traits (8, 9). Because common variants together explain only a modest fraction of the underlying genetic basis of inter-individual variation in major lipid traits, attention has been focused on elucidating the roles of rare variants in lipid phenotypes (10–14). It has been suggested that multiple rare variants may collectively have moderate to strong effects on the phenotypic variation (rare variant common disease hypothesis) (15). Cohen et al. (12) were first to report that multiple rare variants in the ABCA1, APOA1, and LCAT genes contribute to the variation in plasma HDL-C levels in the general population. Likewise, Johansen et al. (13) resequenced the APOA5, GCKR, LPL, and APOB genes that were reported to be associated with hypertriglyceridemia (HTG) in GWASs and found a significant excess of rare variants in individuals with HTG as compared with controls. Here, we sought to further contribute to the current literature by resequencing the lipoprotein lipase gene, a major candidate gene for CHD, in individuals with extreme lipid phenotypes in order to discover and evaluate high penetrance rare and low frequency variants [minor allele frequency (MAF) < 5%] in addition to the common variants (MAF ≥5%).

Human lipoprotein lipase (LPL protein; LPL gene) is encoded by 10 exons; the entire gene spans approximately 30 kb on chromosome 8p22 (16). LPL is a major rate-limiting enzyme that hydrolyzes triglyceride (TG)-rich particles into free fatty acids and glycerol, and it plays a key role in affecting plasma lipid levels.

Abbreviations:  CHD, coronary heart disease; eQTL, expression quantitative trait loci; FDR, false discovery rate; GWAS, genome-wide association study; HDL-C, HDL cholesterol; HTG, hypertriglyceridemia; LD, linkage disequilibrium; MAF, minor allele frequency; NHW, non-Hispanic White; QC, quality control; SKAT, sequence kernel association test; TC, total cholesterol; TF, transcription factor.

To whom correspondence should be addressed.

*Corresponding authors.

This study was supported by National Institutes of Health Grant HL084613.

Manuscript received 22 August 2013 and in revised form 28 October 2013.

Published, JLR Papers in Press, November 8, 2013

DOI 10.1194/jlr.M043265

Copyright © 2014 by the American Society for Biochemistry and Molecular Biology, Inc.

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

Journal of Lipid Research  Volume 55, 2014  85
role in converting very low density lipoprotein (VLDL) to low density lipoprotein (LDL). It is well known that LPL activity and mass are associated with fluctuation TG and high-density lipoprotein cholesterol (HDL-C) levels (17–19). Several common polymorphisms and functional variants in the LPL gene have been identified to be associated with lipid profile and risk of CHD in various studies (20).

In this study, we have comprehensively evaluated the role of LPL genetic variation in relation to plasma lipid levels using a three-step approach: i) SNP discovery by sequencing non-Hispanic White (NHW) individuals having extreme HDL-C/TG levels; ii) genotyping the variants identified in the discovery phase in the total sample of NHWs; and iii) association testing using single-site, haplotype, and rare variant analyses.

MATERIALS AND METHODS

Subjects

The study samples comprised 623 NHWs derived from the San Luis Valley Diabetes Study, a population-based case-control study of type 2 diabetes in the San Luis Valley, Southern Colorado (21, 22). The 623 subjects used in this current study did not have diabetes; a detailed description of this sample set, including the biometric and quantitative data, can be found in Demirci et al. (23). For resequencing the entire LPL gene, we selected 95 individuals who fell in the upper (n = 47) and lower (n = 48) 10th percentile distribution of plasma HDL-C and TG levels (Table 1). The study was approved by the University of Pittsburgh and University of Colorado Denver Institutional Review Boards, and all study participants provided written informed consent.

Lipid measurements

Fasting total cholesterol was determined by esterase-oxidase method (24). Serum HDL-C and triglyceride concentrations were measured by enzymatic procedures described in Harris et al. (25). LDL cholesterol was calculated by using the Friedewald equation if triglyceride levels were less than 400 mg/dl (4.5 mmol/l) (26).

DNA sequencing

The accession number for DNA reference sequence used in PCR primer design and sequence comparison for the LPL gene is NC_00008.10, which was derived from NCBI GenBank (http://www.ncbi.nlm.nih.gov). A total of 37 overlapping sequencing amplicons were sequenced in both directions. Amplicon sizes and primer sequences are given in supplementary Table I. These amplicons cover the entire LPL gene (27,993 bp) as well as 1,196 bp in the 5′ flanking region and 1 kb in the 3′ flanking region resulting in a total of 30,189 bp genomic region covered. Although we designed most of the primers by using Primer 3 software (http://frodo.wi.mit.edu/primer3/), we also used a subset of primers from a previous study (27) that sequenced only a portion of the LPL gene (9.7 kb), from 3′ end of intron 3 to 5′ end of intron 9. PCR conditions are available upon request. DNA samples were extracted from buffy coat using standard DNA extraction procedures. The PCR-amplified DNA samples were sent to a commercial lab (Beckman Coulter Genomics, Danvers, MA) where automated Sanger sequencing was performed on ABI 3730 × 1 DNA Analyzers. Sequencing data was analyzed in our lab using Variant Reporter version 1.0 (Applied Biosystems, Foster City, CA) and Sequencher version 4.8 (Gene Codes Corporation, Ann Arbor, MI).

Genotyping

For genotyping the entire sample of 623 individuals, we used either TaqMan (Applied Biosystems) or iPLEX Gold (Sequenom, San Diego, CA) genotyping methods following manufacturers’ protocols. The 384-well plates containing dried whole genome amplified DNAs were used in both methods. After thermal cycling of custom or premade TaqMan assays, the ABI Prism 7900HT Sequence Detection Systems was used for endpoint fluorescence reading of the 384-well plates. The iPLEX Gold genotyping was performed in the Core laboratories of the University of Pittsburgh. Sequences of primers and probes used for custom TaqMan and iPLEX Gold genotyping are available upon request.

Statistical analyses

Variants identified by sequencing were analyzed by using Haploview (28), (www.broadinstitute.org/haploview) to test the concordance of the genotype distribution with Hardy-Weinberg equilibrium and to determine allele frequencies and their distributions among high and low HDL/TG groups and their linkage disequilibrium (LD) patterns. For those SNPs that were genotyped in the entire sample, the additive linear regression model was used to test for the effects of genotypes on the means of plasma HDL-C, TG, LDL-C, and total cholesterol (TC) levels. HDL-C and TG levels were transformed to natural logarithms using Box-Cox transformation to improve normality. The significant covariates were identified using stepwise regression in both directions. The covariates included in the final model were sex, age, BMI and smoking. The R statistical software package (http://www.r-project.org) was used to perform all computations. A nominal P value of less than 0.05 was considered as suggestive evidence of association. However, we also used false discovery rate (FDR) method (29) to control for multiple testing. The versatile gene-based associations (VEGA) were also performed to assess the relationship between traits and LPL (30). For haplotype association analysis, the generalized linear model (GLM) was used (31). Including too many haplotypes can make above model inefficient and impractical. To reduce the number of haplotypes considered in association

TABLE 1. Characteristics of the resequencing sample (95 NHWs)

| Characteristics | High HDL / Low TG (n = 47) | Low HDL / High TG (n = 48) | P* |
|-----------------|--------------------------|--------------------------|----|
| Male/female (%) | 51/48.9                  | 50/50                    | 0.92 |
| Age (years)     | 55.45 ± 9.8              | 53.95 ± 10.54            | 0.25 |
| BMI (kg/m^2)    | 23.17 ± 3.17             | 27.35 ± 3.90             | <0.001 |
| LDL-C (mg/dl)   | 126.84 ± 46.95           | 125.54 ± 54.97           | 0.90 |
| HDL-C (mg/dl)   | 77.68 ± 13.32            | 31.81 ± 4.37             | <0.001 |
| Triglycerides (mg/dl) | 114.09 ± 60.88   | 240.21 ± 153.22        | <0.001 |
| Total cholesterol (mg/dl) | 227.34 ± 51.76  | 208.81 ± 44.65        | 0.06 |

*P values were calculated based on the original values by using t-test. No covariates were included.
analysis, we used the sliding window, four SNPs per window, and assessed evidence for association within each window. Specifically, a global \(P\) value for testing overall effects of the haplotypes with frequency greater than 0.01 was used to assess the association between the trait and haplotypes in each window. Sliding-window haplotype analysis was performed with the haplo.glm function in the Haplo.stats R package (version 1.5.0).

We analyzed the cumulative effects of uncommon/rare variants by using two methods: burden test (32) and a sequence kernel association test (SKAT-O) (33); SKAT-O has been proposed to be the optimal test for rare variant analysis and exceeded the SKAT and burden tests in several ways. The analyses were performed by using three different minor allele frequency bin thresholds (<1%, <2%, and <5%). The SKAT method was implemented using the “SKAT” R package.

**RESULTS**

**DNA resequencing results**

Complete resequencing of the LPL gene in selected 95 individuals falling in the upper (\(n = 47\)) and lower (\(n = 48\)) 10th percentile of HDL-C/TG distribution identified a total of 176 variants, including 159 single nucleotide substitutions and 17 insertions or deletions (indels) (supplementary Table II and supplementary Fig. I).

Of the 176 variants identified, 148 were previously reported in public databases. Among these 176 variants, 88 had MAF ≥ 0.05; 52 had a MAF between 0.05 and 0.01; and 36 had MAF < 0.01. Seventeen of these variants were indels, and the remaining 159 were single nucleotide substitutions (94 transitions and 65 transversions). The proportions of substitutions were A/G, 23.27%; A/C, 8.33%; A/T, 5.66%; C/T, 35.84%; C/G, 10.06%; and G/T, 13.83%. One hundred forty variants were located in introns and 11 in the flanking regions. The remaining 25 variants were exonic, and only 6 of them were located in translated regions of the exons. Of these 6 exonic coding variants, 3 resulted in nonsynonymous changes: aspartate to asparagine (D9N) in exon 2, asparagine to serine (N291S) in exon 6, and threonine to threonine (T361T) in exon 8. Of the identified 17 indels, only one of the insertions was located in the exonic region affecting the 3′ untranslated region (UTR) in exon 10. The size range of indels was 1–20 bases, except for a 697 nucleotide deletion identified in intron 2.

Of the 28 novel variants that we identified (highlighted in bold in supplementary Table II), 3 had MAF ≥ 0.05, 3 had MAF between 0.05 and 0.01, and 22 had MAF < 0.01. Two of them were located in flanking regions; 24 were in introns; 2 were in UTR of exons 1 and 10, respectively. Twelve of them were transitions; 7 were transversions; and 9 were indels. A previously unreported large deletion (12224_12920del697) was observed in intron 2 (supplementary Table II and supplementary Fig. I).

**Distribution of identified LPL variants in extreme HDL-C/TG groups**

Distribution of the 176 identified variants between the two extreme groups is shown in supplementary Table III.

Of the 88 uncommon or rare variants (MAF < 0.05), 21 were present only in the low HDL-C/low TG group and 22 were present only in the high HDL-C/low TG group; the remaining 45 were present in both groups. Forty of 47 (85.1%) individuals with high HDL-C/low TG carried at least 1 rare variant versus 35 out of 48 (72.9%) individuals with low HDL-C/high TG. Furthermore, 21 of 47 (44.7%) individuals with high HDL-C/low TG had at least 2 rare variants versus 16 out of 48 (33.3%) individuals with low HDL-C/high TG; 14 of 47 (29.7%) individuals with high HDL-C/low TG had at least 3 rare variants versus 7 out of 48 (14.6%) individuals with low HDL-C/high TG.

Among the 25 identified exonic variants, 12 were relatively uncommon or rare (MAF < 0.05). Of these 12 variants, three were found only in the high HDL-C/low TG group [1088G>T (rs8051041) in the UTR of exon 1, 24143C>G (rs328, Ser447×) in exon 9 and 27783A>T (novel variant) in the UTR of exon 10] and another three were found only in low HDL-C/high TG group [958G>A (novel variant) and 27688C>T (rs19121278) in the UTR of exon 1, and 28524C>T (novel variant) in the UTR of exon 10]. The remaining 19 exonic variants were found in both the high and low HDL-C/TG groups.

Of the 17 identified indels, four intronic variants were found only in high HDL-C/low TG group [11888_11889insA (rs149017698), 12224_12920del697 (novel variant), 12878_12889del12 (novel variant), and 12884_12887del4 (novel variant)], whereas one intronic and one 3′ flanking region variants [21125_21128del4 (novel variant) and 29557_29558insA (novel variant)] were found only in low HDL-C/high TG group.

**Linkage disequilibrium and tagger analyses of LPL variants identified in sequencing and follow-up genotyping in the entire sample**

We used linkage disequilibrium (LD) and tagger analysis to identify tagSNPs for the common 88 sequencing variants (MAF ≥ 0.05). By using an \(r^2\) cutoff of 0.9, 43 common tagSNP bins were identified (supplementary Table IV). The SNP genotype data from the HapMap CEU population were used to compare the HapMap tagSNPs with our sequencing-derived tagSNPs. We identified 24 common tagSNP bins (\(r^2 = 0.9\) capturing 48 HapMap SNPs (MAF ≥ 0.05) in the same region; all of them were captured by our sequencing-derived tagSNPs. Thus, the 43 tagSNPs identified in our study were used for subsequent genotyping in the entire sample.

Overall, the following criteria were considered in selecting the sequencing variants for genotyping in the total sample of 623 individuals: i) common tagSNPs (MAF ≥ 0.05, \(r^2 ≥ 0.9\)); ii) all variants located in exons or intron-exon junctions; iii) all uncommon or rare variants (MAF < 0.05) present in two or more individuals included in the sequencing; and iii) suspicious rare variants identified in the sequencing. We selected 86 variants (43 common tagSNPs and 43 uncommon/rare variants) using these criteria. Sixty-five of these variants (35 common tagSNPs and 30 others) were successfully genotyped in the total sample of 623 NHWs using either iPLEX Gold or TaqMan methods; 18 failed genotyping, and 3 suspicious rare variants were not confirmed and thus excluded as...
sequencing artifacts. The genotype call rates and other features of the 65 genotyped variants are shown in supplementary Table V. The discrepancy rate was determined to be 0–0.5% for the genotyped variants based on the random repeats of ~10% of the samples. All genotyped variants were in concordance with Hardy-Weinberg equilibrium expectations. We excluded one SNP (rs343) with genotyping efficiency less than 80%, and so a total of 64 variants (40 common and 24 uncommon/rare based on their frequency in the entire sample) were included in subsequent association analyses. The correlations ($r^2 \geq 0.80$) between these 64 SNPs are shown in supplementary Table VI.

**Gene-based association test results**

Initially we conducted a gene-based association test to evaluate the joint effects of all 64 genotyped variants. This analysis revealed significant association with HDL-C ($P = 0.024$) and with TG ($P = 0.006$). The best SNP associated with HDL-C was rs295 ($P = 0.002$), and for TG, it was rs80181352 ($P = 6.6E-04$) (Table 2). Since gene-based testing revealed associations only with HDL-C and TG, these two traits were further examined by single-site, haplotype, and rare variant analyses.

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

Single-site analyses of 40 common variants with HDL-C and TG levels revealed 22 nominal associations ($P < 0.05$) with either HDL-C or TG (rs1326624, rs270, rs117026536, rs328, rs11570891, rs1059611) or TG (rs8176337, rs80181352, rs248, rs277, rs294, rs312, rs316, rs4922115, rs4921684) or both (rs282, rs286, rs295, rs314, rs320, rs327, rs13702) (Table 3 and supplementary Tables VII and VIII).

Among the seven SNPs that were significantly associated with both traits, rs295 in intron 6 showed the most significant association with both HDL-C ($\beta = 0.039; P = 0.002$) and TG ($\beta = -0.099; P = 7.5E-04$) and this SNP was in LD with four (three intronic and one at 3’UTR) of the remaining six SNPs that also showed associations with both traits (rs314, rs320, rs327 and rs13702 ($r^2 = 0.71-0.74$; Fig. 1). Among these 4 SNPs, rs320 (HindIII polymorphism) has previously been shown to be associated with TG levels (34–36). The other two intronic SNPs (rs282 and rs286) that showed association with both HDL-C ($\beta = 0.032; P = 0.042$ and $\beta = 0.051; P = 0.005$, respectively) and TG ($\beta = -0.095; P = 0.006$ and $\beta = -0.090; P = 0.021$, respectively) were not in LD with rs295 ($r^2 = 0.088$ and $r^2 = 0.36$, respectively) or with each other ($r^2 = 0.01$). However, rs286 was in LD with a coding SNP, rs328 (Ser447×) ($r^2 = 0.81$) that showed a modest association with HDL-C ($P = 0.032$) and a trend for association with TG ($P = 0.088$), as well as with three other SNPs (rs117026536, rs11570891 and rs1059611) that showed nominal significance for HDL-C. Three SNPs (2 intronic and 1 exonic) were associated with only TG with $P < 0.01$: rs80181352 ($P = 6.6E-04$), rs248/EL118E ($P = 0.007$) and rs294 ($P = 0.005$). While rs80181352 and rs248 were in strong LD with each other ($r^2 = 0.88$), their association with TG was independent of the above mentioned SNPs that showed significant associations with both TG and HDL-C. Four other SNPs (rs312, rs316/T361T, rs4922115 and rs4921684) that showed nominal significance for TG were in moderate to high LD ($r^2 = 0.69-0.96$) with rs294. Thus, we have observed five relatively independent ($r^2 < 0.40$) signals with $P < 0.01$ associated with HDL-C and/or TG, including rs80181352, rs282, rs286, rs294, and rs295. As noted above, two of these SNPs (rs80181352 and rs295) were the best associated with TG and HDL-C, respectively, according to the gene-based test. Two additional SNPs, albeit less significant, also showed independent association with TG: rs8176337 ($P = 0.027$) and rs277 ($P = 0.036$, in LD with rs13266204 and rs270 with $r^2 = 0.56-0.81$).

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

We performed the uncommon/rare variant analysis on 24 variants with MAF < 0.05 using burden tests and SKAT-O analysis. The analyses were performed on three sets of uncommon/rare variants using the MAF thresholds of 1%, 2% and 5%. However, our data revealed no significant association with either HDL-C or TG (data not shown). We also performed burden tests and SKAT-O analysis on 12 potentially regulatory variants based on their RegulomeDB scores of 1–5 (discussed below) but again found no significant association with both HDL-C or TG levels (data not shown). Although there was no evidence of cumulative effects of uncommon/rare LPL variants on lipid levels, in single site analysis, we observed three individuals with a novel rare variant (1130G>C; MAF = 0.002) that was associated with TG levels ($P = 0.017; \beta = 0.556$) (see supplementary Table VIII). This variant is located in the 5’UTR and thus may have a functional implication as discussed below.

**Haplotype-based association test results**

The haplotype structure of the successfully genotyped 64 variants were constructed by using the sliding window approach that includes four SNPs in each window and sliding one SNP at a time; $P$-values were calculated based on the comparison to the most common haplotype as a reference (R package-Haplostat). As a result, 61 overlapping sliding windows were constructed and examined for association with HDL-C and TG levels (supplementary Fig. II-A,

### Table 2. Gene-based association test based on all 64 LPL SNPs

| Trait | Chromosome | Gene | Number of SNPs | Test | $P^a$ | Smallest $P^b$ |
|-------|------------|------|----------------|------|-------|---------------|
| HDL-C | 8          | LPL  | 64             | 134.407 | 0.024 | 0.002 (rs295) |
| LDL-C | 8          | LPL  | 64             | 35.7124 | 0.885 | 0.064 (rs292) |
| TG    | 8          | LPL  | 64             | 167.036 | 0.006 | 6.6,10⁻⁴ (rs80181352) |
| TC    | 8          | LPL  | 64             | 34.5964 | 0.925 | 0.177 (rs312) |

$^a$ Gene-based value.

$^b$ Smallest $P$ detected for a single SNP.

88 Journal of Lipid Research Volume 55, 2014
| RefSNP ID | Genotypes | MAF | Location | Adjusted Mean | Beta | P | FDR | Adjusted Mean | Beta | P | FDR | RegulomeDB Score |
|-----------|------------|-----|----------|---------------|------|---|-----|---------------|------|---|-----|-----------------|
| rs13266204 | AA/AG/GG   | 0.214 | Intron 1 | 50.97/50.25/46.11 | -0.027 | 0.033 | 0.211 | 138.07/136.7/150.67 | 0.028 | 0.328 | 0.583 | 6 |
| rs8176337  | CC/CG/GG   | 0.242 | Intron 2 | 50.16/52.23/47.59 | 0.012 | 0.387 | 0.563 | 142.79/131.6/134.54 | -0.065 | 0.027 | 0.124 | No data |
| rs80181352 | GG/GT/TT   | 0.075 | Intron 2 | 50.47/50.56/52.9 | 0.009 | 0.668 | 0.792 | 140.4/124.73/86.8 | -0.148 | 6.610^-4 | 0.024 | 5 |
| rs9248     | AA/GA/AG   | 0.067 | Exon 4 (E118E) | 67.05/50.45/50.53 | 0.019 | 0.375 | 0.559 | 59.35/129.61/139.56 | 0.028 | 0.007 | 0.061 | 5 |
| rs270      | AA/CA/CC   | 0.172 | Intron 6 | 50.29/48.85/51.35 | 0.031 | 0.025 | 0.211 | 133.94/132.87/134.57 | 0.059 | 0.033 | 0.178 | No data |
| rs277      | CC/CT/TT   | 0.196 | Intron 6 | 50.83/49.53/51.13 | 0.020 | 0.135 | 0.362 | 142.79/131.64/134.54 | 0.061 | 0.036 | 0.137 | No data |
| rs282      | CC/GG/CG   | 0.135 | Intron 6 | 50.16/51.58/50.07 | 0.032 | 0.042 | 0.230 | 138.24/128.28/141.24 | -0.095 | 0.006 | 0.061 | 5 |
| rs286      | AA/TA/TT   | 0.095 | Intron 6 | 53.66/53.55/49.92 | 0.051 | 0.005 | 0.133 | 100.36/130.1/140.13 | -0.09 | 0.021 | 0.103 | 5 |
| rs294      | AA/GG/AG   | 0.118 | Intron 6 | 50.35/50.72/57.5 | 0.021 | 0.205 | 0.392 | 140.91/130.55/80.5 | 0.102 | 0.005 | 0.061 | 6 |
| rs295      | AA/CA/CC   | 0.223 | Intron 7 | 49.52/52.37/53.57 | 0.039 | 0.060 | 0.133 | 142.94/133.96/107.17 | 7.5.10^-4 | 0.014 | 0.006 | 0.178 |
| rs312      | CC/GG/CG   | 0.113 | Intron 7 | 57.46/50.86/50.45 | 0.022 | 0.197 | 0.392 | 80.33/131.7/140.37 | -0.095 | 0.010 | 0.069 | No data |
| rs314      | AA/GA/AG   | 0.263 | Intron 7 | 52.48/51.41/49.94 | 0.026 | 0.043 | 0.230 | 121.17/137.5/143.03 | 0.066 | 0.015 | 0.092 | No data |
| rs316      | AA/CA/CC   | 0.113 | Intron 8 (T361T) | 57.58/51.07/50.51 | 0.024 | 0.155 | 0.362 | 81.25/132.13/139.77 | -0.087 | 0.017 | 0.092 | 6 |
| rs17026536 | GG/GT/TT   | 0.104 | Intron 8 | 50.04/52.88/49.83 | 0.034 | 0.048 | 0.235 | 139.51/132.92/113.64 | -0.060 | 0.112 | 0.510 | 6 |
| rs320      | GG/GT/TT   | 0.258 | Intron 8 | 51.97/51.46/49.74 | 0.027 | 0.026 | 0.211 | 128.56/131.55/80.5 | -0.068 | 0.008 | 0.061 | 6 |
| rs327      | GG/GT/TT   | 0.265 | Intron 8 | 52.41/51.15/49.87 | 0.025 | 0.032 | 0.211 | 127.59/132.53/143.84 | -0.069 | 0.007 | 0.061 | 6 |
| rs328      | CC/GC/CG   | 0.103 | Intron 9 (S447×) | 50.05/52.98/50.36 | 0.038 | 0.032 | 0.211 | 139.45/132.99/112.46 | -0.065 | 0.088 | 0.277 | 5 |
| rs1570891  | CC/TC/CT   | 0.107 | Intron 9 | 50.02/52.88/25.79 | 0.039 | 0.024 | 0.211 | 139.3/134.79/104.16 | -0.058 | 0.119 | 0.316 | 5 |
| rs4922115  | AA/AG/AG   | 0.146 | 3′UTR-exon 10 | 57.92/49.61/50.68 | 0.015 | 0.304 | 0.487 | 108.82/134/140.58 | -0.066 | 0.039 | 0.10 | No data |
| rs3702     | CC/CT/TT   | 0.272 | 3′UTR-exon 10 | 52.79/51.67/49.6 | 0.032 | 0.006 | 0.133 | 121.76/133.5/143.62 | -0.074 | 0.005 | 0.061 | 6 |
| rs105611   | CC/TC/CT   | 0.107 | 3′UTR-exon 10 | 52.82/52.75/57.85 | 0.037 | 0.030 | 0.211 | 104.24/156.09/139.14 | -0.050 | 0.176 | 0.401 | 6 |
| rs4921684  | CC/TC/CT   | 0.149 | 3′ flanking | 50.7/49.57/57.85 | 0.013 | 0.360 | 0.559 | 140.4/133.61/180.73 | -0.068 | 0.032 | 0.129 | 5 |

* RegulomeDB scores were generated by using http://regulome.stanford.edu/. Scores represent the following categories/subcategories. 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 footprint + matched TF motif + matched DNase peak; 1c = eQTL + TF binding + matched TF motif + DNase peak; 1d = eQTL + TF binding + matched TF motif + matched DNase footprint + 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 + matched TF motif + 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. Categories 4–6 (minimal binding evidence): 4 = TF binding + DNase peak; 5 = TF binding or DNase peak; 6 = motif hit.
B, respectively). Four-SNP haplotypes were arranged on the X axis by the location from the 5’ to 3’ end of the LPL gene. The threshold for statistical significance ($P = 0.05$) was shown by a horizontal red line. Overall, the haplotype effects were more significant for TG than for HDL-C; 29 significant global $P$ values for TG versus 12 for HDL-C. Specifically, the intron 6-intron 7 region was associated with HDL-C and the intron 6-intron 8 region with TG. Window 37, which includes rs296, rs294, rs295, and rs311 SNPs, revealed the most significant association with HDL-C (global $P = 0.012$) and window 33, which includes rs282, rs283, rs285, and rs286 SNPs, showed the strongest association with TG (global $P = 5E−04$) (supplementary Tables IX and X). Both of these windows harbored variants that showed significant associations with HDL-C and/or TG in single-site analyses.

Functional annotation of significant SNPs

We used the RegulomeDB (37) online database (http://regulome.stanford.edu/) to extract information regarding the possible regulatory effects of SNPs that were significantly associated with lipid traits. The regulome scores for 22 significant SNPs are presented in Table 3. Fifteen of the 22 SNPs were located in LPL genomic sequences that were predicted to have some regulatory implication. The majority of associated SNPs fell into category 5 and category 6, suggesting there was minimal evidence that these variants perturbed the transcription factor (TF) binding sites or disrupted the regulatory motifs, but supporting expression quantitative trait loci (eQTL) data were not available. However, rs316 was classified as 1f within category 1, which had eQTL information in addition to other sources of data. This score indicates that rs316, which is a synonymous variant located in exon 8, is a known cis-eQTL SNP for LPL and lies within a TF binding site or a DNase peak based on the experimental evidence. It is important to note that these scores were generated based on the prediction of the “regulatory” function, and thus, high scores do not rule out the functional significance of these variants; for instance, rs328 (Ser447×) is a well-known functional coding variant and is classified in category 5 of RegulomeDB. Also, the novel rare variant (1130G>C) associated with TG levels appears to be regulatory as indicated by the RegulomeDB score of 2b (likely to affect TF binding) (see supplementary Table II).

DISCUSSION

We have comprehensively examined the role of common and rare genetic variants in the LPL gene in relation to plasma variation in TG and HDL-C levels in a well-characterized, population-based sample of NHWs. In the variant discovery step, we resequenced the entire LPL gene in 95 individuals...
falling at the two extreme ends of the HDL-C/TG distribution and then genotyped common tagSNPs and rare variants in the entire sample of 623 individuals, followed by association analyses. We hypothesized that individuals with extreme lipid phenotypes might have rare and/or low frequency \( LPL \) variants not previously reported in the literature. For genetic associations, we focused on TG and HDL-C because of the established functional role of \( LPL \) in metabolizing TG and its inverse correlation with HDL-C (38, 39) as well as our initial gene-based analysis that revealed significant associations only with HDL-C and TG.

To our knowledge, this is the first study that has sequenced the entire \( LPL \) gene (\( \approx 30 \) kb) in selected individuals with extreme lipid phenotypes from the general population. Previously, Nickerson et al. (27) sequenced a portion of the \( LPL \) gene (9.7 kb), from 3’ end of intron 3 to 5’ end of intron 9, in 71 individuals randomly selected from three racial groups, including European-Americans from Rochester, MN (n = 23), African Americans from Jackson, MS (n = 24), and Europeans from North Karelia, Finland (n = 24). Five uncommon variants (\( \text{MAF} = 0.05 \)) and one common deletion (8538delA; \( \text{MAF} = 0.20 \)) reported by Nickerson et al. (27) in their White sample were not identified in our sample. On the other hand, we identified 17 variants (\( \text{MAF} \leq 0.053 \)) not reported by Nickerson et al. (27). This difference might be due, in part, to the use of different sample selection criteria and/or to the use of different software tools and quality control (QC) criteria for variant calling (Nickerson et al. used PolyPhred whereas we used Variant Reporter and Sequencher). Furthermore, Wright et al. (40) examined all 10 exons and intron-exon boundaries of the \( LPL \) gene, plus 1 kb of the 5’ flanking (promoter) region, and about 300 bases of the 3’ flanking region in 19 Northern Irish individuals with extreme HTG. They identified 42 variants, of which 39 were successfully detected in our sample. The 3 variants (590G>A, 1018G>A, and 345A>C) that we did not see in our sample had \( \text{MAF} \leq 0.05 \) in their Northern Irish sample. Two of these 3 variants were exonic (590G>A and 1018G>A) and were not identified by Nickerson et al. (27), either. We identified all the common variants (\( \text{MAF} \geq 0.05 \)) present in the European descent populations listed in dbSNP build 137. We also identified 28 novel variants not previously reported in public databases. Although majority (80%) of the novel variants were rare (\( \text{MAF} < 0.01 \)), 3 were common indels (12853_12854ins16, 12861_12864delA4, 12810_12829dup20). Altogether, we identified 88 uncommon/rare variants, and interestingly, 50% of them were located in the intron 1-intron 2 region. The excess of low frequency variants in this region deserves further attention.

A total of 64 selected variants (40 common with \( \text{MAF} \geq 0.05 \) and 24 uncommon/rare with \( \text{MAF} < 0.05 \)) that passed genotyping QC in the total set of 623 individuals were included in the association analyses. Our initial gene-based analysis including all 64 genotyped variants confirmed the established role of \( LPL \) in TG and HDL metabolism as it revealed significant associations only with TG (\( P = 0.006 \)) and HDL-C (\( P = 0.024 \)) levels. Gene-based tests for associations can be more powerful than the single-site tests because they consider all variants within a gene rather than individual markers and also take into account LD between the variants (30).

In single-site analysis, we observed 22 nominally significant associations (\( P < 0.05 \)) with either HDL-C or TG or both. Most of the significant SNPs were correlated with each other and/or with 3 known functionally relevant SNPs: rs295 (Ser447×), rs320 (HindIII), and rs13702 (36, 41, 42). Overall, we observed 7 relatively independent signals (\( P < 0.05; \, i^2 < 0.40 \)) represented by the following SNPs: rs80181352 (in LD with exonic rs248); rs282; rs286 (in LD with functional rs328); rs294 (in LD with exonic rs316); rs295 (in LD with functional rs329 and exonic rs13702); rs8176337; and rs277. Of these 7 independent SNPs, 4 (rs80181352, rs282, rs294, and rs277) were never tested before. While no functional data are available for rs8176337 and rs277 in RegulomeDB, rs282 is located in a transcription factor binding site and thus could have regulatory role in affecting the expression of \( LPL \). In the haplotype analysis, majority of the significant haplotypes, including the most significant ones, contained SNPs that also showed significant association in single-site analysis.

The associations of the Ser447× (rs328) and HindIII (rs320) polymorphisms with CHD risk and plasma lipid profile have been reported extensively in the literature (43–49). The Ser447× (rs328) polymorphism located in exon 9 has been shown to increase \( LPL \) enzymatic activity, and the carriers of 447× have lower TG, higher HDL-C levels and reduced risk of CHD (35, 36, 44, 50). We also observed the same associations in our sample with TG and HDL-C. However, rs286 that is in LD with rs328 showed stronger association with both HDL-C (\( P = 0.005 \)) and TG (\( P = 0.021 \)). The HindIII (rs320, G/T) polymorphism located in intron 8 has also been shown to have same effects on TG and HDL-C levels as does rs328, and there has been a debate whether the genetic effects of these polymorphisms are independent from each other (34, 35, 51–54). Our data indicate that the effects of rs320 and rs286 are independent as they are not in moderate or high LD with each other (\( r^2 = 0.32 \)). Furthermore, rs320 is predicted to affect the binding of a transcription factor and so it may be functional by itself (41). We also found that rs320 is in LD with rs295 (\( r^2 = 0.75 \)) and rs13702 (\( r^2 = 0.88 \)) and that the latter two SNPs showed even stronger associations with both HDL-C and TG. Previously, rs295 has shown a significant association with HDL-C in one GWAS (55) and rs13702 was found to be in strong LD with other reported GWAS-significant SNPs (rs326, rs2083637, and rs10105606). Recently, Richardson et al. (42) have shown that rs13702 is a gain-of-function variant that may affect lipid traits due to its suggested role in allele-specific regulation of \( LPL \) by miR-410. Taken together, it appears that both rs320 and rs13702 may account for the observed GWAS associations.

Due to power-related concerns in analyzing rare variants by the single site-approach, different strategies have been suggested to effectively test the contribution of low frequency and rare variants. One method is the burden
test that collapses all rare variants within a region into a single variable and assumes that cumulative impact of rare variants increases as the number of rare minor alleles increases in the region (32). Alternatively, Wu et al. (56) developed a nonburden sequence kernel association test (SKAT) that is more powerful than burden test when the region has various noncausal variants and/or variants with different directional effects. Since each method has its own particular limitations, Lee et al. (33) have proposed the SKAT-O, an optimal test for rare variant analyses that performs better than burden tests and SKAT under different scenarios. Thus, we performed both burden tests and SKAT-O to analyze the association of 24 low frequency and rare LPL variants (MAF < 0.05) genotyped in our entire sample with lipid traits. However, we found no evidence of association of uncommon/rare LPL variants in our sample, which may be due to our small sample size and/or our analysis approach. On the other hand, we observed an association of a novel rare variant (LP11130) with TG (β = 0.556; P = 0.017) in the single-site analysis, which awaits confirmation in independent larger samples. This variant appears to affect the binding of transcription factors as suggested by the RegulomeDB score of 2b (see supplementary Table II for RegulomeDB scores). Previously, Evans et al. (57) have shown that rare LPL variants are frequently seen in patients with hypertriglyceridemia, indicating that rare LPL variants are more likely associated with severe hyperlipidemia rather than with the lipid variation seen in normal population.

This study has a few limitations. The discovery sample used in the resequencing stage was relatively small and we may have missed the identification of certain rare functional variants. Also, we did not have costly sequencing data from our entire sample of 629 individuals; unknown rare variants in nonsequenced individuals could have affected our rare variant analysis results. The other weakness of the study is the uncorrected P-values, which might have given some false-positive associations for a candidate gene already known to be associated with lipid phenotype. However, based on FDR-adjusted P-values, our two top SNPs (rs295 and rs80181352) remained significant (Table 3).

In summary, to the best of our knowledge, this is the first study that has comprehensively examined the role of LPL genetic variation in relation to HDL-C and TG levels. Our gene-based single-site and haplotype analyses support the well-established role of common LPL genetic variation in affecting plasma TG and HDL-C levels. However, the role of uncommon/rare LPL variants in regulation of plasma TG and HDL-C levels remains less clear and warrants additional studies.

REFERENCES

1. Roger, V. L., A. S. Go, D. M. Lloyd-Jones, E. J. Benjamin, J. D. Berry, W. B. Borden, D. M. Bravata, S. Dai, E. S. Ford, C. S. Fox, et al. 2012. Heart disease and stroke statistics–2012 update: a report from the American Heart Association. Circulation. 125: e2-e220.
2. Kannel, W. B., T. R. Dawber, A. Kagan, N. Revovskie, and J. I. Stokes. 1961. Factors of risk in the development of coronary heart disease–six year follow-up experience. The Framingham Study. Ann. Intern. Med. 55: 33–50.
3. Heller, D. A., U. de Faire, N. L. Pedersen, G. Dahlén, and G. E. McClearn. 1993. Genetic and environmental influences on serum lipid levels in twins. N. Engl. J. Med. 328: 1150–1156.
4. Boerwinkle, E., M. O. Souverein, M. W. Tanck, G. K. Hovingh, J. A. Kuivenhoven, R. I. Peters, H. Jansen, P. M. Schiffer, E. E. Van der Wall, P. A. Doevendans, et al. 2006. Common variants of multiple genes that control reverse cholesterol transport together explain only a minor part of the variation of HDL cholesterol levels. Clin. Genet. 69: 263–270.
5. Kathiresan, S., O. Melander, C. Guiducci, A. Surti, N. P. Burt, J. Riedel, G. M. Cooper, C. Reece, 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.
6. Willer, C. J., S. Sanna, A. U. Jackson, A. Scuteri, L. L. Bonnycastle, R. Clarke, S. C. Heath, N. J. Timpoon, 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.
7. Aulchenko, Y. S., S. Ripatti, I. Lindqvist, B. P. Prins, F. W. Asselbergs, H. Snieder, and J. Fu. 2010. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts. Nat. Genet. 42: 357–35.
8. Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou, M. Kosciel, J. P. Pirschuelco, 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.
9. Prins, B. V., L. Lagou, F. W. Asselbergs, H. Snieder, and J. Fu. 2012. Genetics of coronary artery disease: Genome-wide association studies and beyond. Atherosclerosis. 225: 1–10.
10. Schork, N. J., S. S. Murray, K. A. Frazer, and E. J. Topol. 2009. Common vs. rare allele hypotheses for complex diseases. Curr. Opin. Genet. Dev. 19: 212–219.
11. Cohen, J. C., A. Pertsemidis, S. Fähmi, S. Esmail, G. L. Vega, S. M. Grundy, and H. H. Hobbs. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc. Natl. Acad. Sci. USA. 103: 1810–1815.
12. Komaromy, T., M. Nishikawa, A. D. Porteous, C. L. Morgan, L. M. Kivimäki, and A. J. Lusis. 2000. Additional loci for plasma lipoprotein lipase gene in the CEPH sample. Hum. Mol. Genet. 9: 33–40.
13. W. A. Brown, D. M. Breslow, A. Kagan, N. Revovskie, and J. I. Stokes. 1961. Factors of risk in the development of coronary heart disease–six year follow-up experience. The Framingham Study. Ann. Intern. Med. 55: 33–50.
14. Heller, D. A., U. de Faire, N. L. Pedersen, G. Dahlén, and G. E. McClearn. 1993. Genetic and environmental influences on serum lipid levels in twins. N. Engl. J. Med. 328: 1150–1156.
15. Boerwinkle, E., M. O. Souverein, M. W. Tanck, G. K. Hovingh, J. A. Kuivenhoven, R. I. Peters, H. Jansen, P. M. Schiffer, E. E. Van der Wall, P. A. Doevendans, et al. 2006. Common variants of multiple genes that control reverse cholesterol transport together explain only a minor part of the variation of HDL cholesterol levels. Clin. Genet. 69: 263–270.
16. Willer, C. J., S. Sanna, A. U. Jackson, A. Scuteri, L. L. Bonnycastle, R. Clarke, S. C. Heath, N. J. Timpoon, 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.
17. Aulchenko, Y. S., S. Ripatti, I. Lindqvist, B. P. Prins, F. W. Asselbergs, H. Snieder, and J. Fu. 2012. Genetics of coronary artery disease: Genome-wide association studies and beyond. Atherosclerosis. 225: 1–10.
18. Schork, N. J., S. S. Murray, K. A. Frazer, and E. J. Topol. 2009. Common vs. rare allele hypotheses for complex diseases. Curr. Opin. Genet. Dev. 19: 212–219.
19. Cohen, J. C., A. Pertsemidis, S. Fähmi, S. Esmail, G. L. Vega, S. M. Grundy, and H. H. Hobbs. 2006. Multiple rare variants in NPC1L1 associated with reduced sterol absorption and plasma low-density lipoprotein levels. Proc. Natl. Acad. Sci. USA. 103: 1810–1815.
20. Komaromy, T., M. Nishikawa, A. D. Porteous, C. L. Morgan, L. M. Kivimäki, and A. J. Lusis. 2000. Additional loci for plasma lipoprotein lipase gene in the CEPH sample. Hum. Mol. Genet. 9: 33–40.
21. W. A. Brown, D. M. Breslow, A. Kagan, N. Revovskie, and J. I. Stokes. 1961. Factors of risk in the development of coronary heart disease–six year follow-up experience. The Framingham Study. Ann. Intern. Med. 55: 33–50.
22. Heller, D. A., U. de Faire, N. L. Pedersen, G. Dahlén, and G. E. McClearn. 1993. Genetic and environmental influences on serum lipid levels in twins. N. Engl. J. Med. 328: 1150–1156.
23. Boerwinkle, E., M. O. Souverein, M. W. Tanck, G. K. Hovingh, J. A. Kuivenhoven, R. I. Peters, H. Jansen, P. M. Schiffer, E. E. Van der Wall, P. A. Doevendans, et al. 2006. Common variants of multiple genes that control reverse cholesterol transport together explain only a minor part of the variation of HDL cholesterol levels. Clin. Genet. 69: 263–270.
24. Willer, C. J., S. Sanna, A. U. Jackson, A. Scuteri, L. L. Bonnycastle, R. Clarke, S. C. Heath, N. J. Timpoon, 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.
