12-2017

Identity-by-Descent Mapping Identifies Major Locus for Serum Triglycerides in Amerindians Largely Explained by an APOC3 Founder Mutation

Wen-Chi Hsueh

Anup K. Nair

Sayuko Kobes

Peng Chen

Harald H. H. Goring
The University of Texas Rio Grande Valley

See next page for additional authors

HHS Public Access

Author manuscript

Circ Cardiovasc Genet. Author manuscript; available in PMC 2018 December 01.

Published in final edited form as:

Circ Cardiovasc Genet. 2017 December ; 10(6): . doi:10.1161/CIRCGENETICS.117.001809.

Identity-by-Descent Mapping Identifies Major Locus for Serum Triglycerides in Amerindians Largely Explained by an *APOC3* Founder Mutation

Wen-Chi Hsueh, MPH, PhD[1], **Anup K. Nair, PhD**[1], **Sayuko Kobes, BS**[1], **Peng Chen, PhD**[1], **Harald H. H. Göring, PhD**[2], **Toni I. Pollin, PhD, MS**[3], **Alka Malhotra, PhD**[1,*], **William C. Knowler, MD, DrPH**[1], **Leslie J. Baier, PhD**[1], and **Robert L. Hanson, MD, MPH**[1]

[1]Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Phoenix, AZ

[2]South Texas Diabetes and Obesity Institute University of Texas Rio Grande Valley School of Medicine, San Antonio, TX

[3]Departments of Medicine and Epidemiology & Public Health, University of Maryland School of Medicine, Baltimore, MD

Abstract

Background—Identity-by-descent (IBD) mapping using empirical estimates of IBD allele sharing may be useful for studies of complex traits in founder populations, where hidden relationships may augment the inherent genetic information that can be used for localization.

Methods and Results—Through IBD mapping, using ~400,000 SNPs, of serum lipid profiles we identified a major linkage signal for triglycerides (TG) in 1,007 Pima Indians (LOD=9.23, $p=3.5\times10^{-11}$ on chromosome 11q). In subsequent fine-mapping and replication association studies in ~7,500 Amerindians, we determined that this signal reflects effects of a loss-of-function Ala43Thr substitution in *APOC3* (rs147210663) and 3 established functional SNPs in *APOA5*. The association with rs147210663 was particularly strong; each copy of the Thr allele conferred 42% lower TG ($\beta=-0.92\pm0.059$ SD unit, $p=9.6\times10^{-55}$ in 4,668 Pimas and 2,793 Southwest Amerindians combined). The Thr allele is extremely rare in most global populations, but has a frequency of 2.5% in Pimas. We further demonstrated that 3 *APOA5* SNPs with established functional impact could explain the association with the most well-replicated SNP (rs964184) for TG identified by genome-wide association studies (GWAS). Collectively these 4 SNPs account for 6.9% of variation in TG in Pimas (and 4.1% in Southwest Amerindians), and their inclusion in the original linkage model reduced the linkage signal to virtually null.

Conclusions—*APOC3/APOA5* constitutes a major locus for serum triglycerides in Amerindians, especially the Pimas, and these results provide an empirical example for the concept that population-based linkage analysis is a useful strategy to identify complex trait variants.

Correspondence: Wen-Chi Hsueh, PhD, National Institute of Diabetes & Digestive and Kidney Diseases, Building 1, 1550 E Indian School Rd., Phoenix, AZ 85014, wen-chi.hsueh@nih.gov.

*Current affiliation: Illumina Inc., San Diego, CA

DISCLOSURES: None.

Keywords

triglycerides; linkage; replication; American Indians genetics; population studies; causal variants

Genome-wide association studies (GWAS) of unrelated individuals have become popular approaches for identifying susceptibility genes for complex traits. A well-known drawback is that they require very large samples to detect the modest effect sizes often associated with single variants; furthermore, relevant functional variants must be well-captured by the genotyping platform. Analyses of related individuals can have advantages over those of unrelated individuals because they allow for the efficient detection of high-impact variants with a relatively small sample size and increased power for follow-up association studies.[1] Linkage studies, which analyze phenotypic similarity among related individuals with respect to identity by descent (IBD) in a given region, may efficiently detect regions which contain multiple functional variants and they do not require functional variants to be highly concordant with genotyped markers. However, conventional linkage studies have been limited by only analyzing allele sharing among individuals with known relationships. In recent years, methods have been developed to improve calculation of sharing of alleles IBD from dense genotypic data. These methods allow for the IBD calculation among individuals without known relationships (*i.e.*, they are cryptically related). With this approach, one may considerably improve the efficiency of variance component IBD mapping (also known as population-based linkage analysis). This approach may be particularly appealing for studies in founder populations, as members of such populations tend to share longer chromosomal segments IBD with one another.[2] In this study, we performed a population-based genome-wide linkage study (GWLS) of serum lipid levels in 1,024 individuals from a founder population, the Pima Indians residing in the United States, who had previously participated in a GWAS.[3] We were able to pinpoint specific functional variants explaining a very significant linkage signal for serum triglycerides (TG) through a four-stage study, including a GWLS, fine-mapping analyses, replication association studies, and the fitting of a final linkage-and-association model.

METHODS

Study populations

Much of the data were derived from a longitudinal community-based cohort study of type 2 diabetes (T2D) conducted in Arizona, where most of the participants are Pima Indians (the Pima study).[4] For the initial discovery involving GWLS, 1,024 Pima subjects, who had also participated in a GWAS[3] with available lipid data, including total cholesterol, high-density lipoprotein cholesterol (HDL-C), TG and low-density lipoprotein cholesterol (LDL-C) measurements, were included (1,007 had TG measurements).

The first sample for replication studies included 5,491 (4,668 with TG data) additional Pima subjects who were not part of the initial discovery set and did not have GWAS data. Among these participants, 2,713 are full-heritage Pima Indians (defined as self-reported 8/8th Pima heritage) and another 2,778 subjects are, on average, 6/8th American Indian (typically 4/8th Pima and an additional 2/8th from other tribes). A subset of the Pima participants (n=296)

had whole genome sequence (WGS) data available, which were used to profile the genomic variations in this population.

A second set of samples for the replication studies came from the Phoenix extension of FIND (Family Investigation of Nephropathy and Diabetes), a multicenter study designed to identify genes involved in diabetic nephropathy and related traits. Eligible subjects (n=3,189; 2,793 with TG data) had ≥50% Amerindian heritage and most were urban-dwelling Amerindians living in or near Phoenix. Table S1 (in the Data Supplement) shows characteristics of these 3 groups. Studies were approved by the Institutional Review Board of the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), and all participants provided written informed consent. Phenotypic measurements were based on standard protocols and are described in the Data Supplement.

Genotypic data ascertainment

Genotypes used in the linkage analysis (to calculate IBD sharing in the autosomal genome and to assess local IBD) were produced with the Affymetrix 6.0 Human SNP Array (see Methods in the Data Supplement). A total of 398,430 autosomal SNPs passed quality control checks and were used in subsequent analyses. As part of our fine-mapping studies, we used WGS data of 296 Pima Indians for variant discovery and imputation. The WGS data were generated by Illumina (San Diego, CA) at a coverage of 30–40X.

Genotyping of single nucleotide polymorphisms (SNPs) in both replication sets used a variety of genotyping approaches, which are described in detail in the Data Supplement. In addition, 45 selected ancestry informative markers with large differences in allele frequency between Amerindian and European populations[5] were genotyped and used to control for population admixture in replication samples, in which GWAS data were not available.

Statistical analyses

Estimation of the percentage of alleles shared IBD—The execution of variance components linkage analysis of quantitative traits requires information on the alleles shared IBD between 2 individuals. We used the program Beagle[6] and genetic maps from the Hapmap project to carry out IBD estimation at each of ~400,000 SNPs; this program estimates IBD from haplotypic similarity based on a hidden Markov model that takes recombination and linkage disequilibrium (LD) among SNPs into account. More details regarding the parameter settings and calculation are provided in the Data Supplement.

GWLS of 4 lipid traits: All traits were normalized by inverse Gaussian transformation prior to analysis. Linkage analysis was conducted using the principles of the variance-components method developed by Amos[7] and carried out using SAS (SAS Institute, Cary, NC; see Methods in the Data Supplement). For computational efficiency, we carried out linkage analysis at every 50th variant across the genome, after which the maximum LOD (max LOD) was determined. In other words, ~8,000 tests (at ~0.44 cM intervals) were carried out.

Since we used the IBD matrix generated from the Beagle estimates among all 523,776 pairs of individuals in the sample, instead of those from known relationships, the present study is termed a population-based linkage analysis.

Fine-mapping study – association analysis conditional on linkage effects—To identify whether association with specific genetic variant(s) may explain the observed linkage signals, we carried out association analyses conditional on linkage effects (i.e. accounting for random effects of both local and global IBD) to evaluate the association of variants under the linkage peak in the 1,024 subjects used in linkage analyses. The region of interest was defined as the "2-LOD" support interval (i.e., the interval in which the LOD is within 2 units of the max LOD). Sources of genotypic data used in these analyses are described in the Data Supplement. To account for the relatedness among family members in the analyses, we used the measured genotype approach,[8] in which the genotypic effects are incorporated as fixed effects in the mixed model. In an effort to identify genetic variants with distinct effects, variant(s) with stronger association(s) were included as covariate(s) in the next round of analysis, until no significant association was observed (i.e. p value <0.05 corrected for multiple comparisons, see Methods in the Data Supplement).

Replication association analyses of TG—Replication studies were carried out in 2 stages. First, we conducted the replication study in Pima Indians (the Pima sample, n=4,668 after exclusion of those in the GWLS). If any associations were replicated, we conducted a second replication study using 2,793 Amerindian samples from FIND). Detailed statistical approaches and meta-analysis methods are described in the Data Supplement. Because variants examined were from a small region of the genome with a high density of variants where extensive LD was present, we corrected for multiple comparisons using the approach suggested by Moskvina and Schmidt[9] (more details in the Data Supplement). All presented p values for the fine-mapping and replication studies are corrected p values.

Haplotype construction and analyses—Once we identified multiple variants within a small gene with strong association with TG, we carried out haplotype analysis to assess associations conditional on specific allelic backgrounds constituting ≥2 SNPs (see the Data Supplement for details).

Covariates (population admixture estimates, type 2 diabetes status and sex) used in all models are described in the Data Supplement. Population admixture estimates used in the linkage studies were obtained using principal components analysis, and those used in the replication studies were calculated as Amerindian heritage based on 45 ancestry informative markers. Details are described in the Data Supplement.

RESULTS

Empirical IBD estimates

Among the 1,024 subjects in the GWLS, the average estimated IBD across the genome ranged from 0.00001 to 0.57 for 523,776 pairs of subjects. Pairs with no known relationship (98.2% of all pairs) had a mean IBD of 0.02±0.01 (median=0.02). In contrast, the mean IBD for 171 pairs of individuals of white ethnicity genotyped with the same SNP array was 0.0009±0.0007 (median=0.0008), suggesting that average relatedness among Pimas was much greater. A plot showing the estimated empirical IBD sharing by the expected relatedness is shown in the Data Supplement (Figure S1).

Genome-wide linkage studies

Among 4 lipid traits analyzed, significant linkage (max LOD≥3) was identified for TG and HDL-C. The max LOD for TG was 9.23 ($p=3.5\times10^{-11}$) on chromosome 11q23, explaining 10.6% of phenotypic variance (after accounting for effects of covariates), and the max LOD for HDL-C was 3.77 ($p=1.5\times10^{-5}$) on chromosome 1p, explaining 7.5% of the variance (Table S2 and Figure S2 in the Data Supplement). When the same analysis was conducted using only pairs with known self-reported relationships (n=9,664), the max LOD for TG was still observed in the same genomic region, however, with a much reduced significance (LOD=3.24, $p=5.6\times10^{-5}$) (Figure 1A). A similar reduction in LOD was observed for HDL-C, and the max LOD was 2.55 (Figure 1B). We further refined the location of the max LOD by running additional linkage analyses at 5-SNP intervals within the 2-LOD region. This resulted in max LOD of 9.32 (explaining 10.8% of the variance) and 4.05 (explaining 7.9% of the variance), for TG and HDL-C respectively. These refined max LOD locations were used to define genomic regions investigated in subsequent fine-mapping studies.

We also conducted a GWAS of these traits in the same sample, and observed associations for TG with SNPs in 11q23 at genome-wide significance (Figure S3 in the Data Supplement). The strongest association was with rs4417316, an intronic SNP in *ZPR1* ($p=1.8\times10^{-10}$), but rs964184, the most commonly reported GWAS SNP for TG observed in multiple populations, also was strongly associated ($p=4.0\times10^{-8}$). The associations of all SNPs in this region were attenuated to $p>10^{-5}$ after adjusting for the effects of rs964184. Thus, it appeared that the GWAS associations in Pimas largely reflected the effect of the established SNP rs964184. In contrast, linkage analyses of TG conditional on the effect of rs964184 or rs4417316 showed significant residual effect (LOD=6.02 and 4.80, respectively), suggesting the possibility of multiple variants in the region.

Fine-mapping study: association analyses of TG and HDL conditional on linkage effect

We performed imputation, using WGS data from 296 Pimas as the reference, and we conducted fine-mapping of the 2-LOD support interval for both regions by testing association conditional on the peak linkage signal in the same 1,024 samples used in GWLS. We selected the variant with the strongest association and analyzed associations of additional variants conditional on this strongest association; this procedure was repeated until no statistically significant associations were observed. For TG, after 4 rounds of analyses of 3,450 variants in a 1.81Mb region on 11q23, we identified 3 variants with distinct and significant associations (Table 1). The SNP with the strongest association with TG was rs147210663 explaining 6.9% of the variance ($p=1.6\times10^{-13}$) (Table 1). This SNP is in the apolipoprotein C3 gene (*APOC3*) at codon 43 (Ala → Thr substitution, or A43T); the minor allele codes for the Thr residue and has a frequency of 2.6% in GWLS. The SNP with the second strongest association with TG was rs2072560 after adjusting for the effect of rs147210663 (p=0.00028). This SNP resides in intron 3 of the apolipoprotein A5 gene (*APOA5*) but has no known function. The third variant with a significant association was rs11357208 after adjusting for the effects of 2 previous SNPs (p=0.0049). This insertion-deletion variant residesin intron 5 of the SIK family kinase 3 gene (*SIK3*) without any known function. For HDL-C, no significant association was identified with any variant in the 2-LOD support interval on chromosome 1p.

Replication association analyses of TG

To confirm and further characterize the associations identified in the fine-mapping study, we analyzed the 3 distinctly associated TG variants, along with their tags (r^2>0.8 in Pima WGS data) and nearby established TG-associated SNPs selected from published data. These studies were conducted in 2 different samples, namely 1) additional Pimas from the same parent study who were not part of GWLS (n=4,668, the Pima sample) and 2) southwestern urban Amerindians from the FIND sample (n=2,793). We, thus, analyzed 11 SNPs, including 6 established TG-associated SNPs (collectively called "GWAS SNPs", including rs964184, rs3135506, rs651821, rs662799, rs12225230, and rs139961185).[10–19] As shown in Table 1, there was no tag SNP for rs147210663 nor rs11357208. The SNP rs2072560 had 3 tag SNPs (one of them, rs651821, was also identified in several published GWAS reports). As in the fine-mapping study, several rounds of analysis were conducted, with association examined conditional on the strongest variants identified in previous rounds. In the Pima sample, the association with the *APOC3* A43T SNP strongly replicated (rs147210663, $p=7.4\times10^{-48}$), as did the 2nd variant (represented by rs651821, p=0.0012) (Table 2 and Table S3 in the Data Supplement). In addition, we also observed a strong association with a GWAS SNP (rs964184, $p=2.1\times10^{-22}$) distinct from rs147210663. The variant rs11357208 did not show an association distinct from rs147210663, rs964184 and rs651821. However, as rs964184 was in moderate LD with both rs651821 and rs11357208 (r^2=0.20 and 0.45, respectively), this may partially explain the lack of replicated association with rs11357208. In the second replication study (the FIND sample), we also observed distinct associations with rs964184 (1.4×10^{-10}) and rs147210663 ($p=6.3\times10^{-7}$). Of note, the frequency of the Thr allele of rs147210663 in the FIND sample was much lower than that in the Pima sample (1% vs. 2.5%). Taken together, we replicated 2 distinct associations (with rs147210663 and rs964184).

Determining whether known functional variants account for the observed association with rs964184

The "GWAS" SNP rs964184 resides near the 3'UTR of the zinc finger gene (*ZPR1*, also known as *ZNF259*), but evidence for a functional effect of this SNP is lacking. Two "haploblocks" (see methods) encompassing 3 SNPs (rs964184 and rs651821 identified in the replication studies, and rs2072560 identified from fine-mapping analyses) harbor 2 genes: *ZPR1* and *APOA5*. *APOA5* is expressed solely in liver tissues, the key organ for lipid metabolism, whereas *ZPR1* is expressed ubiquitously. *APOA5* has a known role in TG metabolism, and has a SNP (rs651821) significantly associated with TG. Therefore, we extended our association study to investigate if other functional variants in these 2 genes may explain the observed association with rs964184. Based on Pima WGS data, 4 SNPs in *APOA5* were previously documented to be functional[20–23] with a minor allele frequency (MAF) ≥1%: rs2266788, rs3135506, rs651821 and rs662799. We tested associations between TG and these SNPs to examine the extent to which they explained the association of rs964184 with TG. As shown in Figure 2, in the Pima sample, rs964184 (dark bars) accounted for 2.2% of the trait variance with $p=5.7\times10^{-23}$ when no *APOA5* SNP was included in the model (conditioning on the effect of rs147210663 and other covariates). This effect was gradually diminished with the addition of more *APOA5* SNPs. When 3 *APOA5* SNPs (rs651821, rs3135506, rs2266788) were included, they accounted for 2.4% of the

variance ($p=1.2\times10^{-22}$) collectively, whereas the effect of rs964184 was reduced to 0.07% (p=0.015). The addition of rs662799 ($r^2=0.96$ with rs2266788), did not contribute significantly beyond other *APOA5* SNPs. We also conducted similar analyses of all SNPs in the exons or UTRs of *ZPR1* (rs61905116, rs144966144, and rs35120633, identified in our Pima WGS data). The effects of rs964184 on TG remained highly significant after adjusting for all 3 SNPs. In addition, although our study populations had a high T2D prevalence, the genetic associations with these SNPs in *APOC3* and *APOA5* were observed in those with or without diabetes (data not shown). The exclusion of subjects taking antilipidemic medications (~5%) from analyses did not affect any results significantly (data not shown).

Haplotype analyses of APOA5 SNPs

Assessment of the distinct contribution of 3 *APOA5* SNPs is difficult because they are in LD ($r^2=0.04–0.75$, Figure S4 in the Data Supplement) and their statistical associations may be codependent when their effects are assessed individually. Therefore, we carried out analysis of haplotypes composed of these 3 SNPs to better evaluate allelic effects of each of these SNPs relative to the allelic background of other 2. Previous studies have described three common haplotypes at these SNPs: *APOA5*1* containing the TG-lowering allele at all 3 loci (T-Ser-A for rs2266788-rs3135506-rs651821), *APOA5*2* containing the TG-raising allele at rs2266788 and rs651821 (C-Ser-G), and *APOA5*3* containing the TG-raising allele only at rs3135506 (T-Trp-A). In East Asian populations, an additional haplotype (*APOA5**4) containing only one TG-raising allele (at rs651821, T-Ser-G) has been observed.[24] In the Pimas we found that *1, *2 and *3 were common haplotypes and we observed an additional haplotype with frequency of 3.6% which contains the TG-raising allele at rs2266788 only (C-Ser-A, which we call *APOA5*5*). Given the same allelic background at the other 2 SNPs, distinct allelic effects of all 3 *APOA5* SNPs were statistically significant (Table 3). For instance, compared with the allelic background of rs2266788 and rs651821 (reference haplotype: *APOA5*1*), the *APOA5*3* haplotype, which differs only at the minor allele of rs3135506, was still significantly associated with higher TG ($p=1.3\times10^{-12}$). Furthermore, compared with *APOA5*1, APOA5*2*, which differs at rs651821 and rs2266788 was associated with significantly higher TG ($p=2.2\times10^{-23}$). The haplotypic associations were replicated in the FIND sample with the exception of the comparison of *APOA5*2* with *APOA5*5*, whose borderline significance was likely due to the low frequency of *APOA5*5* (0.5%).

Final linkage-and-association model fitting

From replication studies, we identified the *APOC3* A43T SNP (rs147210663) and 3 *APOA5* SNPs (rs2266788, rs3135506 and rs651821), which are well-established as functional, as having significant influences on TG. Their effects all appeared additive (Figure S5 in the Data Supplement). Therefore, we included these 4 SNPs as covariates in the original linkage model (n=1,007 Pimas) and assessed the extent to which these 4 SNPs explain the linkage signal. As shown in Table 4, the *APOC3* 43T SNP alone reduces the LOD from 9.32 to 2.24. When we added 3 *APOA5* SNPs to the model, the collective effect of these 4 SNPs reduced the remaining linkage signal to LOD of 0.08 (residual variance explained=1.4% after accounting for the effects of 4 SNPs and all covariates). In other words, these 4 SNPs virtually explained the linkage signal on 11q23 for TG.

DISCUSSION

Although GWAS in unrelated individuals has been successful in identifying common variants associated with complex traits, family-based studies, such as IBD mapping (linkage) analysis, may have advantages in some situations, such as when the functional variants are not well-captured in the standard GWAS array. In this study, we used an IBD mapping approach in a founder population to identify genomic region(s) harboring substantial genetic effects on lipid levels. Estimating IBD empirically using dense genotypic data, we performed linkage analysis using all possible pairs within our study sample.

Using this approach, we identified a significant linkage signal with LOD=9.23 for TG in our GWLS. Through subsequent fine-mapping and 2 sets of replication studies, we identified 2 SNPs (rs147210663 and rs964184) with significant and distinct associations with TG. One replicated SNP (rs964184) has no known function and is located near the 3'UTR of *ZPR1*. Little evidence has been found to indicate ZPR1's influence on lipid metabolism despite rs964184 being the most widely replicated GWAS SNP for TG. Thus, we focused on variants in the only other candidate, *APOA5*. *APOA5* variants have been extensively studied functionally due to the protein's importance in TG metabolism. From Pima WGS data, we identified 4 SNPs in *APOA5* with strong functional evidence from literature. We showed that the effect of rs964184 could be mostly explained by 3 of these *APOA5* SNPs (the 4th SNP is in almost perfect LD with one of these 3). In other words, we consider it highly likely that rs964184 acts as a marker for the aggregate effect of these 3 SNPs with known effects on APOA5. Collectively, the *APOC3* A43T SNP and the 3 strongly-associated functional SNPs in *APOA5* explained most of the linkage signal.

The *APOC3* A43T SNP had the strongest association with TG, and had an effect size (β= −0.92±0.64 SD unit for the minor allele Thr, which corresponds to a 42% reduction in TG) among the largest reported for any complex trait.[25] Its effect accounted for 4.7% of the variance of TG despite a fairly low frequency of the Thr allele (2.5%) in Pimas. It was also significantly associated with other lipid traits and lipid fractions, including with higher HDL-C and lower TG contained in very low density lipoprotein particles (VLDL-TG) (Table S3 in the Data Supplement). The Thr allele is rare in most other populations, with an MAF of ~1% in the other Southwest Amerindians, and very rare in most non-Amerindians with a collective frequency of 0.3% in phase 3 data from the 1000 Genomes Project (www.1000genomes.org). The A43T SNP was initially identified in two Mayan Indians with low TG and APOC3 levels (where it was denoted as A23T).[26] Strong associations between TG and rare non-synonymous SNPs (nsSNPs) in *APOC3* have been observed in other populations as well, with loss-of-function variants conferring lower TG levels. Such an association was first suggested by a fine-mapping study following a GWAS in the Amish,[27] in whom a premature stop codon (R19X, rs76353203) which is rare in most populations, had frequency of 2.8% and was strongly associated with low TG levels. This same SNP was also identified as strongly associated with low TG levels in a Greek population isolate, in whom its frequency was 1.9%.[28] Subsequently, several large sequencing or candidate gene studies have also provided evidence for strong protective effects of loss-of-function mutations in *APOC3* on TG and cardiovascular diseases;[28–33] these variants included the A43T SNP, but were tested for association in aggregate. The present study shows that the A43T SNP is

relatively common in some Amerindian populations (particularly Pimas), among whom the association achieves genome-wide statistical significance.

APOC3 is a small gene with 4 exons and 297 nucleotides. The protein is synthesized mainly in the liver, and strongly inhibits hepatic uptake of VLDL-C and intermediate density lipoproteins. Intracellularly, it promotes hepatic VLDL-C assembly and secretion.[34] Extracellularly, it attenuates hydrolysis and clearance of triglyceride-rich lipoproteins (TRLs) and impairs the lipolysis of TRLs by inhibiting lipoprotein lipase and the hepatic uptake of TRLs by remnant receptors.[34] Recently, inhibitors of APOC3 synthesis have been shown to reduce hypertriglyceridemia.[35] The sequence of *APOC3* has been extensively determined in large and diverse samples. In the Pimas, there is only 1 nsSNP (A43T) identified in 296 samples. Findings from *in vitro* studies show that this A→T substitution probably alters the structure of APOC3, and leads to loss of its function in promoting the assembly and secretion of triglyceride-rich VLDL from hepatic cells.[36] Also it has been shown that, it has less efficient lipid binding capacity; this leads to faster catabolism of APOC3 and less competition with APOE. This in turn is responsible for enhanced clearance of TG rich lipoproteins and lower plasma TG levels.[26] Given all available statistical, functional, and physiological evidence, and the observation that the A43T SNP has no tag SNP, we consider that this SNP is likely a causal variant for TG.

In our replication studies, the SNP with the second strongest association with TG was rs964184. It is one of the most well-replicated SNPs from many large lipid GWAS, with particularly strong associations for TG.[10, 12–15, 19, 25, 37, 38] Previous studies in Amerindian-derived populations have identified strong associations between TG and variants in this chromosomal region, particularly rs964184.[14, 19, 39] *SIK3* has been suggested as a functional candidate based on evidence for recent natural selection centered on rs139961185.[19] In our study, however, rs139961185 was not associated with TG after conditioning on rs964184 (Table S3 in the Data Supplement). *BUD13* and *ZNF259* have also been implicated, based largely on the strong association with the nearby rs964184.[14, 39] The SNP rs964184 is also near *APOA5*, however, and 4 common *APOA5* SNPs (rs2266788, rs3135506, rs651821, rs6622799) have well-documented functional effects on APOA5. The only nsSNP of these 4 is rs3135506 (S19W) which impairs protein translocation and secretion,[20, 21] whereas rs651821 is located in the promoter region, and rs2266788 is located in the 3' UTR of *APOA5*, a functional target site for a liver-expressed microRNA gene *miR-485-5p*. Our findings provided statistical evidence to implicate 3 functional SNPs (rs651821, rs3135506, and rs2266788) in *APOA5* that explained almost all of the association of rs964184 with TG, contributed distinct effects, and that reduced the original linkage signal to virtually null. It should be noted that our findings are in concordance with functional studies that suggest distinct effects of each of these SNPs (Table 3, Figure 2). These 3 *APOA5* SNPs all have elevated MAFs in Pimas (0.14–0.20, vs. 0.06–0.08 in whites), which resulted in increased statistical power in our study populations. One caveat is that as the 4th putatively functional SNP, rs6622799, is in almost perfect LD with rs2266788 (r^2=0.96), their effects are not statistically distinguishable in our study populations.

Many studies have reported associations of haplotypes composed of these *APOA5* functional variants with TG. In 2001, Pennacchio et al. reported that the *APOA5*2*

haplotype, defined by the minor alleles of 3 SNPs (encompassing multiple functional variants rs662799–rs651821–rs2266788), was associated with elevated TG levels,[40] and similar observations were replicated in multiple populations.[24, 41–43] In some GWAS, these SNPs were individually reported as significant.[11, 17, 44] Our study identified these same SNPs in a systematic approach, and provides additional support for multivariant influences from *APOA5* on TG. *In vitro* studies suggest that all 3 SNPs have functional effects,[22] but separate effects are difficult to observe *in vivo* because of the high degree of LD. With the additional haplotypic information available in the present study, we provide evidence for distinct effects of rs651821, rs3135506 and rs2266788/rs662799 on TG *in vivo*, and associations with the haplotypes *APOA5*2* and *APOA5*3* (representing rs3135506) achieve genome-wide statistical significance. Although these 3 *APOA5* SNPs were only in moderate LD with rs964184 individually (r^2=0.25–0.40, Figure S4 in the Data Supplement), the haplotypes comprised of any minor (TG-raising) allele of *APOA5* SNPs were in strong LD with the minor allele of rs964184 collectively (r^2=0.97) (Figure S6 in the Data Supplement), and, thus, the *APOA5* SNPs can largely account for the association of rs964184 with TG in Amerindians. Based on analysis of data from the 1000 Genomes project (www.1000genomes.org), it is noteworthy that rs964184 is similarly highly concordant with haplotypes containing a TG-raising allele at these *APOA5* SNPs in populations representative of those where rs964184 was identified as a top GWAS SNP (e.g. r^2=0.96 in CEU and GIH, r^2=0.82 in CHB and MXL); thus these *APOA5* SNPs might explain the association between rs964184 and TG in other populations as well. The role of APOA5 in influencing TG levels is supported by recent human sequencing studies showing rare deleterious *APOA5* mutations associated with TG.[45] Given statistical, functional and physiological evidence, we conclude that rs964184 is likely a marker for the collective effects of *APOA5* functional SNPs on TG.

The linkage analysis results provide a context for interrogating whether multiple sources of genetic effect exist in the same genomic region. In our GWAS using the same 1,007 subjects, we observed associations for TG with SNPs on 11q23 at genome-wide statistical significance, but these associations were greatly attenuated after adjustment for the established variant, rs964184. On the other hand, substantial residual linkage remained after adjustment for the effects of rs964184, and this served as the impetus for more detailed investigation of the region. Although computational tools for the IBD calculations required for population-based linkage analyses are available in several software packages, the method has not been widely-used. Our results suggest it can provide complementary information to a standard GWAS approach.

An advantage of the population-based linkage approach is that, in contrast to association analysis, it retains power even when the variants on the genotyping array are not highly concordant with a functional variant. The approach may be especially useful for identifying regions with susceptibility variants not well-captured by the genotyping platform (which will often be of low frequency) or identifying regions containing multiple susceptibility variants. For the 11q23 region identified here as linked with TG, both situations apply. The method uses information from all pairs of individuals in the population, and this remarkably enhances the statistical power, compared with conventional pedigree-based linkage analysis, particularly in populations recently descended from a small number of founders, in which

many individuals without known relationships may share large genomic segments IBD. The approach may be less useful in outbred populations, however. Power of the linkage approach is limited, compared with conventional association analysis, in the situation where a single functional variant, which is well-captured by the genotyping array, drives the association. A further limitation of the present study is that, with the limited sample size of GWLS, only relatively strong effect sizes are detectable; we estimate the detectable effect size is 9.7% of the variance for LOD>3 with 80% power (our linkage signal for TG had an effect size of 10.8% of the variance).

In conclusion, we carried out a population-based GWLS of serum lipids and subsequently identified 4 SNPs with known functional effects in 2 apolipoprotein genes (*APOC3* and *APOA5*) that influence TG levels. These findings suggest that population-based GWLS may provide complementary information to GWAS, particularly in founder populations. Identification of the TG-lowering nsSNP in *APOC3* (A43T) was facilitated by population-specific WGS data, which allowed for accurate imputation of this variant despite its low frequency, and the fact that it was not on the GWAS array (this SNP could not be captured by imputation using the 1000Genome data as the reference panel). The *APOC3* A43T SNP has been established as a loss-of-function variant leading to lower TG based on a small number of previously described individuals and functional studies. By uncovering a founder effect in the Pimas for this SNP, we now provide population level data that unequivocally establish the TG-lowering properties of this SNP. Thus, the linkage signal we detected represents effects of both an established TG variant (rs964184) and a “novel” variant (A43T), not captured by standard GWAS arrays. We also provide evidence for 3 functional *APOA5* SNPs exerting distinct and additive effects on TG. The association between rs964184 and TG has been replicated in multiple populations; our study demonstrates that rs964184 is likely a marker tagging aggregate effects of 3 functional SNPs in *APOA5*, at least in Amerindians. Thus, a single well-replicated GWAS signal can reflect the effects of multiple functional variants. Taken together, the *APOC3* and *APOA5* SNPs account for 6.9% of the variation in TG, an effect which constitutes a major locus in Pimas. Our findings provide an empirical example for the concept that population-based linkage analysis, particularly in founder populations, can be useful for studies of complex traits.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the NIH, and the Moab Linux cluster at the Frederick National Laboratory for Cancer Research.

SOURCES OF FUNDING: This study was supported by the Intramural Research Program of the NIDDK, NIH.

References

1. Rosenthal E, Blue E, Jarvik GP. Next-generation gene discovery for variants of large impact on lipid traits. Curr Opin Lipidol. 2015; 26:114–9. [PubMed: 25636063]

2. Gusev A, Kenny EE, Lowe JK, Salit J, Saxena R, Kathiresan S, et al. DASH: a method for identical-by-descent haplotype mapping uncovers association with recent variation. Am J Hum Genet. 2011; 88:706–17. [PubMed: 21620352]

3. Hanson RL, Muller YL, Kobes S, Guo T, Bian L, Ossowski V, et al. A genome-wide association study in American Indians implicates DNER as a susceptibility locus for type 2 diabetes. Diabetes. 2014; 63:369–76. [PubMed: 24101674]

4. Knowler WC, Pettitt DJ, Saad MF, Charles MA, Nelson RG, Howard BV, et al. Obesity in the Pima Indians: its magnitude and relationship with diabetes. Am J Clin Nutr. 1991; 53:1543S–1551S. [PubMed: 2031485]

5. Tian C, Hinds DA, Shigeta R, Adler SG, Lee A, Pahl MV, et al. A genomewide single-nucleotide-polymorphism panel for Mexican American admixture mapping. Am J Hum Genet. 2007; 80:1014–23. [PubMed: 17557415]

6. Browning SR, Browning BL. High-resolution detection of identity by descent in unrelated individuals. Am J Hum Genet. 2010; 86:526–39. [PubMed: 20303063]

7. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. Am J Hum Genet. 1994; 54:535–543. [PubMed: 8116623]

8. Boerwinkle E, Chakraborty R, Sing CF. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. Ann Hum Genet. 1986; 50(Pt 2):181–194. [PubMed: 3435047]

9. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. Genet Epidemiol. 2008; 32:567–73. [PubMed: 18425821]

10. Kristiansson K, Perola M, Tikkanen E, Kettunen J, Surakka I, Havulinna AS, et al. Genome-wide screen for metabolic syndrome susceptibility Loci reveals strong lipid gene contribution but no evidence for common genetic basis for clustering of metabolic syndrome traits. Circ Cardiovasc Genet. 2012; 5:242–9. [PubMed: 22399527]

11. Wu Y, Waite LL, Jackson AU, Sheu WH, Buyske S, Absher D, et al. Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. 2013; 9:e1003379. [PubMed: 23555291]

12. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466:707–13. [PubMed: 20686565]

13. Braun TR, Been LF, Singhal A, Worsham J, Ralhan S, Wander GS, et al. A replication study of GWAS-derived lipid genes in Asian Indians: the chromosomal region 11q23.3 harbors loci contributing to triglycerides. PLoS One. 2012; 7:e37056. [PubMed: 22623978]

14. Weissglas-Volkov D, Aguilar-Salinas CA, Nikkola E, Deere KA, Cruz-Bautista I, Arellano-Campos O, et al. Genomic study in Mexicans identifies a new locus for triglycerides and refines European lipid loci. J Med Genet. 2013; 50:298–308. [PubMed: 23505323]

15. Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA, Ambite JL, et al. Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. PLoS Genet. 2011; 7:e1002138. [PubMed: 21738485]

16. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al. Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. Am J Hum Genet. 2009; 85:628–42. [PubMed: 19913121]

17. Musunuru K, Lettre G, Young T, Farlow DN, Pirruccello JP, Ejebe KG, et al. Candidate gene association resource (CARe): design, methods, and proof of concept. Circ Cardiovasc Genet. 2010; 3:267–75. [PubMed: 20400780]

18. Rosenthal EA, Ranchalis J, Crosslin DR, Burt A, Brunzell JD, Motulsky AG, et al. Joint linkage and association analysis with exome sequence data implicates SLC25A40 in hypertriglyceridemia. Am J Hum Genet. 2013; 93:1035–45. [PubMed: 24268658]

19. Ko A, Cantor RM, Weissglas-Volkov D, Nikkola E, Reddy PM, Sinsheimer JS, et al. Amerindian-specific regions under positive selection harbour new lipid variants in Latinos. Nat Commun. 2014; 5:3983. [PubMed: 24886709]

20. Talmud PJ, Palmen J, Putt W, Lins L, Humphries SE. Determination of the functionality of common APOA5 polymorphisms. J Biol Chem. 2005; 280:28215–20. [PubMed: 15941721]

21. Ahituv N, Akiyama J, Chapman-Helleboid A, Fruchart J, Pennacchio LA. In vivo characterization of human APOA5 haplotypes. Genomics. 2007; 90:674–9. [PubMed: 17936576]
22. Palmen J, Smith AJ, Dorfmeister B, Putt W, Humphries SE, Talmud PJ. The functional interaction on in vitro gene expression of APOA5 SNPs, defining haplotype APOA52, and their paradoxical association with plasma triglyceride but not plasma apoAV levels. Biochim Biophys Acta. 2008; 1782:447–52. [PubMed: 18395529]
23. Caussy C, Charriere S, Marcais C, Di Filippo M, Sassolas A, Delay M, et al. An APOA5 3' UTR variant associated with plasma triglycerides triggers APOA5 downregulation by creating a functional miR-485-5p binding site. Am J Hum Genet. 2014; 94:129–34. [PubMed: 24387992]
24. Li S, Hu B, Wang Y, Wu D, Jin L, Wang X. Influences of APOA5 variants on plasma triglyceride levels in Uyghur population. PLoS One. 2014; 9:e110258. [PubMed: 25313938]
25. Consortium UK, Walter K, Min JL, Huang J, Crooks L, Memari Y, et al. The UK10K project identifies rare variants in health and disease. Nature. 2015; 526:82–90. [PubMed: 26367797]
26. Liu H, Labeur C, Xu CF, Ferrell R, Lins L, Brasseur R, et al. Characterization of the lipid-binding properties and lipoprotein lipase inhibition of a novel apolipoprotein C-III variant Ala23Thr. Journal of lipid research. 2000; 41:1760–71. [PubMed: 11060345]
27. Pollin TI, Damcott CM, Shen H, Ott SH, Shelton J, Horenstein RB, et al. A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. Science. 2008; 322:1702–5. [PubMed: 19074352]
28. Tachmazidou I, Dedoussis G, Southam L, Farmaki AE, Ritchie GR, Xifara DK, et al. A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. Nat Commun. 2013; 4:2872. [PubMed: 24343240]
29. Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin SY, et al. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. Nat Commun. 2014; 5:4871. [PubMed: 25225788]
30. TG, HDL Working Group of the Exome Sequencing Project NHL. Blood I, Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. N Engl J Med. 2014; 371:22–31. [PubMed: 24941081]
31. Natarajan P, Kohli P, Baber U, Nguyen KD, Sartori S, Reilly DF, et al. Association of APOC3 Loss-of-Function Mutations With Plasma Lipids and Subclinical Atherosclerosis: The Multi-Ethnic BioImage Study. J Am Coll Cardiol. 2015; 66:2053–5. [PubMed: 26516010]
32. Li AH, Morrison AC, Kovar C, Cupples LA, Brody JA, Polfus LM, et al. Analysis of loss-of-function variants and 20 risk factor phenotypes in 8,554 individuals identifies loci influencing chronic disease. Nat Genet. 2015; 47:640–2. [PubMed: 25915599]
33. Helgadottir A, Gretarsdottir S, Thorleifsson G, Hjartarson E, Sigurdsson A, Magnusdottir A, et al. Variants with large effects on blood lipids and the role of cholesterol and triglycerides in coronary disease. Nat Genet. 2016; 48:634–9. [PubMed: 27135400]
34. Mendivil CO, Zheng C, Furtado J, Lel J, Sacks FM. Metabolism of very-low-density lipoprotein and low-density lipoprotein containing apolipoprotein C-III and not other small apolipoproteins. Arteriosclerosis, thrombosis, and vascular biology. 2010; 30:239–45.
35. Gaudet D, Alexander VJ, Baker BF, Brisson D, Tremblay K, Singleton W, et al. Antisense Inhibition of Apolipoprotein C-III in Patients with Hypertriglyceridemia. N Engl J Med. 2015; 373:438–47. [PubMed: 26222559]
36. Sundaram M, Zhong S, Bou Khalil M, Zhou H, Jiang ZG, Zhao Y, et al. Functional analysis of the missense APOC3 mutation Ala23Thr associated with human hypotriglyceridemia. Journal of lipid research. 2010; 51:1524–34. [PubMed: 20097930]
37. Global Lipids Genetics C, Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al. Discovery and refinement of loci associated with lipid levels. Nat Genet. 2013; 45:1274–83. [PubMed: 24097068]
38. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nat Genet. 2008; 40:189–97. [PubMed: 18193044]
39. Parra EJ, Mazurek A, Gignoux CR, Sockell A, Agostino M, Morris AP, et al. Admixture mapping in two Mexican samples identifies significant associations of locus ancestry with triglyceride

levels in the BUD13/ZNF259/APOA5 region and fine mapping points to rs964184 as the main driver of the association signal. PLoS One. 2017; 12:e0172880. [PubMed: 28245265]

40. Pennacchio LA, Olivier M, Hubacek JA, Cohen JC, Cox DR, Fruchart JC, et al. An apolipoprotein influencing triglycerides in humans and mice revealed by comparative sequencing. Science. 2001; 294:169–73. [PubMed: 11588264]
41. Gombojav B, Lee SJ, Kho M, Song YM, Lee K, Sung J. Multiple susceptibility loci at chromosome 11q23.3 are associated with plasma triglyceride in East Asians. Journal of lipid research. 2016; 57:318–24. [PubMed: 26634697]
42. Pennacchio LA, Olivier M, Hubacek JA, Krauss RM, Rubin EM, Cohen JC. Two independent apolipoprotein A5 haplotypes influence human plasma triglyceride levels. Hum Mol Genet. 2002; 11:3031–8. [PubMed: 12417524]
43. Wang J, Ban MR, Kennedy BA, Anand S, Yusuf S, Huff MW, et al. APOA5 genetic variants are markers for classic hyperlipoproteinemia phenotypes and hypertriglyceridemia. Nat Clin Pract Cardiovasc Med. 2008; 5:730–7. [PubMed: 18779834]
44. Service SK, Teslovich TM, Fuchsberger C, Ramensky V, Yajnik P, Koboldt DC, et al. Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. PLoS Genet. 2014; 10:e1004147. [PubMed: 24497850]
45. Do R, Stitziel NO, Won HH, Jorgensen AB, Duga S, Angelica Merlini P, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. Nature. 2015; 518:102–6. [PubMed: 25487149]

Clinical Perspective

Identity-by-descent mapping using empirical estimates of allele sharing between all pairs of individuals may be powerful in founder populations, where hidden relationships may augment the inherent genetic information that can be used for gene localization. We tested the usefulness of this approach by analyzing lipid profiles in 1,024 Pima Indians, a relatively genetically homogeneous population. We identified a major locus for serum triglycerides (p=2.9 10–11 on chromosome 11q). In multi-stage follow-up analyses using ~9,000 subjects, we determined that this signal reflects effects of an Ala43Thr substitution in the APOC3 gene, and 3 established functional genetic variants in the APOA5 gene, collectively accounting for 6.9% of variation of triglyceride levels in Pimas. We further demonstrated that these 3 APOA5 variants could explain the association with the well-established variant for triglycerides levels, rs964184. This study provides a proof of concept that identity-by-descent mapping can be a useful strategy to identify causal variants affecting complex traits. The identification of these genes and specific genetic variants that affect an important risk factor for cardiovascular diseases (CVD) may contribute to the development of novel CVD interventions.

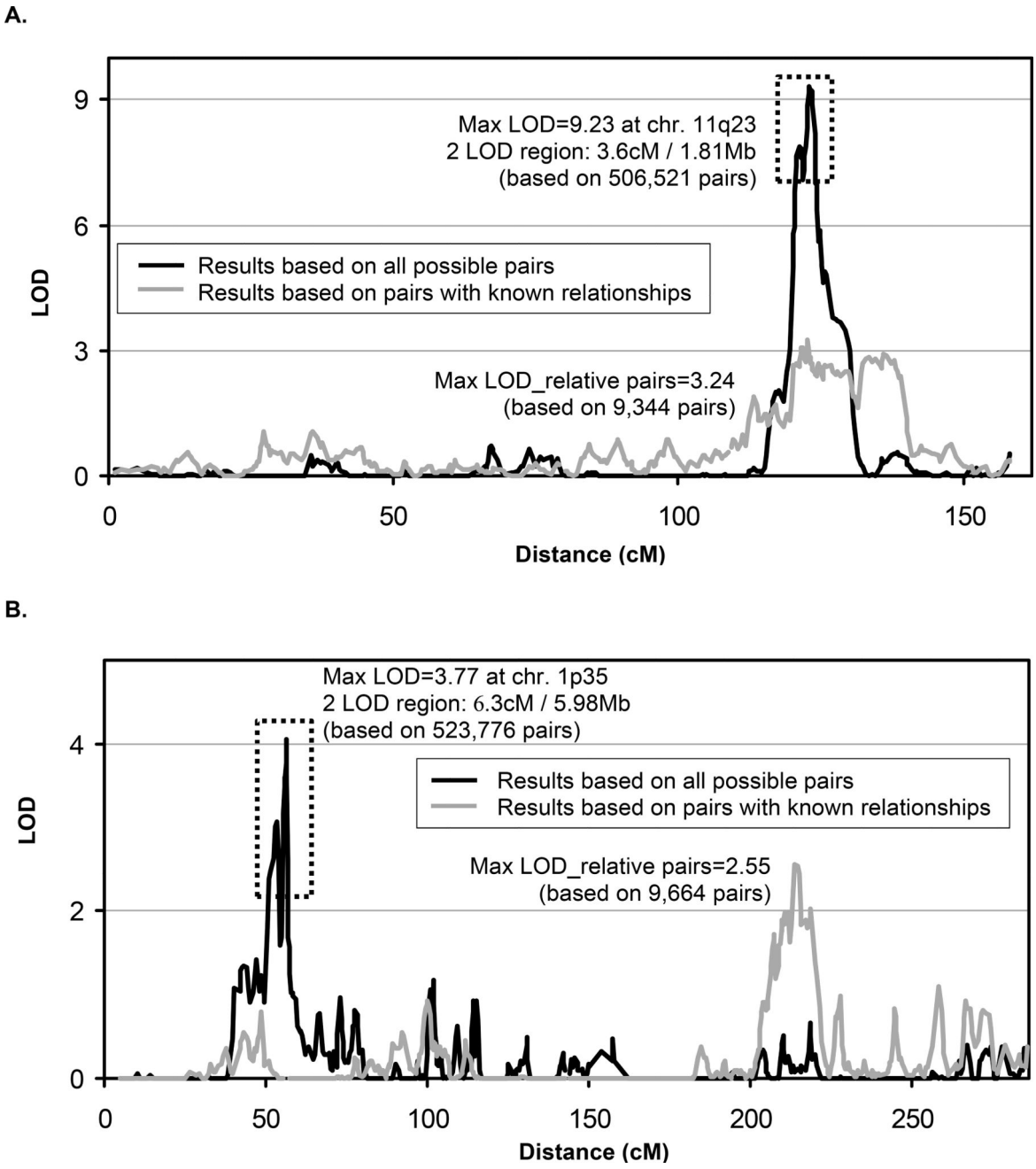

Figure 1.

Linkage study results of serum lipids showing the maximum LOD>3. (A) Results of triglycerides on chromosome 11; (B) Results of HDL-cholesterol on chromosome 1

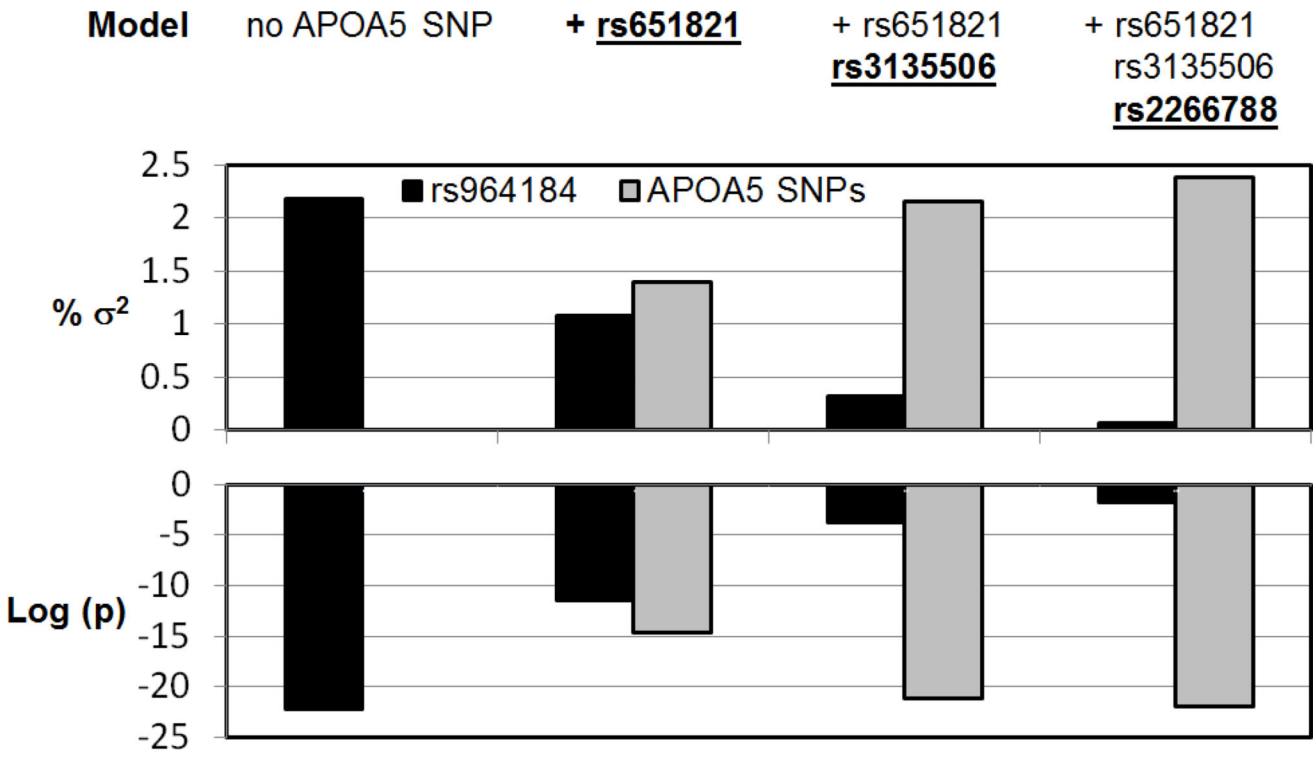

Figure 2.

The impact of *APOA5* SNPs on the association between rs964184 and serum triglycerides in the Pima sample. The top panel shows the % variance (%σ^2) of triglycerides explained by rs964184 (black bars, conditional on any *APOA5* SNPs) and the % σ^2 resulting from the addition of the *APOA5* SNPs (gray bars) to the statistical model. The bottom panel shows the significance of association (log(p)). Covariate effects adjusted included age, sex, diabetes status and rs147210663 genotypes. The order of *APOA5* SNP addition was determined by their association significance conditioning on effects of SNPs with stronger effects. The p values for the *APOA5* SNP associations were calculated with the df equal to the number of SNPs in the model.

Table 1

Three variants with distinct and significant associations with serum triglycerides in the fine-mapping study

rs number	Position (build 37)	Variant type	# tag SNPs*	MAF†	Change in LOD	Remaining LOD	p‡	β± SEM§	variance explained
rs147210663	116701560	nsSNP¶	0	2.6%	−7.08	2.24	1.6×10^{-13}	−1.16±0.14	6.9%
rs2072560	116661826	Intronic SNP	3	12.8%	−1.76	0.48	0.00028	0.31±0.061	2.4%
rs11357208	116784303	Intronic, indel	0	10.6%	−0.48	0	0.0049	0.34±0.075	2.6%

*A tag SNP was defined as being in strong linkage disequilibrium (r^2 >0.8) with the target SNP.

†MAF: minor allele frequency.

‡Corrected for multiple testing.

§For the minor allele, in standard deviation (SD) unit.

¶non-synonymous SNP.

Table 2

SNPs with distinct and significant associations with serum triglycerides in replication studies

Population	Sample size	rs number	Position (build 37)	Gene	Reason	MAF	β*±SEM	variance explained	p†
Pima Indians	4,668	rs147210663	116701560	*APOC3* nsSNP‡	Top hit§	2.5%	−0.93±0.064	4.7%	7.4×10^{-48}
		rs964184	116648917	near the 3'UTR of *ZPR1*	GWAS literature	39.8%	0.20±0.020	2.2%	2.1×10^{-22}
		rs651821	116662579	5'UTR of *APOA5*	2nd hit tag	14.7%	0.12±0.032	0.3%	0.0012
FIND (Southwest Amerindians)	2,793	rs964184	116648917	near the 3'UTR of *ZPR1*	GWAS literature	34.9%	0.22±0.033	2.3%	1.4×10^{-10}
		rs147210663	116701560	*APOC3* nsSNP	Top hit	1.0%	−0.81±0.16	1.4%	6.3×10^{-7}
Pima + FIND samples	7,461	rs147210663	116701560	*APOC3* nsSNP	Top hit	2.0%	−0.92±0.059	3.5%	7.6×10^{-55}
		rs964184	116648917	near the 3'UTR of *ZPR1*	GWAS literature	38.8%	0.21±0.017	2.1%	2.0×10^{-31}
		rs651821	116662579	5'UTR of *APOA5*	2nd hit tag	14.3%	0.10±0.027	0.2%	2.1×10^{-4}

*For the minor allele, in SD unit, 1 SD unit=0.587 and 0.555 in Pima Indian and Southwest Amerindian samples, respectively.

†Corrected for multiple testing.

‡non-synonymous SNP.

§SNPs with significant associations identified from fine-mapping studies.

Table 3

Haplotype analysis results of 3 *APOA5* SNPs* with serum triglycerides

Haplotype	SNP tested for allelic effect	Pima sample (n=4,868)			FIND sample (n=2,794)		
		haplotype frequency (%)	β†±SEM	p	haplotype frequency (%)	β±SEM	p
Reference: *APOA5*1* (T-Ser-A)		60.9	--	--	65.6	--	--
*APOA5*5* (**C**-Ser-A)	rs2266788	3.6	0.09±0.035	0.011	0.5	0.29±0.12	0.015
*APOA5*3* (T-**Trp**-A)	rs3135506	21.2	0.11±0.016	1.3×10^{-12}	20.5	0.11±0.019	7.4×10^{-9}
*APOA5*2* (**C**-Ser-**G**)	rs2266788 + rs651821	14.2	0.19±0.019	2.2×10^{-23}	12.7	0.15±0.023	4.3×10^{-11}
Reference: *APOA5*5* (C-Ser-A)		3.6		--	0.5	--	--
*APOA5*2* (C-Ser-**G**)	rs651821	14.2	0.10±0.038	0.0097	12.7	0.14±0.12	0.24

*SNPs were ordered according to their physical map positions: rs2266788, rs3135506, and rs651821. The major allele for each SNP was coded as "0", and the minor allele coded as "1". There were 4 major haplotypes with a frequency >1%.

†Triglycerides level was presented as SD units. Results adjusted for age, sex, population admixture estimate, diabetes status and rs147210663.

Table 4

Effects of 4 functional SNPs with significant triglyceride associations on the observed linkage signal

Model	Observed LOD	variance due to linkage	variance due to all SNP(s)	p for all SNP effects
Linkage only	9.32	10.8%	----	----
APOC3 SNP effect	2.24	6.2%	6.9%	1.9×10^{-13}
APOC3 + 3 *APOA5* SNP effects	0.08	1.4%	9.1%	2.1×10^{-12} *

* The p value was determined based on a 4-df test.