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

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

Background—Identity-by-descent mapping using empirical estimates of identity-by-descent 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 identity-by-descent mapping, using ≈400 000 single-nucleotide polymorphisms (SNPs), of serum lipid profiles, we identified a major linkage signal for triglycerides in 1007 Pima Indians (LOD=9.23; P=3.5×10^{-11} on chromosome 11q). In subsequent fine-mapping and replication association studies in ≈7500 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 triglycerides (β=−0.92±0.059 SD unit; P=9.6×10^{-55} in 4668 Pimas and 2793 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 triglycerides identified by genome-wide association studies. Collectively, these 4 SNPs account for 6.9% of variation in triglycerides 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. **(*Circ Cardiovasc Genet*. 2017;10:e001809. DOI: 10.1161/CIRCGENETICS.117.001809.)**

Key Words: alleles ■ chromosome ■ genome-wide association study ■ mutation ■ triglycerides

Genome-wide association studies (GWASs) 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 (ie, 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

See Editorial by Aguilar-Salinas et al
See Clinical Perspective

Received May 1, 2017; accepted October 3, 2017.

From the Phoenix Epidemiology and Clinical Research Branch, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, AZ (W.-C.H., A.K.N., S.K., P.C., A.M., W.C.K., L.J.B., R.L.H.); South Texas Diabetes and Obesity Institute, University of Texas Rio Grande Valley School of Medicine, San Antonio (H.H.H.G.); Departments of Medicine and Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore (T.I.P.); and Illumina Inc, San Diego, CA (A.M.).

The Data Supplement is available at http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.117.001809/-/DC1.

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

lipid levels in 1024 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 through a 4-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 mellitus conducted in Arizona, where most of the participants are Pima Indians (the Pima study).[4] For the initial discovery involving GWLS, 1024 Pima subjects, who had also participated in a GWAS[3] with available lipid data, including total cholesterol, high-density lipoprotein cholesterol (HDL-C), triglycerides, and low-density lipoprotein cholesterol measurements, were included (1007 had triglyceride measurements).

The first sample for replication studies included 5491 (4668 with triglyceride data) additional Pima subjects who were not part of the initial discovery set and did not have GWAS data. Among these participants, 2713 are full-heritage Pima Indians (defined as self-reported 8/8th Pima heritage) and another 2778 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=3189; 2793 with triglyceride data) had ≥50% Amerindian heritage, and most were urban-dwelling Amerindians living in or near Phoenix. Table I 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, 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 single-nucleotide polymorphism (SNP) Array (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 to 40×.

Genotyping of 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 perform 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 on the parameter settings and calculation are provided in the Data Supplement.

GWLS of 4 Lipid Traits

All traits were normalized by inverse Gaussian transformation before analysis. Linkage analysis was conducted using the principles of the variance components method developed by Amos[7] and performed using SAS (SAS Institute, Cary, NC; Methods in the Data Supplement). For computational efficiency, we performed linkage analysis at every 50th variant across the genome, after which the maximum logarithm of the odds for linkage (max LOD) was determined. In other words, ≈8000 tests (at ≈0.44 cM intervals) were performed.

Because 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 performed association analyses conditional on linkage effects (ie, accounting for random effects of both local and global IBD) to evaluate the association of variants under the linkage peak in the 1024 subjects used in linkage analyses. The region of interest was defined as the 2-LOD support interval (ie, the interval in which the LOD is within 2 U 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 (ie, P value <0.05 corrected for multiple comparisons; Methods in the Data Supplement).

Replication Association Analyses of Triglycerides

Replication studies were performed in 2 stages. First, we conducted the replication study in Pima Indians (the Pima sample; n=4668 after exclusion of those in the GWLS). If any associations were replicated, we conducted a second replication study using 2793 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 triglycerides, we performed haplotype analysis to assess associations conditional on specific allelic backgrounds constituting ≥2 SNPs (Data Supplement).

Covariates (population admixture estimates, type 2 diabetes mellitus 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 1024 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 Figure I in the Data Supplement.

Genome-Wide Linkage Studies

Among 4 lipid traits analyzed, significant linkage (max LOD≥3) was identified for triglycerides and HDL-C. The max LOD for triglycerides was 9.23 (P=3.5×10^{-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=1.5×10^{-5}) on chromosome 1p, explaining 7.5% of the variance (Table II and Figure II in the Data Supplement). When the same analysis was conducted using only pairs with known self-reported relationships (n=9664), the max LOD for triglycerides was still observed in the same genomic region, however, with a much-reduced significance (LOD=3.24; P=5.6×10^{-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 triglycerides 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 triglycerides with SNPs in 11q23 at genome-wide significance (Figure III in the Data Supplement). The strongest association was with rs4417316, an intronic SNP in *ZPR1* (P=1.8×10^{-10}), but rs964184, the most commonly reported GWAS SNP for triglycerides observed in multiple populations, also was strongly associated (P=4.0×10^{-8}). The associations of all SNPs in this region were attenuated to P >10^{-5} after adjusting for the effects of rs964184. Thus, it seemed that the GWAS associations in Pimas largely reflected the effect of the established SNP rs964184. In contrast, linkage analyses of triglycerides 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 Triglycerides 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 1024 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 triglycerides, after 4 rounds of analyses of 3450 variants in a 1.81 Mb region on 11q23, we identified 3 variants with distinct and significant associations (Table 1). The SNP with the strongest association with triglycerides was rs147210663 explaining 6.9% of the variance (P=1.6×10^{-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

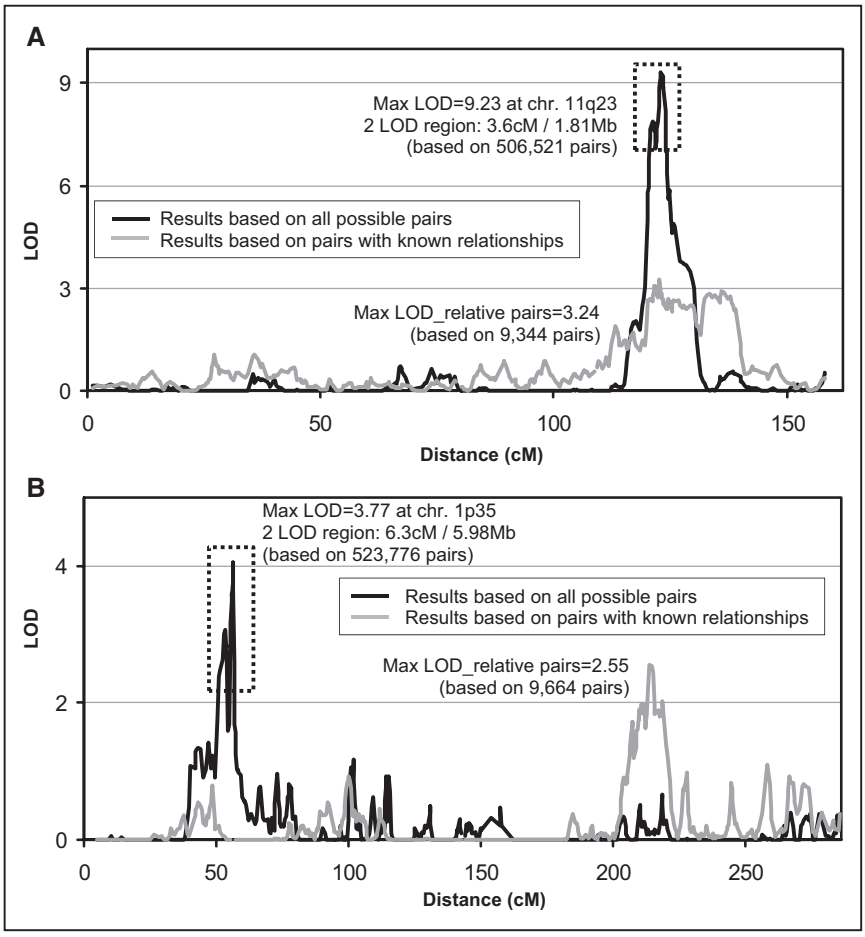


Figure 1. Linkage study results of serum lipids showing the maximum LOD >3. (**A**) Results of triglycerides on chromosome 11; (**B**) Results of high-density lipoproteins cholesterol on chromosome 1.

Table 1. Three Variants With Distinct and Significant Associations With Serum Triglycerides in the Fine-Mapping Study

rs Number	Position (Build 37)	Variant Type	No. of Tag SNPs*	MAF	Change in LOD	Remaining LOD	*P* Value†	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%

MAF indicates minor allele frequency; nsSNP, nonsynonymous SNP; and SNP, single-nucleotide polymorphisms.
*A tag SNP was defined as being in strong linkage disequilibrium (r^2>0.8) with the target SNP.
†Corrected for multiple testing.
‡For the minor allele, in SD unit.

the Thr residue and has a frequency of 2.6% in GWLS. The SNP with the second strongest association with triglycerides 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 resides in 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 Triglycerides

To confirm and further characterize the associations identified in the fine-mapping study, we analyzed the 3 distinctly associated triglyceride variants, along with their tags (r^2>0.8 in Pima WGS data) and nearby established triglyceride-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=4668; the Pima sample) and (2) southwestern urban Amerindians from the FIND sample (n=2793). We, thus, analyzed 11 SNPs, including 6 established triglyceride-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 second variant (represented by rs651821; *P*=0.0012; Table 2; Table III 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% versus 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′untranslated region (UTR) of the zinc finger gene (*ZPR1*, also known as *ZNF259*),

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* Value†
Pima Indians	4668	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*	Second hit tag	14.7%	0.12±0.032	0.3%	0.0012
FIND (Southwest Amerindians)	2793	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	7461	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*	Second hit tag	14.3%	0.10±0.027	0.2%	2.1×10^{-4}

FIND indicates Family Investigation of Nephropathy and Diabetes; GWAS, genome-wide association studies; MAF, minor allele frequency; nsSNP, nonsynonymous SNP; SNP, single-nucleotide polymorphism; and UTR, untranslated region.
*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.
‡SNPs with significant associations identified from fine-mapping studies.

but evidence for a functional effect of this SNP is lacking. Two haploblocks (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 triglyceride metabolism and has an SNP (rs651821) significantly associated with triglycerides. Therefore, we extended our association study to investigate whether other functional variants in these 2 genes may explain the observed association with rs964184. On the basis of Pima WGS data, 4 SNPs in *APOA5* were previously documented to be functional[20–23] with a minor allele frequency ≥1%: rs2266788, rs3135506, rs651821, and rs662799. We tested associations between triglycerides and these SNPs to examine the extent to which they explained the association of rs964184 with triglycerides. As shown in Figure 2, in the Pima sample, rs964184 (dark bars) accounted for 2.2% of the trait variance with *P* value of 5.7×10^{-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 triglycerides remained highly significant after adjusting for all 3 SNPs. In addition, although our study populations had a high type 2 diabetes mellitus prevalence, the genetic associations with these SNPs in *APOC3* and *APOA5* were observed in those with or without diabetes mellitus (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 IV in the Data Supplement), and their statistical associations may be codependent when their effects are assessed individually. Therefore, we performed 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 3 common haplotypes at these SNPs: *APOA5**1 containing the triglyceride-lowering allele at all 3 loci (T-Ser-A for rs2266788-rs3135506-rs651821), *APOA5*2* containing the triglyceride-raising allele at rs2266788 and rs651821 (C-Ser-G), and *APOA5*3* containing the triglyceride-raising allele only at rs3135506 (T-Trp-A). In East Asian populations, an additional haplotype (*APOA5**4) containing only 1 triglyceride-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 triglyceride-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 triglycerides ($P=1.3\times10^{-12}$). Furthermore, compared with *APOA5*1*, *APOA5*2,* which differs at rs651821 and rs2266788, was associated with significantly higher triglycerides ($P=2.2\times10^{-23}$). 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 because of the low frequency of *APOA5*5* (0.5%).

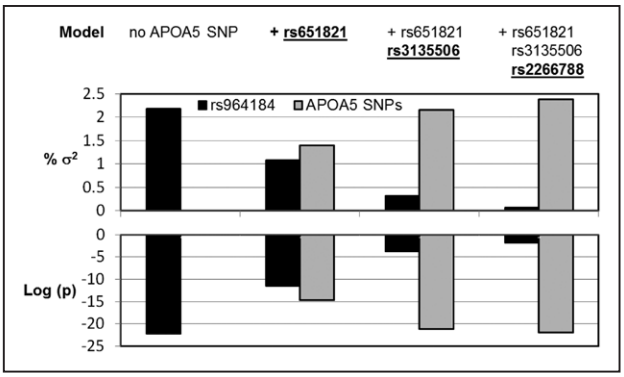


Figure 2. The impact of *APOA5* single-nucleotide polymorphisms (SNPs) on the association between rs964184 and serum triglycerides in the Pima sample. **Top**, the % variance (%σ^2) of triglycerides explained by rs964184 (black bars, conditional on any *APOA5* SNPs) and the %σ^2 resulting from the addition of the *APOA5* SNPs (gray bars) to the statistical model. **Bottom**, The significance of association (log(p)). Covariate effects adjusted included age, sex, diabetes mellitus 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.

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 triglycerides. Their effects all seemed additive (Figure V in the Data Supplement). Therefore, we included these 4 SNPs as covariates in the original linkage model (n=1007 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 triglycerides.

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

Table 3. Haplotype Analysis Results of 3 *APOA5* SNPs* With Serum Triglycerides

Haplotype	SNP Tested for Allelic Effect	Pima Sample (n=4868)			FIND Sample (n=2794)		
		Haplotype Frequency (%)	β†±SEM	*P* Value	Haplotype Frequency (%)	β±SEM	*P* Value
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

FIND indicates Family Investigation of Nephropathy and Diabetes; and SNP, single-nucleotide polymorphism.

*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 mellitus status, and rs147210663.

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 of 9.23 for triglycerides 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 triglycerides. 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 influence on lipid metabolism despite rs964184 being the most widely replicated GWAS SNP for triglycerides. Thus, we focused on variants in the only other candidate, *APOA5*. *APOA5* variants have been extensively studied functionally because of the protein's importance in triglyceride 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 fourth SNP is in almost perfect LD with 1 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 triglycerides and had an effect size (β=–0.92±0.64 SD unit for the minor allele Thr, which corresponds to a 42% reduction in triglycerides) among the largest reported for any complex trait.[25] Its effect accounted for 4.7% of the variance of triglycerides 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 triglycerides contained in very-low-density lipoprotein particles (Table III in the Data Supplement). The Thr allele is rare in most other populations, with a minor allele frequency 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 2 Mayan Indians with low triglycerides and APOC3 levels (where it was denoted as A23T).[26] Strong associations between triglycerides and rare nonsynonymous SNPs in *APOC3* have been observed in other populations as well, with loss-of-function variants conferring lower triglyceride 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 triglyceride levels. This same SNP was also identified as strongly associated with low triglyceride 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 triglycerides 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.

Table 4. Effects of 4 Functional SNPs With Significant Triglyceride Associations on the Observed Linkage Signal

Model	Observed LOD	Variance Because of Linkage	Variance Because of All SNP(s)	*P* Value 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}*

SNP indicates single-nucleotide polymorphism.

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

APOC3 is a small gene with 4 exons and 297 nucleotides. The protein is synthesized mainly in the liver and strongly inhibits its hepatic uptake of very-low-density lipoprotein cholesterol and intermediate-density lipoproteins. Intracellularly, it promotes hepatic very-low-density lipoprotein cholesterol assembly and secretion.[34] Extracellularly, it attenuates hydrolysis and clearance of triglyceride-rich lipoproteins and impairs the lipolysis of triglyceride-rich lipoproteins by inhibiting lipoprotein lipase

and the hepatic uptake of triglyceride-rich lipoproteins 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 nonsynonymous SNP (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 very-low-density lipoprotein 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 triglyceride-rich lipoproteins and lower plasma triglyceride 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 triglycerides.

In our replication studies, the SNP with the second strongest association with triglycerides was rs964184. It is one of the most well-replicated SNPs from many large lipid GWAS, with particularly strong associations for triglycerides.[10,12–15,19,25,37,38] Previous studies in Amerindian-derived populations have identified strong associations between triglycerides 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 triglycerides after conditioning on rs964184 (Table III 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, and rs6622799) have well-documented functional effects on APOA5. The only nonsynonymous SNP 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 triglycerides, 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 minor allele frequencies in Pimas (0.14–0.20 versus 0.06–0.08 in whites), which resulted in increased statistical power in our study populations. One caveat is that as the fourth 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 triglycerides. In 2001, Pennacchio et al[40] 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 triglyceride levels, 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 provided additional support for multivariant influences from *APOA5* on triglycerides. 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 triglycerides 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 IV in the Data Supplement), the haplotypes comprised any minor (triglyceride-raising) allele of *APOA5* SNPs were in strong LD with the minor allele of rs964184 collectively (r^2=0.97; Figure VI in the Data Supplement), and, thus, the *APOA5* SNPs can largely account for the association of rs964184 with triglycerides in Amerindians. On the basis of analysis of data from the 1000 Genomes project (www.1000genomes.org), it is noteworthy that rs964184 is similarly highly concordant with haplotypes containing a triglyceride-raising allele at these *APOA5* SNPs in populations representative of those where rs964184 was identified as a top GWAS SNP (eg, r^2=0.96 in European ancestry in Utah and Gujarati Indian in Houston; r^2=0.82 in Chinese in Beijing and Mexican ancestry in Los Angeles); thus, these *APOA5* SNPs might explain the association between rs964184 and triglycerides in other populations as well. The role of APOA5 in influencing triglyceride levels is supported by recent human sequencing studies showing rare deleterious *APOA5* mutations associated with triglycerides.[45] Given statistical, functional, and physiological evidence, we conclude that rs964184 is likely a marker for the collective effects of *APOA5* functional SNPs on triglycerides.

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 1007 subjects, we observed associations for triglycerides 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 that 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 triglycerides, 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 that the detectable effect size is 9.7% of the variance for LOD >3 with 80% power (our linkage signal for triglycerides had an effect size of 10.8% of the variance).

In conclusion, we performed 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 triglyceride levels. These findings suggest that population-based GWLS may provide complementary information to GWAS, particularly in founder populations. Identification of the triglyceride-lowering nonsynonymous SNP 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 1000 Genome data as the reference panel). The *APOC3* A43T SNP has been established as a loss-of-function variant leading to lower triglycerides 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 triglyceride-lowering properties of this SNP. Thus, the linkage signal we detected represents effects of both an established triglyceride 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 triglycerides. The association between rs964184 and triglycerides 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 triglycerides, 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.

Acknowledgments

This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health 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 National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health.

Disclosures

None.

References

1. Rosenthal E, Blue E, Jarvik GP. Next-generation gene discovery for variants of large impact on lipid traits. *Curr Opin Lipidol*. 2015;26:114–119. doi: 10.1097/MOL.0000000000000156.
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;88:706–717. doi: 10.1016/j.ajhg.2011.04.023.
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;63:369–376. doi: 10.2337/db13-0416.
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(suppl 6):1543S–1551S.
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;80:1014–1023.
6. Browning SR, Browning BL. High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet*. 2010;86:526–539. doi: 10.1016/j.ajhg.2010.02.021.
7. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet*. 1994;54:535–543.
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.
9. Moskvina V, Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol*. 2008;32:567–573. doi: 10.1002/gepi.20331.
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;5:242–249. doi: 10.1161/CIRCGENETICS.111.961482.
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. doi: 10.1371/journal.pgen.1003379.
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;466:707–713. doi: 10.1038/nature09270.
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. doi: 10.1371/journal.pone.0037056.
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. doi: 10.1136/jmedgenet-2012-101461.
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. doi: 10.1371/journal.pgen.1002138.
16. Talmud PJ, Drenos F, Shah S, Shah T, Palmen J, Verzilli C, et al; ASCOT Investigators; NORDIL Investigators; BRIGHT Consortium. Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. *Am J Hum Genet*. 2009;85:628–642. doi: 10.1016/j.ajhg.2009.10.014.
17. Musunuru K, Lettre G, Young T, Farlow DN, Pirruccello JP, Ejebe KG, et al; NHLBI Candidate Gene Association Resource. Candidate gene association resource (CARe): design, methods, and proof of concept. *Circ Cardiovasc Genet*. 2010;3:267–275. doi: 10.1161/CIRCGENETICS.109.882696.
18. Rosenthal EA, Ranchalis J, Crosslin DR, Burt A, Brunzell JD, Motulsky AG, et al; NHLBI GO Exome Sequencing Project. Joint linkage and association analysis with exome sequence data implicates SLC25A40 in hypertriglyceridemia. *Am J Hum Genet*. 2013;93:1035–1045. doi: 10.1016/j.ajhg.2013.10.019.
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. doi: 10.1038/ncomms4983.
20. Talmud PJ, Palmen J, Putt W, Lins L, Humphries SE. Determination of the functionality of common APOA5 polymorphisms. *J Biol Chem*. 2005;280:28215–28220. doi: 10.1074/jbc.M502144200.
21. Ahituv N, Akiyama J, Chapman-Helleboid A, Fruchart J, Pennacchio LA. In vivo characterization of human APOA5 haplotypes. *Genomics*. 2007;90:674–679. doi: 10.1016/j.ygeno.2007.08.003.
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;1782:447–452. doi: 10.1016/j.bbadis.2008.03.003.
23. Caussy C, Charrière S, Marçais 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;94:129–134. doi: 10.1016/j.ajhg.2013.12.001.
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. doi: 10.1371/journal.pone.0110258.
25. Walter K, Min JL, Huang J, Crooks L, Memari Y, McCarthy S, et al; UK10K Consortium. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526:82–90. doi: 10.1038/nature14962.
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. *J Lipid Res*. 2000;41:1760–1771.
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;322:1702–1705. doi: 10.1126/science.1161524.
28. Tachmazidou I, Dedoussis G, Southam L, Farmaki AE, Ritchie GR, Xifara DK, et al; UK10K Consortium. A rare functional cardioprotective APOC3 variant has risen in frequency in distinct population isolates. *Nat Commun*. 2013;4:2872. doi: 10.1038/ncomms3872.
29. Timpson NJ, Walter K, Min JL, Tachmazidou I, Malerba G, Shin SY, et al; UK1OK Consortium Members; UK1OK Consortium Members. A rare variant in APOC3 is associated with plasma triglyceride and VLDL levels in Europeans. *Nat Commun*. 2014;5:4871. doi: 10.1038/ncomms5871.
30. Blood I, Crosby J, Peloso GM, Auer PL, et al; TG, HDL Working Group of the Exome Sequencing Project NHL. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med*. 2014;371:22–31.
31. Natarajan P, Kohli P, Baber U, Nguyen KH, 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;66:2053–2055. doi: 10.1016/j.jacc.2015.08.866.
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;47:640–642. doi: 10.1038/ng.3270.
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;48:634–639. doi: 10.1038/ng.3561.
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. *Arterioscler Thromb Vasc Biol*. 2010;30:239–245. doi: 10.1161/ATVBAHA.109.197830.
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;373:438–447. doi: 10.1056/NEJMoa1400283.
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. *J Lipid Res*. 2010;51:1524–1534. doi: 10.1194/jlr.M005108.
37. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, et al; Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat Genet*. 2013;45:1274–1283. doi: 10.1038/ng.2797.
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;40:189–197. doi: 10.1038/ng.75.
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. doi: 10.1371/journal.pone.0172880.
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;294:169–173. doi: 10.1126/science.1064852.
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. *J Lipid Res*. 2016;57:318–324. doi: 10.1194/jlr.P063461.
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;11:3031–3038.
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;5:730–737. doi: 10.1038/ncpcardio1326.
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. doi: 10.1371/journal.pgen.1004147.
45. Do R, Stitziel NO, Won HH, Jørgensen AB, Duga S, Angelica Merlini P, et al; NHLBI Exome Sequencing Project. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015;518:102–106. doi: 10.1038/nature13917.

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 1024 Pima Indians, a relatively genetically homogeneous population. We identified a major locus for serum triglycerides (P=2.9×10^{-11} on chromosome 11q). In multistage follow-up analyses using ≈9000 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 triglyceride 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 may contribute to the development of novel cardiovascular diseases interventions.

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. Göring, Toni I. Pollin, Alka Malhotra, William C. Knowler, Leslie J. Baier and Robert L. Hanson

Circ Cardiovasc Genet. 2017;10:
doi: 10.1161/CIRCGENETICS.117.001809
Circulation: Cardiovascular Genetics is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231

Print ISSN: 1942-325X. Online ISSN: 1942-3268

SUPPLEMENTAL MATERIAL

Supplemental Methods

Phenotypic measurements

Individuals in the Pima study were community residents ≥ 15 years old; individuals in the FIND study were urban Amerindians (≥50% heritage by self-report) who were ≥18 years old. In both studies participants were included irrespective of health status. For the Pima study, we selected data from the time point from the last available examination prior to 2005 because the self-reported use of antilipidemic medication in this population was very limited (4.6% of subjects) during this time.[1] Use of antilipidemic medicines was also limited in the FIND participants (5.1% of subjects). Over 85% of those using antilipidemic medicines were taking statins; fibrate use was uncommon. Measurement methods of lipid levels were the same for both the Pima and FIND studies, and have been previously described.[2] Briefly, serum samples were collected 1–5 days before measurements. Total serum cholesterol was determined with a colorimetric method from 1965 to March 1992 and with an enzymatic method subsequently. Serum TG and HDL-C concentrations have been measured since 1993 by enzymatic methods. LDL-C was estimated with the Friedewald formula.[3] Hypertriglyceridemia was defined if TG was ≥1.69 mmol/L. In addition, measurements of 3 lipoprotein fractions (very low density lipoprotein cholesterol or VLDL-C, LDL-TG, and VLDL-TG) were available in a subset (n=206).[4]

In the Pima study, measurements of height and weight were performed by medically trained personnel to calculate body mass index (BMI). T2D was determined according to 1997 American Diabetes Association criteria based on results from an oral glucose tolerance test (or OGTT, i.e. fasting plasma glucose ≥7.0 mmol/l, 2-h plasma glucose concentration ≥11.1 mmol/l) or medical record reviews. In FIND participants, T2D was defined based on a previous medical

diagnosis, hemoglobin $A1_C$ ≥6.5% or fasting plasma glucose concentration ≥7.0 mmol/l (2010 American Diabetes Association criteria).

Genotypic data ascertainment

Genotype data 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 (Affymetrix, Santa Clara, CA) using the BIRDSEED algorithm, as described previously.[5] SNPs were excluded under any of 4 conditions: (1) >15% of missing genotype calls, (2) genotype frequencies diverged from Hardy-Weinberg expectations (P <0.001), (3) concordance among 100 duplicate samples <97%, or (4) the minor allele frequency (MAF) <5%. Genotyping of SNPs in both replication sets was performed by BeadXpress system (Illumina, San Diego, CA), Taqman genotyping assays (Applied Biosystems, Carlsbad, CA), KASP based assays (LGC, Middlesex, UK), or a custom SNP Array designed to capture common variants in the Pima Indian population (Affymetrix, Santa Clara, CA), all according to manufacturers' protocols. Genotype quality control required a call rate >95%, no deviation from HWE (P<0.001) and a discrepancy rate of <2.5% for blind duplicates (>100 for each sample set). Genotypes of all carriers of *APOC3* A43T were verified by direct sequencing.

Statistical analyses

Estimation of the percentage of alleles shared identical-by-descent matrix. The execution of variance components linkage analysis of quantitative traits requires information on the alleles shared identical-by-descent (IBD) between 2 individuals. Traditionally, IBD was estimated between pairs of individuals in a pedigree based on self-reported relationships. In recent years, several new methods have been developed to estimate IBD more accurately using high-density genetic data, and these methods are applicable even in pairs of individuals without known relationships.[6-9] We used the program Beagle[10] and genetic maps from the Hapmap project

(http://hapmap.ncbi.nlm.nih.gov/, based on Phase II and Build 37 data) to carry out IBD estimation at each of ~400,000 SNPs in our dataset. The program Beagle takes the LD among variants into account for the IBD estimates. By using high-density genetic data (e.g. SNP data used for GWAS), probabilities for phased haplotypes are calculated and used as the basis to estimate IBD at any given locus. The average relatedness for any given pair was calculated as the genome-wide average of local IBD. As our Pima samples had the characteristics of a founder population, the recommended default setting for the scale parameter[11] was not optimal; thus we set the scale parameter at 10, which we determined by simulation to produce more accurate estimates for a sample of this size derived from a founder population. Beagle calculates the probability that a pair of individuals share ≥1 allele IBD at each SNP (*i.e.*, ignoring bilineal sharing); this value was converted to proportion IBD sharing by multiplying by ½. For known full sibling pairs (who are the major pair type with bilineal sharing), the multiplication was by ⅔ instead of ½, as given that they share ≥ 1 allele, they will on average share 1 allele ⅔ of the time and 2 alleles ⅓ of the time (*i.e.,* the expected proportion of alleles shared IBD conditional on sharing ≥1 allele =0.5*⅔ + 1.0*⅓ = ⅔).

Genome-wide linkage analysis of 4 lipid traits. Linkage analysis was conducted using the principles of the variance-components method developed by Amos.[12] Details of our approach have been described in detail previously.[13-15] Briefly, a linear mixed model is fitted to estimate fixed effects, representing the intercept and covariate effects, and three components of variance: an additive "monogenic" component (σ^2_M) that estimates effects of a locus in the region of interest, a "polygenic" component (σ^2_G) that incorporates overall relatedness, and an "environmental" component (σ^2_E) that incorporates effects unique to the individual. Thus, the variance-covariance matrix for the trait (Ω) among all individuals in the sample is modeled as:

$$\Omega = \Pi\sigma^2_M + \Phi\sigma^2_G + I\sigma^2_E,$$

where Π is a matrix of the IBD estimates between pairs of individuals at the location of interest, Φ is a matrix of the genome-wide average IBD, and I is an identity matrix. The null hypothesis of no linkage was assessed by comparing the full model to one in which the additive monogenic effect was constrained to 0, and the models were compared using a likelihood ratio test.[16] The logarithm of the odds score (LOD) for linkage was calculated by dividing the likelihood ratio test by $2\times\log_e(10)$.

As the variance component analysis can be sensitive to departures from a normal distribution of the trait, all 4 lipid traits were normalized by inverse Gaussian transformation for the linkage analysis. The same analysis of the trait using its natural logarithm transformation did not substantially alter results. The model was fit with the "PROC MIXED" function of SAS (SAS Institute, Cary, NC). To estimate the power of this approach in the present sample, we conducted simulations in which trait data were generated under the assumption that a randomly selected SNP was a functional variant with a specified effect on the trait variance.

Fine-mapping study –association analysis conditional on linkage effects. We used the GWAS data from the Affymetrix 6.0 array and the WGS data of 296 Pima subjects as the reference panel to impute genotypes of all variants (both SNPs and insertion-deletion variants) not available from the Affymetrix 6.0 array. Imputation was performed with MINIMAC.[17] All directly genotyped variants and imputed variants (either nsSNP, with minor allele frequency (MAF) ≥1% or any other variant with MAF ≥5%) with imputation r^2 >0.3 were analyzed for association. A total of 3,450 and 7,377 variants were analyzed for fine-mapping of the genome-wide significant TG and the HDL-C linkage locus, respectively. All analyses were conducted using SAS.

Replication association analyses of TG. The association between genotypes and TG was determined with linear regression modeling (additive model), where homozygotes for the major allele, heterozygotes, and homozygotes for the minor allele were coded to a numeric variable

for genotype (0, 1, and 2). To avoid the reduction in sample size resulting from missing data at multiple loci, we inferred missing genotypic data. To accomplish this, we calculated the probability of each of the 3 possible genotypes for each individual with missing data from the genotypes in the individual's relatives using MLINK,[18] and used these probabilities to construct the genotypic score. For the Pima sample, the model was fitted using a mixed model procedure to account for sibship. For the FIND sample, a linear regression model was used. To account further for cryptic relatedness in FIND (where family data are less extensive), the genomic control procedure was used, based on 42 randomly-selected SNPs.[19] Finally, meta-analysis for the Pima and FIND sample was conducted by the inverse variance method[20] to evaluate the association effects based on the largest sample. TG values were transformed by their natural logarithms (ln) to reduce skewness. We also included tag SNPs (r^2>0.8) of variants found to have significant associations with TG in the fine-mapping study in the replication study. The tag SNPs and haploblocks (default definition) were identified using Haploview.[21]

Correction for multiple testing in association analyses. We corrected for multiple comparisons accounting for linkage disequilibrium among SNPs by calculating the effective number of independent comparisons as suggested by Moskvina and Schmidt.[22] This method estimates the effective number of independent tests based on the pairwise correlation matrix between markers (i.e. the maximum absolute pairwise correlation between a given marker and all other markers in a defined window) and the desired overall type I error rate. By this method, for example, the 3,450 variants tested in the fine-mapping study of TG represented 718 effectively "independent" tests. In the replication study using the Pima sample, 11 variants tested were equivalent to 7.6 "independent" tests. Observed p values were corrected with a Bonferroni correction using this factor accordingly. All presented p values for the fine-mapping and replication studies are corrected p values.

Haplotype construction and analyses. Haplotype frequencies for pairs of variants were calculated in all Pima and FIND subjects with the Estimating Haplotypes (EH) program.[23, 24] D′ was calculated as a measure of allelic association, and r^2 as a measure of concordance. Association between traits and individual haplotypes were examined with a modification of the zero-recombinant haplotyping procedure.[25]

Estimates of population admixture. In the genome-wide and fine-mapping linkage studies, this estimate was calculated as the first principal component (PC1) in the principal component analysis of GWAS SNPs. For all samples used in the replication studies, as there were no SNP array data available, we typed 45 ancestry informative markers (AIMs) recommended by Tian et al.[26] The proportion of Amerindian heritage (% AI heritage) was estimated using a maximum likelihood method proposed by Hanis et al.[27] The correlation between PC1 and % AI heritage was high (r=0.87). In other words, the population admixture estimate for each subject used in the linkage studies was done based on GWAS SNP data, and that used in the replication studies was obtained based on 45 AIMs.

Covariates. Covariates used in all models included age, sex, T2D status at the time of the lipid measurement, and population admixture estimates. Participants taking antilipidemic medicines were included in the analyses presented (without adjustment), but linkage and association analyses were repeated excluding those taking antilipidemic medicines with similar results.

TABLE AND FIGURE LEGENDS

Table S1. Characteristics of 3 sets of study samples

Table S2. Maximum LOD from genome wide linkage studies of 4 serum lipid traits

Table S3. Results for 4 rounds of fine-mapping studies of 11 variants in 4,668 Pima Indians (Pima sample)

Table S4. The associations between rs147210663, rs964184, 3 APOA5 SNPs and triglycerides-related traits combining Pima and FIND sample

Figure S1. Empirically estimated IBD sharing by expected IBD sharing in the sample used for linkage analysis of serum lipids

Figure S2. Genome-wide linkage study results for 4 serum lipid traits

Figure S3. Genome-wide association study results for four serum lipid traits using the Affymetrix 6.0 array

Figure S4. Linkage disequilibrium pattern (r^2) between 6 SNPs in 4,636 Pima samples

Figure S5. Raw levels of serum triglycerides by genotypes of 4 SNPs

Figure S6. rs964184 vs. *APOA5* 3-SNP haplotypes and frequencies

Table S1. Characteristics of 3 sets of study samples

Pima Indian subjects (n = 1,024) used for genome-wide linkage analysis

characteristics	Men (n = 458, 44.7%)	Women (n = 566, 55.3%)
Age (years)	41.1 ± 13.6	42.9 ± 13.8
BMI (kg/m^2)	34.0 ± 8.2	36.9 ± 9.1
% T2D	37.6%	52.5%
Total cholesterol (mmol/L)	4.76 ± 0.99	4.69 ± 1.10
HDL cholesterol (mmol/L)	1.20 ± 0.41	1.26 ± 0.37
LDL cholesterol (mmol/L)	2.79 ± 0.81	2.63 ± 0.80
Triglycerides (mmol/L)	1.76 ± 1.48	1.70 ± 1.21

Replication Set 1 - 5,491 additional Pima Indians (Pima sample)

characteristics	Men (n = 2,295, 41.8%)	Women (n = 3,196, 58.2%)
Age	34.8 ± 14.5	36.9 ± 15.4
BMI	33.3 ± 8.0	35.5 ± 8.8
% T2D	31.6%	39.1%
% Amerindian heritage	0.88 ± 0.16	0.90 ± 0.14
Total cholesterol (mmol/L)	4.69 ± 1.02	4.51 ± 0.98
HDL cholesterol (mmol/L)	1.16 ± 0.37	1.23 ± 0.35
LDL cholesterol (mmol/L)	2.79 ± 0.84	2.58 ± 0.77
Triglycerides (mmol/L)	1.67 ± 1.47	1.51 ± 1.09
Hypertriglyceridemia*	34.7%	30.6%

Replication Set 2 - 3,189 southwestern Native Americans (FIND sample)

characteristics	Men (n = 1,576, 49.4%)	Women (n = 1,613, 50.6%)
Age	36.5 ± 12.4	36.7 ± 13.3
BMI	30.1 ± 6.5	32.8 ± 7.8
% T2D	24.2%	31.6%
% Amerindian heritage	0.80 ± 0.16	0.81 ± 0.16
Total cholesterol (mmol/L)	4.73 ± 1.07	4.50 ± 1.09
HDL cholesterol (mmol/L)	1.30 ± 0.45	1.25 ± 0.37
LDL cholesterol (mmol/L)	2.68 ± 0.80	2.49 ± 0.75
Triglycerides (mmol/L)	1.69 ± 1.31	1.61 ± 1.79
Hypertriglyceridemia*	36.0%	32.1%

* Hypertriglyceridemia was defined as having a serum triglyceride levels ≥ 1.69 mmol/L.

Table S2. Maximum LOD from genome wide linkage studies of 4 serum lipid traits*

Trait	Total heritability	n	Chromosome	LOD	p	variance explained
Total cholesterol	48%	1,023	15q	2.75	1.8×10^{-4}	7.7%
Triglycerides	36%	1,007	11q	9.23	3.5×10^{-11}	10.6%
HDL-cholesterol	37%	1,024	1p	3.77	1.5×10^{-5}	7.5%
LDL-cholesterol	51%	970	15q	2.12	8.9×10^{-4}	6.8%

* The distributions of all 4 traits were normalized for the linkage analyses.

Table S3. Results for 4 rounds of fine-mapping studies of 11 variants in 4,668 Pima Indians (Pima sample)

rs number	Position*	Gene	Variant type	Reason for testing	p† from Round 1 analysis	p† from Round 2 analysis	p† from Round 3 analysis	p† from Round 4 analysis
rs2075295	116628401	*BUD13*	Intronic	Tag SNP for rs2072560	2.5×10^{-8}	8.8×10^{-6}	0.091	1.00
rs3825041	116631707	*BUD13*	Intronic	Tag SNP for rs2072560	1.4×10^{-16}	1.4×10^{-13}	0.004	1.00
rs964184	116648917	*ZPR1*	Near 3' UTR	GWAS SNP	5.4×10^{-30}	2.1×10^{-22}‡	covariate	covariate
rs2072560	116661826	*APOA5*	Intronic	The 2nd strongest SNP from find-mapping study	2.9×10^{-8}	9.5×10^{-7}	0.100	1.00
rs3135506	116662407	*APOA5*	Missense SNP, S19W	GWAS SNP	4.6×10^{-6}	4.1×10^{-4}	0.041	1.00
rs651821	116662579	*APOA5*	5' UTR	Tag SNP for rs2072560 and a GWAS SNP	1.2×10^{-17}	1.5×10^{-14}	0.0012‡	covariate
rs662799	116663707	*APOA5*	Promoter	GWAS SNP	1.9×10^{-17}	6.6×10^{-14}	0.027	1.00
rs147210663	116701560	*APOC3*	Missense SNP, A43T	The strongest SNP from find-mapping study	7.4×10^{-48}‡	covariate	covariate	covariate
rs12225230	116728630	*SIK3*	Missense SNP, P917R	GWAS SNP	3.1×10^{-6}	4.4×10^{-4}	0.63	1.00
rs11357208	116784304	*SIK3*	Indel, Intronic	The 3rd strongest SNP from find-mapping study	1.2×10^{-6}	1.8×10^{-4}	0.79	1.00
rs139961185	116807343	*SIK3*	Intronic	GWAS SNP	2.6×10^{-9}	1.1×10^{-7}	1.00	1.00

* Build 37 position on chromosome 11.

† Corrected for multiple testing.

‡ The SNP with the strongest p value in a given round of analysis.

Table S4. The associations between rs147210663, rs964184, 3 *APOA5* SNPs and triglycerides-related traits combining Pima and FIND samples

A. Results for quantitative traits

Trait	SNP	n	β*±s.e	variance explained	p
Ln(triglycerides)	rs147210663	7,297	-0.922 ± 0.059	3.53%	9.7×10^{-55}
	rs964184	7,057	0.208 ± 0.017	2.26%	8.6×10^{-33}
	APOA5 SNPs†	6,487	0.142 ± 0.012	2.41%	1.2×10^{-31}
Total cholesterol	rs147210663	8,413	-0.265 ± 0.058	0.23%	4.7×10^{-6}
	rs964184	8,153	0.101 ± 0.017	0.47%	4.5×10^{-9}
	APOA5 SNPs	7,500	0.065 ± 0.012	0.45%	6.7×10^{-8}
HDL cholesterol	rs147210663	8,414	0.437 ± 0.059	0.67%	9.2×10^{-14}
	rs964184	8,154	-0.085 ± 0.018	0.28%	1.6×10^{-6}
	APOA5 SNPs	7,501	-0.045 ± 0.012	0.18%	3.7×10^{-4}
LDL cholesterol	rs147210663	7,822	-0.089 ± 0.055	0.01%	0.101
	rs964184	7,734	0.052 ± 0 .021	0.15%	0.013
	APOA5 SNPs	7,089	0.030 ± 0.011	0.13%	0.006
VLDL-C‡	rs147210663	206	-1.11 ± 0.240	10.2%	3.9×10^{-6}
	rs964184	206	0.29 ± 0.095	1.67%	0.0022
	APOA5 SNPs	191	0.21 ± 0.069	3.46%	0.0019
VLDL–triglycerides	rs147210663	206	-1.26 ± 0.231	14.5%	5.6×10^{-8}
	rs964184	206	0.42 ± 0.092	7.17%	3.8×10^{-6}
	APOA5 SNPs	191	0.34 ± 0.067	9.57%	5.0×10^{-7}
LDL–triglycerides	rs147210663	206	-0.019 ± 0.0051	7.20%	0.00017
	rs964184	206	0.0071 ± 0.0019	6.04%	3.8×10^{-6}
	APOA5 SNPs	191	0.0060 ± 0.0014	8.43%	1.6×10^{-5}

B. Results for a qualitative trait

Trait	SNP	n	OR§ (95% CI)	p
Hypertriglyceridemia¶	rs147210663	8,039	0.14 (0.09, 0.22)	4.5×10^{-18}
	rs964184	7,945	1.53 (1.42, 1.65)	5.0×10^{-31}
	APOA5 SNPs	7,281	1.32 (1.25, 1.39)	2.0×10^{-27}

* Based on an additive genetic model, for the effect of the minor allele, in SD unit.

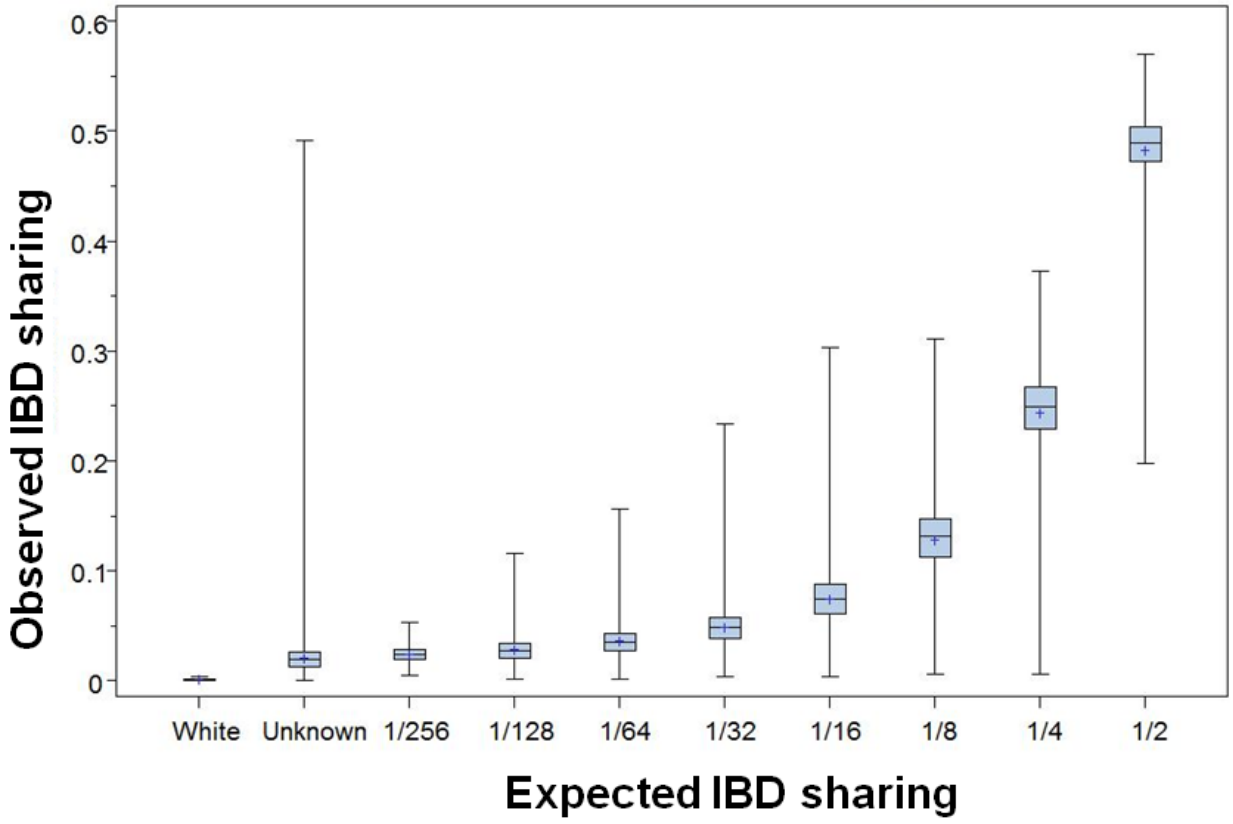
† Effects expressed per copy of any minor allele of these 3 functional SNPs in *APOA5*: rs2266788, rs3135506, rs662799.

‡ Very low density lipoprotein cholesterol.

§ OR = odds ratio.

¶ Hypertriglyceridemia defined as having serum triglyceride levels ≥ 1.69 mmol/L.

Figure S1. Empirically estimated IBD sharing by expected IBD sharing in the sample used for linkage analysis of serum lipids



IBD sharing refers to the percentage of allele shared identical by decent. N = 1,024 subjects, or 523,776 pairs; 9,664 pairs (1.8%) with known relationship. 98.2% pairs had no known relationship. Data are shown as a box plot, with the error bars representing the range, the thicker bars representing the 25th and 75th centiles, the horizontal bar representing the median and the + sign representing the mean. IBD sharing estimates for whites (19 individuals) were calculated with ibdscale = 1 (which was determined to give comparable estimates in 19 Pimas to those obtained in the larger set of 1,024 Pimas).

Figure S2. Genome-wide linkage study results for 4 serum lipid traits

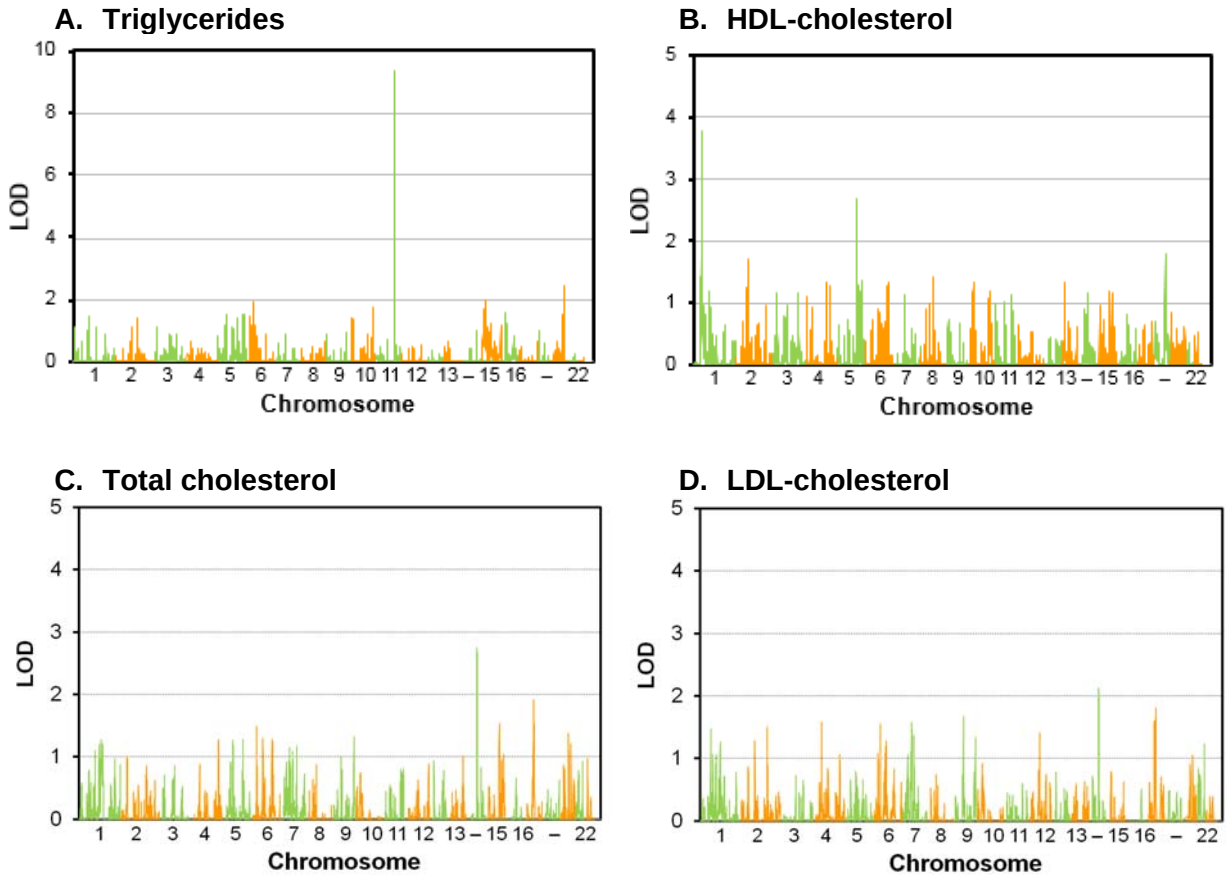
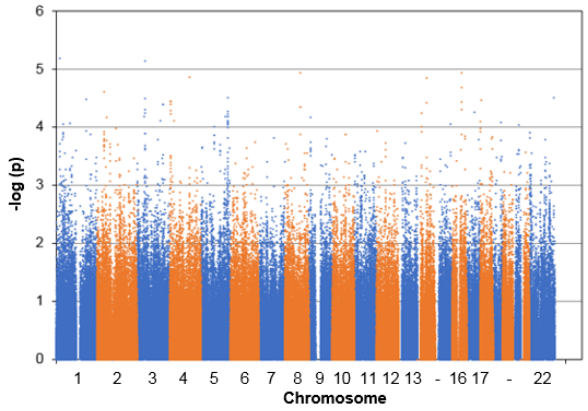
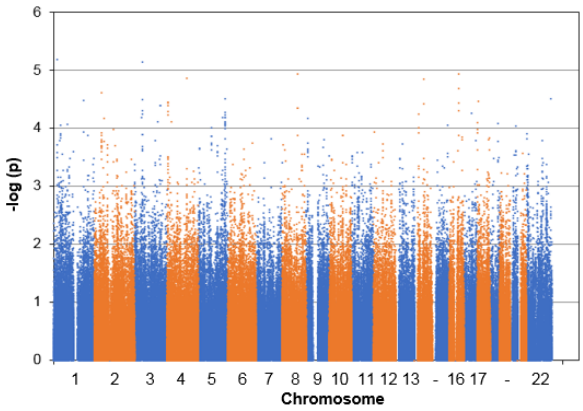


Figure S3. Genome-wide association study results for four serum lipid traits using the Affymetrix 6.0 array

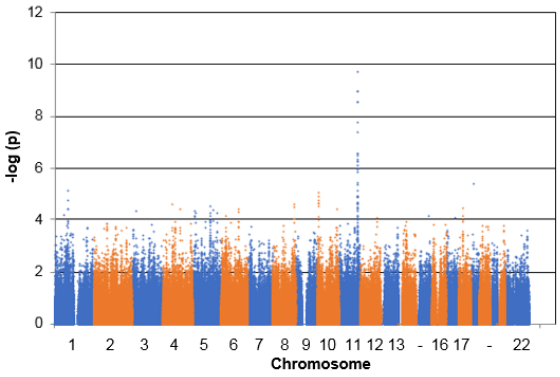
A. Total cholesterol



B. HDL cholesterol



C. Triglycerides



D. LDL cholesterol

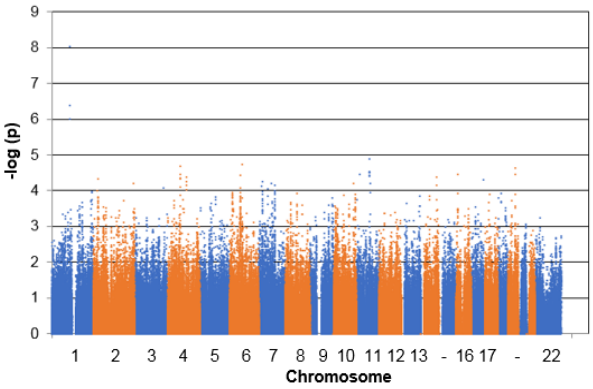


Figure S4. Linkage disequilibrium pattern (r^2) between 6 SNPs in 4,636 Pima samples

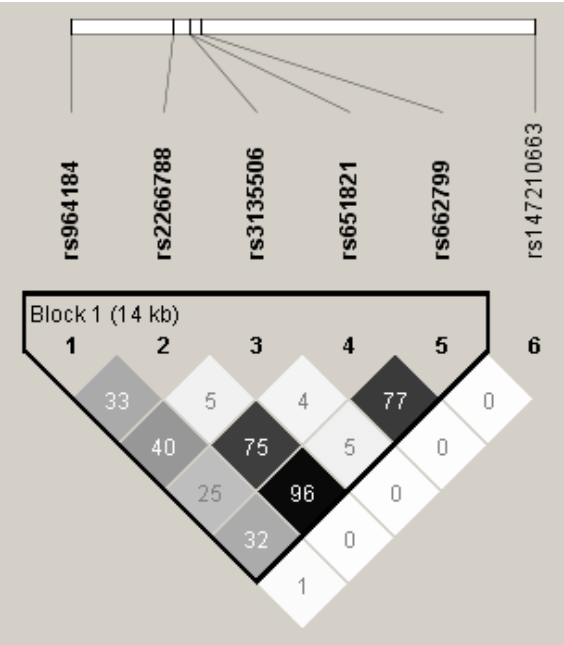
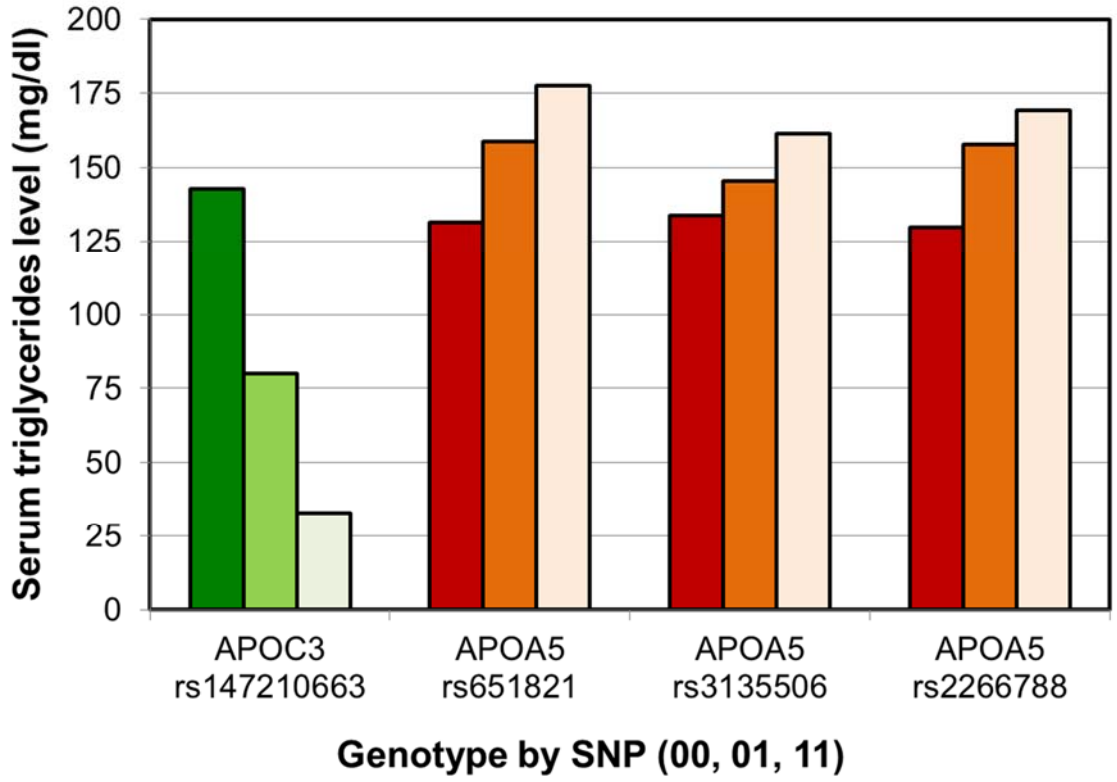


Figure S5. Raw levels of serum triglycerides by genotypes of 4 SNPs in Pima Indians



Bars of a dark color represent mean levels of plasma triglycerides among homozygotes of the major allele (genotype 00); bars of an intermediate color represent mean levels of serum triglycerides among heterozygotes of the major allele (genotype 01); and bars of a light color represent mean levels of serum triglycerides among homozygotes of the minor allele (genotype 11).

Figure S6. rs964184 vs. *APOA5* 3-SNP haplotypes and frequencies

rs964184	rs2266788	rs3135506	rs651821	rs6622799	Frequency
G	T	S	T	A	60.6%
C	T	W	T	A	20.8%
C	C	S	C	G	14.0%
C	C	S	T	G	3.3%

These results were calculated in the Pima sample (n = 4,636).

Supplemental References

1. Looker HC, Krakoff J, Andre V, Kobus K, Nelson RG, Knowler WC, et al. Secular trends in treatment and control of type 2 diabetes in an American Indian population: a 30-year longitudinal study. *Diabetes Care*. 2010;33:2383-9.
2. Imperatore G, Knowler WC, Pettitt DJ, Kobes S, Fuller JH, Bennett PH, et al. A locus influencing total serum cholesterol on chromosome 19p: results from an autosomal genomic scan of serum lipid concentrations in Pima Indians. *Arteriosclerosis, thrombosis, and vascular biology*. 2000;20:2651-6.
3. Friedewald WT, Levy RI and Fredrickson DS. Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. *Clin Chem*. 1972;18:499-502.
4. Fagot-Campagna A, Narayan KM, Hanson RL, Imperatore G, Howard BV, Nelson RG, et al. Plasma lipoproteins and incidence of non-insulin-dependent diabetes mellitus in Pima Indians: protective effect of HDL cholesterol in women. *Atherosclerosis*. 1997;128:113-9.
5. Malhotra A, Kobes S, Knowler WC, Baier LJ, Bogardus C and Hanson RL. A genome-wide association study of BMI in American Indians. *Obesity (Silver Spring)*. 2011;19:2102-6.
6. Browning BL and Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009;84:210-23.
7. Bercovici S, Meek C, Wexler Y and Geiger D. Estimating genome-wide IBD sharing from SNP data via an efficient hidden Markov model of LD with application to gene mapping. *Bioinformatics*. 2010;26:i175-82.
8. Han L and Abney M. Using identity by descent estimation with dense genotype data to detect positive selection. *Eur J Hum Genet*. 2013;21:205-11.

9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559-75.
10. Browning SR and Browning BL. High-resolution detection of identity by descent in unrelated individuals. *Am J Hum Genet*. 2010;86:526-39.
11. Browning BL and Browning SR. Improving the accuracy and efficiency of identity-by-descent detection in population data. *Genetics*. 2013;194:459-71.
12. Amos CI. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am J Hum Genet*. 1994;54:535-543.
13. Hanson RL, Ehm MG, Pettitt DJ, Prochazka M, Thompson DB, Timberlake D, et al. An autosomal genomic scan for loci linked to type II diabetes mellitus and body-mass index in Pima Indians. *Am J Hum Genet*. 1998;63:1130-1138.
14. Pratley RE, Thompson DB, Prochazka M, Baier L, Mott D, Ravussin E, et al. An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. *J Clin Invest*. 1998;101:1757-1764.
15. Hanson RL and Knowler WC. Quantitative trait linkage studies of diabetes-related traits. *Curr Diab Rep*. 2003;3:176-83.
16. Hopper JL and Mathews JD. Extensions to multivariate normal models for pedigree analysis. *Ann Hum Genet*. 1982;46 (4):373-383.
17. Fuchsberger C, Abecasis GR and Hinds DA. minimac2: faster genotype imputation. *Bioinformatics*. 2015;31:782-4.
18. Lathrop GM and Lalouel JM. Easy calculations of lod scores and genetic risks on small computers. *Am J Hum Genet*. 1984;36:460-5.
19. Devlin B, Roeder K and Bacanu SA. Unbiased methods for population-based association studies. *Genet Epidemiol*. 2001;21:273-84.

20. Petitti DB. Decision analysis and cost-effectiveness analysis: methods for quantitative synthesis in medicine. In: D. B. Petitti, ed. *Meta-Analysis* Oxford, UK: Oxford University Press; 2000: 94-118.
21. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ and Altshuler D. Efficiency and power in genetic association studies. *Nat Genet*. 2005;37:1217-23.
22. Moskvina V and Schmidt KM. On multiple-testing correction in genome-wide association studies. *Genet Epidemiol*. 2008;32:567-73.
23. Ott J. *Analysis of human genetic linkage.* Revised ed. Baltmore: Johns Hopkins University Press; 1991.
24. Terwilliger JDaO, J. Linkage disequilibrium between alleles at marker loci *Handbook of Human Genetic Linkage*; 1994.
25. Vozarova de Courten B, Hanson RL, Funahashi T, Lindsay RS, Matsuzawa Y, Tanaka S, et al. Common polymorphisms in the adiponectin gene ACDC are not associated with diabetes in Pima Indians. *Diabetes*. 2005;54:284-9.
26. 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.
27. Hanis CL, Chakraborty R, Ferrell RE and Schull WJ. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. *Am J Phys Anthropol*. 1986;70:433-41.