RESEARCH ARTICLE

Genetic Epidemiology

OFFICIAL JOURNAL
INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY
www.geneticepi.org

Genome-Wide Family-Based Linkage Analysis of Exome Chip Variants and Cardiometabolic Risk

Jacklyn N. Hellwege,[1,2,3] Nicholette D. Palmer,[2,3,4] Laura M. Raffield,[1,2,3] Maggie C.Y. Ng,[2,3] Gregory A. Hawkins,[3] Jirong Long,[5] Carlos Lorenzo,[6] Jill M. Norris,[7] Y.-D. Ida Chen,[8] Elizabeth K. Speliotes,[9,10] Jerome I. Rotter,[8] Carl D. Langefeld,[11] Lynne E. Wagenknecht,[11] and Donald W. Bowden[2,3,4]*

[1]*Molecular Genetics and Genomics Program, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America;* [2]*Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America;* [3]*Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America;* [4]*Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America;* [5]*Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America;* [6]*Department of Medicine, University of Texas Health Science Center, San Antonio, Texas, United States of America;* [7]*Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado, United States of America;* [8]*Institute for Translational Genomics and Population Sciences and Department of Pediatrics, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, United States of America;* [9]*Department of Internal Medicine, Division of Gastroenterology, University of Michigan, Ann Arbor, Michigan, United States of America;* [10]*Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan, United States of America;* [11]*Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America*

Received 22 November 2013; Revised 14 February 2014; accepted revised manuscript 28 February 2014.
Published online 9 April 2014 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/gepi.21801

ABSTRACT: **Linkage analysis of complex traits has had limited success in identifying trait-influencing loci. Recently, coding variants have been implicated as the basis for some biomedical associations. We tested whether coding variants are the basis for linkage peaks of complex traits in 42 African-American (n = 596) and 90 Hispanic (n = 1,414) families in the Insulin Resistance Atherosclerosis Family Study (IRASFS) using Illumina HumanExome Beadchips. A total of 92,157 variants in African Americans (34%) and 81,559 (31%) in Hispanics were polymorphic and tested using two-point linkage and association analyses with 37 cardiometabolic phenotypes. In African Americans 77 LOD scores greater than 3 were observed. The highest LOD score was 4.91 with the *APOE* SNP rs7412 (MAF = 0.13) with plasma apolipoprotein B (ApoB). This SNP was associated with ApoB (P-value = 4×10^{-19}) and accounted for 16.2% of the variance in African Americans. In Hispanic families, 104 LOD scores were greater than 3. The strongest evidence of linkage (LOD = 4.29) was with rs5882 (MAF = 0.46) in *CETP* with HDL. *CETP* variants were strongly associated with HDL ($0.00049 < P$-value $< 4.6 \times 10^{-12}$), accounting for up to 4.5% of the variance. These loci have previously been shown to have effects on the biomedical traits evaluated here. Thus, evidence of strong linkage in this genome wide survey of primarily coding variants was uncommon. Loci with strong evidence of linkage was characterized by large contributions to the variance, and, in these cases, are common variants. Less compelling evidence of linkage and association was observed with additional loci that may require larger family sets to confirm.**
Genet Epidemiol 38:345–352, 2014. © 2014 Wiley Periodicals, Inc.

KEY WORDS: **Hispanic; African American; genetic variance**

Introduction

Family-based linkage analysis has been highly successful in the study of Mendelian disorders. With few exceptions [Hugot et al., 2001; Musunuru et al., 2010; Ogura et al., 2001], linkage analysis has been largely unsuccessful in contributing insights into the origin of complex traits or common diseases. In contrast, Genome-Wide Association Studies (GWAS) have been highly successful in identifying trait or disease associated common genetic variants. However, in most cases, GWAS have identified variants of relatively small effect sizes [Bodmer and Bonilla, 2008; Bowden, 2011] that contribute minimally to the variance in disease or quantitative traits. Investigators have speculated that low frequency variants, especially previously untested coding variants, will have larger effect sizes and contribute meaningfully to the variance in complex traits and common diseases [Kiezun et al., 2012]. Next-generation sequencing has facilitated the generation of extensive exome-sequencing data resources and made it possible to investigate the impact of coding variants.

In a prior exome sequencing analysis we have shown that a low frequency (1% minor allele frequency; MAF) coding

Supporting Information is available in the online issue at wileyonlinelibrary.com.

*Correspondence to: Donald W. Bowden, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA. E-mail: dbowden@wakehealth.edu

Table 1. Demographic summaries of IRASFS Hispanic and African-American cohorts

	IRASFS African Americans				IRASFS Hispanics			
Trait	*N*				*N*			
Samples	596				1,414			
Pedigrees	42				90			
Pedigree size	Mean 14.2				Mean 15.7			
	N	*Mean*	*Range*	*SD*	*N*	*Mean*	*Range*	*SD*
Age (yrs)	573	42.76	18–80	13.98	1,262	42.75	18–81	14.58
Gender	596	60.2% F			1,414	58.3% F		
BMI (kg/m^2)	570	29.96	15–55	6.79	1,252	28.88	16–58	6.13
ApoB (mg/dL)	565	89.74	34–191	24.48	1,152	89.08	33–201	22.69
HDL (mg/dL)	574	47.48	21–103	12.85	1,253	43.82	18–125	13.05
LDL (mg/dL)	571	114.91	31–273	34.05	1,242	109.17	31–218	30.86
Total cholesterol (mg/dL)	574	178.66	99–350	38.62	1,254	177.83	74–348	37.30
Triglycerides (mg/dL)	574	81.59	13–888	70.36	1,251	124.20	18–836	84.38

variant in the *ADIPOQ* gene was the basis for a linkage peak in Hispanic families in the Insulin Resistance Atherosclerosis Family Study (IRASFS) [Bowden et al., 2010]. Here, we have performed an extensive analysis of family-based linkage and combined association analysis of 37 cardiometabolic phenotypes with data from exome chip analysis of 130 Hispanic and African-American families from the IRASFS to test the hypothesis that linkage analysis of coding variants can identify trait defining genetic variations.

Methods

Samples

The samples used in this study include both the Hispanic and African-American cohorts from the Insulin Resistance and Atherosclerosis Family Study (IRASFS). Ascertainment and examination of these cohorts have been previously described in detail [Henkin et al., 2003]. Briefly, subjects were ascertained on the basis of large family size in San Luis Valley, Colorado (Hispanics), San Antonio, Texas (Hispanics), and Los Angeles, California (African Americans). The Hispanic cohort consisted of 1,414 individuals from 90 families, while the African-American sample has 596 individuals from 42 families (Table 1). These subjects were extensively phenotyped, including a frequently sampled intravenous glucose tolerance test (FSIGT), measures of blood lipids and inflammatory markers, anthropomorphic measures, and fat deposition measured by computed tomography (CT) scan (Supplementary Table S1). IRB approval was obtained at all clinical and analysis sites and all participants provided informed consent.

Exome Chip Genotyping and Quality Control

Genotyping was performed at the Wake Forest Center for Genomics and Personalized Medicine Research using the Illumina HumanExome BeadChip v1.0 (596 African Americans and 552 Hispanics) and v1.1 (862 Hispanics). Version 1.0 of the exome chip contained 247,870 variants, while version 1.1 contained 242,901. Over 92% of the variants were in protein coding regions. The chip also included 4761 single nucleotide polymorphisms (SNPs) previously associated with a diverse range of traits in prior GWAS (summary of exome chip design: http://genome.sph.umich.edu/wiki/Exome_Chip_Design). Additional markers included coverage of the HLA region, ancestry informative markers, and SNPs for identity by descent determination. Relevant to the study reported here, the variants on the exome chips were chosen based on their appearance in more than one individual and more than one cohort and thus are likely to be observed in the general population.

Sample and autosomal SNP call rates for both the African-American and Hispanic datasets were ≥99%, and SNPs with poor cluster separation (<0.35) were excluded. Mendelian errors were identified using PedCheck [O'Connell and Weeks, 1998] and resolved by removing conflicting genotypes. Hardy–Weinberg Equilibrium (HWE) was assessed for the variants on the exome chip using the program PLINK [Purcell et al., 2007] in only the unrelated samples (N = 39 and 229, African Americans and Hispanics, respectively), in order to reduce biases introduced by familial allele frequencies. In Hispanics, 7 variants had significant ($P < 1 \times 10^{-5}$) deviation from HWE (minimum P-value = 7.2×10^{-20}). African-American variants were all in accordance with HWE ($P > 1 \times 10^{-5}$).

Individual SNP Genotyping and Quality Control

Targeted genotyping of variants in *APOE* (chromosome 19), the neighboring gene *APOC1*, and *APOB* (chromosome 2) was also performed, including *APOE* variant rs429358, which in combination with rs7412 (contained on the exome chip) defined *APOE* isoforms $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$. In addition to rs429358, SNPs were chosen from the NHLBI ESP GO (https://esp.gs.washington.edu/drupal/) [Tennessen et al., 2012] and 1000 Genomes [Abecasis et al., 2012] databases to more thoroughly cover the *APOE/APOC1* and *APOB* regions. Coding variants seen in at least two African-American samples from these resources were prioritized for

genotyping, as well as variants predicted to impact splicing or causing frame-shift changes. Genotyping of 43 SNPs was carried out using the Sequenom Mass Array system (Sequenom; San Diego, CA). Blind duplicates (n = 22) were genotyped for quality control purposes and had a concordance rate of 99.6%. SNP and plate call rate thresholds were set at >90%. PedCheck [O'Connell and Weeks, 1998] was used to identify Mendelian inconsistencies, which were resolved by removal of conflicting genotypes. These SNPs were combined with the *APOE-APOC1* and *APOB* variants contained on the exome chip for further analysis.

Statistical Analysis

Marker-specific identity-by-descent (IBD) probabilities were computed using the Monte Carlo method implemented in Sequential Oligogenic Linkage Analysis Routines (SOLAR) [Almasy and Blangero, 1998]. Trait variables were transformed to approximate normal distributions when necessary. Two-point linkage and measured genotype association analyses were also performed using the variance components method implemented in SOLAR, with age, gender, recruitment center (in the Hispanic cohorts), ancestry proportions (one principal component for African Americans, ADMIXTURE [Alexander et al., 2009] variables 2–4 for Hispanics), and BMI as covariates. The measured genotype analysis, which accounts for the nonindependence of family members, involves incorporating each variant separately in a model as a measured covariate (the number of copies of the minor allele) evaluating genotype-specific differences in the trait means. The proportion of variance among the mean trait values for the genotypes to the total phenotypic variance was estimated as the contribution of the locus to variation in the trait.

Admixture in the Hispanic sample was estimated using an ADMIXTURE [Alexander et al., 2009] analysis from genome-wide SNP data (N = 1,034 samples) pruned for LD from the Illumina OmniExpress as part of the Genetic and Epidemiologic Predictors of Glucose Homeostasis Measures in Hispanics (GUARDIAN) Consortium [data not shown]. All HapMap II populations were included (CEU, YRI, CHB, JPT) using a five population model. Admixture variables 2, 3, 4 represented variance within the IRASFS cohort while 1 and 5 did not vary within our sample. Admixture in the African-American sample was estimated using principal components from 39 ancestry informative markers (AIMs) and including HapMap CEU and YRI samples for comparison [Palmer et al., 2010]. BMI was omitted as a covariate when assessing BMI for both association and linkage analyses, but was included for all other measures of adiposity and body composition.

Phasing of the *APOE* haplotypes was carried out manually based on the assumption that none of the individuals carried the extremely rare ε3r isoform, in which case the SNPs are completely predictive and phasing can be inferred.

Allelic interaction between all *ApoB* SNPs and the ApoE ε2 isoform (coded 0, 1, or 2 for number of copies of the ε2 isoform) was assessed using PLINK [Purcell et al., 2007]. This analysis was exploratory in nature and did not account for family structure.

Results

The goal of this study was to investigate the ability of family-based linkage analysis to identify coding variants linked to complex traits. Two-point linkage was chosen as the primary family-based linkage analysis tool. Multipoint linkage of variants in close proximity and consequent moderate to high linkage disequilibrium artificially inflates evidence of cosegregation. In the simplest scenario, a trait-defining variant will show evidence of linkage in the simple two-point analysis approach. The combination of two-point linkage with exome chip variants provides a comprehensive genome-wide assessment of coding and other SNPs observed across these American populations, i.e., are not private mutations. Specifically, we expected to observe evidence of linkage at a SNP that directly influences the cardiometabolic trait. Given the design of the exome chip was centered on coding variants, this SNP would likely have an annotation reflecting the role of the variant.

Two-point linkage analysis was performed using SOLAR on a total of 132 families with average family size of 14.2 (African American) and 15.7 (Hispanic) (Table 1). Table 1 also summarizes sample characteristics relevant to this analysis (demographic, biometric, and lipid levels). Overall the sample was representative of a population cross-section of adults in these populations.

African American Two-Point Linkage

A total of 92,157 (34%) variants in African Americans, were polymorphic and passed quality control metrics, including Mendelian error checking. Of these, 84% (77,526) are in coding regions, leaving 14,623 SNPs from intergenic regions (Table 2). Assessment of HWE in this sample set indicated that no variants deviated from the expected ($P < 1.0 \times 10^{-5}$). The two-point linkage analysis of the exome chip variants with 37 cardiometabolic phenotypes (Supplementary Table S1) yielded a total of nearly 3.5 million LOD scores computed in African Americans. To summarize the overall results (Table 2), there were a total of 19,210 LOD scores greater than 1, 1,127 greater than 2, and 77 greater

Table 2. Descriptive summary of exome chip content broken down by LOD score in IRASFS African Americans and Hispanics

African American	Exome Chip	LOD>1	LOD>2	LOD>3
N	92,157	19,210	1,127	77
Genic	77,526	14,745	832	57
Missense	–	10,422	569	29
Nonsense	–	61	5	0
Synonymous	–	645	33	3
Hispanic	**Exome Chip**	**LOD>1**	**LOD>2**	**LOD>3**
N	81,559	18,840	1,216	104
Genic	66,920	14,309	891	67
Missense	–	9,805	574	40
Nonsense	–	83	7	0
Synonymous	–	650	60	4

Table 3. *APOE* locus linkage analysis and association results with ApoB levels from exome chip and follow-up single SNP genotyping analysis in African Americans

SNP	Position[a]	Gene	Annotation	MAF[b]	LOD	*P*-value	β	Variance explained
rs1081101*	45408076		5′	0.10	0.00	0.31	0.18	0.001
rs405509*	45408838		5′ Promoter	0.29	0.67	**0.0004**	0.31	0.039
rs440446*	45409167	*APOE*	Intron	0.22	**1.53**	0.20	0.13	0.010
19:45409946*	45409946	*APOE*	Intron	0.013	0.044	**0.045**	–0.72	0.008
rs769449	45410002	*APOE*	Synonymous	0.018	0.66	0.054	0.46	0.009
rs769451*	45410910	*APOE*	Intron	0.007[c]	0.00	0.80	–0.092	0.000
rs11083750*	45411857	*APOE*	P102R	0.025	0.00	0.57	–0.17	0.001
rs429358*	45411941	*APOE*	R130C	0.21	**2.54**	**5.4E-06**	0.43	0.046
rs769455	45412040	*APOE*	R163C	0.018	0.00	0.50	–0.25	0.001
rs121918394*	45412043	*APOE*	K164Q	0.002[c]	0.094	0.33	–1.1	0.000
rs7412	45412079	*APOE*	R176C	0.12	**4.91**	**4.4E-19**	–1.1	0.162
rs439401	45414451		Intergenic	0.17	**1.26**	0.71	0.036	0.002
rs445925	45415640		Intergenic	0.28	**3.58**	**0.020**	–0.2	0.018
rs72654452*	45418086	*APOC1*	Intron	0.009	0.43	**0.0088**	–1.1	0.010
rs72654456	45419555	*APOC1*	K56R	0.009	**1.78**	**1.7E-06**	–2.1	0.044
19:45419605*	45419605	*APOC1*	Intron	0.002[c]	0.00	0.59	–0.58	0.001
rs4420638	45422946		3′	0.17	**3.34**	0.22	0.11	0.004
ApoE Isoform				*0.099*	***4.44***	***6.6E-13***	*0.32*	***0.236***

*Indicates SNPs that were genotyped using the Sequenom platform.
[a] Positions determined based on hg19.
[b] Minor allele frequencies were computed using a subset of unrelated individuals.
[c] Unless indicated.
Bold indicates *P*-values less than 0.05 and LOD scores greater than 1. Shading indicates the two SNPs that together determine the ApoE isoform (bottom, italicized). Haplotype frequencies for the ApoE isoform are as follows: ε2 0.099, ε3 0.677, ε4 0.224.

than 3, with a maximum identified LOD score of 4.91 with rs7412, a SNP at the *APOE* locus. In addition to linkage analysis, conventional SNP association analysis was performed accounting for family structure using SOLAR. A summary of LOD scores greater than 3.0 is shown in Supplementary Table S2 that lists both two-point LOD score and additive *P*-value for association. With the exception of rs7412 in the *APOE* locus (*P*-value = 4.4×10^{-19}), *P*-values for association in these linked variants are either nonsignificant ($P > 0.05$), or, at best, suggest nominal ($P > 0.0001$) association.

African Americans: Apolipoprotein B Levels and *APOE*

The maximum LOD score in African Americans was with the *APOE* SNP rs7412 (MAF = 0.12) and Apolipoprotein B (ApoB) levels. ApoB is the primary protein component of low-density lipoprotein particles. The rs7412 variant had a *P*-value for association with ApoB of 4.4×10^{-19} (Table 3). The estimated effect corresponds to a reduction of ≈22 mg/dL in ApoB levels per minor allele of rs7412. The estimated proportion of variance of ApoB levels as attributable to rs7412 was 16.2%. This coding variant represents an arginine to cysteine amino acid change, which also is one of the two components defining the much-studied ApoE isoforms. In particular, this SNP is indicative of the ε2 isoform, which has often been implicated as protective of cardiovascular changes, as well as Alzheimer's disease.

African Americans: Additional *APOE*, *APOC1*, and *APOB* Genotyping

Additional *APOE* and *APOC1* variants (absent from the exome chip) including rs429358, the other SNP determinant of the canonical ApoE isoforms, were also genotyped and data analyzed as summarized in Table 3. No other genotyped SNP showed stronger evidence of linkage or association than rs7412, including rs429538 (MAF = 0.21). There was very little linkage disequilibrium across the *APOE* region, with only one variant having an r^2 value of at least 0.40 with one of the isoform-determining SNPs (rs7412 and rs445925; Supplementary Fig. S1). Combining rs7412 and rs429358 into *APOE* haplotypes resulted in LOD = 4.40 and *P*-value for association of 6.6×10^{-13} accounting for an estimated 23.6% of the variance in ApoB (Table 3). Figure 1 summarizes the means of each haplogenotype for ApoE isoforms. Haplogenotypic means were 51.4 mg/dL for homozygous ε2, 89.2 mg/dL for homozygous ε3, and 102.2 mg/dL for homozygous ε4. The figure shows the clear trend for increasing numbers of ε2 alleles and decreasing ApoB. The statistically significant linkage and association (LOD>3; *P*-value $< 5 \times 10^{-7}$) results for the *APOE-APOC1* locus were driven largely by the *APOE* haplotype based on analyses conditioning on *APOE* (Supplementary Table S3). In addition, since *APOE* and ApoB proteins are both components of lipoprotein particles, allelic interaction between *APOE* ε2 isoform and *APOB* variants were assessed for their contribution to the evidence of linkage and association. Although some *APOB* variants are associated with ApoB levels, (Supplementary Table S4) they explain little variance in ApoB levels and there was limited evidence of *APOB* × *APOE* interaction (Supplementary Table S5),

Twopoint Linkage Analysis in Hispanic Americans

The Hispanic set resulted in 81,559 (31%) polymorphic SNPs passing quality control (Table 2). Coding SNPs represented 82% of the polymorphic SNPs, with 14,639 additional intergenic SNPs remaining for analysis. In this sample set seven SNPs deviated from expected HWE

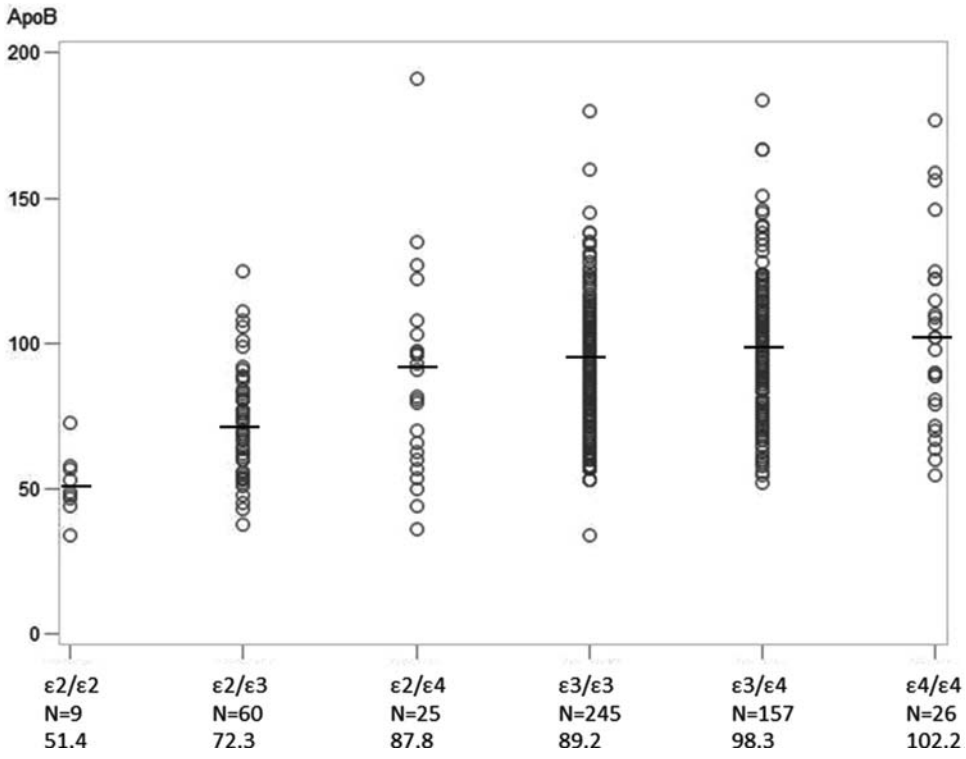

Figure 1. Haplotypic means of ApoB levels (mg/dL) by ApoE isoform status in IRASFS African Americans. Number of samples and mean ApoB values (also noted by the horizontal line) are indicated below each isoform type.

Equilibrium ($P < 1.0 \times 10^{-5}$). The linkage results for the Hispanic sample included 18,840 LOD scores greater than 1, 1,216 greater than 2 and 104 greater than 3 (Table 2). A previously identified linkage peak on chromosome 3 for adiponectin [Bowden et al., 2010] was overrepresented in the strong linkage results, which are presented in Supplementary Table S6, alongside corresponding *P*-values for association for those variants. Excluding the chromosome 3 locus, the most biologically interesting result was evidence of linkage with HDL levels on chromosome 16 with a variant in the *CETP* gene (rs5882, LOD = 4.29). The CETP protein is known to facilitate exchange of cholesterol esters for triglycerides between HDL and triglyceride-rich lipoprotein remnants.

Hispanic Americans: *CETP*

In addition to rs5882 there were 11 additional variants in and near *CETP* on the exome beadchip. The result of linkage and association with these variants are summarized in Table 4. Only the initial exome chip variant, rs5882, showed strong evidence of linkage with the next highest LOD score being less than 2. The *CETP* locus has extensive evidence of association with HDL with the strongest association (P-value = 4.6×10^{-12}) seen with a promoter SNP rs247616. Multiple other SNPs were also strongly associated, with a range of *P*-values of less than 10^{-6}. In contrast to the *APOE* results, the most strongly linked SNP rs5882 was more nominally associated with HDL levels ($P = 0.00049$). The most significantly associated SNP in this gene region (rs247616) accounted for 4.5% of the variance in HDL levels in this population while the linked SNP rs5882 explained an estimated 1.2% (Table 4).

To gain a better picture of the complex pattern of linkage and association with HDL in the *CETP* locus analyses conditioning on rs5882 (the strongest linkage) and rs247616 (the most associated) were performed (Table 5). Conditioning on rs5882 reduced, but did not eliminate evidence of linkage in SNPs with prior nominal evidence of linkage (e.g., rs9939224) suggesting rs5882 alone does not explain the linkage evidence. Conditional analysis of the association data, however, suggested that rs247616 (or a SNP in high LD) largely accounted for evidence of association with *P*-values substantially reduced and only modest evidence of residual variance. It is notable that rs247616 is part of an LD block (Supplementary Fig. S2). Although there was substantial LD across the promoter region, there was less between the coding regions of the gene (in particular, the linked SNP rs5882) and the promoter and upstream areas of significant association (P-value $< 6 \times 10^{-7}$). The genotypic means of rs247616 and rs5882 follow an additive pattern and are presented in Supplementary Table S7.

Table 4. Analysis of exome chip results at the CETP region in Hispanics

SNP	Position	Gene	Annotation	MAF[a]	LOD	*P*-value	β	Variance explained
rs9989419	56985139		5′	0.32	**1.35**	**5.8E-04**	–0.043	0.013
rs173539	56988044		5′	0.27	**1.78**	**3.4E-10**	0.081	0.039
rs247616	56989590		5′ Promoter	0.27	**1.71**	**4.6E-12**	0.09	0.045
rs3764261	56993324		5′ Promoter	0.27	**1.69**	**6.1E-12**	0.09	0.044
rs1800775	56995236		5′ Promoter	0.49	0.98	**1.1E-05**	–0.053	0.023
rs9939224	57002732	*CETP*	Intron	0.22	**1.53**	**1.1E-05**	–0.064	0.019
rs34716057	57003846	*CETP*	R154W	0.0022	0.22	0.38	0.078	0.004
rs1532624	57005479	*CETP*	Intron	0.35	**1.79**	**1.0E-07**	0.064	0.025
rs201438792	57005908	*CETP*	S221R	0.0022	0.00	0.78	0.047	0.000
rs7499892	57006590	*CETP*	Intron	0.19	**1.18**	**7.0E-06**	–0.068	0.018
rs5880	57015091	*CETP*	A390P	0.077	0.07	**0.0044**	–0.064	0.006
rs5882	57016092	*CETP*	V422I	0.46	**4.29**	**4.9E-04**	0.042	0.012

[a] MAF as computed using a subset of unrelated individuals.
Boldface text indicates *P*-values less than 0.05 as well as LOD scores greater than 1. Position is based on hg19.

Table 5. Conditional analyses of rs5882 and rs247616 CETP variants on HDL levels

		Unconditioned		Conditioning on rs5882				Conditioning on rs247616			
SNP	Position	LOD	*P*-value	LOD	*P*-value	β	Variance explained	LOD	*P*-value	β	Variance explained
rs9989419	56985139	**1.35**	**5.8E-04**	0.81	**0.015**	–0.032	0.007	0.11	0.45	–0.010	0.0015
rs173539	56988044	**1.78**	**3.4E-10**	**1.11**	**9.2E-09**	0.076	0.033	0.02	**0.021**	–0.144	0.0014
rs247616	56989590	**1.71**	**4.6E-12**	**1.01**	**1.6E-10**	0.085	0.039				
rs3764261	56993324	**1.69**	**6.1E-12**	0.99	**2.3E-10**	0.085	0.038	0.00	0.91	0.012	2.5E-05
rs1800775	56995236	0.98	**1.1E-05**	0.55	**0.0024**	0.043	0.012	0.05	0.55	0.009	0.0012
rs9939224	57002732	**1.53**	**1.1E-05**	**1.29**	**9.4E-04**	–0.052	0.011	**1.25**	**0.014**	–0.037	0.006
rs34716057	57003846	0.22	0.38	0.29	0.22	0.109	0.005	0.31	0.35	0.082	0.004
rs1532624	57005479	**1.79**	**1.0E-07**	**1.06**	**1.0E-05**	0.057	0.017	0.04	0.83	0.004	1.1E-06
rs201438792	57005908	0.00	0.78	0.00	0.96	0.0085	2.5E-06	0.00	0.68	0.066	0.000
rs7499892	57006590	**1.18**	**7.0E-06**	0.98	**6.1E-04**	–0.056	0.010	0.79	**0.0069**	–0.042	0.0050
rs5880	57015091	0.07	**0.0044**	0.00	**0.046**	–0.047	0.003	0.10	**0.014**	–0.054	0.004
rs5882	57016092	**4.29**	**4.9E-04**					**1.90**	**0.037**	0.025	0.006

Position as determined from hg19. Boldface text indicates *P*-values less than 0.05 as well as LOD scores greater than 1.

Discussion

In this study, we have performed analysis of family-based linkage and association analysis of exome chip data from 130 Hispanic-American and African-American families from the IRASFS with 37 cardiometabolic phenotypes. The goal of the study was to test the hypothesis that linkage analysis of coding variants can identify trait defining genetic variations. Based on the exome chip design, this analysis tested primarily coding variants that were distributed in the general population. It is noteworthy that it has been argued previously that linkage is only effective when the variant or locus of interest is contributing a large proportion of the variance of the disease or trait [Risch and Merikangas, 1996; Risch, 2000]. The results here, especially the two highlighted linkages with *APOE* and *CETP* are consistent with this model.

APOE and *CETP* had the strongest evidence of linkage. SNPs at these loci were strongly associated with the linked traits (*P*-values for the top associated SNPs of 4.4×10^{-19} for the *APOE* locus and 4.6×10^{-12} at the *CETP* locus; Tables 3 and 4). Association *P*-values for other SNPs on the chip with LODs greater than 3 were more nominal (*P*-value >0.0001). Given the large number of tests, we suspect that many of the other LOD scores greater than 3 may not represent true linkages. The two most prominent linkage results were highly consistent with prior genetic studies [Anuurad et al., 2009; Chasman et al., 2008; Edmondson et al., 2011; Lu et al., 2008; Ozturk et al., 2010] and were the focus of more detailed analysis.

The *APOE* SNP rs7412 in African Americans provided the most striking linkage and association results with the ApoB phenotype. The association with ApoB appeared to be driven by the *APOE* locus rather than the influence of the ApoB coding locus (*APOB)* (Supplementary Tables S4 and S5). *APOE* has been an intensively studied genetic locus from multiple perspectives including a well-documented association with lipid levels such as LDL cholesterol [Anand et al., 2009; Chasman et al., 2006, 2008, 2012; Coram et al., 2013; Klos et al., 2008; Teslovich et al., 2010; Waterworth et al., 2010]. However, association with ApoB levels has been noted only in a small number of studies with SNPs in the region, and the isoforms have only had moderate study in this context [Chasman et al., 2008, 2012; Ozturk et al., 2010; Surakka et al., 2012]. Particularly, Anuurad et al. showed a marked difference in ApoB levels in African Americans, but not Caucasians by ApoE isoform in 232 African Americans and 326 Caucasians; $P < 0.05$ [Anuurad et al., 2009]. In addition to ApoB we have assessed the linkage and association with other lipid traits (Supplementary Table S8), and found association with LDL ($P = 7.5 \times 10^{-10}$), total cholesterol ($P = 1.4 \times 10^{-6}$) and modest evidence of linkage and no compelling association with HDL and triglycerides.

In this African-American sample, the *APOE* isoforms explain over 23% of the variance in ApoB levels (Table 3) that contrasts with an estimated 13% of the variance for LDL (Supplementary Table S8). *APOE* was not linked to ApoB in the larger Hispanic cohort (rs7412; LOD = 0.15) but was associated (rs7412; P = 5.7E-06; Supplementary Table S9). rs7412 and *APOE* isoforms explained a comparatively modest amount of the variance of ApoB and LDL (rs7412, 3.3% and 3.2% respectively; *APOE*, 2.8% and 2.8%, respectively). Prior estimates of *APOE* contributions to LDL variance range from 3.3% to 4.1% in non-African derived populations (24–28).

The association of the *CETP* locus with HDL levels, seen here in the Hispanic sample, is well documented. CETP facilitates exchange of cholesterol esters for triglycerides between HDL and triglyceride-rich lipoprotein remnants. *CETP* has previously been associated with HDL levels in numerous studies, albeit with different index SNPs [Chasman et al., 2008; Edmondson et al., 2011; Lu et al., 2008; Papp et al., 2012; Reddy et al., 2012; Teslovich et al., 2010; Waterworth et al., 2010; Wu et al., 2013] in which collections of various SNPs at the *CETP* locus in various ethnic groups were estimated to account for approximately 3–4% of the variance in HDL. Contribution to variance however has not been analyzed previously in a Hispanic sample, though our proportion of variance attributable to the *CETP* locus (~5.5%) is comparable to those seen in other ethnic groups.

Results of analysis with the *CETP* locus in African Americans were qualitatively different from the Hispanics (Supplementary Table S10). In African Americans, only 4 of the 15 polymorphic variants in this region were nominally associated (P-values ranging from 0.001 to 0.05). The strongest association (P-value = 0.0010) here was with an intronic *CETP* variant, rs7499892, which explained 2.3% of the HDL variance in this African-American cohort with no evidence of linkage to HDL. This variant was monomorphic in the Hispanic sample, and was therefore not available for linkage and association testing.

We focused here on two previously well-documented loci, *APOE* and *CETP*, to assess performance of family-based linkage in an exome-wide analysis. There were, however, 68 other LOD scores greater than 1 and 11 LOD scores greater than 2 at SNPs that also have association P-values down to 1×10^{-5}. These loci bear further evaluation. A number of the nominal linkages (e.g., LOD > 3) were also nominally associated ($0.0001 < P < 0.05$; Supplementary Tables S2 and S4). Testing of whether these more nominal linkages represent true effects will require additional data from other family studies. In addition, this study was based on exome chip genotyping and therefore does not touch on the possibility that private mutations, i.e., coding variants in a single lineage, may contribute to overall variance in traits. This will be possible to test when the planned exome sequencing of the IRASFS Hispanic families is completed. Another path of examination in these family-based cohorts would be the examination of single families for rare variants with high impact on a trait. While possibly productive, the number of computed LOD scores for 132 families with >20,000 SNPs suggests it will be challenging to identify real linkage in a background of a potentially very high proportion of false positives.

The results of this study emphasize the utility of exome chips and linkage analysis in detecting relevant trait-associated variants. It is also of note that the cohorts examined were small (n = 600–1400) relative to extremely large GWAS. Despite this limitation, the family-based design of these cohorts allowed for the use of linkage analysis in conjunction with association, which was successful in identifying high-impact variants. The *APOE* and *CETP* results add to the prior identification of a high impact variant in the *ADIPOQ* locus [Bowden et al., 2010]. Thus, linkage and association were observed in multiple scenarios: an uncommon variant (1% MAF, *ADIPOQ*), a relatively common variant (12% MAF, *APOE*), and a combination of common variants (*CETP*). The characteristic that these have in common is a substantial contribution to the variance in the trait. In summary, linkage in families remains a powerful tool for discovery in human genetics, especially when combined with conventional association analysis.

Acknowledgments

This work was supported by the grant R01 HG007112 from the National Human Genome Research Institute (D.W.B. and C.D.L.). A subset of the IRASFS exome chips were contributed with funds from the Department of Internal Medicine at the University of Michigan (E.K.S.). Computational support was provided by the Center for Public Health Genomics at Wake Forest School of Medicine.

References

Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA, Consortium GP. 2012. An integrated map of genetic variation from 1,092 human genomes. *Nature* 491(7422):56–65.

Alexander DH, Novembre J, Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 19(9):1655–1664.

Almasy L, Blangero J. 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 62(5):1198–1211.

Anand SS, Xie C, Paré G, Montpetit A, Rangarajan S, McQueen MJ, Cordell HJ, Keavney B, Yusuf S, Hudson TJ and others. 2009. Genetic variants associated with myocardial infarction risk factors in over 8000 individuals from five ethnic groups: the INTERHEART genetics study. *Circ Cardiovasc Genet* 2(1):16–25.

Anuurad E, Yamasaki M, Shachter N, Pearson TA, Berglund L. 2009. ApoE and ApoC-I polymorphisms: association of genotype with cardiovascular disease phenotype in African Americans. *J Lipid Res* 50(7):1472–1478.

Bodmer W, Bonilla C. 2008. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40(6):695–701.

Bowden DW. 2011. Will family studies return to prominence in human genetics and genomics? Rare variants and linkage analysis of complex traits. *Genes Genomics* 33(1):1–8.

Bowden DW, An SS, Palmer ND, Brown WM, Norris JM, Haffner SM, Hawkins GA, Guo X, Rotter JI, Chen YD and others. 2010. Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Hum Mol Genet* 19(20):4112–4120.

Chasman DI, Giulianini F, MacFadyen J, Barratt BJ, Nyberg F, Ridker PM. 2012. Genetic determinants of statin-induced low-density lipoprotein cholesterol reduction: the justification for the use of statins in prevention: an intervention trial evaluating rosuvastatin (JUPITER) trial. *Circ Cardiovasc Genet* 5(2):257–264.

Chasman DI, Kozlowski P, Zee RY, Kwiatkowski DJ, Ridker PM. 2006. Qualitative and quantitative effects of APOE genetic variation on plasma C-reactive protein, LDL-cholesterol, and apoE protein. *Genes Immun* 7(3):211–219.

Chasman DI, Paré G, Zee RY, Parker AN, Cook NR, Buring JE, Kwiatkowski DJ, Rose LM, Smith JD, Williams PT and others. 2008. Genetic loci associated with plasma concentration of low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, triglycerides, apolipoprotein A1, and Apolipoprotein B among 6382 white women in genome-wide analysis with replication. *Circ Cardiovasc Genet* 1(1):21–30.

Coram MA, Duan Q, Hoffmann TJ, Thornton T, Knowles JW, Johnson NA, Ochs-Balcom HM, Donlon TA, Martin LW, Eaton CB and others. 2013. Genome-wide characterization of shared and distinct genetic components that influence blood lipid levels in ethnically diverse human populations. *Am J Hum Genet* 92(6):904–916.

Edmondson AC, Braund PS, Stylianou IM, Khera AV, Nelson CP, Wolfe ML, Derohannessian SL, Keating BJ, Qu L, He J and others. 2011. Dense genotyping of candidate gene loci identifies variants associated with high-density lipoprotein cholesterol. *Circ Cardiovasc Genet* 4(2):145–155.

Henkin L, Bergman RN, Bowden DW, Ellsworth DL, Haffner SM, Langefeld CD, Mitchell BD, Norris JM, Rewers M, Saad MF and others. 2003. Genetic epidemiology of insulin resistance and visceral adiposity. The IRAS family study design and methods. *Ann Epidemiol* 13(4):211–217.

Hugot JP, Chamaillard M, Zouali H, Lesage S, Cézard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M and others. 2001. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 411(6837):599–603.

Kiezun A, Garimella K, Do R, Stitziel NO, Neale BM, McLaren PJ, Gupta N, Sklar P, Sullivan PF, Moran JL and others. 2012. Exome sequencing and the genetic basis of complex traits. *Nat Genet* 44(6):623–630.

Klos K, Shimmin L, Ballantyne C, Boerwinkle E, Clark A, Coresh J, Hanis C, Liu K, Sayre S, Hixson J. 2008. APOE/C1/C4/C2 hepatic control region polymorphism influences plasma apoE and LDL cholesterol levels. *Hum Mol Genet* 17(13):2039–2046.

Lu Y, Dollé ME, Imholz S, van 't Slot R, Verschuren WM, Wijmenga C, Feskens EJ, Boer JM. 2008. Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. *J Lipid Res* 49(12):2582–2589.

Musunuru K, Pirruccello JP, Do R, Peloso GM, Guiducci C, Sougnez C, Garimella KV, Fisher S, Abreu J, Barry AJ and others. 2010. Exome sequencing, ANGPTL3 mutations, and familial combined hypolipidemia. *N Engl J Med* 363(23):2220–2227.

O'Connell JR, Weeks DE. 1998. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63(1):259–266.

Ogura Y, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH and others. 2001. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 411(6837):603–606.

Ozturk Z, Enkhmaa B, Shachter NS, Berglund L, Anuurad E. 2010. Integrated role of two apoliprotein E polymorphisms on apolipoprotein B levels and coronary artery disease in a biethnic population. *Metab Syndr Relat Disord* 8(6):531–538.

Palmer ND, Mychaleckyj JC, Langefeld CD, Ziegler JT, Williams AH, Bryer-Ash M, Bowden DW. 2010. Evaluation of DLG2 as a positional candidate for disposition index in African-Americans from the IRAS Family Study. *Diabetes Res Clin Pract* 87(1):69–76.

Papp AC, Pinsonneault JK, Wang D, Newman LC, Gong Y, Johnson JA, Pepine CJ, Kumari M, Hingorani AD, Talmud PJ and others. 2012. Cholesteryl ester transfer protein (CETP) polymorphisms affect mRNA splicing, HDL levels, and sex-dependent cardiovascular risk. *PLoS One* 7(3):e31930.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ and others. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3):559–575.

Reddy MV, Iatan I, Weissglas-Volkov D, Nikkola E, Haas BE, Juvonen M, Ruel I, Ruel MJ, Sinsheimer JS, Genest J and others. 2012. Exome sequencing identifies 2 rare variants for low high-density lipoprotein cholesterol in an extended family. *Circ Cardiovasc Genet* 5(5):538–546.

Risch N, Merikangas K. 1996. The future of genetic studies of complex human diseases. *Science* 273(5281):1516–1517.

Risch NJ. 2000. Searching for genetic determinants in the new millennium. *Nature* 405(6788):847–856.

Surakka I, Whitfield JB, Perola M, Visscher PM, Montgomery GW, Falchi M, Willemsen G, de Geus EJ, Magnusson PK, Christensen K and others. 2012. A genome-wide association study of monozygotic twin-pairs suggests a locus related to variability of serum high-density lipoprotein cholesterol. *Twin Res Hum Genet* 15(6):691–699.

Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, McGee S, Do R, Liu X, Jun G and others. 2012. Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* 337(6090):64–69.

Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, Pirruccello JP, Ripatti S, Chasman DI, Willer CJ and others. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* 466(7307):707–713.

Waterworth DM, Ricketts SL, Song K, Chen L, Zhao JH, Ripatti S, Aulchenko YS, Zhang W, Yuan X, Lim N and others. 2010. Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler Thromb Vasc Biol* 30(11):2264–2276.

Wu Y, Waite LL, Jackson AU, Sheu WH, Buyske S, Absher D, Arnett DK, Boerwinkle E, Bonnycastle LL, Carty CL and others. 2013. Trans-ethnic fine-mapping of lipid Loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. *PLoS Genet* 9(3):e1003379.