Genome Scan for Quantitative Trait Loci Linked to High-Density Lipoprotein Cholesterol
The NHLBI Family Heart Study

James M. Peacock, Donna K. Arnett, Larry D. Atwood, Richard H. Myers, Hilary Coon, Stephen S. Rich, Michael A. Province, Gerardo Heiss, on behalf of the Investigators of the NHLBI Family Heart Study

Abstract—We conducted a genome-wide linkage scan for quantitative trait loci influencing total HDL-cholesterol (HDL-C) concentration in a sample of 1027 whites from 101 families participating in the NHLBI Family Heart Study. To maximize the relative contribution of genetic components of variance to the total variance of HDL-C, the HDL-C phenotype was adjusted for age, age², body mass index, and Family Heart Study field center, and standardized HDL-C residuals were created separately for men and women. All analyses were completed by the variance components method, as implemented in the program GENEHUNTER using 383 anonymous markers typed at the NHLBI Mammalian Genotyping Service in Marshfield, Wis. Evidence for linkage of residual HDL-C was detected near marker D5S1470 at location 39.9 cM from the p-terminal of chromosome 5 (LOD=3.64). Suggestive linkage was detected near marker D13S1493 at location 27.5 cM on chromosome 13 (LOD=2.36). We conclude that at least 1 genomic region is likely to harbor a gene that influences interindividual variation in HDL cholesterol. (Arterioscler Thromb Vasc Biol. 2001; 21:1823-1828.)

Key Words: HDL cholesterol ■ genome scan ■ anonymous marker linkage ■ quantitative trait loci ■ genetic epidemiology

Low plasma HDL-cholesterol (HDL-C) concentration is a major risk factor for cardiovascular disease, especially coronary heart disease (CHD), which remains the largest source of morbidity and mortality in the United States and other industrialized countries. Recent studies show that HDL particles influence the atherosclerotic process in many ways, including by reducing oxidation of LDL particles and inhibiting various coagulation pathways. But it is their role in the reverse cholesterol transport pathway that is believed to be most important. During reverse cholesterol transport, HDL particles remove cholesterol from peripheral tissues and transport it to the liver for either repackaging or excretion through the bile. This process depends on many enzymes to move the cholesterol molecules from the peripheral tissues to the liver. Despite the large number of control points involved in the movement of cholesterol by HDL particles, relatively few environmental predictors have been identified from large prospective studies.

At least 9 twin studies3–11 and 14 family studies5,9,12–23 have estimated the heritability of serum HDL-C. Heritability estimates ranged from 24% to 83%, with most studies in the 40% to 60% range. Variability in heritability estimates across studies could be due to a number of factors, including differences in study sample sizes, statistical methods, ages of the study participants, covariate adjustment, reliability of the HDL-C measurement, and whether only twins or families were studied. Despite these differences, most studies found HDL-C to be a highly heritable characteristic, warranting further research into the genetic sources of phenotypic variation.

The demonstration that a heritable trait has a multimodal distribution pattern lends further support to a significant genetic component to the variation in the trait. Segregation analysis has been used to determine the probable mode of inheritance of a disease or trait, as well as to estimate the residual polygenic heritability. At least 9 studies24–32 have applied segregation analysis to the detection of major gene loci that influence variation in HDL-C. After adjustment for environmental covariates, major gene effects were detected in some studies, but the estimated allele frequency for low
HDL-C was very small (range 0.003 to 0.086). These low-frequency alleles would explain little of the total variability in HDL-C. However, environmental modifiers of the major gene may obscure its detection unless gene-environment interactions are incorporated into the model. Mahaney et al. developed a model that included genotype-specific sex and covariate effects and age-by-sex interactions and found a major gene effect for high HDL-C with a high-frequency allele (0.84). Consistent with these findings, all studies indicate a significant residual polygenic heritability (0.30 to 0.60), similar to results from twin and family studies.

These heritability and segregation results prompted the initiation of anonymous marker genome scans to identify genomic regions that exhibit linkage to genetic variation in HDL-C. Currently, there are 5 published reports of genome-wide scans for identification of quantitative trait loci (QTLs) that contribute to variation in either total HDL-C or HDL-C subfractions. Two of these studies were conducted in large Mexican-American families and found significant linkage of HDL-C and various HDL-C subclass phenotypes to new genomic regions. Using a 10-cM map, investigators in the San Antonio Family Diabetes Study mapped total HDL-C (logarithm of the odds [LOD] 3.6) to chromosome 9p (between markers D9S925 and D9S1121). Investigators in the San Antonio Family Heart Study used a 15-cM map and found significant linkage of the HDL-C subfraction (LOD = 4.87) to chromosome 8q (near marker D8S1128) and to a location near the hepatic lipase gene on chromosome 15 (LOD = 3.26). Several other LOD scores >1.9, considered “suggestive” of linkage, warrant further investigation. Using a denser map (≈6 cM), Newman et al. found suggestive linkage of HDL-C to location 45.3 cM on chromosome 5p in 522 adult individuals from a genetically isolated population in Canada (P = 0.02). In addition, Kort et al. found a QTL on chromosome 11q23 linked to the low HDL-C phenotype in large Utah pedigrees (LOD = 3.5). Imperatore et al. reported suggestive linkage of a location near marker D3S2053 on chromosome 3q (LOD = 2.64). A recently published genome scan of several lipid phenotypes in the Rochester Family Heart Study found no evidence for linkage of HDL-C to any autosomal location. The authors did report suggestive linkage of apolipoprotein (apo) A-II to both chromosome 4 (LOD = 2.35 at 169 cM) and chromosome 5 (LOD = 2.13 at 79 cM). The goal of the present study was to conduct genomic scans with a dense marker map and a larger sample size than the previously reported studies as part of the National Heart, Lung, and Blood Institute (NHLBI) Family Heart Study (FHS).

**Methods**

The NHLBI FHS is a multicenter study incorporating participants from 4 population-based cohorts: the Framingham Heart Study, Utah Health Family Tree Study, and the Minneapolis, Minn, and Forsyth County, NC, field centers of the Atherosclerosis Risk in Communities (ARIC) Study. The primary goal of the FHS is to identify genetic and environmental predictors of CHD in randomly sampled families and families determined to be at high risk for CHD.

**Phenotyping**

Participants (n = 5348) were asked to fast the night before attending a 3.5-hour clinic examination that included measurements of anthropometry, blood pressure, and pulmonary function, as well as a carotid artery ultrasound and an ECG. Blood was drawn in all participants for standard laboratory tests (lipids, hemostatic factors, insulin, glucose, blood chemistries, and hematology) and for long-term storage. Interviews and questionnaires assessing lifestyle, medical, personal, and reproductive histories; medication use; food frequency intake; physical exercise; and psychosocial characteristics were conducted. An additional 627 remote site examinations were conducted for participants unable to visit the clinic. Although not all of the clinic components could be completed during remote site visits, blood was also collected from individuals who visited the remote sites. Blood samples from both clinic visits and remote site examinations (total N = 5975) were stored for additional phenotyping and genotyping at a later date.

The laboratory at Fairview-University Medical Center at the University of Minnesota performed all blood assays on samples from 5975 participants. HDL-C was measured by COBAS FARA (Roche Diagnostic Systems) after precipitation of the other lipoprotein fractions by dextran sulfate. The laboratory variability of the HDL-C assay was calculated in 299 blind replicates, yielding a very high reliability coefficient of 0.98. Body mass index (BMI) was calculated as weight (in kilograms) divided by height (in meters) squared.

**Genotyping:** Mammalian Genotyping Service Sample

A total of 1027 white individuals from 101 families were genotyped for the entire genome by the NHLBI Mammalian Genotyping Service (MGS) in Marshfield, Wis. Highly polymorphic autosomal markers (n = 383) from the CHLC Screening Set 10 were used, with an average distance of ~9.5 cM between markers. These 101 families were chosen because they were the largest FHS pedigrees containing at least 1 individual above the age-sex specific 80th percentile of the Individual Risk Score for CHD (a score derived for each individual using sex-specific proportional hazards models to predict the age of onset of CHD, based on carotid artery intima-media thickness, lipid, BMI, blood pressure, hypertension, diabetes, and CHD family risk score). In general, these pedigrees include marker data on a proband, his/her spouse, his/her siblings, and his/her children. The marker heterozygosity for this sample was 76%. Fluorescent dyes were incorporated into the FHS DNA samples during amplification by a standardized polymerase chain reaction protocol. The DNA fragments were then separated by electrophoresis on a 6% acrylamide gel with 100 lanes. A Scanning Fluorescence Detector (SCAFUD) was used to read the alleles, and software automatically scored each individual genotype. Detailed information on the genotyping methods, instruments, and software used can be found at the Marshfield Laboratory World Wide Web site.

Marker inconsistencies were evaluated in 2 steps with the genetics programs ASPEX, PEDSYS, and MAPMAKER/SIBS. Reported pedigree relationships were validated with the marker data. Although these steps ensure that the genotypic data will be consistent with the reported relationships as a whole, there may be isolated markers that remain inconsistent with the reported family structure. To resolve these inconsistencies, either a single subject in a pedigree or the entire family would be set to “missing” for that genetic marker. After quality control processing at the FHS Coordinating Center, the average missing data rate across all MGS markers was 2.8 ± 2.2%, and 91.4% of these markers were complete for at least 95% of participants. At least 95% of the anonymous markers were complete for 93.4% of all subjects (n = 959). Because of the limited number of genotypes scored by MGS, no phantom replicates were provided; however, internal quality control procedures were followed at the MGS Laboratory.

**Statistical Analysis**

To increase the relative contribution of genetic components of variance to the total variance of HDL-C, we have removed the effects of important environmental predictors of HDL-C. Adjustment for these covariates was achieved by generating a nested series of sex-specific residual phenotypes. These residual phenotypes were created by regressing covariates on HDL-C and retaining those
covariates when they explained a significant portion of the total phenotypic variance. An HDL-C residual phenotype was created separately for males and females and included the covariates age, age², FHS field center, and BMI. These covariates explained ~11% of the trait variation in men and 13% in women. This sex-specific residual phenotype was approximately normally distributed.

Quantitative-trait linkage analyses were performed with version 2.1 of the program GENEHUNTER.⁴⁷,⁴⁸ All analyses were completed with a variance-components model to examine evidence for linkage of a QTL for total HDL-C with an autosomal genome-wide map of 383 anonymous markers. GENEHUNTER estimates the amount of variance in a quantitative trait that can be attributed to a QTL at that marker position. Maximum likelihood values for the mean trait value, additive and dominant variance components for the QTL, additive and dominant variance components due to other, unlinked loci, and an environmental variance were calculated. We tested the hypothesis that the genetic variance due to the QTL equals zero by comparing the log likelihoods of a model in which the observed variances between relatives was determined as a function of the identity-by-descent relationships at a given marker locus with a model in which the QTL variance components were constrained to zero. The difference in these 2 log likelihoods is similar to the classic LOD score in linkage analysis. The multipoint identity-by-descent method used the proportion of alleles shared identical by descent at genotyped loci to estimate identity-by-descent sharing at 5 evenly spaced points between adjacent markers along a chromosome for each nuclear family. Allele frequencies were derived from the entire sample genotyped by MGS.

Because of computational limitations, GENEHUNTER was unable to use all data from every family genotyped for the MGS marker set. As a result, we were only able to perform analyses on families with 24 max bits or fewer (max bits are defined as twice the number of nonfounders [NF] minus the number of founders [F]). To meet these constraints, we modified our pedigrees in 2 ways. First, we divided the 4 largest families into 8 new families. Three additional families were “trimmed” by removing 9 genotyped individuals from the analysis based on their relative contribution to the power to detect linkage. These individuals were chosen from the largest sibships within a pedigree and were either missing covariate information or were closest to the sex-specific raw HDL-C mean value. Table 1 shows the pedigree size distribution after modification. All pedigrees included at least 5 individuals, with a mean pedigree size of 9.7. The 9 individuals deleted represented 1% of our total genotyped sample.

Results

Table 2 illustrates the mean values and distributions for the phenotype and related covariates in this sample of whites from the NHLBI FHS. HDL-C levels were normally distributed for both males (mean 42.3 mg/dL) and females (mean 53.7 mg/dL). The mean age was 50.9 years for males and 51.9 years for females. The distribution of BMI was similar for males (mean 28.1 kg/m²) and females (mean 28.5 kg/m²).

Residual heritability of HDL-C was estimated at 53%. Multipoint linkage analyses were performed across all 22 autosomes for sex-specific and age-, age²-, FHS field center-, and BMI-adjusted total HDL-C. As shown in the Figure, only 1 LOD score approached genome-wide significance,⁹,⁵⁰ occurring at location 39.9 cM on chromosome 5p (LOD = 3.64, \( P_2 = 0.0002 \)) with a 1-LOD support interval of 33.6 to 53.7 cM. This LOD score peak is in the 9-cM region between markers D5S2845 and D5S1470. Both markers are trinucleotide repeats and have high heterozygosities of 0.68 and 0.83. A significant portion of the genetic variance (63%) was partitioned to a dominant QTL. When only additive genetic variances were modeled, the maximum LOD score was attenuated to 2.95. Location 27.5 cM on chromosome 13p (LOD = 2.36, \( P_2 = 0.004 \)) produced the second-highest LOD score peak in our data. The 1-LOD support interval for this possible QTL was 20.4 to 32.8 cM and was between markers D13S1493 and GATA86H01. However, when only the additive genetic variance was modeled, there was no evidence for linkage at this location (LOD = 0.16). Additional QTLs met the nominal significance level (approximate LOD score >1.30, \( P_2 = 0.05 \)) on chromosomes 1, 4, 6, and 8 (Table 3).
TABLE 3. Chromosomal Locations With Nominally Significant* Linkage to Adjusted† Plasma HDL-C in the NHLBI FHS

<table>
<thead>
<tr>
<th>Chromosome</th>
<th>Location, cM</th>
<th>Nearest Marker</th>
<th>Maximum LOD</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>107.6</td>
<td>D1S1728</td>
<td>1.59</td>
<td>0.026</td>
</tr>
<tr>
<td>4</td>
<td>180.76</td>
<td>G4S2417</td>
<td>1.33</td>
<td>0.046</td>
</tr>
<tr>
<td>5</td>
<td>39.89</td>
<td>D5S1470</td>
<td>3.64</td>
<td>0.0002</td>
</tr>
<tr>
<td>6</td>
<td>126.87</td>
<td>D6S1040</td>
<td>1.49</td>
<td>0.032</td>
</tr>
<tr>
<td>8</td>
<td>77.89</td>
<td>D8S1113</td>
<td>1.33</td>
<td>0.046</td>
</tr>
<tr>
<td>13</td>
<td>27.54</td>
<td>D13S1493</td>
<td>2.36</td>
<td>0.004</td>
</tr>
</tbody>
</table>

*Nominal significance (P<0.05) with 2 degrees of freedom — LOD 1.30.
†Sex-specific age-, age², FHS field center, BMI-adjusted residual phenotypes.

Discussion

This genome-wide linkage result was found in a sample composed primarily of 3-generation pedigrees selected for genotyping because of their large size and the presence of at least 1 individual at high risk for CHD. Finding evidence for a QTL in multiple populations is an important consideration when making determinations about the likelihood of a false-positive result. Our LOD score peak of 3.64 at location 39.9 cM on chromosome 5p is very close to a reported suggestive linkage detected in a genetically isolated population of Hutterites in Alberta, Canada. Newman et al.35 found suggestive linkage (P=0.02) of HDL-C to a QTL at location 45.3 cM on chromosome 5p. Our peak is also 30 cM away from a suggestive linkage (LOD=2.13) to apoAII reported in multigenerational pedigrees from the Rochester Family Heart Study.38 ApoAII is 1 of 2 important structural components of HDL molecules and is highly correlated with HDL-C. However, these researchers found no evidence for linkage of HDL-C to any autosomal location (maximum genome-wide LOD=1.09). We know of no other reports of linkage of total HDL-C or HDL-C subfractions to this location.

The Rochester Family Heart Study researchers reported another suggestive linkage of apoAII to location 169 cM on chromosome 5q (LOD=2.35). We also found evidence of a QTL affecting HDL-C only 11 cM telomeric to this location at 180.8 cM (LOD=1.33, P=0.046). We found no overlap of our linkage results with either report from the San Antonio Family Heart Study investigators.\(^{33,34}\) or with those by Imperatore et al.\(^{37}\) or Kort et al.\(^{36}\)

HDL-C is a complex trait that likely has important polygenic and environmental determinants. The HDL-C literature includes multiple studies indicating a strong heritable component, even after adjustment for important environmental predictors. As new candidates are described, investigators rush to genotype their study families’ DNA to replicate the reported findings, often with little success. The heterogeneity of study populations and often the small sample sizes of many linkage studies have led to a confused literature, with important linkage findings in one study not replicated in many others. In the case of HDL-C, previously reported findings were from very different populations in North America, including Mexican-Americans in Texas,33,34 the genetically isolated Hutterites in Canada,35 Pima Indians in Arizona,37 and whites in Minnesota.38 Although the sample sizes were generally large (range 415 to 1484), the family structures varied from sibships\(^{37}\) to extended pedigrees of varying complexities\(^{33,34,38}\) to a very large inbred kindred.35 Also, the covariate adjustment strategies ranged from minimal in the Pima Indians\(^{38}\) and Rochester Family Heart Study\(^{37}\) to complete in the San Antonio Family Heart Study.\(^{34}\) Our covariate adjustment most closely resembled that of the Rochester Family Heart Study investigators.

Even the most consistent candidate-gene/HDL-C–pheno-
type linkages are thought to play relatively small roles in variation of HDL-C at the population level. The well-established apoA-I, lipoprotein lipase, apoC-II, lecithin:cho-
esterol acyltransferase, and cholesteryl ester transfer protein candidate genes are thought to explain <2% of the total phenotypic variation in the United States.51 Simulations performed on the FHS family structures indicate that variance-components methods would be capable of detecting QTLs accounting for >10% of the phenotypic variance, so it is not surprising that our LOD score peaks do not overlap with known candidate genes.

As already mentioned, the metabolism of HDL-C is a complex process with multiple structural and enzymatic proteins exerting major influences on the ability of HDL to remove cholesterol from the bloodstream. As such, many candidate genes that significantly affect HDL-C metabolism have been reported in the literature. None of the LOD score peaks listed in Table 3 corresponds to any of these candidate genes. It is to be expected that complex traits, such as HDL-C, are not primarily determined by 1 or even just a few genes. At the same time, it is unlikely that there is 1 unknown gene that solely regulates the observed genetic variability in HDL-C. A review of the literature points to no obvious candidate genes in the general location of the QTL detected on chromosome 5.

In short, these results indicate linkage of a new genomic region to genetic variability in HDL-C. A broad peak with a maximum LOD score of 3.64 (P=0.0002) is localized to a region near 40 cM on the short arm of chromosome 5. This location is very close to a QTL identified in a sample of Alberta Hutterites. As far as we know, there are no other published reports of this possible QTL. Taken together, the results of all of these genome scans indicate many loci influencing HDL-C. Subtle differences in study samples and marker sets may contribute to the identification of different QTLs and to what appears to be a lack of replication across studies. Anonymous marker genome scans such as this one are largely hypothesis-generating exercises. Although the QTLs identified in different studies may not be the same, the use of genome-wide methods such as these may make it possible to identify multiple-component QTLs that contribute to the polygenic HDL-C trait. As more genome scans are published, the overlapping regions and positional candidate genes should become the focus of more intensive efforts to identify novel mechanisms that influence variability in HDL-C.

Acknowledgments

Support was provided in part by the NHLBI cooperative agreement grants U01 HL56563, U01 HL56564, U01 HL56565, U01 HL56566, U01 HL56567, U01 HL56568, and U01 HL56569. We wish to thank the investigators, staff, and participants in the NHLBI FHS. We also thank the University of Minnesota Supercomputing Institute for...
References


Genome Scan for Quantitative Trait Loci Linked to High-Density Lipoprotein Cholesterol: The NHLBI Family Heart Study
on behalf of the Investigators of the NHLBI Family Heart Study

doi: 10.1161/hq1101.097804
Arteriosclerosis, Thrombosis, and Vascular Biology is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2001 American Heart Association, Inc. All rights reserved.
Print ISSN: 1079-5642. Online ISSN: 1524-4636

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://atvb.ahajournals.org/content/21/11/1823

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Arteriosclerosis, Thrombosis, and Vascular Biology can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Arteriosclerosis, Thrombosis, and Vascular Biology is online at:
http://atvb.ahajournals.org/subscriptions/