Admixture mapping of quantitative trait loci for blood lipids in African-Americans

Analabha Basu1,*, Hua Tang3, Cora E. Lewis5, Kari North6, J. David Curb7, Thomas Quertermous4, Thomas H. Mosley8, Eric Boerwinkle9, Xiaofeng Zhu10 and Neil J. Risch1,2,11,*

1Institute for Human Genetics and 2Department of Epidemiology and Biostatistics, University of California San Francisco, 513 Parnassus Avenue, Room 901F HSW, San Francisco, CA 94143, USA, 3Department of Genetics, Stanford University, Palo Alto, CA, USA, 4Department of Medicine, University of Alabama, Birmingham, UK, 5Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA, 6Department of Geriatric Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, HI, USA, 7Department of Mississippi Medical Center, Jackson, MS, USA, 8School of Public Health, University of Texas Health Science Center, Houston, TX, USA, 9Department of Epidemiology and Biostatistics, Case Western Reserve University School of Medicine, Cleveland, OH, USA and 11Kaiser Permanente Division of Research, Oakland, CA, USA

Blood lipid levels, including low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C) and triglycerides (TG), are highly heritable traits and major risk factors for atherosclerotic cardiovascular disease (CVD). Using individual ancestry estimates at marker locations across the genome, we present a novel quantitative admixture mapping analysis of all three lipid traits in a large sample of African-Americans from the Family Blood Pressure Program. Regression analysis was performed with both total and marker-location-specific European ancestry as explanatory variables, along with demographic covariates. Robust permutation analysis was used to assess statistical significance. Overall European ancestry was significantly correlated with HDL-C (negatively) and TG (positively), but not with LDL-C. We found strong evidence for a novel locus underlying HDL-C on chromosome 8q, which correlated negatively with European ancestry ($P = .0014$); the same location also showed positive correlation of European ancestry with TG levels. A region on chromosome 14q also showed significant negative correlation between HDL-C levels and European ancestry. On chromosome 15q, a suggestive negative correlation of European ancestry with TG and positive correlation with HDL-C was observed. Results with LDL-C were less significant overall. We also found significant evidence for genome-wide ancestry effects underlying the joint distribution of HDL-C and TG, not fully explained by the locus on chromosome 8. Our results are consistent with a genetic contribution to and may explain the healthier HDL-C and TG profiles found in Blacks versus Whites. The identified regions provide locations for follow-up studies of genetic variants underlying lipid variation in African-Americans and possibly other populations.

INTRODUCTION

Cardiovascular disease (CVD) is a leading cause of morbidity and mortality, making it a major public health concern worldwide (1). The etiology of the disease is multifactorial, with established genetic and environmental risk factors (2–5). Atherosclerosis, which is characterized by lipid accumulation, inflammatory response, cell death and fibrosis in arterial walls, accounts for 75% of all deaths from CVD (6). There is compelling evidence of involvement of blood lipid concentrations in atherosclerosis and CVD (7–9). Blood lipids are complex traits influenced by both genetic and environmental factors (2–5,10). Heritability estimates for lipid concentrations vary [~40–75% for high-density lipoprotein cholesterol

*To whom correspondence should be addressed. Tel: +1 4154761129; Fax: +1 4154762956; Email: basu@humgen.ucsf.edu (A.B.) and rischn@humgen.ucsf.edu (N.J.R.)

© The Author 2009. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
(HDL-C), \(\sim 40–65\%\) for low-density lipoprotein cholesterol (LDL-C) and \(\sim 35–50\%\) for triglycerides (TG)] (11), yet are consistently high for all lipid traits.

The genetic dissection of these complex traits has long been a major challenge, and there have been many studies aimed at finding specific loci for these phenotypes. Various study designs have been employed, ranging from large-scale genome-wide association (12–15) to candidate gene (16–20). Recently, Willer et al. (15), in a large meta-analysis combining genome-wide scans with \(\sim 9000\) individuals of European descent in stage 1 and \(\sim 11000\) individuals of European descent in stage 2, found convincing evidence for 11 loci (genes) previously implicated in lipid metabolism as well as compelling evidence for several newly identified loci (two for LDL-C, three for TG, one for HDL-C and one region encompassing several genes associated with both LDL-C and TG).

At the same time, there are known ethnic trends in lipid levels (22,23). For example, after adjusting for socioeconomic factors, African-Americans generally have healthier lipid levels (lower TG and higher HDL-C) than people of European descent (16,24,25), whereas both have healthier lipid levels than South Asians (26,27). This underscores the importance of undertaking large-scale mapping studies of populations with distinct ethnic backgrounds.

New world admixed populations provide unique opportunities for locating trait loci by genetic admixture mapping (28–33). The African-American population of the United States is typically characterized by admixture of European and African ancestral genomes in different proportions with some spatial variation (34–36). Studies have correlated genome-wide African ancestry in African-Americans with healthier lipid levels (37). In this report, we present results of an admixture mapping analysis examining the correlation of lipid levels (LDL-C, HDL-C and TG) with estimated genome-wide ancestry proportions as well as ancestry estimates at each of 284 marker loci distributed around the genome in 1044 unrelated African-Americans subjects from the Family Blood Pressure Program (FBPP), in a search for locus-specific effects.

RESULTS

Sample demographics for the 1044 individuals are given in Table 1. There were 698 females and 346 males. The subjects from GENOA were on average older (average age 59.6) than South Asians (26,27). This underscores the importance of undertaking large-scale mapping studies of populations with distinct ethnic backgrounds.

Table 1. Demographic characteristics of unrelated study participants: Family Blood Pressure Program

<table>
<thead>
<tr>
<th></th>
<th>GENOA</th>
<th>HyperGEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of individuals</td>
<td>102</td>
<td>247</td>
</tr>
<tr>
<td>BMI male</td>
<td>29.25 ± 4.9</td>
<td>31.77 ± 6.9</td>
</tr>
<tr>
<td>BMI female</td>
<td>31.24 ± 5.2</td>
<td>33.28 ± 7.1</td>
</tr>
<tr>
<td>Age male</td>
<td>61.83 ± 8.3</td>
<td>60.67 ± 10.8</td>
</tr>
<tr>
<td>Age female</td>
<td>58.77 ± 9.6</td>
<td>59.67 ± 10.8</td>
</tr>
<tr>
<td>HDL-C male</td>
<td>52.96 ± 18.6</td>
<td>60.37 ± 18.8</td>
</tr>
<tr>
<td>HDL-C female</td>
<td>58.77 ± 9.6</td>
<td>59.67 ± 10.8</td>
</tr>
<tr>
<td>LDL-C male</td>
<td>110.06 ± 43.4</td>
<td>121.01 ± 38.1</td>
</tr>
<tr>
<td>LDL-C female</td>
<td>112.07 ± 46.8</td>
<td>120.97 ± 36.9</td>
</tr>
<tr>
<td>TG male</td>
<td>145.12 ± 84.2</td>
<td>142.57 ± 67.8</td>
</tr>
<tr>
<td>TG female</td>
<td>15.57</td>
<td>15.57</td>
</tr>
<tr>
<td>European ancestry (%)</td>
<td>14.2</td>
<td>14.2</td>
</tr>
</tbody>
</table>

Table 2. Analysis of variance results on LDL, HDL and TG

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>F-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>LDL-C</td>
<td>0</td>
<td>30.18</td>
<td>4.99 × 10⁻⁸</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>14.41</td>
<td>1.56 × 10⁻⁴</td>
</tr>
<tr>
<td>Sex</td>
<td>0</td>
<td>0.37</td>
<td>0.54</td>
</tr>
<tr>
<td>Network</td>
<td>0</td>
<td>6.31</td>
<td>0.012</td>
</tr>
<tr>
<td>HDL-C</td>
<td>0</td>
<td>39.04</td>
<td>6.05 × 10⁻¹⁰</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>31.96</td>
<td>2.00 × 10⁻⁸</td>
</tr>
<tr>
<td>Sex</td>
<td>0</td>
<td>78.39</td>
<td>&lt;2.2 × 10⁻¹⁶</td>
</tr>
<tr>
<td>Network</td>
<td>0</td>
<td>0.20</td>
<td>0.66</td>
</tr>
<tr>
<td>TG</td>
<td>0</td>
<td>18.03</td>
<td>2.37 × 10⁻⁵</td>
</tr>
<tr>
<td>Age</td>
<td>0</td>
<td>74.66</td>
<td>&lt;2.2 × 10⁻¹⁶</td>
</tr>
<tr>
<td>Sex</td>
<td>0</td>
<td>2.84</td>
<td>0.09</td>
</tr>
<tr>
<td>Network</td>
<td>0</td>
<td>53.44</td>
<td>5.31 × 10⁻¹³</td>
</tr>
</tbody>
</table>

According to normal distribution theory, we expected 14 \(\sim 5\%\) of 284 marker positions to have an absolute standardized regression coefficient \(|r_j|\) value > 1.96, or seven in network, the average body mass index (BMI), HDL-C and LDL-C were generally higher in females than in males, whereas TG levels showed slightly reversed tendencies. The male and female subjects from GENOA had higher HDL-C and TG levels than their corresponding counterparts from HyperGEN. There was little variation in BMI between networks. Average European ancestry varied modestly between the two networks, as was previously noted (36).

LDL-C levels were modestly correlated with HDL-C and TG levels (correlation coefficient \(-0.13\) with HDL-C and \(0.19\) with TG). HDL-C and TG, as observed in other studies, showed a stronger inverse correlation \((-0.28)\). The correlation between the traits was virtually unaffected by the transformations. By analysis of variance, all three lipid traits were strongly associated with BMI and age (Table 2). HDL-C was strongly associated with sex. TG and LDL-C differed significantly between networks.

HDL-C was not found to be correlated with overall European individual ancestry (IA) \((P = 0.80)\). On the other hand, HDL-C was strongly negatively correlated with European IA \((t = -2.76, P = 0.0059)\) and TG was strongly positively correlated with European IA \((t = 3.29, P = 0.0010)\).

A Q–Q plot of the \(r_j\) values (standardized regression coefficients) for the 284 loci for LDL-C, TG and HDL-C showed a reasonably good fit to normality (Fig. 1) except for some outlier points, in particular at the low (negative) end of the distribution for HDL-C.
The two most promising locations for TG included one extended region positively correlated with European ancestry on chromosome 8q and another negatively correlated with European ancestry on chromosome 15q. On chromosome 15q, the peak $Z$-score of $-2.75$ occurred at 90.0 cM (marker D15S652), whereas two neighboring markers also had $Z$-scores less than $-2$. On chromosome 8q, an extended region of 87 cM stretching from 77.9 to 164.5 cM contained seven markers with $Z$-scores $>2$. Two separate peaks occurred in this interval, one at 135.1 cM with a $Z$-score of 2.58 (marker D8S1179) and the other at 77.9–82.3 cM (markers D8S1113 and D8S1136) with a $Z$-score of 2.31. We note that this is precisely the same location harboring the most significant result for HDL-C (marker D8S1136).

To determine the statistical significance of our findings, we employed a permutation test (see Materials and Methods). For LDL-C and TG, none of the markers was significant at $P < 0.05$. For HDL-C, however, in 5000 permutations, the minimum of 284 $r_I$ scores crossed $-4.13$ only seven times, $-3.01$ seventy-three times and $-2.99$ ninety times. Hence, the result associated with the marker D8S1136 is highly significant on a genome-wide level ($P = 0.0014$), whereas that for D8S1113 and GATA193A07 are also formally significant ($P < 0.02$).

We also looked at possible significant interactions of BMI, age, sex, network, usage of lipid-lowering medication, and hypertension status with local ancestry at the significant genomic marker locations for all three lipid traits and did not find any. The $Z$-scores for HDL-C and TG across chromosomes 8 and 14 are plotted in Figure 2, which also gives a visual impression of the inverse relationship between the two traits. We also examined the possibility of outlier points driving the $Z$-scores for the three significant loci (Supplementary Material, Fig. S1) but found no evidence of outliers.

Furthermore, as another check on the robustness of our conclusions, we examined results of multiple additional analyses based on different randomly selected sets of unrelated individuals from the African-American families in our study. The results for these subsets were essentially identical to those we obtained in the original analysis, indicating that our results were not dependent on the original selection of unrelated individuals from these families.

Our results also indicated an overall genome-wide ancestry deviation in opposite directions for HDL-C and TG. To determine whether multiple loci are influential in the joint distribution of these lipid traits, we looked at the correlation between the regression coefficients of the 284 markers for the two lipids. If a common set of genes is responsible, for example, for both high HDL-C and low TG values, then we would observe a large negative correlation between the ancestry regression coefficients.

For the 284 markers, the correlation between the regression coefficients for HDL-C and TG was $-0.55$. To determine statistical significance of this correlation, we performed 5000 permutations (as described in Materials and Methods), and for each permutation calculated the pairwise correlation between the ancestry regression coefficients for the two lipid traits. Out of 5000 permutations, the correlation exceeded $-0.55$ only 70 times, which corresponds to a $P$-value of 0.014. After removing the two chromosomes eight markers with highly

![Figure 1](https://academic.oup.com/hmg/article-abstract/18/11/2091/2527385) Q–Q plot comparing $r_I$ with a standard normal distribution and the corresponding histograms for the trait LDL-C, HDL-C and TG. Each tail. However, we found an excess of points in the tails in each of the three distributions. For LDL-C, there were seven $r_I$ values $> 1.96$ as expected but 10 in the left tail (i.e. values less than $-1.96$) (Supplementary Material, Table S1). For HDL-C, there were 11 $r_I$ values $> 1.96$ (i.e. in the right tail) and 10 in the left tail (less than $-1.96$) (Supplementary Material, Table S2). For TG, there were 12 $r_I$ values $> 1.96$ (i.e. in the right tail) and eight in the left tail (Supplementary Material, Table S3).

For LDL-C, the two most promising chromosomal regions were at 3q and 4q, both showing negative correlation with European ancestry. On 3q, five markers covering 64 cM had $Z$-scores less than $-2.0$, with a peak of $-2.35$ at 188.3 cM (marker D3S2427). On chromosome 4q, three markers had $Z$-scores less than $-2.0$, with a peak of $-2.33$ at 78 cM (marker D4S2367).

For HDL-C, two significant regions on chromosomes 8q and 14q showed negative correlation with European ancestry, whereas a region on chromosome 9q showed positive correlation with European ancestry. Chromosome 8q had four markers covering 27 cM with $Z$-scores less than $-2.0$, with a peak of $-4.13$ at 82.3 cM (marker D8S1136). On chromosome 14q, three markers had $Z$-scores less than $-2$, with a peak of $-3.00$ at 95.9 cM (marker GATA193A07). On chromosome 9q, there were three markers with $Z$-scores $> 2$, with a peak of 2.69 at 110.9 cM.
negative $r_I$ scores for HDL-C and highly positive $r_I$ scores for TG from the analysis, the correlation declined only to $-0.53$, which was still statistically significant ($P = 0.028$). After removal of the markers on chromosomes 8q and 15q, the correlation still trended in the same direction but was no longer significant ($P = .063$). To assess whether the regressions of HDL-C and TG on total European IA could be attributed solely to ancestry at the peak location on chromosome 8q, we performed a regression analysis of HDL-C (and TG) including both 8q locus-specific European ancestry and total European ancestry in the model. For both lipids, total European IA was no longer significant (and regression coefficients were close to 0), once the locus-specific European ancestry was included in the model. Hence, the overall ancestry effect we observed for both lipids could be explained simply by a locus on chromosome 8q, although the situation may be more complex, with other loci contributing.

**DISCUSSION**

Numerous genes have been directly implicated in the variation of lipid traits such as apolipoproteins (APOA1, APOA2, APOA4, APOE), ATP-binding cassette sub-family A member 1 (ABCA1), lecithin cholesterol acyltransferase (LCAT) and lipoprotein lipase (LPL), among others (16,38–43). All these genes have been extensively studied and well-characterized biochemically. With the exception of APOE, variants in the above-mentioned genes and a few others have been found mostly in Mendelian disorders involving lipid metabolism. Identifying more common and/or less penetrant variants affecting lipid levels has been a more challenging endeavor.

Among the three lipid traits we studied, LDL-C had the least significant results. Although no points for LDL-C were formally significant, we did note a modest increase in markers correlated with excess African ancestry on chromosomes 3q and 4q. The region on chromosome 3q was previously identified in a linkage study with 470 subjects in 10 pedigrees with a maximum LOD score of 4.1 (44), but not in other linkage and association studies. On chromosome 4q, the marker D4S2367, which had the second lowest $r_I$ value, is close to the SNP rs10518072. This SNP was associated with LDL-C levels in the Framingham Heart genome-wide association study ($P = 1.4 \times 10^{-4}$) (13) but was not reported by Willer et al. (15) study.

In contrast, the results for HDL-C were more significant. Most of the markers in the extreme left tail (negative correlation with European ancestry) are on chromosomes 8q and 14q. The peak region on chromosome 8q (at 82 cM), our most significant finding overall, also showed evidence of positive correlation of European ancestry with TG. Deviations in the other tail of the Z-score distribution were more subtle and less significant. Three consecutive markers on chromosome 9q showed up in this tail, with a peak between 104 and 111 cM.

Of interest, the region on chromosomes 9q22–q32 that we identified harbors the ABC1 (ATP-binding cassette) gene at 9q31. ABC1 has long been associated with studies involving HDL-C and is regarded as a candidate gene for HDL-C deficiency and other related phenotypes (45–48). Mutations in ABC1 can cause Tangier disease, a genetic disorder of cholesterol characterized by very low levels of HDL-C (46,47,49–52). Hence, ABC1 would be a candidate for an HDL-C quantitative trait locus (QTL) in our analysis as well.

The three markers in our study, which were clearly significant after permutation testing, are D8S1131 and D8S1136 from the region 8q11.23–q21.11 and GATA193A07 from the region 14q24.1–q31. The region 8q11.23–q21.11 is the location of the CYP7A1 gene at 8q12.1. This gene encodes a member of the cytochrome P450 superfamily of enzymes which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. This region also harbors the candidate gene CYP7B1 (at 8q12.3; 65.6 Mb from pter) between the markers D8S1113 and D8S1136. Little other evidence in the literature implicates this region of 8q in HDL-C or TG levels, suggesting our results may be indicating a novel locus. In a study of lipid levels among 113 Hispanic families, the GATA193A07 marker on chromosome 14 showed modest evidence of linkage for TG (two point LOD score 2.12) and for the ratio of TG to HDL-C (two point LOD score 2.46) (52). Weak evidence of linkage to the locus was also obtained in a genome scan of QTLs influencing HDL-C levels (19,53).
Analysis of TG did not reveal any formally significant loci, although there was an overall excess of observations in the right tail of the Z-score distribution. This excess was due to a preponderance of loci on chromosome 8. Three of these loci overlapped with our most significant region for HDL-C at 8q11–q21, whereas four other markers mapped to a non-overlapping, more distal location between 8q23 and q24. We do note, however, that rs7007075, the SNP most strongly associated ($P < 7.7 \times 10^{-6}$) with TG levels in the Framingham Heart Genome-wide association study, lies in this region. The SNP rs17321515 (8q24.13) at 3' downstream of the gene TRIB1 has been one of the most significant findings ($P=4 \times 10^{-17}$) of genome-wide association studies (12). This study also reported five independently associated SNPs in this region, all of which lie within the two consecutive markers D8S592 at 8q24.11 and D8S1179 at 8q24.13, included in our study. The $r_I$ value corresponding to the marker D8S1179 (2.59) and D8S592 (2.48) for TG are the highest and the third highest that we found in our study (Supplementary Material, Table S3).

As noted by us and others, the various lipid fractions are modestly correlated within individuals. The largest correlation is the negative association between HDL-C and TG. Prior studies have suggested this correlation is, in part, genetically determined (44,54–62). Thus, it is appropriate in any genomic analysis to search for genes underlying the joint distribution of these traits.

In our admixture mapping analysis, we therefore searched for locations potentially underlying multiple traits. Examination of the joint distribution of regression coefficients for HDL-C and TG did indeed demonstrate a correlation ($-0.55$) that was statistically significant based on a permutation analysis. The excess correlation over expected was due partially but not exclusively to markers on chromosomes 8q and 15q, suggesting a possible multi-genic contribution to this correlation.

Admixture mapping analysis is complementary to other gene mapping approaches such as QTL linkage analysis and genome-wide association studies. The power of the method depends both on a measurable effect influencing the trait of interest and on a sizeable difference in allele frequency between ancestral populations, in this case Europeans and Africans. Some of our more significant findings overlapped with prior linkage and genome-wide association studies, and these indicate chromosome regions for more detailed study in follow-up. However, the degree of overlap among these approaches depends on a number of factors, including the underlying allele frequencies and degree of linkage disequilibrium between trait-associated alleles and marker SNPs.

The power of admixture mapping is also dependent upon the accuracy of the estimates of locus-specific ancestry. Ancestry informativeness of the marker set that we have used is on an average 50% and never falls $< 40\%$ (33).

Of note, the direction of our significant findings for HDL-C and TG and lack of significant findings for LDL-C are consistent with a genetic contribution to the known ethnic difference for HDL-C and TG. Specifically, the regression coefficient we obtained for European IA on (untransformed) HDL-C ($-8.18$) can fully explain the observed Black–White differences in HDL-C. If we assume that African-Americans have 15% European ancestry, on average (36), then the expected difference between Whites and African-Americans would be 85% of $-8.18$ or $-6.95$. In our sample, the standard deviation of HDL-C was 16.80, so the projected difference of $-6.95$ units corresponds $\sim 0.4$ standard deviations, a figure comparable to the ethnic difference observed in two large studies (24,25). Similarly, the regression coefficient of European IA on (untransformed) TG levels was 64.47. The standard deviation of TG in our sample was 76.36. Therefore, the predicted ethnic difference from our analysis would be 85% of 64.47 = 54.80, which corresponds $\sim 0.7$ standard deviations, again comparable to the differences observed for US Blacks and Whites (24,25).

To our knowledge, this is the largest quantitative admixture mapping effort of blood lipid levels in terms of sample size and marker locus involvement. Overall, our findings are encouraging and provide regions for follow-up analyses of genes influencing HDL-C and TG, and possibly LDL-C in these and other African-American individuals.

**MATERIALS AND METHODS**

**Subjects**

The FBPP is a large multicenter genetic study of high blood pressure and related traits in multiple racial/ethnic groups, including European-Americans, African-Americans, Mexican Americans and Asians and Asian Americans (63). It includes four component networks: GenNet, GENOA, HyperGEN and SAPHIRE. GenNet, GENOA and HyperGEN who independently collected samples from European-American and African-American families.

All the individuals we included in this study were unrelated self-identified African-Americans from the field centers of GENOA and HyperGEN for whom fasting plasma lipid concentrations were available. To maximize the number of unrelated individuals in our sample, whenever possible we selected unrelated founder individuals; otherwise we chose at random one individual per family. Our final sample of 1044 individuals consisted of 349 individuals sampled by the GENOA network and 695 individuals sampled by the HyperGEN network. Out of these 1044 individuals, 32 did not have their LDL-C on record, so all our findings for LDL-C are based on 1012 individuals, whereas those for HDL-C and TG are based on 1044 individuals. There were 57 individuals (52 for LDL-C) who were noted to be taking lipid-lowering medication. The two networks included in this study measured fasting plasma lipids using standard enzymatic methods in laboratories participating in the CDC standardization program. All individuals who participated in the FBPP gave informed consent; the Institutional Review Board at each clinic site approved all protocols, and a Certificate of Confidentiality was obtained from the Federal Government for this study.

**Genotyping**

DNA was extracted from whole blood by standard methods by each of the four FBPP networks and was sent to the US National Heart, Lung and Blood Institute’s Mammalian genotyping service in Marshfield, Wisconsin, for genotyping.
Statistical analysis

We used the computer program Structure (64,65) to estimate genome-wide, as well as site-specific IAs in all African-American participants. The linkage model was used, with genetic distance between markers specified according to the Marshfield map. In each analysis, the Markov Chain Monte Carlo algorithm was run for 100,000 steps of burn-in followed by another 100,000 steps. We assumed a model with two ancestral populations. We included 1378 unrelated non-Hispanic White participants from the FBPP to represent the European ancestral population as well as 127 unrelated sub-Saharan African individuals from the Human Genome Diversity Project to represent the African ancestral population (66). The latter set of individuals had been genotyped at more than 300 short tandem repeat at the time of our analysis, and we included genotypes at 284 markers which were overlapping with the FBPP genotypes.

The distributions of all three lipid traits for the 1044 (1012 for LDL-C) unrelated individuals were non-normal due to positive skewness. The best normalization of the LDL-C data was obtained using the Box–Cox power transformation (67) with parameter (λ) 0.69, whereas the value of λ was −0.33 and −0.14, respectively, for HDL-C and TG.

We first calculated regression coefficients for each of the transformed lipid traits on overall European IA, allowing for covariates. Then, for each individual at each locus, a European ancestry deviation \( \beta_l \) was defined as the estimated ancestry at locus \( l \) minus the background ancestry estimated from the genome-wide markers for that individual, as previously described (68). The variable \( \beta_l \) was then used as the primary independent variable in a linear regression model with the quantitative trait value of LDL-C or HDL-C or TG as the dependent variable. Log–log transformed BMI, age, sex and network were included as covariates in this analysis, if significant for any of the traits, possibly due to the small sample size.

The use of lipid-lowering medication was not found to be significant for any of the traits, possibly due to the small sample size. The interaction structure of the markers and the correlation structure of the covariates data remained intact. This procedure preserved the correlation structure of the markers and the correlation structure of the traits and covariates, but dissociated the relationship between the markers and phenotypes. For each permuted dataset, we performed the same regression analysis of the traits on excess ancestry at each marker location, as was done for the original data, and obtained the most extreme values (positive and negative) of \( r_l \) (the Z-score statistics). Five thousand permutations were performed. To derive \( P \)-values adjusted for multiple testing, we determined the percentage of times out of 5000 permutations that an observed value of \( r_l \) was exceeded in the permuted data analysis.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS


REFERENCES

of plasma high density lipoprotein cholesterol by genetic and environmental factors. 


