Title

Genome-wide Characterization of Shared and Distinct Genetic Components that Influence Blood Lipid Levels in Ethnically Diverse Human Populations

Authors

Coram, Marc A
Duan, Qing
Hoffmann, Thomas J
et al.

ARTICLE

Genome-wide Characterization of Shared and Distinct Genetic Components that Influence Blood Lipid Levels in Ethnically Diverse Human Populations

Marc A. Coram,[1] Qing Duan,[2] Thomas J. Hoffmann,[3,4] Timothy Thornton,[5] Joshua W. Knowles,[6] Nicholas A. Johnson,[7,21] Heather M. Ochs-Balcom,[8] Timothy A. Donlon,[9,10] Lisa W. Martin,[11] Charles B. Eaton,[12,13] Jennifer G. Robinson,[14] Neil J. Risch,[3,4,15] Xiaofeng Zhu,[16] Charles Kooperberg,[17] Yun Li,[2,18] Alex P. Reiner,[19,*] and Hua Tang[20,*]

Blood lipid concentrations are heritable risk factors associated with atherosclerosis and cardiovascular diseases. Lipid traits exhibit considerable variation among populations of distinct ancestral origin as well as between individuals within a population. We performed association analyses to identify genetic loci influencing lipid concentrations in African American and Hispanic American women in the Women's Health Initiative SNP Health Association Resource. We validated one African-specific high-density lipoprotein cholesterol locus at *CD36* as well as 14 known lipid loci that have been previously implicated in studies of European populations. Moreover, we demonstrate striking similarities in genetic architecture (loci influencing the trait, direction and magnitude of genetic effects, and proportions of phenotypic variation explained) of lipid traits across populations. In particular, we found that a disproportionate fraction of lipid variation in African Americans and Hispanic Americans can be attributed to genomic loci exhibiting statistical evidence of association in Europeans, even though the precise genes and variants remain unknown. At the same time, we found substantial allelic heterogeneity within shared loci, characterized both by population-specific rare variants and variants shared among multiple populations that occur at disparate frequencies. The allelic heterogeneity emphasizes the importance of including diverse populations in future genetic association studies of complex traits such as lipids; furthermore, the overlap in lipid loci across populations of diverse ancestral origin argues that additional knowledge can be gleaned from multiple populations.

Introduction

Plasma concentrations of lipoproteins (low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL] cholesterol, and triglycerides [TG] are heritable risk factors for atherosclerosis and cardiovascular diseases (CVDs).[1,2] These lipid concentrations vary substantially between individuals as well as between populations.[3] For example, mean TG levels are highest among Hispanic American populations; despite their higher CVD mortality, African Americans tend to have higher HDL levels and lower TG levels compared to whites.[4–7] The heritability estimates of blood lipids vary across studies and by population but are consistently high: 40%–80% for both LDL and HDL and 30%–50% for TG.[8–10] A recent meta-analysis of more than 100,000 individuals of European ancestry identified 95 loci significantly associated with blood lipids.[11] Follow-up studies indicated that a majority of these loci are "ethnically transferrable," in the sense that they show statistical association and consistent direction of genetic effects in non-European populations.[3,11,12] Specifically, of the 36 LDL loci, 44 HDL loci, and 30 TG loci examined in African American participants, 33, 37, and 24 loci, respectively, showed the same direction of association as in the European cohorts; even higher "replication" rates were observed in East Asians, South Asians, and Hispanics, suggesting a shared genetic contribution to lipid variability among human populations. However, our understanding of the ethnically shared and distinct components of the genetic architecture that underlie blood lipid concentrations is incomplete. Non-replication could arise as a result of ethnic-specific causal variants, different linkage disequilibrium (LD) patterns surrounding the same causal variants, interaction effects with a distinct genetic background, or simply lack of statistical

CrossMark

power. Furthermore, the genetic architecture underlying a trait depends not only on the direction and magnitude of genetic effects, but also on factors such as frequencies of the risk variants. In other words, replication of variants across populations does not automatically imply that these variants constitute a substantial genetic component that explains phenotype variation in all populations.

Characterizing the shared and distinct genetic components that underlie complex traits among human populations has significant public health impact. Such an understanding regarding lipids levels can provide valuable insights on the genetic contribution to ethnic health disparities related to atherosclerosis and CVD across different ethnicities. However, a genome-wide assessment of the relative contribution of panethnic and population-specific genetic components to any complex trait is analytically challenging because our knowledge of the genetic architecture within each population is limited. Here we present a series of genome-wide analyses that consider various aspects of the genetic architecture of blood lipid levels, using genotype and phenotype data from the Women's Health Initiative SNP Health Association Resource (WHI-SHARe), which includes 8,153 African American (AA) and 3,587 Hispanic American (HA) participants. By applying a mixed effects model that makes use of the full reference GWAS (instead of just the top SNPs), we demonstrate that a large overlapping genetic component influences individual variation of lipid levels across these populations, including many loci that do not reach genome-wide significance level in any population to date. At the same time, genotype association and admixture mapping analyses reveal important population-specific lipid loci and variants. Our findings argue that an effective approach for elucidating the genetic architecture of complex traits is to combine evidence across multiple populations. The analytic approach we use is applicable to studies of other heritable complex traits.

Material and Methods

Our analytic approach consists of three components. (1) A discovery GWAS in a cohort of 8,153 WHI-SHARe AA samples and 3,587 HA samples. Validation analysis of variants not previously implicated in a large European GWAS was performed in an independent cohort of 7,138 AAs in the Candidate Gene Association Resource (CARe). (2) Genome-wide assessment of transethnic overlap in the genetic architecture underlying lipid variation was performed with the WHI-SHARe AA and HA cohorts by examining statistical significance, allelic effect, and proportion of variance explained by subsets of the genome. (3) Characterization of genetic variation that could account for population-level lipid differences by admixture mapping in WHI-SHARe AAs and conditional analyses in regions showing admixture signal (i.e., association between local ancestry and trait).

Study Subjects

The WHI is a U.S.-wide study focusing on common health issues in postmenopausal women. A total of 161,808 postmenopausal women aged 50–79 years old were recruited, including 12,151 self-identified AAs and 5,469 self-identified HAs. Fasting blood samples were collected at the baseline clinic visit by venipuncture. Clinical information was collected by self-report and physical examination. All participants provided written informed consent as approved by local Human Subjects Committees. Details of the study design and cohort characteristics have been previously described.[13] An independent cohort of 7,138 AA participants from the NHLBI Candidate-gene Association Resource (CARe) Study was used to validate those SNP-trait associations identified in WHI, which have not been reported in previous GWASs.[11]

Genotyping QC and Biomarker Measurement
Genotyping and QC
A cohort of 8,515 self-identified AA and 3,642 self-identified HA participants from WHI, who had consented to genetic research, were selected for WHI SHARe (n = 12,157) and genotyped on the Affymetrix 6.0 array. Genotype quality control criteria included call rate, concordance rates for blinded and unblinded duplicates, and sex discrepancy. Furthermore, individuals whose genetic ancestries differ from self-reported ethnicities and one individual from each close relative pair were excluded. In total, 11,740 individuals passed all genotype and sample QC criteria (8,153 AA, 3,587 HA).[14,15] Details of the QC procedures have been described in previous WHI-SHARe studies.[16,17] Sample sizes for each stage of the study are displayed in Figure S1 available online; demographic and lipid trait variables are summarized in Table S1.

Lipid Measurements
HDL, LDL, and TG measurements were performed at the University of Minnesota by standard biochemical methods on the Roche Modular P Chemistry analyzer (Roche Diagnostics): HDL was measured in serum by the HDL-C plus third generation direct method; TG was measured in serum by Triglyceride GB reagent, and total cholesterol (TC) was measured in serum by a cholesterol oxidase method. The accuracy and precision of the lipid assays were regularly monitored with the CDC/NHLBI Lipid Standardization Program to control for any potential drift over time. LDL was calculated in serum specimens having a TG value < 400 mg/dl according to the formula of Friedewald et al.[18] Based on the LDL-lowering effects of statins, we estimated the pretreatment LDL value for individuals on lipid-lowering medication by dividing treated LDL values by 0.75. All analyses described here are based on analyzing the pretreatment LDL, although the results are largely consistent when treated LDL values were analyzed. For all association analyses, LDL values greater than 300 mg/dl and TG values greater than 650 md/dl were excluded. TG values were log-transformed. HDL values were also log-transformed to better satisfy the normality assumption for the linear and linear mixed effect models, although post hoc analyses indicated that the transformed and untransformed trait values yielded qualitatively similar results, and in particular the identical set of genome-wide significant loci.

Genotype Imputation
Imputed genotypes were examined in regions where at least one genotyped SNP achieved a p < 10^{-6} and in regions showing ancestry association in WHI AAs (p < 7 × 10^{-6}). GWAS data in the WHI AAs were prephased via MaCH, with options "–states 200" and "–rounds 50."[19] The imputation reference panel was derived from the 2012-02-14 release of the 1000 Genomes Project, which included 246 Africans/African Americans, 379 Europeans, 181 Americans, and 286 Asians, using minimac with default parameter settings.[20] We note that genome-wide imputation to 1000

Genomes Project data is in fact available for WHI-SHARe. However, we chose not to test all imputed markers at the beginning of the study based on a power consideration. The rationale is that, for untyped risk variants with common frequencies, with high probability, there would be a typed variant in moderate LD in the vicinity reaching the relaxed threshold of $p < 10^{-6}$. Additionally, this relaxed threshold can identify regions harboring untyped rare variants, which are well tagged by a SNP on the array. On the other hand, for regions harboring very rare variants that are not tagged by any markers on the array, our sample size is probably underpowered even if we tested the imputed genotype. For this reason, we did not expect an exhaustive scan of all imputed SNPs to be fruitful.

Population Structure and Ancestry Estimation

Population Structure and Genome-level Ancestry
Principal component analysis (PCA) was performed for AAs and HAs combined, using Eigenstrat[21] at 178,101 markers that were in common between our samples and the reference panels. We also determined individual ancestral proportions by using Frappe from 656,852 autosomal markers.[22] For both of these calculations we included 475 publicly available samples from ancestral populations (YRI and CEU from HapMap and East Asian and Native Americans from the Human Genome Diversity Project).[23]

Local Ancestry Estimation
For each AA individual in the sample, locus-specific ancestry (probabilities of whether an individual has 0, 1, or 2 alleles of African ancestry at each locus) was estimated with program SABER+, an extension of the SABER algorithm.[24] In brief, SABER+ uses a graphical model approach to adaptively capture local haplotype structure within each ancestral population, and thereby more accurately accounts for background linkage disequilibrium (LD). In the current analysis, phased haplotype data from the HapMap3 CEU and YRI individuals were augmented as the reference panels. In simulation studies, SABER+ has an error rate of less than 2%, similar in accuracy to another commonly used method, HapMix.[25] Based on analysis of simulated and real AA genotypes, the correlation between local-ancestry estimates produced by the two methods is greater than 0.98 (N.A.J., unpublished data).

GWAS Analysis

The overall GWAS strategy was as follows: we initially tested genotyped SNPs in WHI AAs and HAs separately, with $\alpha = 5 \times 10^{-8}$ as the threshold for genome-wide significance. In regions showing suggestive evidence (at least one SNP with $p < 10^{-6}$), imputed genotypes were tested. To validate SNPs not previously associated with lipid traits, we combined results from WHI AAs and CARe AAs by using the program METAL to perform sample-size weighted meta-analysis.[26] Association analysis in CARe used individual-level genotype data and corrected for both PC1 and subcohort membership. Genome-wide association (GWA) analysis was performed under an additive genetic model using linear regression adjusted for covariates and implemented in PLINK v.1.05.[14] Analyses were conducted separately for AAs and HAs. To correct for population stratification, the genome-wide European ancestry proportions, which have a correlation of 0.99 with PC1, were adjusted as covariates for AAs (PC2–PC10 showed no evidence of association with any of the lipid traits). For HAs, the first four principal components were adjusted as covariates (PC5–PC10 showed no evidence of association with any of the lipid traits). Age, age^2, BMI, and smoking history were included as covariates for all lipid

traits; additionally, fasting status was adjusted for LDL. The same regression model was used to test the imputed allelic dosage at each SNP, via MACH2QTL.[19] The same threshold of 5×10^{-8} was used to declare genome-wide statistical significance for the imputed genotypes, although we note that the significant regions identified remained the same when a more stringent threshold of 2.5×10^{-8} was adopted.[27]

Comparison of p Values and Genetic Effects between Populations

We characterized the shared genetic architecture between Europeans, AAs, and HAs by enrichment in statistical significance (p values), correlation in estimated genetic effects, and the proportions of variance explained by subsets of genomes. To delineate the enrichment in statistical significance, we asked whether SNPs showing suggestive evidence in Europeans (defined by $p < 10^{-5}$) tend to have small p values in AAs and HAs. The p values in Europeans were taken from one of the largest lipid meta-analysis to date.[11] The p values for WHI AAs and HAs were obtained from GWAS based on the single-marker additive model.

To compare genetic effects across populations, we again defined candidate lipid-associated SNPs based on the p values in the European GWAS ($p < 10^{-5}$) and computed correlation coefficients between the estimates in WHI AAs and HAs. Because all WHI participants were genotyped with the same platform and all lipid biomarkers were measured in a consistent manner, this analysis avoids potential artifacts resulting from nonoverlapping SNPs, allele flipping, and assay batch effects. Furthermore, because the SNPs were chosen based on the p values in a GWAS whose participants do not overlap with WHI, the estimated genetic effects in AAs and HAs do not suffer from the typical bias due to selecting the most significant SNPs (i.e., winner's curse).[28]

Estimation of Phenotypic Variance Explained by Subsets of Genomes

We developed a method, termed population overlap in genetic architecture (POGA), which tests the hypothesis that loci identified in one population explain a large proportion of the phenotypic variance in other populations. POGA is an extension of the mixed effects model of a polygenic trait, introduced by Yang and Visscher.[29] Under this approach, we prioritized the genomes into regions based on association evidence in Europeans and estimated the phenotypic variance explained in AAs and HAs, using a mixed effects model. We reasoned that if the genetic architecture overlaps between populations, loci showing the strongest evidence of association in one population would account for a disproportionate amount of phenotypic variation in another population.

POGA Algorithm
For each SNP genotyped in WHI, we assigned a priority score (denoted by u_m) as the smallest p values within a 20 kb neighborhood based on the European GWAS.[11] We then grouped the genome into 22 nested subsets, such that the first subset included regions around SNPs that reached genome-wide significance in the European GWAS ($p < 5 \times 10^{-8}$); the second subset included SNPs with the top 1% priority scores; the third and subsequent subsets incrementally included the 5%, 10%, … 100% of the markers according to u_m. For each subset s = 1,…,22, we computed the proportions of phenotypic variance explained with program GCTA (v.1.02) and the AI-REML algorithm.[30] The model is:

$$Y_i = \alpha + \beta Q_i + G_i^s + H_i^s + \varepsilon_i,$$

where Q are the fixed effect covariates (age and PCs) and G_i^s and H_i^s are the random effects representing the genetic effects resulting from markers included and excluded by set s respectively. The inclusion of H^s in the model is necessary because markers in set s can be in LD with flanking regions not in the set, and thus a model that includes G^s alone tends to overestimate the variance explained just by the genomic regions.

Genetic Relationship Matrix

In a standard application of GCTA, a genetic relationship matrix (GRM) is computed as the covariance matrix of the standardized genotype matrix, $w_{im} = (g_{im} - 2p_m)/\sqrt{2p_m(1-p_m)}$, where $g_{im} \in (0, 1, 2)$ represents the original genotype and p_m represents the allele frequencies. We have previously shown that relationship coefficients estimated this way can be biased in an admixed population, and we have developed a *reap* estimator, which does not suffer from such bias.[31] In brief, the *reap* estimator standardizes the genotype by an "admixture-adjusted" allele frequency: $w_{im} = (g_{im} - 2p_{im})/\sqrt{2p_{im}(1-p_{im})}$ where p_{im} is the expected allele frequencies given the individual's genome-wide ancestry proportions. For a given subset of markers, s, we computed the GRM by using the *reap* estimator, for markers included in the subset (G^s) and not included in the subset (H^s), and supplied these two GRM into GCTA. For each subset, denote the proportions of variance attributed to G_i^s as h_s^2. As an informal test of significance, we compared the estimated h_s^2 with a simple null model, under which we expect that the phenotypic variance explained by a randomly selected region is proportional to its coverage of the genome. Thus, a set of regions that includes $x\%$ of the genome would explain $h^2x\%$ of the phenotypic variance, where h^2 is the variance explained by the entire genome.

A Scenario with No Overlap in Genetic Architecture

To verify that the expected h_s^2, for a set of randomly selected regions, is proportional to the proportion of the genome included, we performed the following permutation experiment. First, the European p value list was permuted in a way that largely preserves the correlation structure between neighboring SNPs. The p values were sorted by chromosome and base pair location. The 22 autosomes were concatenated, circularized, and then cut at a random location. The resulting list of p values was then mapped to the original SNP positions, starting with the p-term of chromosome 1 and ending with the last SNP position on chromosome 22. This permuted list of p values were then used as if they were p values from a European GWAS study, and the genomes were regrouped based on these permuted p values. GCTA analyses were repeated and h_s^2 estimated for these randomly partitioned genomes. This permutation procedure was performed for each HDL, LDL, and TG p value list and used to estimate the proportion of LDL variance explained.

Admixture Mapping

With the estimated local ancestry, we performed an admixture mapping analysis in AAs to detect variants present at different frequencies among the European and African ancestral populations. In this analysis, we regressed lipid levels on locus-specific ancestry, adjusting for the same covariates as in the GWAS. The critical value for genome-wide significance level of admixture mapping is substantially lower than for the genotype test, because the recent admixing history gives rise to extensive correlation in local ancestry. Based on previous theoretical analysis and simulation results, a nominal p value of 7×10^{-6} yielded a genome-wide type I error of 0.05.[32] We chose not to perform admixture mapping in the HA sample because we expect such an analysis is severely underpowered as a result of the lack of availability of an appropriate Native American reference panel (which impacts estimation accuracy) and the smaller HA sample size (compared to the AA sample).

Conditional Analysis under Admixture Mapping Peaks

Genetic regions showing significant association with local ancestry tend to be broad because of the recent admixing history in AAs. To refine these admixture mapping regions and in an effort to reveal genes or variants that contribute to population-level trait differences, we performed conditional analyses with all typed and imputed SNPs in each region. We reason that variants that "explain" an admixture mapping peak should meet two criteria. First, these variants should show suggestive association with the trait conditioning on the genome-wide ancestry; and second, these variants should substantially reduce the local ancestry-trait association. To test the first criterion, we required the variants to have GWAS $p < 10^{-5}$. In genes where multiple rare variants (defined by an allele frequency of less than 1% in AAs) show association with blood lipids, we also evaluated gene-based association by using either dosage (sum of rare alleles) or an indicator defined as dosage > 0. To assess the second criterion, we used a joint regression model that includes both local ancestry and SNP genotype in addition to all covariates adjusted in the GWAS and required that the p value for local ancestry to be attenuated (i.e., less significant) by at least 100-fold compared to that in the model without the SNP genotype. In regions where multiple SNPs or genes meet both criteria, we performed a step-wise regression to nominate a set of SNPs that may jointly explain the local-ancestry association.

Results

Genome-wide Association Analysis

Genotype association analyses in AAs identified seven, five, and four loci significantly associated with HDL, LDL, and TG, respectively (Table 1 and Figure S1). For HDL, these included *CD36* (MIM 173510), *PPP1R3B* (MIM 610541), *LPL* (MIM 609708), *CETP* (MIM 118470), *LOC55908*, the *APOA/APOC* gene cluster (*APOA1, APOC3, APOA4, APOA5* [MIM 107680, MIM 107720, MIM 107690, MIM 606368]), and an intergenic locus on 21q22; for LDL, *PCSK9* (MIM 607786), *APOB* (MIM 107730), *ABCG8* (MIM 605460), *APOE* (MIM 107741), and *LDLR* (MIM 606945); and for TG, *LPL*, *APOA/APOC*, *APOC1* (MIM 107710), and *GCKR* (MIM 600842). The inflation factor for genomic control in AAs was 1.061, 1.057, and 1.046 for HDL, LDL, and TG, respectively, indicating adequate adjustment of population stratification.

With the exception of *CD36* and the locus on 21q22, all other loci are in proximity to previously implicated regions in GWASs in Europeans.[11] *CD36* is a scavenger receptor that binds long-chain fatty acids and lipoproteins; genetic variants in *CD36* have been associated with protection from various components of the metabolic syndrome (MetS) in AAs,[33,34] and CD36-deficient individuals have been observed to have higher HDL compared to controls.[35] Recently, association between a nonsynonymous

Chr	Index SNP in the Region[a]	Pos (hg18)	Minor/Major Allele	Minor Allele Frequency	Candidate Gene	MIM	Beta[b]	p Value	Trait
Genotype Association									
1	rs17111684	55398136	A/G	0.120	PCSK9	607786	−9.01	2.40×10^{-17}	LDL
2	rs12713956	21095010	G/A	0.183	APOB	107730	−4.86	3.74×10^{-08}	LDL
2	rs4665972[c]	27451601	T/C	0.123	GCKR	600842	0.065	1.05×10^{-08}	TG
2	rs4245791	43927935	G/A	0.143	ABCG8	605460	5.97	1.24×10^{-09}	LDL
7	rs2366858	80178558	C/A	0.173	CD36	173510	0.0325	5.59×10^{-10}	HDL
8	rs1461729	9224652	T/C	0.116	PPP1R3B	610541	−0.0355	7.39×10^{-09}	HDL
8	rs326	19863719	T/C	0.469	LPL	609708	0.0221	1.23×10^{-08}	HDL
8	rs326	19863719	T/C	0.469	LPL	609708	−0.0410	1.02×10^{-08}	TG
11	rs6589566	116157633	C/T	0.0176	APOA/APOC	107680, 107720, 107690, 606368	0.2066	4.99×10^{-14}	TG
11	chr11: 116,799,496[c]	116304706	C/A	0.0016	APOA/APOC	107680, 107720, 107690, 606368	0.409	1.08×10^{-12}	HDL
16	rs247617	55548217	A/C	0.262	CETP	118470	0.0619	1.48×10^{-44}	HDL
19	rs17249141[c]	11061008	T/C	0.0126	LDLR	606945	−32.93	2.43×10^{-17}	LDL
19	rs12979813	11203703	T/C	0.495	LOC55908	NA	−0.0235	1.99×10^{-09}	HDL
19	rs1160985	50095252	C/T	0.365	APOE	107741	6.772	1.87×10^{-21}	LDL
19	rs12721054	50114427	G/A	0.1137	APOC1	107710	−0.101	2.86×10^{-19}	TG
21	rs13046373	30982361	C/T	0.391			−0.0226	2.26×10^{-08}	HDL
Admixture Mapping									
9	rs10818782	98.5–101.3 Mb					0.0508	5.57×10^{-07}	HDL
11	rs11217785	118.6–122.1 Mb			UBASH3B	609201	0.0526	2.82×10^{-07}	HDL
1	rs1889209	54.8–55.5 Mb			PCSK9	607786	−8.50	2.11×10^{-06}	LDL

n = 7,917; 7,861; and 7,918 for HDL, LDL, and TG, respectively.
[a]SNPs with the lowest p value at a locus.
[b]For genotype association, the direction of the regression coefficient represents the effect of each extra minor allele. For admixture mapping, the direction of the regression coefficient represents the effect of an additional African-derived allele.
[c]Loci where no genotyped SNP reaches genome-wide significance (5×10^{-8}) but at least one genotyped SNP reaches a $p < 10^{-6}$ and at least one imputed SNP reaches $p < 5 \times 10^{-8}$ (see Methods).

SNP in CD36 (rs3211938, p.Tyr325Ter) and HDL was reported in a meta-analysis in CARe.[12] The same SNP achieved a $p < 7 \times 10^{-9}$ in WHI and $p < 1.45 \times 10^{-19}$ in the meta-analysis that included AA participants in WHI and CARe. The HDL-increasing allele (G) occurs at a frequency of 0.28 in YRI and is essentially absent in Europeans and East Asians. At 21q22, two SNPs reached genome-wide significance in WHI AAs for HDL. This region has not been previously implicated in lipid genetics and was not replicated in CARe (meta $p = 4.10 \times 10^{-6}$); therefore, we did not pursue this locus further in this study.

In the WHI HA cohort, APOA/APOC and CETP reached genome-wide significant association with HDL and GCKR, LPL, and APOA/APOC were found significantly associated with TG. The number of HA participants is much lower compared to AA participants in WHI; therefore, we expect, a priori, that the power of GWAS is much lower in HAs (Table 2 and Figure S2).

Genetic Variants Associated with Lipid Traits Overlap across Ethnicities and Have Correlated Allelic Effects

Previous studies of traits such as lipids, height, BMI, and coronary heart disease have found that genetic risk factors identified in populations of European descent often show consistent direction of effect in non-European populations.[3] In WHI, a majority of lipid loci reaching the genome-wide significance level in AAs and all loci found in HAs overlapped with loci discovered in European populations.[11] Furthermore, we observed a strong enrichment of small p values in the WHI cohorts among those SNPs showing significance or suggestive evidence of association in the European GWAS ($p < 10^{-5}$) (Figure 1); as a comparison, the p value distribution for all SNPs appears roughly uniform (Figure S3). This suggests that, with a sufficient sample size, a large fraction of loci influencing lipid traits in Europeans would probably reach statistical significance in AAs and HAs.

Table 2. Genome-wide Significant Regions in WHI HAs

Chr	Index SNP in the Region[a]	Pos (hg 18)	Minor/Major Allele	Minor Allele Frequency	Candidate Gene	MIM	Beta[b]	p Value	Trait
2	rs780094	27594741	C/T	0.358	*GCKR*	600842	0.0688	7.35×10^{-09}	TG
8	rs17410962	19892360	G/A	0.112	*LPL*	609708	−0.1064	7.35×10^{-09}	TG
11	rs964184	116154127	G/C	0.248	*APOA/APOC*	107680, 107720, 107690, 606368	−0.0459	2.81×10^{-12}	HDL
11	rs964184	116154127	G/C	0.248	*APOA/APOC*	107680, 107720, 107690, 606368	0.1567	3.66×10^{-33}	TG
16	rs247617	55548217	C/A	0.298	*CETP*	118470	0.0509	3.48×10^{-16}	HDL

n = 3,506; 3,425; and 3,506 for HDL, LDL, and TG, respectively.
[a]SNPs with the lowest p value at a locus.
[b]The direction of the regression coefficient represents the effect of each extra minor allele.

Although previous studies have focused on consistency of the direction of genetic effects between populations, it is desirable to assess whether the risk variants have genetic effects of similar magnitude because this quantity plays an important role in individual disease risk prediction. Furthermore, the correlation in allelic effects is a more informative test because it also considers the strength of correlation. A caveat in this analysis is that, in order to avoid bias resulting from regression to the mean (or winner's curse), the effects should be estimated from a sample that is independent of the sample used to select the risk variants. Therefore, we defined lipid-associated SNPs based on the results from the largest European GWAS to date[11] and compared the estimated genetic effects per allele, or allelic effect, in WHI AAs and HAs. We examined SNPs with $p \leq 10^{-5}$ in Europeans for HDL, LDL, and TG, respectively. For HDL, 992 SNPs were considered and the correlation of the estimated allelic effects in AAs and HAs was 0.498; for LDL, 786 SNPs were considered and the correlation was 0.429; for TG, 810 SNPs were considered and the correlation was 0.602 (Figures 2A–2C). As a negative control, we repeated this analysis on SNPs with p values greater than 0.01 in Europeans. As expected, estimated allelic effects in AAs and HAs were essentially uncorrelated (0.0113, 0.00241, and 0.00519 for HDL, LDL, and TG, respectively). Because some SNPs included in this analysis may be in LD, the observed correlation between AAs and HAs may be overestimated. To eliminate this possibility, Figure 2D displays the estimated genetic effects in HAs versus AAs for 92 index SNPs from distinct loci implicated in the European GWAS for HDL, LDL, and TG. The Pearson correlation between the estimated effects was 0.69 (p = 1.11 × 10^{-14}), despite the fact that only 34 and 40 of the 92 SNPs had nominal p values less than 0.05 in AAs and HAs, respectively. These results suggest that the degree of overlap in risk loci between populations exceeds the set of SNPs that replicate on the basis of a predefined p value or significance threshold; therefore, combining association evidence across populations is likely to increase the overall efficiency of the ability to identify new trait loci.

Quantifying "Overlapping Heritability"

We next asked whether loci associated in European populations contribute substantially to the genetic architecture of lipid traits in AAs, in the sense of phenotypic variance explained. The approach we developed, termed population overlap in genetic architecture (POGA), is based on an extension of the mixed effects model of a polygenic trait, introduced by Yang and Visscher.[29] The rationale of POGA is intuitive: if the "important" risk loci overlap between populations, then we expect that the loci showing the strongest evidence of association in one population will account for a substantial proportion of the phenotypic variation in another population. This approach has two features. First it makes use of the complete list of p values from a reference GWAS instead of just the top SNPs; and second, this method defines "overlap" more broadly to accommodate allelic heterogeneity and unknown LD patterns.

Figure 3 displays the proportion of phenotypic variance in AAs explained by G^s when the markers in s expand from regions that met genome-wide significance levels in Europeans to the entire genome, and will be referred to as a POGA plot. The entire genome explains 28.5%, 30.5%, and 18.8% of the additive variance for HDL, LDL, and TG, respectively, of which nearly one-third can be attributed to the top 1% of the genome showing strongest evidence of association in Europeans and more than half can be attributed to the top 10% of the genome. The excess variance explained by the top regions is statistically significant in the sense that the 95% confidence intervals (with the standard error estimated by GCTA) do not include the expected values under the null model. In HAs, too, genomic regions showing association in Europeans account for disproportionate phenotypic variance, although the uncertainties associated with the estimates are greater because of the smaller sample size (Figure S5). In contrast, when we divided the genomic regions according to a permuted list of p values, the proportion of variance increased roughly linearly as a proportion of the genome included; not a single point significantly deviated from the expectation under the null model in the POGA plot (Figure S4).
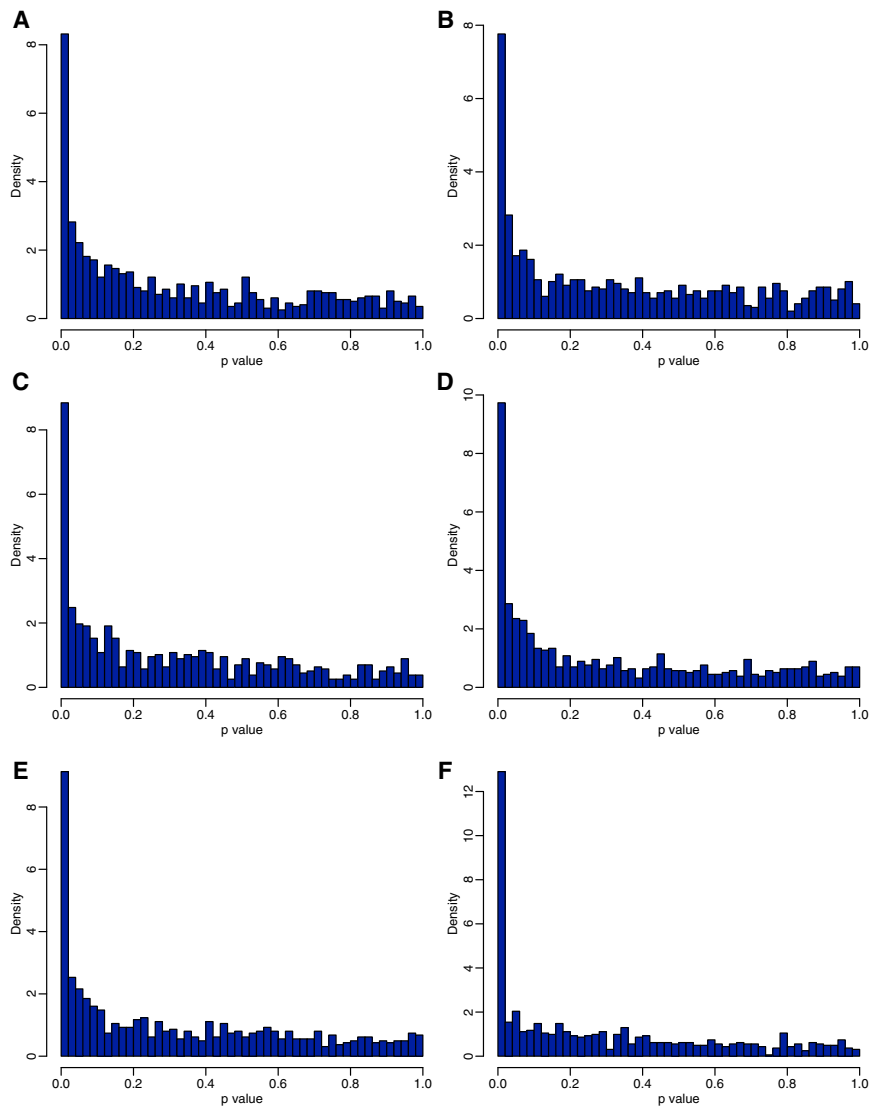
Figure 1. Enrichment for Small p Values among SNPs that Are Significantly or Suggestively Associated in European GWASs
(A and B) High-density lipoprotein (HDL) cholesterol in African Americans (AAs) (A) and in Hispanic Americans (HAs) (B).
(C and D) Low-density lipoprotein (LDL) cholesterol in AAs (C) and in HAs (D).
(E and F) Triglycerides (TG) in AAs (E) and in HAs (F).

wide significance threshold of 7×10^{-6}.[32] In the genome-wide admixture scan, local African ancestry at 9q22 (p = 2.82×10^{-7}) and at 11q23 (p = 5.57×10^{-7}) was associated with increased HDL (Table 1, Figure S6). Together, the ancestry at these two loci explains 0.6% of variation in HDL, and genome-wide ancestry was no longer significantly associated with the trait upon adjusting for local ancestry at the two loci. At 1p32, local African ancestry was associated with decreased LDL (p = 2.11×10^{-6}), explaining <0.3% of phenotypic variation. Somewhat surprisingly, no locus reached the genome-wide significance level for TG, despite the strong correlation between the genome-wide African ancestry and TG.

Association between local ancestry and a trait can arise because the region harbors population-specific risk variants or because it harbors risk variants shared among populations that occur at disparate allele frequencies. Thus, regions showing local ancestry association are excellent candidates for identifying the genetic basis that underlies trait differences between ethnicities. Below we describe further fine-scale characterization of each admixture-mapping peak. These analyses aim to identify genes or variants that are associated with lipids and explain the local ancestry-trait association. We note a caveat that, because admixture generates extensive linkage disequilibrium, these markers may simply be in LD with the true causal variants.

Ancestry Association and Admixture Mapping in AAs
We next investigated factors that underlie ethnic differences in lipid levels. Adjusting for the same set of covariates as used in the AA GWAS, genome-wide African ancestry was associated with increased HDL (β = 0.10, p = 4.26×10^{-7}) and decreased TG (β = −0.33, p < 10^{-16}) and was not significantly associated with LDL (p = 0.178). This pattern is consistent with a previous analysis of an independent AA cohort in the Family Blood Pressure Program.[36,37] The phenotypic variance attributed to genome-wide ancestry is very low: less than 0.3% for HDL and less than 1% for TG. Comparing HAs and AAs in WHI, HAs have significantly lower HDL (β = −0.10, p < 2×10^{-16}), lower LDL (β = −3.56, p = 5.08×10^{-5}), and higher TG (β = 0.37, p < 2×10^{-16}).

Admixture mapping is an effective approach for identifying loci with strong effects and disparate allele frequencies between ancestral populations. Because the recent admixing history induces strong correlation in admixture mapping test statistics, we adopted a genome-

Admixture Association with LDL at 1q32
In the region of the 1q32 admixture signal for LDL (UCSC Genome Browser hg18: 54.8–55.5 Mb), genotype association analyses revealed three low-frequency nonsynonymous variants in *PCSK9* associated with decreased LDL (rs28362286 [p.Cys679Ter], rs28362263 [p.Ala443Thr], and rs28362261 [p.Asn425Ser]); the first two variants were associated with decreased LDL in another AA
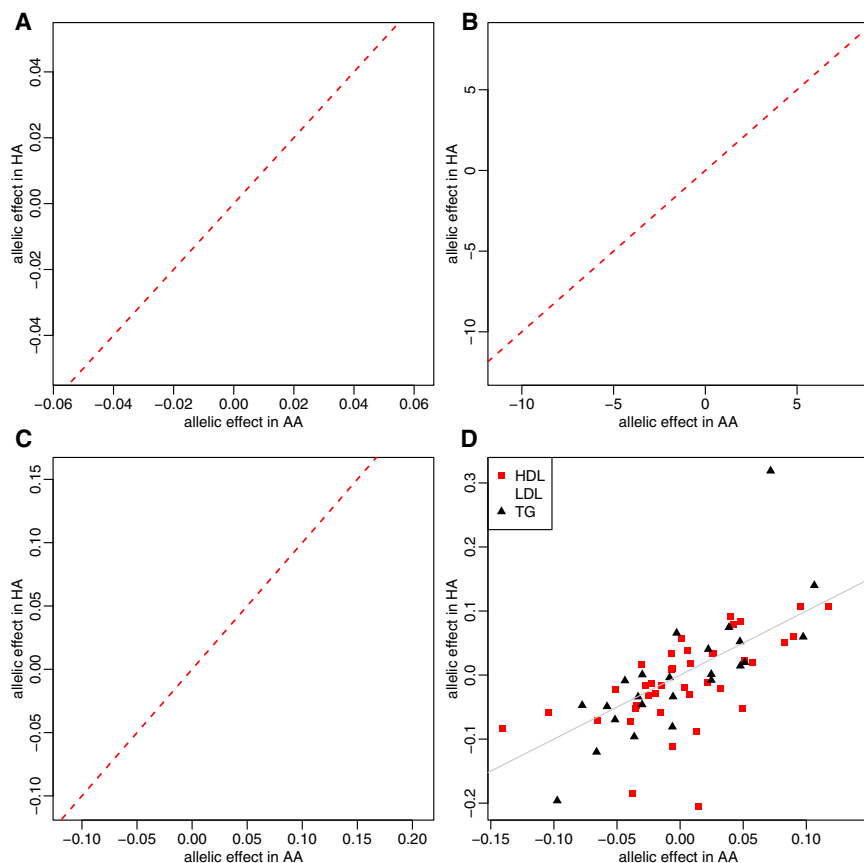
Figure 2. Genetic Effects of Candidate Lipid Variants Are Correlated between African Americans and Hispanic Americans
(A–C) SNPs are those with $p < 10^{-5}$ in a European GWAS. The *x* axis represents the estimated allelic effects in African Americans (AAs) and the *y* axis represents the estimated allelic effects of the corresponding SNPs in Hispanic Americans (HAs).
(A) High-density lipoprotein (HDL) cholesterol.
(B) Low-density lipoprotein (LDL) cholesterol.
(C) Triglycerides (TG).
(D) Genetic effects of 92 SNPs representing independent loci; SNPs are the best surrogate index SNPs on Affy 6.0 arrays defined by Telosvich et al.[11] The estimated allelic effects in (D) are in the unit of standard deviation of the phenotype.

CORO2A, rs1537960 in *GABBR2* (MIM 607340), and rs4149310 in *ABCA1* (Table S3). At rs1537960 and rs751800, the HDL-decreasing alleles are essentially absent in YRI, whereas the HDL-decreasing allele at rs4149310 occurs in both CEU and YRI but at a much higher frequency in CEU. Intriguingly, previous transcriptomic studies have identified rs751800, a SNP in the 3′ UTR of *CORO2A*, as a *cis*-eQTL in monocytes;[41] based on data generated from the ENCODE project, this SNP is categorized as likely to affect transcription factor binding (category 1b in RegulomeDB v.1.0).[42] On the other hand, rs751800 was only weakly associated with HDL in the CARe cohort ($p = 0.05$) and showed no association in Europeans.[11] We note that rs4149310 also shows suggestive association in European populations ($p = 8.46 \times 10^{-6}$); in contrast, the *ABCA1* SNP with the strongest HDL association in Europeans, rs1883025, had similar frequencies in YRI and CEU and was not significantly associated with HDL in AAs ($p = 0.215$). Thus, rs1883025 cannot explain the HDL-ancestry association. In summary, the ancestry-HDL association on 9q22 can be attributed to variants in *CORO2A*, *GABBR2*, and *ABCA1*. Although *ABCA1* appears to influence HDL in both Europeans and AAs, the risk variants are either distinct in the two populations or have very different allele frequencies.

cohort,[38] and the third was validated in CARe (meta-analysis $p = 1.43 \times 10^{-10}$). Conditioning on the dosage or carrier status of these three variants substantially reduced the local ancestry association ($p = 1.15 \times 10^{-2}$). Curiously, we note that a nonsynonymous variant in *PCSK9*, rs505151 (p.Gly670Glu), was associated with LDL, but the allelic effect was opposite to the local ancestry effect. The minor allele (G) at this SNP was associated with higher LDL ($p = 1.09 \times 10^{-6}$ in WHI, 4.31×10^{-09} in CARe, and 2.42×10^{-12} in Europeans[11]); this allele occurs in YRI at a frequency of 0.336 but at a much lower frequency of 0.031 in Europeans. Indeed, a joint analysis that included local ancestry and rs505151 showed stronger LDL association with both variables (Table S2).

Admixture Association with HDL at 9q22
Local African ancestry at 9q22 (UCSC Genome Browser hg18: 98.5–101.4 Mb) was associated with increased HDL. From a biological standpoint, the nearest gene with a plausible role in lipid metabolism is *CORO2A* (MIM 602159), which encodes coronin 2A. Coronin 2A (CORO2A) is a nuclear receptor corepressor (NCoR) exchange factor that interacts with cholesterol-sensing liver X receptor to derepress inflammatory genes in macrophages.[39,40] The peak local ancestry association is 5 Mb away from *ABCA1* (MIM 600046), a known HDL locus identified in Europeans. The ancestry association signal vanished upon conditioning on three SNPs: rs751800 in

Admixture Association with HDL at 11q23
Local African ancestry at chromosomal region 11q23 (UCSC Genome Browser hg18: 118.6–122.1 Mb) was also associated with increased HDL. The region is flanked at either end by two HDL loci identified in Europeans, *UBASH3B* (MIM 609201) and the *APOA/APOC* gene cluster. At the *APOA/APOC* locus, 13 SNPs were associated with HDL ($p < 5 \times 10^{-8}$) in WHI AAs. The minor alleles were
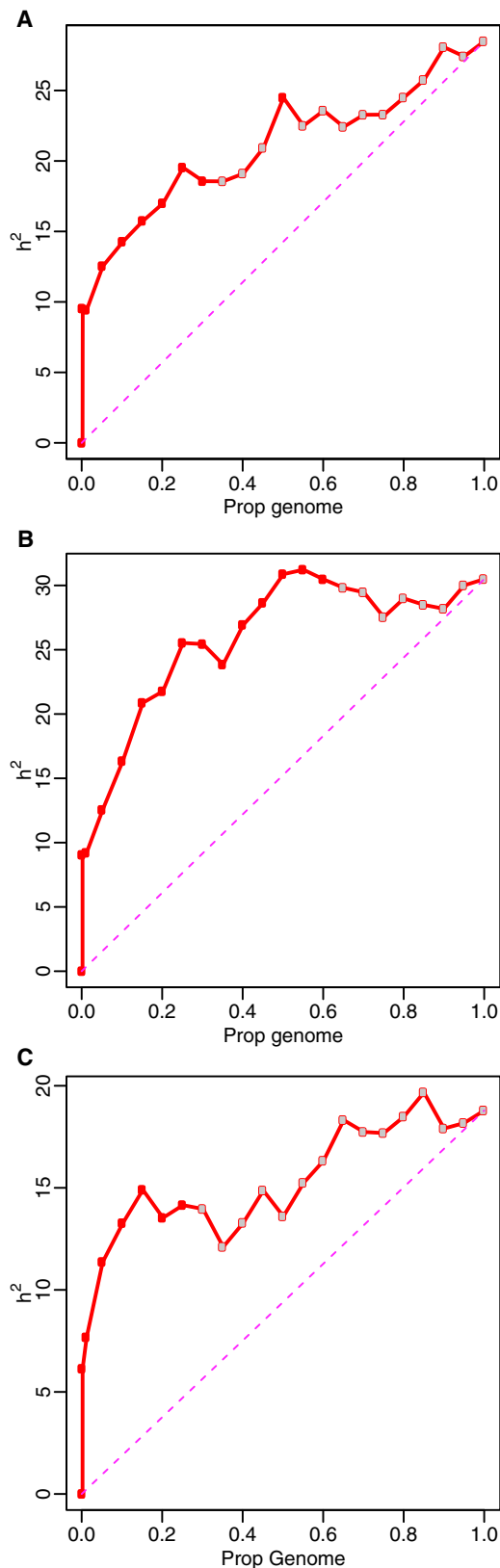
Figure 3. Overlap in Genetic Architecture in AAs
Genomic regions are ranked by the SNP with the strongest association evidence in a European GWAS. Proportions of the phenotypic variance explained in AAs (y axis) by the top x% of the genome (x axis) are estimated with a mixed effect model. Red points indicates the 95% confidence interval excludes the expected value (magenta line) under the null model that x% randomly selected genome explains $h^2 x$% of phenotypic variance, where h^2 is the variance explained by the entire genome. When the null cannot be excluded, the point estimates are drawn with gray points.
(A) High-density lipoprotein (HDL) cholesterol.
(B) Low-density lipoprotein (LDL) cholesterol.
(C) Triglycerides (TG).

each associated with decreased HDL and are all very rare (<1% in AAs). Collectively, the number of rare variants carried by WHI AA individuals was positively correlated with African ancestry ($p = 0.02$) and strongly correlated with HDL ($p = 6.78 \times 10^{-14}$). We performed conditional analysis of the local-ancestry-HDL association, adjusting for either single markers, combined dosage across all markers, or a 0/1 variable indicating whether an individual harbors any of the rare variants, but none of these variables explained the ancestry peak (Table S4). At *UBASH3B*, the most significant SNP in Europeans (rs7115089) and the SNP showing the strongest association in AAs (rs7107934 [$p = 2.49 \times 10^{-7}$]), both occur at similar frequencies in YRI and CEU and thus cannot explain the local ancestry association. No other variants were found in the 2 Mb region around *UBASH3B* that could explain the HDL admixture association signal. Because the local ancestry association at 11q23 cannot be attributed to variants in *APOA/APOC* or *UBASH3B*, the region probably harbors unrecognized HDL-influencing loci. A plausible candidate is *ABCG4* (MIM 607784), which has been shown to promote cholesterol efflux to HDL-like particles.[43,44] However, single marker association analysis in *ABCG4* did not identify significant genotype-HDL association.

Discussion

In this study, we have demonstrated a substantial overlap in genes that contribute to the variation in plasma lipid levels among human populations. At the same time, substantial allelic heterogeneity has been observed within the shared loci, which contributes to the ethnic variation in lipid levels.

A Shared Genetic Basis of Lipid Phenotypes among Human Populations
Previous studies have found that variants associated with lipid traits in Europeans can be "replicated" in non-European populations by using criteria of relaxed statistical stringency and consistent direction of genetic effects.[3,11] By using several approaches, we demonstrate evidence for even greater shared genetic components beyond the known replicated loci. First, we find a strong enrichment of small p values, in both AAs and HAs, among SNPs that show suggestive evidence of association in Europeans ($p < 10^{-5}$) (Figure 1). With an increased sample size, the fraction of variants that can be replicated in non-European populations is likely to increase. Second, despite some

well-established examples in which a genetic variant confers striking population-specific risk,[45,46] the estimated genetic effects in AAs and HAs, at loci chosen based on the association evidence in Europeans, show strong correlations in both direction and magnitude (Figure 2). This is even more striking considering the moderate sample sizes of AAs and HAs; thus the interethnic correlation in SNP effects will probably be strengthened with larger sample size and reduced sampling errors. Finally, POGA plots demonstrate that the portion of the human genome that contributes importantly to lipid variation in Europeans also contributes substantially to phenotypic variation in non-Europeans: for the majority of lipid traits we examined in AAs and HAs, more than half of the additive phenotypic variance explained by the entire genome can be attributed to the 10% of the genome showing the strongest evidence of association in Europeans (Figures 3 and S5). This last approach is particularly useful because it defines an overlapping genetic architecture more broadly by taking into account the possibility that, although a locus influences the trait in multiple populations, the precise variants may not be identical. Furthermore, a variant discovered in a GWAS may simply be a tagging SNP for a functional variant; hence, even when a causal variant is shared between populations, association with the tagging SNP may not replicate across populations because of different LD patterns. Estimating the proportion of phenotypic variation that is explained by a part of the genome accommodates both true allelic heterogeneity and population-specific LD patterns and has an intuitive interpretation as the contribution by a set of candidate loci to phenotypic variance. Again, the degree of overlap in genetic basis represents a lower bound because of the limited sample sizes.

Genetic Factors that Underlie Lipid Trait Differences between Populations

Although the distribution of plasma-lipid concentration overlaps considerably between populations, these traits also vary between populations. Because shared ancestry can confound genetic and nongenetic risk factors, it is difficult to quantify the genetic and nongenetic contribution to population differences in traits such as lipid levels. However, genetic association and admixture mapping analyses identified a number of genetic factors that underlie the ethnic differences. Furthermore, these examples illustrate that both common and rare variants can contribute to population differences. As expected, rare variants that contribute to population differences tend to be population specific, exemplified by multiple African-specific variants in *PCSK9* (associated with LDL) and in *APOA/APOC* (associated with HDL). However, we also identified common variants that underlie population differences: the African-specific *CD36* variant, Tyr325Ter, has a frequency of 0.28 in YRI; the HDL-increasing allele at rs4149310 in *ABCA1* has a frequency of 0.75 and 0.14 in YRI and CEU, respectively, and explains a large proportion of the

ancestry association at 9q22 with HDL (Table S2). It is also interesting to note that *PCSK9*, *APOA/APOC*, and *ABCA1* all play a role in lipid genetics in both Europeans and AAs, yet the precise alleles within each locus differ between the populations. This supports a view that lipid loci are largely shared among populations, but the allelic structure within a locus has been shaped by population history and thus can exhibit considerable heterogeneity. The presence of allelic heterogeneity has also been demonstrated in a recent transethnic study of lipid trait via the Metabochip.[47]

Admixture mapping has successfully detected loci that contribute to phenotypic diversity between populations. In most admixture mapping examples, where the genetic factors that give rise to the ancestry association are known, the ancestry-phenotype association can be largely attributed to a single variant with disparate allele frequencies between populations. Examples include *SLC24A5* (MIM 609802) for skin pigmentation, *DARC* (MIM 613665) null allele for white blood cells, and *APOL1* (MIM 603743) for kidney diseases; the striking genetic differentiation at the implicated loci signifies selective sweep.[48–50] In contrast, of the three loci showing ancestry-trait association for lipid levels (1p32, 9q22, and 11q23), none could be attributed to single common variants. Instead, two or more variants explain the ancestry association in an additive fashion in each region. Furthermore, in at least one instance (*PCSK9*), we find African-specific coding variants that are associated with either increased or decreased LDL. Our interpretation of these observations is that the evolutionary dynamics of lipid traits in humans are influenced by a combination of forces and are in contrast to the strong directional adaptation that diversifies, for example, skin pigmentation. Rather, the many low-frequency coding variants may have been subjected to weak purifying selection,[51] whereas the more common variants may have been driven to disparate population allele frequencies by other mechanisms, including balancing selection and random drift.[52]

Implication for Future GWASs

Racial and ethnic minorities constitute a growing proportion of the US population and suffer disproportionately higher rates of CVD.[53] The abundance of population-specific variants that underlie lipid traits highlights the importance of including individuals of diverse ethnic background in future GWASs. At the same time, the identification of a large fraction of ethnically shared trait loci suggests that further insights into the genetic mechanisms that underlie lipid traits can be gained by continued study of multiple ethnicities simultaneously. Based on published GWASs that focus on populations of European decent, a majority of complex traits and diseases are influenced by a large number of loci, each conferring a modest risk. Studies with sample size exceeding 100,000 are nonetheless underpowered, as indicated by the substantial fraction of "missing heritability."[54] Thus, the scarcity of cohorts

that represent AAs, HAs, and other minority populations poses a challenge for understanding the genetic basis of complex traits in these populations. Take coronary heart disease (CAD) as an example: the CARDIoGRAM GWAS cohort numbered 82,000 Europeans,[55] compared to the largest published AA cohort of 8,000.[56] Moreover, an increasing proportion of the world's populations do not fall into conventional ethnic categories (e.g. individuals with mixed African and South Asian ancestry). In contrast to studying each ethnic population separately, a multi-ethnic approach that integrates evidence across populations will probably be more efficient, both for gene discovery and for individual risk prediction. One analytic approach has been proposed and takes the form of a "trans-ethnic meta-analysis;" it is desirable to extend this approach to accommodate allelic heterogeneity and differential LD patterns.[57,58] The ability to harness information across multiple populations will provide a more complete portrait of the genetic bases that underlie complex traits in human species, which will in turn allow all people to benefit from the new paradigm of personalized medicine.

Supplemental Data

Supplemental Data include six figures and four tables and can be found with this article online at http://www.cell.com/AJHG/.

Acknowledgments

Web Resources

The URLs for data presented herein are as follows:

1000 Genomes, http://browser.1000genomes.org
GCTA, http://www.complextraitgenomics.com/software/gcta/
Human Genome Diversity Project, http://hagsc.org/hgdp/files.html
Imputation Reference Panel, http://www.sph.umich.edu/csg/yli/mach/download/1000G.2012-02-14.html
International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/
RegulomeDB, http://RegulomeDB.org/
SABER+, http://med.stanford.edu/tanglab/software/saber.html
UCSC Genome Browser, http://genome.ucsc.edu

References

1. Namboodiri, K.K., Kaplan, E.B., Heuch, I., Elston, R.C., Green, P.P., Rao, D.C., Laskarzewski, P., Glueck, C.J., and Rifkind, B.M. (1985). The Collaborative Lipid Research Clinics Family Study: biological and cultural determinants of familial resemblance for plasma lipids and lipoproteins. Genet. Epidemiol. 2, 227–254.

2. Kannel, W.B., Dawber, T.R., Kagan, A., Revotskie, N., and Stokes, J., 3rd. (1961). Factors of risk in the development of coronary heart disease—six year follow-up experience. The Framingham Study. Ann. Intern. Med. 55, 33–50.

3. Dumitrescu, L., Carty, C.L., Taylor, K., Schumacher, F.R., Hindorff, L.A., Ambite, J.L., Anderson, G., Best, L.G., Brown-Gentry, K., Bůžková, P., et al. (2011). Genetic determinants of lipid traits in diverse populations from the population architecture using genomics and epidemiology (PAGE) study. PLoS Genet. 7, e1002138.

4. LaRosa, J.C., and Brown, C.D. (2005). Cardiovascular risk factors in minorities. Am. J. Med. 118, 1314–1322.

5. Metcalf, P.A., Sharrett, A.R., Folsom, A.R., Duncan, B.B., Patsch, W., Hutchinson, R.G., Szklo, M., Davis, C.E., and Tyroler, H.A. (1998). African American-white differences in lipids, lipoproteins, and apolipoproteins, by educational attainment, among middle-aged adults: the Atherosclerosis Risk in Communities Study. Am. J. Epidemiol. 148, 750–760.

6. Gartside, P.S., Khoury, P., and Glueck, C.J. (1984). Determinants of high-density lipoprotein cholesterol in blacks and whites: the second National Health and Nutrition Examination Survey. Am. Heart J. 108, 641–653.

7. Wei, M., Mitchell, B.D., Haffner, S.M., and Stern, M.P. (1996). Effects of cigarette smoking, diabetes, high cholesterol, and hypertension on all-cause mortality and cardiovascular disease mortality in Mexican Americans. The San Antonio Heart Study. Am. J. Epidemiol. 144, 1058–1065.

8. Beekman, M., Heijmans, B.T., Martin, N.G., Pedersen, N.L., Whitfield, J.B., DeFaire, U., van Baal, G.C., Snieder, H., Vogler, G.P., Slagboom, P.E., and Boomsma, D.I. (2002). Heritabilities of apolipoprotein and lipid levels in three countries. Twin Res. 5, 87–97.

9. Weiss, L.A., Pan, L., Abney, M., and Ober, C. (2006). The sex-specific genetic architecture of quantitative traits in humans. Nat. Genet. 38, 218–222.

10. Luo, B.F., Du, L., Li, J.X., Pan, B.Y., Xu, J.M., Chen, J., Yin, X.Y., Ren, Y., and Zhang, F. (2010). Heritability of metabolic syndrome traits among healthy younger adults: a population based study in China. J. Med. Genet. 47, 415–420.

11. Teslovich, T.M., Musunuru, K., Smith, A.V., Edmondson, A.C., Stylianou, I.M., Koseki, M., Pirruccello, J.P., Ripatti, S., Chasman, D.I., Willer, C.J., et al. (2010). Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466, 707–713.

12. Elbers, C.C., Guo, Y., Tragante, V., van Iperen, E.P., Lanktree, M.B., Castillo, B.A., Chen, F., Yanek, L.R., Wojczynski, M.K., Li, Y.R., et al. (2012). Gene-centric meta-analysis of lipid traits

in African, East Asian and Hispanic populations. PLoS ONE *7*, e50198.

13. Hays, J., Hunt, J.R., Hubbell, F.A., Anderson, G.L., Limacher, M., Allen, C., and Rossouw, J.E. (2003). The Women's Health Initiative recruitment methods and results. Ann. Epidemiol. *13*(9, Suppl), S18–S77.

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

15. Thornton, T., and McPeek, M.S. (2010). ROADTRIPS: case-control association testing with partially or completely unknown population and pedigree structure. Am. J. Hum. Genet. *86*, 172–184.

16. Reiner, A.P., Beleza, S., Franceschini, N., Auer, P.L., Robinson, J.G., Kooperberg, C., Peters, U., and Tang, H. (2012). Genome-wide association and population genetic analysis of C-reactive protein in African American and Hispanic American women. Am. J. Hum. Genet. *91*, 502–512.

17. Carty, C.L., Johnson, N.A., Hutter, C.M., Reiner, A.P., Peters, U., Tang, H., and Kooperberg, C. (2012). Genome-wide association study of body height in African Americans: the Women's Health Initiative SNP Health Association Resource (SHARe). Hum. Mol. Genet. *21*, 711–720.

18. Friedewald, W.T., Levy, R.I., and Fredrickson, D.S. (1972). Estimation of the concentration of low-density lipoprotein cholesterol in plasma, without use of the preparative ultracentrifuge. Clin. Chem. *18*, 499–502.

19. Li, Y., Willer, C.J., Ding, J., Scheet, P., and Abecasis, G.R. (2010). MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genet. Epidemiol. *34*, 816–834.

20. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J., and Abecasis, G.R. (2012). Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. Nat. Genet. *44*, 955–959.

21. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A., and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

22. Tang, H., Peng, J., Wang, P., and Risch, N.J. (2005). Estimation of individual admixture: analytical and study design considerations. Genet. Epidemiol. *28*, 289–301.

23. Li, J.Z., Absher, D.M., Tang, H., Southwick, A.M., Casto, A.M., Ramachandran, S., Cann, H.M., Barsh, G.S., Feldman, M., Cavalli-Sforza, L.L., and Myers, R.M. (2008). Worldwide human relationships inferred from genome-wide patterns of variation. Science *319*, 1100–1104.

24. Tang, H., Coram, M., Wang, P., Zhu, X., and Risch, N. (2006). Reconstructing genetic ancestry blocks in admixed individuals. Am. J. Hum. Genet. *79*, 1–12.

25. Price, A.L., Tandon, A., Patterson, N., Barnes, K.C., Rafaels, N., Ruczinski, I., Beaty, T.H., Mathias, R., Reich, D., and Myers, S. (2009). Sensitive detection of chromosomal segments of distinct ancestry in admixed populations. PLoS Genet. *5*, e1000519.

26. Willer, C.J., Li, Y., and Abecasis, G.R. (2010). METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics *26*, 2190–2191.

27. Pe'er, I., Yelensky, R., Altshuler, D., and Daly, M.J. (2008). Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genet. Epidemiol. *32*, 381–385.

28. Zollner, S., and Pritchard, J.K. (2007). Overcoming the winner's curse: estimating penetrance parameters from case-control data. Am. J. Hum. Genet. *80*, 605–615.

29. Yang, J., Benyamin, B., McEvoy, B.P., Gordon, S., Henders, A.K., Nyholt, D.R., Madden, P.A., Heath, A.C., Martin, N.G., Montgomery, G.W., et al. (2010). Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. *42*, 565–569.

30. Yang, J., Lee, S.H., Goddard, M.E., and Visscher, P.M. (2011). GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. *88*, 76–82.

31. Thornton, T., Tang, H., Hoffmann, T.J., Ochs-Balcom, H.M., Caan, B.J., and Risch, N. (2012). Estimating kinship in admixed populations. Am. J. Hum. Genet. *91*, 122–138.

32. Tang, H., Siegmund, D.O., Johnson, N.A., Romieu, I., and London, S.J. (2010). Joint testing of genotype and ancestry association in admixed families. Genet. Epidemiol. *34*, 783–791.

33. Love-Gregory, L., Sherva, R., Sun, L., Wasson, J., Schappe, T., Doria, A., Rao, D.C., Hunt, S.C., Klein, S., Neuman, R.J., et al. (2008). Variants in the CD36 gene associate with the metabolic syndrome and high-density lipoprotein cholesterol. Hum. Mol. Genet. *17*, 1695–1704.

34. Love-Gregory, L., Sherva, R., Schappe, T., Qi, J.S., McCrea, J., Klein, S., Connelly, M.A., and Abumrad, N.A. (2011). Common CD36 SNPs reduce protein expression and may contribute to a protective atherogenic profile. Hum. Mol. Genet. *20*, 193–201.

35. Furuhashi, M., Ura, N., Nakata, T., and Shimamoto, K. (2003). Insulin sensitivity and lipid metabolism in human CD36 deficiency. Diabetes Care *26*, 471–474.

36. Basu, A., Tang, H., Lewis, C.E., North, K., Curb, J.D., Quertermous, T., Mosley, T.H., Boerwinkle, E., Zhu, X., and Risch, N.J. (2009). Admixture mapping of quantitative trait loci for blood lipids in African-Americans. Hum. Mol. Genet. *18*, 2091–2098.

37. Deo, R.C., Reich, D., Tandon, A., Akylbekova, E., Patterson, N., Waliszewska, A., Kathiresan, S., Sarpong, D., Taylor, H.A., Jr., and Wilson, J.G. (2009). Genetic differences between the determinants of lipid profile phenotypes in African and European Americans: the Jackson Heart Study. PLoS Genet. *5*, e1000342.

38. Huang, C.C., Fornage, M., Lloyd-Jones, D.M., Wei, G.S., Boerwinkle, E., and Liu, K. (2009). Longitudinal association of PCSK9 sequence variations with low-density lipoprotein cholesterol levels: the Coronary Artery Risk Development in Young Adults Study. Circ. Cardiovasc. Genet. *2*, 354–361.

39. Huang, W., Ghisletti, S., Saijo, K., Gandhi, M., Aouadi, M., Tesz, G.J., Zhang, D.X., Yao, J., Czech, M.P., Goode, B.L., et al. (2011). Coronin 2A mediates actin-dependent de-repression of inflammatory response genes. Nature *470*, 414–418.

40. Zelcer, N., and Tontonoz, P. (2006). Liver X receptors as integrators of metabolic and inflammatory signaling. J. Clin. Invest. *116*, 607–614.

41. Zeller, T., Wild, P., Szymczak, S., Rotival, M., Schillert, A., Castagne, R., Maouche, S., Germain, M., Lackner, K., Rossmann, H., et al. (2010). Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. PLoS ONE *5*, e10693.

42. Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C.,

Weng, S., et al. (2012). Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. *22*, 1790–1797.

43. Wang, N., Lan, D., Chen, W., Matsuura, F., and Tall, A.R. (2004). ATP-binding cassette transporters G1 and G4 mediate cellular cholesterol efflux to high-density lipoproteins. Proc. Natl. Acad. Sci. USA *101*, 9774–9779.

44. Wang, N., Yvan-Charvet, L., Lütjohann, D., Mulder, M., Vanmierlo, T., Kim, T.W., and Tall, A.R. (2008). ATP-binding cassette transporters G1 and G4 mediate cholesterol and desmosterol efflux to HDL and regulate sterol accumulation in the brain. FASEB J. *22*, 1073–1082.

45. Farrer, L.A., Cupples, L.A., Haines, J.L., Hyman, B., Kukull, W.A., Mayeux, R., Myers, R.H., Pericak-Vance, M.A., Risch, N., and van Duijn, C.M.; APOE and Alzheimer Disease Meta Analysis Consortium. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. JAMA *278*, 1349–1356.

46. Helgadottir, A., Manolescu, A., Helgason, A., Thorleifsson, G., Thorsteinsdottir, U., Gudbjartsson, D.F., Gretarsdottir, S., Magnusson, K.P., Gudmundsson, G., Hicks, A., et al. (2006). A variant of the gene encoding leukotriene A4 hydrolase confers ethnicity-specific risk of myocardial infarction. Nat. Genet. *38*, 68–74.

47. Wu, Y., Waite, L.L., Jackson, A.U., Sheu, W.H., Buyske, S., Absher, D., Arnett, D.K., Boerwinkle, E., Bonnycastle, L.L., Carty, C.L., et al. (2013). Trans-ethnic fine-mapping of lipid loci identifies population-specific signals and allelic heterogeneity that increases the trait variance explained. PLoS Genet. *9*, e1003379.

48. Reiner, A.P., Lettre, G., Nalls, M.A., Ganesh, S.K., Mathias, R., Austin, M.A., Dean, E., Arepalli, S., Britton, A., Chen, Z., et al. (2011). Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). PLoS Genet. *7*, e1002108.

49. Lamason, R.L., Mohideen, M.A., Mest, J.R., Wong, A.C., Norton, H.L., Aros, M.C., Jurynec, M.J., Mao, X., Humphreville, V.R., Humbert, J.E., et al. (2005). SLC24A5, a putative cation exchanger, affects pigmentation in zebrafish and humans. Science *310*, 1782–1786.

50. Genovese, G., Friedman, D.J., Ross, M.D., Lecordier, L., Uzureau, P., Freedman, B.I., Bowden, D.W., Langefeld, C.D., Oleksyk, T.K., Uscinski Knob, A.L., et al. (2010). Association of trypanolytic ApoL1 variants with kidney disease in African Americans. Science *329*, 841–845.

51. Tennessen, J.A., Bigham, A.W., O'Connor, T.D., Fu, W., Kenny, E.E., Gravel, S., McGee, S., Do, R., Liu, X., Jun, G., et al.; Broad GO; Seattle GO; NHLBI Exome Sequencing Project. (2012). Evolution and functional impact of rare coding variation from deep sequencing of human exomes. Science *337*, 64–69.

52. Charlesworth, B., Nordborg, M., and Charlesworth, D. (1997). The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet. Res. *70*, 155–174.

53. Mensah, G.A., Mokdad, A.H., Ford, E.S., Greenlund, K.J., and Croft, J.B. (2005). State of disparities in cardiovascular health in the United States. Circulation *111*, 1233–1241.

54. Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., et al. (2009). Finding the missing heritability of complex diseases. Nature *461*, 747–753.

55. Schunkert, H., König, I.R., Kathiresan, S., Reilly, M.P., Assimes, T.L., Holm, H., Preuss, M., Stewart, A.F., Barbalic, M., Gieger, C., et al.; Cardiogenics; CARDIoGRAM Consortium. (2011). Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. Nat. Genet. *43*, 333–338.

56. Lettre, G., Palmer, C.D., Young, T., Ejebe, K.G., Allayee, H., Benjamin, E.J., Bennett, F., Bowden, D.W., Chakravarti, A., Dreisbach, A., et al. (2011). Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARe Project. PLoS Genet. *7*, e1001300.

57. Morris, A.P. (2011). Transethnic meta-analysis of genomewide association studies. Genet. Epidemiol. *35*, 809–822.

58. Dastani, Z., Hivert, M.F., Timpson, N., Perry, J.R., Yuan, X., Scott, R.A., Henneman, P., Heid, I.M., Kizer, J.R., Lyytikäinen, L.P., et al.; DIAGRAM+ Consortium; MAGIC Consortium; GLGC Investigators; MuTHER Consortium; DIAGRAM Consortium; GIANT Consortium; Global B Pgen Consortium; Procardis Consortium; MAGIC investigators; GLGC Consortium. (2012). Novel loci for adiponectin levels and their influence on type 2 diabetes and metabolic traits: a multiethnic meta-analysis of 45,891 individuals. PLoS Genet. *8*, e1002607.