Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study

Logan Dumitrescu 1, Cara L. Carty 2, Kira Taylor 3, Fredrick R. Schumacher 4, Lucia A. Hindorff 5, José L. Ambite 6, Garnet Anderson 2, Lyle G. Best 7, Kristin Brown-Gentry 1, Petra Bůžková 8, Christopher S. Carlson 2, Barbara Cochran 9, Shelley A. Cole 10, Richard B. Devereux 11, Dave Duggan 12, Charles B. Eaton 13, MyrIAM ForNage 14, 15, Nora Franceschini 19, Jeff Haessler 2, Barbara V. Howard 16, Karen C. Johnson 17, Sandra Laston 10, Laurence N. Kolonel 18, Elisa T. Lee 19, Jean W. MacCluer 10, Teri A. Manolio 5, Sarah A. Pendergrass 1, Miguel Quibrera 20, Ralph V. Shohet 21, Lynne R. Wilkens 15, Christopher A. Haiman 4, Loïc Le Marchand 18, Steven Buyske 22, Charles KoopErb 23, Kari E. North 3, 23, Dana C. Crawford 1, 24

1 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, 2 Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington, United States of America, 3 Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina, United States of America, 4 Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, United States of America, 5 Office of Population Genomics, National Human Genome Research Institute, Bethesda, Maryland, United States of America, 6 Information Sciences Institute, University of Southern California, Los Angeles, California, United States of America, 7 Missouri Breaks Industries Research, Timber Lake, South Dakota, United States of America, 8 Department of Biostatistics, University of Washington, Seattle, Washington, United States of America, 9 Sponsored Programs, Baylor College of Medicine, Houston, Texas, United States of America, 10 Department of Genetics, National Human Genome Research Institute, Bethesda, Maryland, United States of America, 11 Department of Medicine, Weill Cornell Medical College, New York, New York, United States of America, 12 The Translational Genomics Research Institute, Phoenix, Arizona, United States of America, 13 Department of Family Medicine and Community Health, Alpert Medical School of Brown University School of Medicine, Providence, Rhode Island, United States of America, 14 Institute of Molecular Medicine, University of Texas Health Sciences Center at Houston, Texas, United States of America, 15 Division of Epidemiology, School of Public Health, University of Texas Health Sciences Center, Houston, Texas, United States of America, 16 Medstar Research Institute, Washington, D.C., United States of America, 17 Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis, Tennessee, United States of America, 18 Epidemiology Program, University of Hawaii Cancer Center, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, 19 University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, United States of America, 20 School of Public Health, University of North Carolina, Chapel Hill, North Carolina, United States of America, 21 Center of Cardiovascular Research, Department of Medicine, John A. Burns School of Medicine, University of Hawaii, Honolulu, Hawaii, United States of America, 22 Department of Statistics and Biostatistics, Rutgers University, Piscataway, New Jersey, United States of America, 23 Carolina Center for Genome Sciences, University of North Carolina, Chapel Hill, North Carolina, United States of America, 24 Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee, United States of America

Abstract

For the past five years, genome-wide association studies (GWAS) have identified hundreds of common variants associated with human diseases and traits, including high-density lipoprotein cholesterol (HDL-C), low-density lipoprotein cholesterol (LDL-C), and triglyceride (TG) levels. Approximately 95 loci associated with lipid levels have been identified primarily among populations of European ancestry. The Population Architecture using Genomics and Epidemiology (PAGE) study was established in 2008 to characterize GWAS-identified variants in diverse population-based studies. We genotyped 49 GWAS-identified SNPs associated with one or more lipid traits in at least two PAGE studies and across six racial/ethnic groups. We performed a meta-analysis testing for SNP associations with fasting HDL-C, LDL-C, and ln(TG) levels in self-identified European American (~20,000), African American (~9,000), American Indian (~6,000), Mexican American/Hispanic (~2,500), Japanese/East Asian (~690), and Pacific Islander/Native Hawaiian (~175) adults, regardless of lipid-lowering medication use. We replicated 55 of 60 (92%) SNP associations tested in European Americans at p<0.05. Despite sufficient power, we were unable to replicate ABCA1 rs149268 and rs1883025, CETP rs1864163, and T7C9H rs471364 previously associated with HDL-C and MAFB rs6102059 previously associated with LDL-C. Based on significance (p<0.05) and consistent direction of effect, a majority of replicated genotype-phenotype associations for HDL-C, LDL-C, and ln(TG) in European Americans generalized to African Americans (48%, 61%, and 57%), American Indians (45%, 64%, and 77%), and Mexican Americans/Hispanics (57%, 56%, and 86%). Overall, 16 associations generalized across all three populations. For the associations that did not generalize, differences in effect sizes, allele frequencies, and linkage disequilibrium offer clues to the next generation of association studies for these traits.

Citation: Dumitrescu L, Carty CL, Taylor K, Schumacher FR, Hindorff LA, et al. (2011) Genetic Determinants of Lipid Traits in Diverse Populations from the Population Architecture using Genomics and Epidemiology (PAGE) Study. PLoS Genet 7(6): e1002138. doi:10.1371/journal.pgen.1002138

Editor: Greg Gibson, Georgia Institute of Technology, United States of America

Received December 30, 2010; Accepted April 30, 2011; Published June 30, 2011

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: The Population Architecture Using Genomics and Epidemiology (PAGE) program is funded by the National Human Genome Research Institute (NHGRI), supported by U01HG004803 (CALiCo), U01HG004798 (EAGLE), U01HG004802 (MEC), U01HG004790 (WHI), and U01HG004801 (Coordinating Center). Genotyping...
services for select EAGLE NHANES III SNPs presented here were also provided by the Johns Hopkins University under federal contract number (N01 -HV-48195) from NHBLI. EAGLE study participants derive from the National Health and Nutrition Examination Surveys (NHANES), and these studies are supported by the Centers for Disease Control and Prevention. The MEC study is funded through the National Cancer Institute (R01 CA36032, R01 CA14851, R01 CA57692, and U01 CA98758). The WHI program is funded by the National Heart, Lung, and Blood Institute (NHLBI) (NHLBI) and by the U.S. Department of Health and Human Services through contracts N01WH22110, 21452, 21200-2, 21056-5, 21088-9, 21113-13, 21115, 32118-32119, 32122, 42107-26, 42121-32, and 44421. Funding support for individual CALiCo studies is as follows: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by NHLBI contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022. The Coronary Artery Risk Development in Young Adults (CARDIA) study is supported by the following NHLBI contracts: N01-HC-59095, N01-HC-48047, N01-HC-48048, N01-HC-48049, N01-HC-48050, N01-HC-45134, N01-HC-05187, and N01-HC-45205. The Cardiovascular Health Study (CHS) is supported by contracts N01-HC-85079 through N01-HC-85086, N01-HC-35129, N01-HC-15103, N01-HC-55222, N01-HC-75150, N01-HC-45132, grants U01 HL080295 and R01 HL087652 from the NHLBI, with additional contribution from the National Institute of Neurological Disorders and Stroke. CHS GWAS DNA handling and genotyping was supported in part by National Center for Research Resources grant M01-RR00425 to the Cedars-Sinai General Clinical Research Center Genotyping core and National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Strong Heart Study (SHS) is supported by NHLBI grants U01 HL65520, U01 HL61462, U01 HL41652, U01 HL41654, and U01 HL65521. Assistance with phenotype harmonization, SNP selection and annotation, data cleaning, data management, integration and dissemination, and general study coordination was provided by the PAGE Coordinating Center (U01HG004801-01). The National Institutes of Mental Health also contributes to the support for the Coordinating Center. NHGRI collaborators (LAH and TAM) assisted in the study design, analysis, and preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: crawford@chgr.mc.vanderbilt.edu

Introduction

Since its introduction in 2005, the genome-wide association study (GWAS) design has become a powerful tool in human genetics to identify single nucleotide polymorphisms (SNPs) associated with common diseases or traits using an experimental design that does not require a priori biological knowledge. As of September 2010, greater than 1,000 SNPs across the genome have been reported as genome-wide significant ($p<5 \times 10^{-8}$) for 165 traits [1]. An early analysis of the GWAS-reported SNPs demonstrated that most identified variants were intergenic or intronic [2], suggesting either novel biology or that the functional variant has yet to be found.

While GWAS have been successful in identifying novel associations, there are several limitations. First, the majority of GWAS have been conducted in populations of European descent. There are several GWAS in populations of Asian-descent, and GWAS are just emerging for other populations such as African Americans [3–20], Mexican Americans/Hispanics [9,20–26], and American Indians [27]. It is possible that novel associations await discovery in these populations given the differing linkage disequilibrium (LD) patterns when compared with populations of European-descent [28]. Second, much work is needed to test SNPs discovered in case-control studies in more population-based, representative cohorts to determine whether the associations generalize. Data on generalization will inform future fine-mapping [29] and discovery studies as well as provide clues to whether GWAS-identified SNPs are simply tagSNPs or are more likely to be true functional SNP[s].

A major goal of the Population Architecture using Genomics and Epidemiology (PAGE) study is to determine whether GWAS-identified variants generalize to diverse groups drawn from population-based studies [30]. Generalization is defined here as a significant association ($p<0.05$, uncorrected for multiple testing) in a non-European population and a direction of genetic effect in the same direction as that of European Americans. In PAGE, variants identified in GWAS and well replicated in multiple studies are chosen for targeted genotyping in hundreds to thousands of European Americans (~20,000), African Americans (~9,000), American Indians (~6,000), Mexican Americans/Hispanics (~2,500), Japanese/East Asians (~690), and Native Hawaiians/Pacific Islanders (~175). All samples are linked to extensive demographic, health, and exposure data, making the PAGE study a rich resource for post-discovery generalization and characterization for common diseases and traits.

We present here PAGE study data on the replication and generalization for 49 SNPs associated with three common lipid traits: low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), and triglycerides. Each of these three traits has numerous GWAS published in European ancestry individuals [30–43] but only a handful published in other populations (such as Asians [44] and Micronesians [45]). Additional data are just now emerging from large sample sizes of diverse populations for generalization [32,46–51] and fine-mapping [52] of these lipid GWAS-identified SNPs. We demonstrate that the majority of the targeted GWAS-identified SNPs replicate in European Americans in PAGE and that many generalize to diverse populations. Both power and LD are explored as explanations of non-generalization, highlighting the complexities involved in properly interpreting results of even robust genetic associations such as these.

Results

Study population characteristics

The PAGE study sites are diverse across multiple variables (Table 1 and Table S1). Together, the PAGE study consists of several populations: European Americans, African Americans, Mexican Americans/Hispanics, American Indians, Japanese/East Asians, and Native Hawaiians/Pacific Islanders. All PAGE study sites except WHI ascertained both men and women. Participant age varies widely across PAGE. For example, CHS ascertained on average older adults (median age = 74 and 72 years for European and African Americans, respectively), CARDIA ascertained younger adults (median age = 26 and 24.5 years for European and African Americans, respectively), and NHANES ascertained all ages of adults (18 years to 90 years; median age = 51, 39, and 40 years for European, African, and Mexican Americans, respectively). In addition to demographic differences, lifestyles and health differed across the PAGE study sites by population, including lipid lowering medication use and current smoking status. More Japanese participants ascertained by MEC reported lipid lowering medication use compared with other populations ascertained by other PAGE study sites: 38.3% versus <5–10%. American Indians from the Dakotas reported more smoking (42.2–47.8%) than other American Indians (25–33%) or other PAGE study site populations (6.3% to 35.3%). The differences in demographics, lifestyle, and health characteristics observed across the PAGE study sites and populations are reflected in the three traits studied here (Table S1). Given the diversity observed across the PAGE study sites, we performed all tests of association for HDL-C, LDL-C, and triglycerides unadjusted, minimally adjusted (for age and sex), and adjusted for various demographic, lifestyle, and health variables.
Allele frequencies

Coded allele frequencies are presented in Table 2, Table 3, Table 4 and in Figure S1, by population. We calculated the Pearson correlation coefficient (r) and FST between European American coded allele frequencies and all other groups. The highest correlation was observed in the comparison with Mexican American coded allele frequencies and all other groups. We also investigated if the effect of these variants generalizes to non-European descent populations, including African Americans, American Indians, and Mexican Americans/Hispanics. A majority of these GWAS–identified associations replicated in our European American cohort. However, the ability of associations to generalize across other racial/ethnic populations varied greatly, indicating that some of these GWAS–identified variants may not be functional and are more likely to be in linkage disequilibrium with the functional variant(s).

Table 1. Characteristics of PAGE study populations.

| Type of Study | EAGLE | MEC | WHI | CALiCo | ARIC | CARDIA | CHS | SHS |
|--------------|-------|-----|-----|--------|------|--------|-----|-----|
| Focus of Cohort | Cross-sectional | Nested Case Control | Cohort and Clinical Trials | Longitudinal | Longitudinal | Longitudinal | Longitudinal | Longitudinal |
| Years Collected | 1991–1994, 1999–2002 | 1993–1996 | 1993–1998 | 1987–2007 | 1986–2006 | 1989–1999 | 1988–present |
| Median Age | 43 | 67 | 63 | 54 | 25 | 73 | 47 |
| Age Range | 18–90 | 48–86 | 50–79 | 44–66 | 18–35 | 64–96 | 14–93 |
| % Women | 54 | 36 | 100 | 57 | 56 | 62 | 59.3 |
| Race/Ethnicity (nmax) | European Americans | 3,999 | 317 | 4,688 | 11,178 | 2,134 | 2,787 | – |
| African Americans | 1,896 | 552 | 1,840 | 3,770 | 2,035 | 550 | – |
| American Indians | – | – | 113 | – | – | – | 6,021 |
| Mexican Americans | 2,361 | 299 | 762 | – | – | – | – |
| Japanese/East Asian | – | 576 | 251 | – | – | – | – |
| Native Hawaiian/Pacific Islander | – | 87 | 113 | – | – | – | – |

*Only controls (cancer-free participants) from the overall nested case-control study were included in this lipid study.

Epidemiologic Architecture for Genes Linked to Environment (EAGLE); Multietnic Cohort (MEC); Women's Health Initiative (WHI); Causal Variants Across the Life Course (CALiCo); Atherosclerosis Risk in Communities (ARIC); Coronary Artery Risk Development in Young Adults (CARDIA); Cardiovascular Health Study (CHS); Strong Heart Study (SHS).

doi:10.1371/journal.pgen.1002138.t001
Table 2. Meta-analysis of GWAS–identified HDL-C SNPs.

| SNP          | Nearest Gene | European Americans (nmax = 25,167) | African Americans (nmax = 10,436) | American Indians (nmax = 6,134) | Mexican Americans and Hispanics (nmax = 3,371) |
|--------------|--------------|------------------------------------|-----------------------------------|--------------------------------|-----------------------------------------------|
|              | CAF (SE)     | P-value                            | CAF (SE)                          | CAF (SE)                        | CAF (SE)                                      |
| rs2144300    | G0.098 (0.10) | 3.35E-03                           | 0.048 (0.21)                      | 0.46                            | 0.31                                          |
| rs1714738    | T0.121 (0.21) | 1.66E-03                           | 0.099                            | 0.46                            | 0.31                                          |
| rs2197089    | C0.090       | 2.29 (0.10)                        | 0.39                             | 0.46                            | 0.31                                          |
| rs1986981    | A0.556       | 0.96 (0.14)                        | 0.84                             | 0.46                            | 0.31                                          |
| rs471386     | A0.012       | 1.30 (0.20)                        | 0.12                             | 0.46                            | 0.31                                          |
| rs198581     | A0.012       | 1.30 (0.20)                        | 0.12                             | 0.46                            | 0.31                                          |
| rs198525     | T0.044       | 0.84 (0.38)                        | 0.25                             | 0.46                            | 0.31                                          |
| rs2972980    | G0.093       | 1.51 (0.20)                        | 0.84                             | 0.46                            | 0.31                                          |
| rs941484     | C0.086       | 1.57 (0.20)                        | 0.80                             | 0.46                            | 0.31                                          |
| rs135506     | C0.006       | 1.08 (0.31)                        | 0.06                             | 0.46                            | 0.31                                          |
| rs2338104    | C0.046       | 0.49 (0.14)                        | 0.27                             | 0.72                            | 0.31                                          |
| rs7679       | T0.820       | 0.95 (0.21)                        | 0.68                             | 0.30                            | 0.60                                          |

Coded allele (CA), coded allele frequency (CAF), beta coefficient (b); standard error (SE); data not available (–); generalized (G); yes (Y); no (N). Generalization is defined here as a significant association (p < 0.05) and a similar direction of effect (b) compared with European Americans for the same test of association, across all racial/ethnic populations.
Table 3. Meta-analysis of GWAS–identified LDL-C SNPs.

| SNP    | Nearest Gene       | CA     | European Americans (n<sub>max</sub> = 21,986) | African Americans (n<sub>max</sub> = 9,328) | American Indians (n<sub>max</sub> = 6,144) | Mexican Americans and Hispanics (n<sub>max</sub> = 2,532) | G    |
|--------|--------------------|--------|-----------------------------------------------|---------------------------------------------|--------------------------------------------|-------------------------------------------------|------|
|        |                    |        | CA                                             | P-value                                     | CA                                         | P-value                                         | P-value |
| rs11206510 | PCSK9              | T      | 0.81                                           | 1.98 (0.45)                                 | 0.86                                       | 0.09 (0.84)                                     | 0.01   |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.96   |
| rs11591147 | PCSK9              | T      | 0.02                                           | -16.92 (1.42)                               | 1.00E-32                                   | -22.64 (5.21)                                   | 1.14E-05 |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.01   |
| rs646776   | CELSR2/PSRC1/SORT1 | A      | 0.78                                           | 5.74 (0.44)                                 | 1.44E-37                                   | 0.65                                           | 4.46 (0.63) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 1.48E-12 |
| rs599839   | CELSR2/PSRC1/SORT1 | A      | 0.77                                           | 5.67 (0.45)                                 | 3.61E-36                                   | 0.28                                           | 1.60 (0.72) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 2.67E-02 |
| rs693      | APO8               | T      | 0.50                                           | 3.45 (0.36)                                 | 3.38E-21                                   | 0.24                                           | 1.60 (0.69) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 2.04E-02 |
| rs562338   | APO8               | T      | 0.19                                           | -5.52 (0.45)                                | 1.05E-33                                   | -2.54 (0.59)                                   | 1.57E-5 |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.09   |
| rs754523   | APO8               | T      | 0.68                                           | -3.64 (0.40)                                | 3.4E-19                                    | 0.78                                           | -2.12 (0.76) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 5.52E-03 |
| rs6544713  | ABCG8              | T      | 0.31                                           | 2.98 (0.42)                                 | 1.17E-12                                   | 0.17                                           | 1.49 (0.74) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 4.45E-02 |
| rs12654264 | HMGCR              | A      | 0.62                                           | -2.66 (0.37)                                | 6.56E-13                                   | 0.67                                           | -2.02 (0.61) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 9.39E-04 |
| rs1501908  | TIMD4              | G      | 0.64                                           | 1.23 (0.44)                                 | 4.96E-03                                   | 0.37                                           | 1.31 (0.64) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 4.18E-02 |
| rs2650000  | HNF1A              | T      | 0.35                                           | 1.20 (0.39)                                 | 2.33E-03                                   | 0.12                                           | 0.15 (0.97) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.88   |
| rs6511720  | LDLR               | T      | 0.12                                           | -7.32 (0.52)                                | 2.99E-44                                   | 0.13                                           | -8.10 (0.80) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 7.05E-24 |
| rs222671   | LDLR               | T      | 0.12                                           | -5.83 (0.97)                                | 1.96E-09                                   | 0.04                                           | -6.62 (2.94) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 2.43E-02 |
| rs1696148  | CILP2/PBX4/NCAN1   | T      | 0.08                                           | -2.88 (0.66)                                | 1.40E-05                                   | 0.15                                           | 0.77 (0.80) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.34   |
| rs4803750  | BCL3               | A      | 0.93                                           | 5.57 (0.92)                                 | 1.37E-09                                   | 0.92                                           | 1.52 (1.77) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.39   |
| rs10402271 | APOE/C1/C4         | T      | 0.67                                           | -2.27 (0.49)                                | 3.86E-06                                   | 0.84                                           | -1.38 (1.28) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.28   |
| rs4420638  | APOE/C1/C4         | A      | 0.82                                           | -5.34 (0.51)                                | 2.16E-25                                   | 0.79                                           | 0.16 (0.92) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.87   |
| rs2207650  | TOMM40             | A      | 0.88                                           | -4.77 (1.23)                                | 1.14E-04                                   | 0.87                                           | -2.26 (2.40) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.35   |
| rs6102059  | MAFB               | T      | 0.30                                           | -0.41 (0.52)                                | 0.42                                       | 0.43                                           | -0.77 (0.88) |
|         |                    |        | CA                                             |                                             |                                             |                                                | 0.38   |

Coded allele (CA); coded allele frequency (CAF); beta coefficient (β); standard error (SE); data not available (–); generalized (G); yes (Y); no (N). Generalization is defined here as a significant association (p < 0.05) and a similar direction of effect (β) compared with European Americans for the same test of association, across all racial/ethnic populations.

doi:10.1371/journal.pgen.1002138.t003
Table 4. Meta-analysis of GWAS–identified Triglyceride SNPs.

| SNP          | Nearest Gene              | CA | European Americans (nmax = 24,258) | African Americans (nmax = 9,844) | American Indians (nmax = 6,157) | Mexican Americans and Hispanics (nmax = 2,973) | G |
|--------------|---------------------------|----|-----------------------------------|----------------------------------|---------------------------------|---------------------------------------------|---|
|              |                           |    | CAF | β (SE) | P-value | CAF | β (SE) | P-value | CAF | β (SE) | P-value | CAF | β (SE) | P-value |
| rs1748195    | ANGPTL3                   | C  | 0.66 | 0.03 (0.01) | 1.93E-07 | 0.35 | 0.01 (0.01) | 0.19 | 0.61 | 0.16 (0.07) | 2.44E-02 | 0.60 | 0.04 (0.01) | 1.17E-02 | N |
| rs1260326    | GCKR                      | T  | 0.42 | 0.05 (0.01) | 6.44E-13 | 0.16 | 0.05 (0.02) | 9.98E-04 | 0.28 | 0.15 (0.09) | 8.52E-02 | 0.33 | 0.06 (0.02) | 1.97E-04 | N |
| rs78094      | GCKR                      | A  | 0.46 | 0.06 (0.01) | 1.69E-32 | 0.18 | 0.02 (0.01) | 2.91E-02 | 0.25 | 0.04 (0.01) | 3.23E-03 | 0.33 | 0.06 (0.02) | 1.13E-03 | Y |
| rs17145738   | MLXIPL                    | T  | 0.12 | -0.07 (0.01) | 5.71E-24 | 0.09 | -0.03 (0.01) | 2.53E-02 | 0.08 | -0.07 (0.02) | 2.30E-04 | 0.07 | -0.09 (0.03) | 7.40E-04 | Y |
| rs328        | LPL                       | C  | 0.90 | 0.09 (0.01) | 4.16E-30 | 0.93 | 0.08 (0.02) | 2.62E-08 | 0.97 | 0.09 (0.03) | 4.83E-03 | 0.93 | 0.09 (0.03) | 6.31E-04 | Y |
| rs2197089    | LPL                       | T  | 0.55 | -0.03 (0.01) | 4.97E-15 | 0.78 | -0.01 (0.01) | 7.45E-02 | 0.41 | -0.05 (0.01) | 2.57E-06 | 0.48 | -0.05 (0.01) | 4.01E-04 | N |
| rs2954029    | TRIB1                     | A  | 0.54 | 0.05 (0.01) | 1.13E-04 | 0.68 | -0.01 (0.02) | 0.46 | -  | -                    | -            | 0.62 | 0.06 (0.02) | 9.28E-04 | N |
| rs174547     | FADS1                     | T  | 0.66 | -0.03 (0.01) | 3.82E-10 | 0.91 | -0.05 (0.01) | 3.73E-04 | 0.21 | -0.06 (0.02) | 1.10E-04 | 0.39 | -0.05 (0.02) | 1.51E-03 | Y |
| rs28927680   | APOA1/C3/A4/A5 gene cluster | C  | 0.93 | -0.12 (0.01) | 2.88E-38 | 0.84 | <0.001 (0.01) | 0.95 | 0.83 | -0.13 (0.01) | 6.33E-19 | 0.86 | -0.08 (0.02) | 2.15E-05 | N |
| rs966184     | APOA1/C3/A4/A5 gene cluster | G  | 0.86 | -0.14 (0.01) | 1.91E-59 | 0.80 | -0.02 (0.01) | 4.87E-02 | 0.78 | -0.17 (0.07) | 1.43E-02 | 0.72 | -0.14 (0.02) | 1.04E-19 | Y |
| rs3135506    | APOA1/C3/A4/A5 gene cluster | C  | 0.06 | 0.13 (0.01) | 2.59E-33 | 0.06 | 0.11 (0.02) | 2.06E-10 | 0.17 | 0.13 (0.01) | 4.28E-20 | 0.14 | 0.13 (0.02) | 3.08E-08 | Y |
| rs4775041    | LIPC                      | C  | 0.29 | 0.01 (0.01) | 3.15E-02 | 0.14 | 0.03 (0.01) | 4.29E-03 | 0.21 | 0.02 (0.01) | 5.15E-02 | 0.18 | 0.01 (0.02) | 0.58  | N |
| rs16996148   | CILP2/PBX4/NCAN           | T  | 0.08 | -0.04 (0.01) | 3.91E-05 | 0.15 | <0.001 (0.01) | 0.77 | 0.04 | -0.07 (0.03) | 8.86E-03 | 0.06 | -0.06 (0.03) | 2.69E-02 | N |
| rs7679       | PLTP                      | T  | 0.82 | -0.02 (0.01) | 2.84E-02 | 0.96 | -0.01 (0.02) | 0.61 | 0.94 | -2.0E-03 (0.02) | 0.93 | 0.89 | -0.03 (0.03) | 0.31  | N |

Coded allele (CA); coded allele frequency (CAF); beta coefficient (β); standard error (SE); data not available (–); generalized (G); yes (Y); no (N). Generalization is defined here as a significant association (p < 0.05) and a similar direction of effect (β) compared with European Americans for the same test of association, across all racial/ethnic populations.

doi:10.1371/journal.pgen.1002138.t004
respectively, across European American populations collected by individual PAGE study sites (Table S2). For HDL-C, 23 of the 27 (85%) SNPs tested were associated at $p < 0.05$ assuming an additive genetic model and adjusting for age and sex (Figure 1 and Table 2). The four SNPs that did not replicate at this liberal significance threshold were rs471364 ($TTC39B$), rs1883025 ($ABCA1$), rs4149268 ($ABCA1$), and rs1864163 ($CETP$), all of which are intronic (Table S2). For LDL-C, only one (intergenic $MIAF$ rs6102059) of the 19 SNPs tested was not significantly associated at $p < 0.05$ (Figure 1 and Table 3). Finally, for ln(TG), all 14 SNPs tested were associated at $p < 0.05$ (Figure 1 and Table 4).
Associations in non-European–descent populations

Of the associations that did not replicate in the European-descent populations from PAGE, four out of five had sufficient power (>80%) to detect the previously reported effect size: TTYC39B rs471364 (>99% power; HDL-C), CETP rs1864163 (80% power; HDL-C), MAFB rs6109259 (>90% power; LDL-C), and ABCA1 rs4149268 (99% power; HDL-C). ABCA1 rs1883025, which did not replicate the expected association with HDL-C, did not have sufficient power to detect the reported effect size (68% power; n = 3,865).

We then compared the genetic effect sizes reported in the literature to the genetic effect sizes estimated from the meta-analysis of these population-based studies. We observed that the majority of the point estimates of effect size (β) were smaller than previously reported estimates. Using the HDL-C association results as an example, 15 out of the 23 (65%) significant associations had effect estimates smaller than published effect estimates. We caution, however, that we did not formally test for significant differences between estimates and that these smaller effect estimates may or may not be significantly different than the published reports. However, it is interesting to note that 11 of our effect estimates differed from previous reports by more than 25%, including two HDL-C associations whose effect sizes differed by 50% or more from those in the literature (ANGPTLI rs2967605 and MLXIPL rs17145738; Table 2 and Table S2).

Associations in non-European–descent populations

We meta-analyzed tests of association performed in African Americans for the same 27, 19, and 14 SNPs previously associated with HDL-C, LDL-C, and/or triglycerides in populations of European-descent. For all three traits studied, assuming an additive genetic model and adjusting for age and sex, approximately half of the tested GWAS-identified SNPs were associated at p<0.05: 12/27 (44%) for HDL-C, 11/19 (58%) for LDL-C, and 8/14 (57%) for ln(TG) (Figure 1, Figure S3, Table 2, Table 3, Table 4, Table 5). The majority of SNPs that failed to replicate in the meta-analysis for European Americans also failed to associate in the meta-analysis for African Americans. Interestingly, one SNP (CETP rs1864163) was significantly associated with HDL-C in African Americans (n = 451; CAF = 0.27; β = −2.79; p = 6.19×10⁻³) but not in European Americans (n = 291; CAF = 0.23; β = −2.07; p = 0.13).

Other populations that were examined for select SNPs included American Indians, Mexican Americans/Hispanics, Japanese/East Asians, and Native Hawaiians/Pacific Islanders. Among American Indians, 9/21 (43%), 10/14 (71%), and 10/13 (77%) of the SNPs tested for association with HDL-C, LDL-C, and ln(TG), respectively, were associated at the liberal significance threshold of p<0.05. For Mexican Americans/Hispanics, 14/27 (52%), 10/19 (53%), and 12/14 (86%) SNPs were significantly associated at p<0.05 with HDL-C, LDL-C, and ln(TG), respectively. Despite a small sample size, intronic CETP rs1864163 was significantly associated with HDL-C in Mexican Americans/Hispanics (n = 263; CAF = 0.28; β = −2.98; p = 1.78×10⁻³) but not in European Americans (n = 291; CAF = 0.27; β = −2.07; p = 0.13), although the size and the direction of effect were similar. Venn diagrams representing the overlap of significant associations across the four major PAGE populations are presented in Figure S3.

The sample sizes for Japanese/East Asians and Native Hawaiians/Pacific Islanders are considerably smaller compared with the other populations examined. Despite the lower power to detect associations, significant associations were observed for both groups at a liberal significance threshold of p<0.05. Among the 26, 18, and 13 SNPs tested for associations with HDL-C, LDL-C, and ln(TG), respectively, there were nine (35%), three (17%), and three (23%) SNPs significantly associated in the combined Japanese/East Asian group.

For Native Hawaiians/Pacific Islanders, the group with the smallest sample size considered here, one SNP each was associated with HDL-C (APOA1/C3/A4/A5 gene cluster rs28927680) and LDL-C (APOB rs734523) out of the 24 and 18 SNPs tested for association, respectively. Three out of 12 SNPs tested for an association with ln(TG) were associated at p<0.05 (PLTP rs73679, MLXIPL rs17145738, and APOA1/C3/A4/A5 gene cluster rs209 27680), with the latter at a significance of p<10⁻¹⁹.

Generalization across non-European–descent populations

For the 55 SNP-trait associations that replicated in European Americans, we determined which associations generalized across all

### Table 5. Observed versus expected number of significant associations, by trait and population.

| Trait | Race/Ethnicity | # of Total Tests of Association* | # of Observed Significant Associations | # of Expected Significant Associations | P-value |
|-------|---------------|-------------------------------|--------------------------------------|--------------------------------------|---------|
| HDL-C | AA            | 23                            | 11                                   | 17.3                                 | 0.01    |
|       | Al            | 20                            | 9                                    | 14.4                                 | 0.01    |
|       | MA/H          | 23                            | 13                                   | 13.8                                 | 0.83    |
| LDL-C | AA            | 18                            | 11                                   | 14.7                                 | 0.03    |
|       | Al            | 14                            | 10                                   | 11.9                                 | 0.15    |
|       | MA/H          | 18                            | 10                                   | 10.6                                 | 0.81    |
| ln(TG) | AA          | 14                            | 8                                    | 11.9                                 | 0.01    |
|        | Al            | 13                            | 10                                   | 8.4                                  | 0.56    |
|        | MA/H          | 14                            | 12                                   | 10.4                                 | 0.54    |

The expected number of significant associations was based on power calculations assuming an additive genetic model and liberal significance threshold (0.05) in each racial/ethnic group for each test of association. We further assumed the observed genetic effect size (β) from PAGE European Americans and the observed allele frequency, sample sizes, and trait mean/standard deviations from each non-European American population.

*Only includes associations that replicated in EA.

P-values were based on a one-sample binomial test.

African Americans (AA); American Indians (AI); Mexican Americans/Hispanics (MA/H); number (n).

doi:10.1371/journal.pgen.1002138.t005
four of our largest populations (European Americans, African Americans, American Indians, and Mexican Americans/Hispanics). Generalization was based on two criteria: 1) level of significance (i.e. p-value) and 2) direction of effect (i.e. positive or negative beta). SNPs that were significantly associated at p<0.05 and had the same direction of effect as European Americans in all populations studied were considered to have generalized. For HDL-C, five SNPs (CEP T rs3704261, LPL rs6386091, LIPC rs4773041, LPL rs2197089, and APOA1/C3/A4/A5 gene cluster rs135506) met these criteria, and two SNPs (LCAT rs2271293 and LPL rs328) were associated in three groups and trended towards significance in a fourth group (p = 0.06 and p = 0.07 in Mexican Americans/Hispanics and American Indians, respectively; Table 2).

For LDL-C, six SNPs generalized across all four groups, if genotyped: APOB rs562338, CERSR2/PSRC1/SORT1 rs599039 and rs646776, PCSK9 rs11591147, HMGCR rs12654264, and LDLR rs2228671 (Table 3). Similarly for ln(TG), six SNPs were significantly associated across the four largest populations: APOA1/C3/A4/A5 gene cluster rs3135506 and rs780094, LPL rs328, MLXIPs rs174573, and FADS1 rs174347. In addition, for ln(TG), two SNPs (LPL rs2197089 and GCKR rs1260726) were associated in three groups and trended towards significance in a fourth group (p = 0.07 in African Americans and p = 0.09 in American Indians, respectively). Among the 17 SNPs that generalized across the largest groups among the three lipid traits, only four (24%) were either nonsense (rs328) or missense SNPs (rs3135506, rs11591147, and rs1260726; Table S2).

Power
Based on our definition of generalization, several SNPs discovered and replicated in European-descent populations failed to generalize to other populations. There are several possible explanations for non-generalization, including power. To further investigate potential lack of power, we first performed post-hoc power calculations assuming an additive genetic model and liberal significance threshold (0.05) in each racial/ethnic group for each test of association. In these power calculations, we further assumed the observed genetic effect size (beta) from PAGE European Americans and the observed allele frequency, sample sizes, and trait mean/standard deviation from each non-European American population. By adding the power of all tested loci, we estimated the number of expected significant associations and compared this to the number of observed significant associations (Table 5).

In general, the number of expected significant associations was greater than the number observed. African Americans consistently had fewer significant associations (11, 11, and 8 for HDL-C, LDL-C, and ln(TG), respectively) than expected (17.3, 14.7, and 11.9 for HDL-C, LDL-C, and ln(TG), respectively) based on power, regardless of the lipid trait being tested. More specifically, we were powered to detect in American Indians 17 of the 25 associations that replicated in European Americans but failed to generalize to African Americans.

Compared to African Americans, differences between the observed and the expected number of associations for American Indians and Mexican Americans/Hispanics were less extreme. In fact, for ln(TG), more significant associations were detected in these two populations than the PAGE study was powered to detect (8.4 and 10.4 expected; 10 and 12 observed for American Indians and Mexican Americans/Hispanics, respectively; Table 5). We were powered to detect in American Indians nine of the 18 associations that replicated in European Americans but did not generalize to American Indians. Similarly, we were powered to detect in Mexican Americans/Hispanics eight of the 20 associations that replicated in European Americans but failed to generalize to Mexican Americans/Hispanics.

Linkage disequilibrium
To examine whether LD can account for the lack of generalization of the properly powered tests of association in African Americans, we examined LD patterns in HapMap Europeans (CEU) and West Africans (YRI) as well as those published in the literature for the genotyped SNPs and surrounding variation. For APOA1/C3/A4/A5 rs28927680, previous studies in European-descent populations have noted that this SNP is in strong LD (r2 = 0.98) with missense APOA5 rs3135506 [12]. APOA1/C3/A4/A5 rs964184 is also in moderate LD with missense rs3135506 (r2 = 0.510 in CEU). However, neither rs28927680 nor rs964184 are in LD with missense rs3135506 (r2 = 0.039 and r2 = 0.048) in YRI. Furthermore, APOA5 rs3135506 is associated with HDL-C in European Americans, African Americans, Mexican Americans/Hispanics, and American Indians (Table 1 and Table 2). Generalization of rs3135506 coupled with non-generalization and differences in YRI LD patterns for rs28927680 and rs964184 suggest that APOA5 rs3135506 is either the putative functional SNP for the association with HDL-C or in LD with the functional SNP. Although the exact mechanism is not yet known, molecular modeling [53] as well as in vitro [53] and in vivo [54,55] studies support the epidemiologic evidence that rs3135506 is functional.

Other interpretations of LD patterns are more difficult. For example, CETP rs9898419, which failed to generalize in African Americans for HDL-C despite sufficient power, is not in strong LD with obvious functional SNPs in CEU within 50 kb flanking the genotyped SNP. The strongest pair-wise LD (r2 = 0.251) consists of intergenic and intronic SNPs, and these same SNPs have weak LD (r2<0.03) or are not found in YRI. Similarly, LIPC rs261332 associated with HDL-C levels in European Americans but failed to generalize in African Americans. LIPC rs261332 is in strong LD (r2>0.80 in CEU) with SNPs in the 5’ flanking region of LIPC, but not in LD with these same SNPs in YRI (r2<0.15).

Adjustments for exposures and co-morbidities
Genetic variations in isolation are not the sole determinants of lipid trait distributions. Many environmental exposures and demographic variables are associated with lipid traits. To account for these variables, we meta-analyzed all tests of association for HDL-C, LDL-C, and ln(TG) adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use. Adjustment for these additional covariates did not appreciably alter the results compared with the models minimally adjusted for age and sex (Figures S4, S5, S6). Inclusion of previous myocardial infarction as a variable to the fully adjusted model also did not appreciably alter the results compared with the minimally adjusted models (Figures S4, S5, S6).

Effect of including versus excluding by medication use
All analyses presented thus far include fasting adult participants regardless of lipid lowering medication use. Many GWAS conducted for the lipid traits excluded participants on lipid lowering medication [40,42,43] given that these medications substantially alter the upper extreme of the normal LDL-C distribution associated with a genetic profile found in a general population. Exclusion of these participants would preclude these meta-analyses from fully describing the extent and strength of associations relevant to these traits in a population-based setting. However, if genetic
variation is associated with lipid concentrations and medication use lowers lipid concentrations, inclusion of participants on lipid lowering medications could bias associations towards the null. As a sensitivity analysis, WHI used detailed medication data available on a subset of participants, and performed the tests of association for HDL-C, LDL-C, and ln(TG) excluding and including participants on lipid lowering medication with the latter adjusted for medication usage using average effects estimated in Wu et al [56] for specific drug classes. Figure S7 suggests that both the point estimates and the confidence intervals of the genetic effects are similar for this female-only study whether participants are excluded or included and adjusted for medication use.

We also performed a second sensitivity analysis tests of association excluding participants on lipid lowering medication for all models. As detailed in Figures S8, S9, S10, excluding participants on lipid lowering medication usage does not appreciably alter the results, with the possible exception of LDL-C associations in Japanese/East Asians. More specifically, two SNPs (rs11206510 and rs1501908) became significantly associated with LDL-C after excluding participants on medications while two other SNPs (rs562338 and rs6544713) were no longer significantly associated (Figure S9). The difference in significance for these four tests of association may be related to lipid lowering medication use; however, it is more likely due to statistical fluctuations from small samples sizes ($n_{Mexican} = 690$; $n_{Native Hawaiian} = 467$). Also of note, use of lipid-lowering medications was low (<10%) in the ARIC, CHS, NHANES, and WHI studies since the majority of study recruitment occurred before the introduction or widespread use of the recent generation of lipid-lowering medications. Medication use was higher in the MEC study (20–38% depending on the population), which contributed the majority of Japanese/East Asian samples.

**Discussion**

We have performed an extensive replication and generalization effort for HDL-C, LDL-C, and TG GWAS-identified SNPs. The PAGE study consists of six racial/ethnic groups: European American, African American, Mexican American/Hispanic, American Indian, Japanese/East Asian, and Native Hawaiian/Pacific Islander, with population-specific sample sizes ranging from ~100 to >20,000 for any one test of association. Although power to detect associations varied across the lipid traits and populations, we observed general patterns worth noting for future genetic epidemiological studies.

**Replication in European-descent populations**

Perhaps not unexpectedly, we were able to replicate most reported associations in European Americans. Regardless of significance, all but one of the tested SNPs had effect estimates in the same direction as the previously reported association from the literature. *FADS1* rs174547, which was significantly associated with decreased ln(TG) in this meta-analysis for European Americans, was associated with increased TG in European Americans from the Framingham Heart Study ($n = 7,423$) [43]. HDL-C had proportionally (15%) the greatest number of SNPs that failed to replicate in European Americans compared with LDL-C (5%) and TG (0%) despite the fact that we had sufficient power to detect the reported genetic effect size for many of these tests. *TTC39B* rs471364 was not associated with HDL-C levels despite a sample size of 18,089 and >99% power to detect the reported effect size. Neither *ABCA1* rs149268 nor rs1883025 was associated with HDL-C, although the latter test of association was underpowered (68%; $n = 3,965$). Finally, as previously discussed, *CEP1* rs1864163 was not associated with HDL-C in this European American dataset although we had 80% power to detect the reported genetic effect size. For LDL-C, only *MAFB* rs6102059 was not associated despite >90% power to detect the reported effect size.

The reasons for non-replication in this European American dataset for properly powered tests of association are unclear. It is possible that we have overestimated our power to detect associations. The “winner’s curse” and inflated genetic effect estimates from initial discovery are well known [57,58]. Indeed, for the five SNPs that did not replicate in this meta-analysis for European Americans, the association was described in only one GWAS each despite the fact that numerous GWAS [31,33–43] and a large meta-analysis [32] for these three traits have been conducted in populations of European-descent. The meta-analysis recently reported by Teslovich et al [32] did report significant associations between *TTC39B* rs581080 for HDL-C and *MAFB* rs2902940 for LDL-C. *TTC39B* rs581080 is in moderate linkage disequilibrium (LD) with rs471364 ($r^2 = 0.49$ in CEU HapMap), but *MAFB* rs2902940 is not in LD with rs6102059 ($r^2 = 0.03$ in HapMap CEU).

A second possibility for our observed non-replication is heterogeneity among the PAGE studies. Because it is important to understand the degree to which associations are consistent across individual studies, we compared directions of effect (betas) across PAGE study sites for each test of association (Figures S11, S12, S13) and performed tests of heterogeneity. Association results for *TTC39B* rs471364, which meta-analysis result for HDL-C in European Americans was insignificant, had significant evidence for heterogeneity across studies ($\chi^2_{heterogeneity} = 0.048; I^2 = 58.25\%$). In four of the five PAGE study sites, the association between this SNP and HDL-C had consistent directions of effect; however, only one test of association was significant in European Americans ($p = 0.005$ in EAGLE; Figure S11). Only two other association results had evidence for heterogeneity among European Americans: *FADS1* rs174547 for HDL-C ($\chi^2_{heterogeneity} = 0.006; I^2 = 75.73\%$) and *PCSK9* rs11206510 for LDL-C ($\chi^2_{heterogeneity} = 0.048; I^2 = 55.34\%$). However, for both of these loci, the tests of association were significant in European Americans and had similar directions of effect in all but one of the PAGE study sites (Figures S11 and S12).

**Generalization to non-European populations**

When taking into account power, significance, and direction of effect, most SNPs discovered in European Americans generalized to African Americans, Mexican Americans, and American Indians. Of note are the eleven tests of association significant in European Americans that did not generalize to African Americans despite having adequate power. Given that GWAS products are a mixture of tagSNPs and functional SNPs, it is likely that discovery in European Americans represents tagSNPs rather than the true functional SNP. Because linkage disequilibrium patterns differ across populations, tagSNPs genotyped directly in populations of non-European descent may not recapitulate the association observed in European-descent populations depending on the pattern of LD. The association of HDL-C and nonsynonymous rs3135506 versus tagSNPs rs28927680 in the *APOA1/C3/A4/A5* gene cluster in this analysis is an example of the effects of LD and the ability to generalize across populations.

Evoking LD as an explanation for lack of generalization is appealing, but it does have limitations given that the functional SNP is not often obvious. All tests of association that did not generalize to African Americans had evidence of LD differences between CEU and YRI using the HapMap data. However, most of these SNPs are located in the intergenic and intronic regions.
Further fine-mapping in both the discovery population as well as other diverse populations will be needed along with a better understanding of genetic variation and its relationship to biological function to identify the true functional SNPs for these traits.

Among the five putative functional SNPs genotyped (nonsynonymous rs11591147, rs1260326, rs13135506, and rs1800961 and nonsense rs328), all five replicated in populations of European descent, and three of the five generalized to populations of non-European descent. One putative functional SNP that did not replicate across populations was INPAA rs1800961, likely due to low power because of the very low minor allele frequency in all subpopulations (0.0065 to 0.0398). Both the direction and magnitude of effect, however, were consistent across groups. GCKR rs1260326 did not generalize to all populations of non-European descent but did generalize in three of the four populations tested and trended towards significance in American Indians (p = 0.085; Table 4).

Limitations and strengths

The major strengths and limitations of the PAGE study for lipids are sample size and diversity. The largest sample size is for samples of European-descent (~20,000), followed by African Americans and American Indians. The sample sizes for Mexican Americans, Japanese/East Asians, and Pacific Islanders/Native Hawaiians are smaller and consequently underpowered for tests of association as estimated from genetic effect sizes in the published European-descent discovery studies. Also, not all SNPs were genotyped in all PAGE studies, further affecting the power of the meta-analyses.

An additional limitation is the lack of data related to lipid lowering medication. Ideally, all analyses would be adjusted for use of lipid lowering medication based on the type and dose of medication. In most PAGE studies, these data were not available and in many, use was low at baseline when blood samples were obtained. As we demonstrate in Supplementary material, inclusion of participants using lipid-lowering medication did not appreciably alter the results of the meta-analysis when compared with excluding these participants. While this finding may be useful for future studies, we caution that the majority of participants in this study were not on lipid lowering medications.

In general, the cohorts and surveys included in PAGE are diverse with regard to demographics, genetic ancestry, lifestyle, health, and environmental exposure. Despite this diversity, very few tests of association from the meta-analysis exhibited evidence of heterogeneity.

Conclusions

Overall, the majority of GWAS-identified SNPs for HDL-C, LDL-C, and TG replicated in European Americans and generalized to non-European-descent populations. These results suggest that the genotyped SNP either tags the functional SNP(s) common across these populations or that the genotyped SNP represents the risk SNP directly. SNPs that replicated in European Americans but did not generalize in the largest non-European-descent populations, despite adequate power, could represent priority associations that require fine-mapping and re-sequencing to identify the functional variant(s).

Materials and Methods

Study populations and phenotypes

All studies were approved by Institutional Review Boards at their respective sites (details are given in Text S1). PAGE study samples were drawn from four large population-based studies or consortia: EAGLE (Epidemiologic Architecture for Genes Linked to Environment), based on three National Health and Nutrition Examination Surveys (NHANES) [59-61], the Multiethnic Cohort (MEC) [62], the Women’s Health Initiative (WHI) [63,64], and Causal Variants Across the Life Course (CALCo), a consortium of several cohort studies: Atherosclerosis Risk in Communities Study (ARIC) [65], Coronary Artery Risk in Young Adults (CARDIA) [66], Cardiovascular Health Study (CHS) [67], Strong Heart Family Study (SHFS) [68], and Strong Heart Cohort Study (SHS) [69] (Table 1). The PAGE study design is detailed in Matise et al [30].

SNP selection and genotyping

All SNPs considered for genotyping were previously associated with HDL-C, LDL-C, and/or triglycerides in published (as of 2008) candidate gene and genome-wide association studies. A total of 52 SNPs were targeted for genotyping by two or more PAGE study sites. There is no overlap between samples used in this study and samples used in GWAS from which the SNPs were selected. The 52 targeted variants are located in or nearby 32 different genes/gene regions, with 12 of the gene/gene regions represented by two or more SNPs. Five SNPs are nonsynonymous, one SNP is a nonsense variant, and two SNPs are synonymous; the remainder are located in introns, flanking, or intergenic regions. The full list of targeted SNPs, their locations, and their previously associated lipid trait can be found in Table S2.

Cohorts and surveys were genotyped using either commercially available genotyping arrays (Affymetrix 6.0, Illumina 370CNV BeadChip), custom mid- and low-throughput assays (TaqMan, Sequenom, Illumina GoldenGate or BeadXpress), or a combination thereof. Quality control was implemented at each study site independently. In addition to site-specific quality control, all PAGE study sites genotyped 360 DNA samples from the International HapMap Project and submitted these data to the PAGE Coordinating Center for concordance statistics [71]. Study specific genotyping details are described in Text S1. Of the 52 targeted SNPs, three (CETP rs1800775, APOE rs49358, and APOE rs7412) failed at all PAGE study sites that attempted genotyping; therefore, a total of 49 SNPs were tested in this analysis.

Statistical methods

All tests of association were performed by each PAGE study site using the same analysis protocol prior to meta-analysis. The study protocol excluded participants <18 years of age as well as non-fasting samples (defined here as <8 hours). When triglyceride level was the dependent variable, participants with >1,000 mg/dl were excluded from analyses. Triglyceride (TG) levels were natural-log transformed (ln) prior to analysis.

Linear regression was performed for fasting adults regardless of lipid lowering medication use with HDL-C, LDL-C, or ln(TG) as the dependent variable and a SNP as the independent variable, assuming an additive genetic model, stratified by race/ethnicity. The coded allele is reported in Table 2, Table 3, Table 4. The beta estimate is per additional copy of the coded allele. For each SNP, four models were considered: 1) unadjusted, 2) adjusted for age (continuous in years) and sex, 3) adjusted for age, body mass index (continuous in kg/m²),
current smoking (yes/no; binary), type 2 diabetes (yes/no; binary), post-menopausal status (yes/no for females only; binary), and current hormone use (yes/no for females only; binary), and for 4 adjusted for age, body mass index, current smoking, type 2 diabetes, post-menopausal status, current hormone use, and previous myocardial infarction (yes/no; binary). All PAGE study sites (except for WHI, which is female only) stratified models 3 and 4 by sex given the sex-specific variables (post-menopausal status and hormone use) prior to meta-analysis. Select PAGE study sites also included study site or site of ascertainment as a covariate in all models. Results from Model 2 (adjusted for age and sex) are reported in the main text while results from Models 1, 3, and 4 are presented in Figures S4, S5, S6. Model 2 excluding participants on lipid-lowering medications are presented in Figures S8, S9, S10.

Meta-analyses, using a fixed-effects inverse-variance weighted approach and tests for effect size heterogeneity across studies, were performed using METAL [72]. P-values were not adjusted for multiple testing, and association results were plotted using Synthesis-View [73,74], where indicated. Power calculations were performed using Quanto [75,76] assuming unrelated participants, an additive genetic model, the published effect size from European-descent populations listed in Table S1, and the population-specific allele frequencies listed in Table 2, Table 3, Table 4. Linkage disequilibrium was calculated using HapMap European (CEU) and West African (YRI) data accessed through the Genome Variation Server. FST was calculated using the Weir and Cockerham algorithm [77]. Aggregate data from the meta-analysis as well as individual tests of association from each PAGE study site will be made available via dbGaP [30,78].

Web resources

NHGRI GWAS Catalog (www.genome.gov/GWAStudies).

Genome Variation Server (pga.gs.washington.edu).

Synthesis-View (http://chgr.mc.vanderbilt.edu/ritchielab/method.php?method=Synthesisview).

Supporting Information

Figure S1 Coded allele frequency, by population. The coded allele frequency (CAF) is plotted for each of the 49 SNPs by population using Synthesis-View [73,74]. The populations include European Americans (EA), African Americans (AA), Mexican Americans/Hispanics (MA/H), American Indians (AI), Japanese/East Asians (J/EA), and Native Hawaiians/Pacific Islanders (NH/PI).

Figure S2 Coded allele frequency across PAGE study sites, by population. The coded allele frequency (CAF) is plotted for each of the 49 SNPs by population using Synthesis-View [73,74]. The studies include: Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk in Young Adults (CARDIA), Cardiovascular Heart Study (CHS), Epidemiologic Architecture for Genes Linked to Environment (EAGLE), Multiethnic Cohort (MEC), Women’s Health Initiative (WHI), Strong Heart Community Study (SHCS), and Strong Heart Family Study (SHFS) in Arizona (AZ), Oklahoma (OK) and South Dakota (SD).

Figure S3 Venn diagrams representing the overlap of significant associations (p<0.05) across the four major PAGE populations (European Americans, African Americans, Native Americans, and Mexican Americans/Hispanics, for the three lipid traits (LDL-C, HDL-C, and TG).

Figure S4 Comparison of unadjusted, minimally adjusted, adjusted models for HDL-C, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2 (adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with HDL-C. Meta-analysis was performed, and p-values (–log10 transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance (p<10^{-8}). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p = 0.05.

Figure S5 Comparison of unadjusted, minimally adjusted, adjusted models for LDL-C, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2 (adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with LDL-C. Meta-analysis was performed, and p-values (–log10 transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance (p<10^{-8}). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p = 0.05.

Figure S6 Comparison of unadjusted, minimally adjusted, adjusted models for triglyceride concentrations, by population. Results of tests of association for four regression models are plotted: model 1 (unadjusted), model 2 (adjusted for age and sex; and site of ascertainment for select PAGE studies), model 3 (adjusted for age, sex, body mass index, current smoking, type 2 diabetes, post-menopausal status, and current hormone use), and model 4 (model 3 with the addition of previous myocardial infarction). Each SNP was tested for an association with triglycerides. Meta-analysis was performed, and p-values (–log10 transformed) of the meta-analysis are plotted along the y-axis. SNP location is given on the x-axis. Each triangle represents a meta-analysis p-value for each population. Models are color coded. Large triangles represent p-values at or smaller than genome-wide significance (p<10^{-8}). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p = 0.05.

Figure S7 Comparison of genetic effect estimates when participants are excluded or included based on medication use with adjustments in WHI. Genetic effect estimates (β) and 95% confidence interval are plotted for each SNP tested for an association. The tests of association were performed on fasting European Americans adjusted for age and sex and excluding participants on lipid lowering medication (blue), including all participants regardless of medication use (green), and all participants on lipid lowering medications are presented in Figures S8, S9, S10.
lowering medication, adjusted for the average HDL-C, LDL-C, and ln(TG) effects estimated by Wu et al [87].

(DOCX)

**Figure S8** HDL-C and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication (Include) versus fasting adults not on lipid lowering medication (Exclude). All tests of association results shown here are minimally adjusted for age and sex.

(DOCX)

**Figure S9** LDL-C and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication versus fasting adults not on lipid lowering medication. All tests of association results shown here are minimally adjusted for age and sex.

(DOCX)

**Figure S10** Transformed triglycerides and the effects of lipid lowering medication use on genetic associations, by population. Comparison of genetic effects and significance when tests of association are performed within fasting adults regardless of lipid lowering medication versus fasting adults not on lipid lowering medication. All tests of association results shown here are minimally adjusted for age and sex.

(DOCX)

**Figure S11** Comparison of HDL-C associations across PAGE study sites, by population. Results of tests of association for the various PAGE study sites are plotted (where available) along with meta-analysis results (META): Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk in Young Adults (CARDIA), Cardiovascular Heart Study (CHS), Epidemiologic Architecture for Genes Linked to Environment (EAGLE), Multiethnic Cohort (MEC), Women’s Health Initiative (WHI), Strong Heart Community Study (SHCS), and Strong Heart Family Study (SHFS) in Arizona (AZ), Oklahoma (OK) and South Dakota (SD). Each SNP was tested for an association with log-transformed triglyceride levels, adjusted for age and sex (Model 2), including fasting adults on lipid lowering medications. SNP location is given on the x-axis and p-values (-log10 transformed) are plotted along the y-axis. Each triangle represents a p-value for each PAGE study. PAGE study sites are color coded. Large triangles represent p-values at or smaller than genome-wide significance (p<10^{-8}). The direction of the arrows corresponds to the direction of the beta coefficient. The exact beta coefficients are reported on the bottom panel. The significance threshold is indicated by the red bar at p=0.05.

(DOCX)

**Table S1** Study characteristics by PAGE study and population. Descriptive statistics for fasting (≥8 hours) adults (≥18 years of age) are expressed as percentage, median, and standard deviation (SD) for each variable.

(DOCX)

**Table S2** List of candidate gene and GWAS-identified SNPs targeted for genotyping in PAGE. For each SNP (denoted by rs number), we list the chromosomal and genomic location, the putative function of the SNP (based on SNP location) and the nearest gene, the number of PAGE studies that genotyped the SNP, the trait associated with the SNP based on the literature, the effect allele and effect size based on the literature, and the reference for these data.

(DOCX)

**Text S1** Study descriptions.

(DOCX)

**Acknowledgments**

The contents of this paper are solely the responsibility of the authors and do not necessarily represent the official views of the NIH. The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention. The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the views of the Indian Health Service. The PAGE consortium thanks the staff and participants of all PAGE studies for their important contributions. The complete list of PAGE members can be found at http://www.pagestudy.org.

The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: http://www.whiscience.org/publications/WHI_investigators_shortlist.pdf.

EAGLE would like to thank Dr. Geraldine McQuillan and Jody McLean for their help in accessing the Genetic NHANES data. EAGLE would also like to thank Dr. William Bush and Justin Giles for their help in
Author Contributions
Conceived and designed the experiments: LD CLC KT FR S LAH PB. Performed the experiments: DD BC. Analyzed the data: LD CLC KT FR S PB. Contributed reagents/materials/analysis tools: JLA GA LG BC SBD CBF HJ KCJ SL LNK ET LJM SAP RVS LRW CAH LLM BVH. Wrote the paper: LD DC.
54. Talmud PJ, Palmen J, Putt W, Lins L, Humphries SE (2005) Determination of
   the Functionality of Common APOA5 Polymorphisms. J Biol Chem 280:
   28215–28220.
55. Vaessen SFC, Sierts JA, Kuivenhoven JA, Schaap FG (2009) The Effect of
   Lipid-Lowering Medications: Genome-Wide Linkage Analysis of
   Lipids in a Japanese Population. J Med Genet 46: 370–374.
56. Wu J, Province M, Coon H, Hunt S, Eckfeldt J, et al. (2007) An Investigation
   of Locus-Specific Effects from Genomewide Scans. Am J Hum Genet 69:
   674–679.
57. Goring HH, Terwilliger JD, Blangero J (2001) Large Upward Bias in
   Estimation of Locus-Specific Effects from Genomewide Scans. Am J Hum
   Genet 69: 1357–1369.
58. Zöller S, Pritchard JK (2007) Overcoming the Winner’s Curse: Estimating
   Penetrance Parameters from Case-Control Data. Am J Hum Genet 80:
   605–615.
59. Centers for Disease Control and Prevention (2010) National Health and
   Nutrition Examination Survey (NHANES) DNA Samples: Guidelines for
   Proposals to Use Samples and Cost Schedule. Federal Register 75:
   32191–32193.
60. Centers for Disease Control and Prevention (2004) Plan and Operation of
   the Third National Health and Nutrition Examination Survey, 1988–94. Bethesda,
   MD.
61. Centers for Disease Control and Prevention (CDC) NCHSN (2002) U.S.
   Department of Health and Human Services, Hyattsville, MD.
62. Kolonel LN, Alhoushler D, Henderson BE (2004) The Multicentric Cohort
   Study: exploring genes, lifestyle and cancer risk. Nat Rev Cancer 4: 519–527.
63. (1998) Design of the Women’s Health Initiative Clinical Trial and Observational
   Study. Controlled Clinical Trials 19: 61–109.
64. Anderson GL, Mannon J, Wallace R, Lund R, Hall D, et al. (2003) Implementation
   of the women’s health initiative study design. Annals of Epidemiology 13:
   85–817.
65. The ARIC Investigators (1989) The Atherosclerosis Risk in Communities (ARIC)
   Study: design and objectives. Am J Epidemiol 129: 667–702.
66. Friedman GD, Cutter GR, Donahue RP, Hughes GH, Holluy SB, et al. (1988)
   CARDIA: Study design, recruitment and some characteristics of the examined
   subjects. J Clin Epidemiol 41: 1105–1116.
67. Fried LP, Borhani NO, Ezenwait P, Furberg CD, Gardin JM, et al. (1991) The
   Cardiovascular Health Study: design and rationale. Ann Epidemiol 3:
   263–276.
68. North KE, Howard BV, Wolsy TK, Best LG, Lee ET, et al. (2003) Genetic and
   Environmental Contributions to Cardiovascular Disease Risk in American
   Indians. Am J Epidemiol 157: 305–314.
69. Lee ET, Wolsy TK, Fabrizi B, Cowan LD, Le NA, et al. (1990) The Strong
   Heart Study. A study of cardiovascular disease in American Indians: design
   and methods. Am J Epidemiol 132: 1141–1155.
70. Friedewald WT, Levy RJ, Fredrickson DS (1972) Estimation of the
   Concentration of Low-Density Lipoprotein Cholesterol in Plasma, without Use of
   the Preparative Ultracentrifuge. Clin Chem 18: 499–501.
71. Matsis T, Amabile JL, Baysse S, Cole SA, Crawford DC, Haiman C, Heiss H,
   Kooperberg C, Le Marchand L, Manolio TA, et al. (2010).
72. Willer CJ, Li Y, Abecasis GR (2010) METAL: fast and efficient meta-analysis of
   multi-phenotype data and meta-analysis. BioData Mining 3: 10.
73. Pendergrass S, Dudek S, Roden DM, Crawford DC, Ritchie MD (2011) Visual
   Integration of Results from BioVU using Synthesis View. Pacific Symposium on
   Biocomputing, pp 265–276.
74. Pendergrass SA, Dudek SM, Crawford DC, Ritchie MD (2010) Synthesis-View:
   Visualization and Interpretation of SNP Association Results for Multi-Cohort,
   Multi-Phenotype Data and Meta-Analysis. BioData Mining 3: 10.
75. Gaedelmann W, Morrison J QUANTO 1.1: A Computer Program for Power and
   Sample-Size Calculations for Genetic-Epidemiology Studies.
76. Gaedelmann WJ (2002) Sample Size Requirements for Association Studies of
   Gene-Gene Interaction. Am J Epidemiol 155: 478–484.
77. Weir BS, Cockermun CC (1984) Estimating F-Statistics for the Analysis of
   Population Structure. Evolution 38: 1358–1370.
78. Mallin JD, Fido M, Jin Y, Kimura M, Tryka K, et al. (2007) The NCBI dbGaP
   database of genotypes and phenotypes. Nat Genet 39: 1181–1186.