Consistent Directions of Effect for Established Type 2 Diabetes Risk Variants Across Populations: The Population Architecture using Genomics and Epidemiology (PAGE) Consortium

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

| Published Version | doi:10.2337/db11-1296 |
| Citable link | http://nrs.harvard.edu/urn-3:HUL.InstRepos:11708555 |
| Terms of Use | This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA |
Consistent Directions of Effect for Established Type 2 Diabetes Risk Variants Across Populations

The Population Architecture using Genomics and Epidemiology (PAGE) Consortium

Christopher A. Haiman,1 Megan D. Fesinmeyer,2 Kylee L. Spencer,3 Petra Buzkóva,4 V. Saroja Voruganti,5 Peggy Wan,1 Jeff Haessler,2 Nora Franceschini,6 Kristine R. Monroe,1 Barbara V. Howard,1 Rebecca D. Jackson,8 Jose C. Florez,9,10,11 Laurence N. Kolonel,12 Steven Buysys,13 Robert J. Goodloe,3 Simin Liu,14 JoAnn E. Manson,9,15 James B. Meigs,9,16 Kevin Waters,1 Kenneth J. Mukamal,17 Sarah A. Pendergrass,3 Peter Shrader,16 Lynne R. Wilkens,12 Lucia A. Hindorff,18 Jose Luis Ambite,19 Kari E. North,6 Ulrike Peters,2 Dana C. Crawford,3 Loic Le Marchand,12 and James S. Pankow20

Common genetic risk variants for type 2 diabetes (T2D) have primarily been identified in populations of European and Asian ancestry. We tested whether the direction of association with 20 T2D risk variants generalizes across six major racial/ethnic groups in the U.S. as part of the Population Architecture using Genomics and Epidemiology Consortium (16,235 diabetes case and 46,122 control subjects from six major U.S. populations). The percentage of positive (OR >1) for putative risk allele associations ranged from 69% in American Indians to 100% in European Americans. Of the nine variants where we observed significant heterogeneity of effect by racial/ethnic group (P heterogeneity < 0.05), eight were positively associated with risk (OR >1) in at least five groups. The marked directional consistency of association observed for most genetic variants across populations implies a shared functional common variant in each region. Fine-mapping of all loci will be required to reveal markers of risk that are important within and across populations. 

Over the past decade, genome-wide association studies (GWAS) and candidate gene association studies have been successful in identifying common risk variants for type 2 diabetes (T2D) (1–15). The loci revealed have provided insight into the genetic basis of this common disease, as well as biological pathways important in its pathogenesis. Most of these previously reported risk variants were identified in very large studies or meta-analyses conducted among populations of European and Asian ancestry and have been associated with modest increases in T2D risk (per-allele odds ratios [ORs] between 1.1 and 1.4) (12). Subsequent testing of these well-established variants in other racial and ethnic groups has been limited (12,16–24), and most of the studies have been undersized and underpowered to provide reliable risk estimates and clarity regarding generalizability of the associations in non-European populations. Aggregating results from multiple studies conducted among racially and ethnically diverse populations is one approach to amass an adequate sample size for replicating these modest genetic associations and extend our understanding of T2D genetics to non-European populations. As part of the Population Architecture using Genomics and Epidemiology (PAGE) Consortium, we have tested 20 validated risk variants for association with T2D. These 20 variants represent 18 risk regions and were examined in as many as 16,235 diabetes case and 46,122 control subjects from six major U.S. populations (European Americans, African Americans, Hispanics, East Asians, Native Hawaiians, and American Indians) from six large population-based studies.

RESEARCH DESIGN AND METHODS

The PAGE Consortium consists of large ongoing population-based studies or consortia (25). The following studies are included in the current study: from the CALCo (Causal Variants Across the Life Course) consortium, ARIC (the...
Atherosclerosis Risk in Communities Study) (26), CHS (Cardiovascular Health Study) (27), and SHS (Strong Heart Study) (28,29); EAGLE (Epidemiologic Architecture of Genes Linked to Environment, based on three National Health and Nutrition Examination Surveys [NHANES]) (30–33); MEC (The Multiracial Cohort) (34) and PAGE (genital schistosomiasis); and 1000 Genomes Project. Detailed information about each study can be found in Supplementary Data.

**Diabetes case and control definitions.** To facilitate harmonization of diabetes case definitions across studies, data-collection methods were reviewed and compared between studies. All studies collected self-reported information on previous diagnosis by a physician or medical professional and use of medication for diabetes. However, different sets of questions were used to assess whether the diagnosis was based on fasting blood glucose levels, which more specifically define uncontrolled or undiagnosed T2D. In order to incorporate the T2D information across studies, two case definitions were allowed: self-report and exam based. To be classified as a case subject according to the self-report definition, participants had to report both a previous diagnosis of diabetes and use of medication to treat diabetes. To be classified as a control subject (self-report), participants had to report neither previous diagnosis nor use of diabetes medications. To be classified as a case subject according to the exam-based definition, participants had to either meet the self-report case definition or have a fasting (≥8 h) blood glucose ≥126 mg/dL. To be classified as a control subject (exam based), participants had to be classified as a control subject per the self-report definition and have a fasting blood glucose <126 mg/dL. Both prevalent and incident cases were included. For both definitions, self-reported diabetes diagnoses before age 30 were excluded. Sensitivity analyses in the ARIC study suggested that the magnitude of association between candidate variants and T2D did not differ systematically according to the case definitions we applied (Supplementary Data). Additional study-specific details on the data-collection methods and case definitions can be found in the Supplementary Data.

A total of 16,235 diabetes case and 46,122 control subjects were included in this study (case and control subjects, respectively, by study: ARIC, 1,348/10,078; CHS, 850/4,488; SHS, 1,575/1,249; MEC, 6,298/9,980; EAGLE/NHANES, 1,029/1,502; and WHI, 5,120/14,925). None of these studies was involved in the initial discovery efforts of these T2D risk loci. The data from the MEC have previously been reported (37).

**Genotyping.** The 20 variants evaluated in the current study were selected from 18 genomic regions found to be significantly associated with risk of T2D in studies published as of September 2009 (Supplementary Table 1). In the CDKN2A/CDKN2B and KCNQ1 regions, more than one variant was investigated, as many of the index signals identified in the initial GWAS populations are not perfectly correlated. An additional variant, rs8050136, at the CDKN2A/CDKN2B locus, was also evaluated in the HapMap samples in more than one laboratory was assessed by genotyping 360 HapMap samples from populations most relevant to American Indian (SHS) or East Asian (MEC) samples or samples in EAGLE. As Asians, and 2% in Native Hawaiians (<1% in Hispanics and Europeans) and interferes with genotyping assays, thus resulting in genotype misclassification. Genetic markers that distinguish the major ancestral populations (African, European, and Asian) were available in three studies. For ARIC, principal components of ancestry were derived from 200,000 SNPs genotyped on a custom array. For WHI (all populations) and MEC (African Americans and Native Hawaiians), ~100 ancestry-informative markers were used in a principal-components analysis to assess major axes of variation (38,39). For a subset of the MEC Latins, principal components were derived from markers on the Illumina 2.5M array. The ancestry information was not available for the majority of the American Indian (SHS) or East Asian (MEC) samples or samples in EAGLE.

**Statistical analysis.** β values and SEs for each variant were obtained by unconditional logistic regression or Cox proportional hazards regression. For each variant, the allele tested was the allele that was associated with increased risk in previous studies. In each study, models were run separately for each racial/ethnic population and adjusted for sex, age (continuous), and BMI (continuous) unless otherwise noted. The overlap of the ancestry information was not available for the majority of the American Indian (SHS) or East Asian (MEC) samples or samples in EAGLE.

This selection was nonrandom, therefore, analyses in WHI incorporated inverse probability weighting to account for sampling. For SHS, models were also run separately for each center. Information on genetic ancestry was available for a large number of European Americans (~64%), African Americans (~85%), Hispanics (65%), and Native Hawaiians (~33%). Results were similar after adjustment for population structure in all populations except for five SNPs in Native Hawaiians and four SNPs in Africans, where log ORs changed by >20% and P values changed by more than one order of magnitude in either direction (Supplementary Table 2). For each ethnic group, a pooled estimate was calculated using a fixed-effects model and analyzed with the weighted mean effect size.  The value of the log OR. A combined estimate across ethnic groups was calculated using a random-effects model. We tested also for heterogeneity by study and by race using the Q statistic. For Native Hawaiians (MEC), we used the results adjusted for genetic ancestry. Similarly, for Latinos results are presented for MEC and WHI, as no ancestry information was available in EAGLE. All reported P values were derived from two-sided statistical tests. A P value <0.05 was used to declare an association as statistically significant. For each SNP in each racial/ethnic population, we estimated the statistical power to detect the previously reported relative risks in discovery populations of European or Asian ancestry (40) (Supplementary Table 1).

**RESULTS**

The descriptive characteristics of case and control subjects by racial/ethnic group and study are presented in Table 1. The mean age of case or control subjects ranged across studies from 47.1 (EAGLE, African American control subjects) to 73.0 (CHS, European American case subjects and African American control subjects). Both men and women were represented in each study except for WHI, which included only women. Case subjects were consistently heavier than control subjects in each study and population (Table 1).

We found no significant association with the first principal component (a measure of European admixture) and T2D risk in African Americans (in ARIC, MEC, or WHI). In Native Hawaiians, the first principal component is a measure of European admixture (and ancestry) and was significantly inversely associated with T2D risk (P = 3.2 × 10^-5) (Supplementary Fig. 1). In Native Hawaiians, the significance of the association with three variants, which were all more common in Native Hawaiians than European Americans, diminished after adjustment for stratification (rs10010131, WFS1; rs7754840, CDKAL1; and rs864745, JAZF1). In contrast, the variants at TCFTL72 (rs7903146) and KCNQ1 (rs2237887) became nominally significant. The observation of larger β values for TCFTL72 and KCNQ1 variants after adjustment for stratification is consistent with negative confounding due to lower risk allele frequencies in Native Hawaiians compared with European Americans (Supplementary Table 1) and an inverse association of European ancestry and T2D risk in this population. Similarly, in Hispanics the first principal component, which is also a measure of European admixture (and ancestry) in this population, was significantly associated with lower T2D risk (P = 2.1 × 10^-12 in the MEC) (Supplementary Fig. 2). Adjustment for the first principal component in Hispanics increased the OR and degree of statistical significance for three SNPs that were all less common, although marginally, in Hispanics than in European Americans (rs2237887, KCNQ1; rs4402960, JAZF1; and rs864745, TCFTL72) and diminished significance for rs864745 (JAZF1), which is more common in Hispanics than in European Americans.

For the most part, the risk allele frequencies of each population tracked with the risk allele frequency of European Americans (Supplementary Fig. 3). Effect estimates were >1 for 69–100% of the SNPs across populations (average: 85%) (Fig. 1). Three variants were significantly associated (P < 0.05) with risk in at least four groups (rs4402960, JAZF1; rs864745, JAZF1; and rs7903146, TCFTL72), and of the 17 SNPs evaluated in five or more populations, positive associations were observed with 13 SNPs (OR >1) in at least five groups (Fig. 1). Of the 108
We examined 20 validated risk variants for T2D, representing 18 risk regions, in as many as 16,235 diabetes case and 46,122 control subjects from six major population groups. The vast majority of the variants were positively associated with risk in the European Americans and African Americans. However, adjustment for population structure using principal components did affect the association with several variants for Native Hawaiians and Hispanics. Native Hawaiians are highly admixed with the three main groups being Polynesian, Asian, and European. The first few principal components capture European admixture, with European ancestry lower in Hawaiian case subjects than in control subjects (41). Therefore, adjustment for European admixture reduced the strength of association for some of the variants that were more common in Polynesians and increased the strength of some of the variants more common in Europeans. Similar differences were noted for some SNPs after principal-components adjustment in Hispanics. Unfortunately, ancestry-informative markers were not available to address the issue of population stratification in the admixed American Indian populations.

The marked directional consistency of association for most genetic variants across populations implies a shared functional common variant in each region. This general pattern of consistency provides little support for the "synthetic association" model (42), which suggests that GWAS signals with common alleles are due to rare alleles, many of which are likely to be ethnically distinct. The inability to replicate associations with variants in populations where statistical power is sufficient may highlight loci for which fine-mapping may be helpful. For example, in African Americans, power was high (≥94%) to detect significant estimated effects (total number of tests: SNP × population), 91 had ORs >1 (84%). Removing European Americans, the population in which most of the original signals were reported, only reduced this percentage to 80%. We observed significant heterogeneity of effect by racial/ethnic group for nine SNPs (P_{heterogeneity} < 0.05). However, aside from rs7901581 at TSPAN8, eight of these variants (at THADA, IGF2BP2, WFS1, CDKAL1, CDKN2A/CDKN2B [rs2383208], TCF7L2, KCNQ1 [rs2373895], and KCNJ11) were positively associated with risk (OR >1) in at least five populations (Fig. 1). Thus, even for variants that displayed evidence of significant heterogeneity across population, the direction of effect was generally consistent in the majority of the populations.

We did not find evidence of substantial confounding by population stratification in European Americans or African Americans. However, adjustment for population structure using principal components did affect the association with several variants for Native Hawaiians and Hispanics. Native Hawaiians are highly admixed with the three main groups being Polynesian, Asian, and European. The first few principal components capture European admixture, with European ancestry lower in Hawaiian case subjects than in control subjects (41). Therefore, adjustment for European admixture reduced the strength of association for some of the variants that were more common in Polynesians and increased the strength of some of the variants more common in Europeans. Similar differences were noted for some SNPs after principal-components adjustment in Hispanics. Unfortunately, ancestry-informative markers were not available to address the issue of population stratification in the admixed American Indian populations.

The marked directional consistency of association for most genetic variants across populations implies a shared functional common variant in each region. This general pattern of consistency provides little support for the "synthetic association" model (42), which suggests that GWAS signals with common alleles are due to rare alleles, many of which are likely to be ethnically distinct. The inability to replicate associations with variants in populations where statistical power is sufficient may highlight loci for which fine-mapping may be helpful. For example, in African Americans, power was high (≥94%) to detect significant estimated effects (total number of tests: SNP × population), 91 had ORs >1 (84%). Removing European Americans, the population in which most of the original signals were reported, only reduced this percentage to 80%. We observed significant heterogeneity of effect by racial/ethnic group for nine SNPs (P_{heterogeneity} < 0.05). However, aside from rs7901581 at TSPAN8, eight of these variants (at THADA, IGF2BP2, WFS1, CDKAL1, CDKN2A/CDKN2B [rs2383208], TCF7L2, KCNQ1 [rs2373895], and KCNJ11) were positively associated with risk (OR >1) in at least five populations (Fig. 1). Thus, even for variants that displayed evidence of significant heterogeneity across population, the direction of effect was generally consistent in the majority of the populations.
associations, with the index variants at five loci (WFS1, HHEX, CDKN2A/B, THADA, and KCNQ1) that were found to be significantly associated with risk in at least one of the other non-European populations. The lack of a statistically significant association in African Americans at these loci could be because the risk allele is relatively invariant in populations of African ancestry or low linkage disequilibrium between the index signal and the functional allele. Fine-mapping of these loci, and others such as TCF7L2 in American Indians, where we observed no evidence of a significant association (OR 1.08 [95% CI 0.90–1.29]) despite >99% power and despite the suggestion that rs7903146 is the biologically functional variant in African Americans (43) and in genomic studies of open chromatin (44), should be of high priority to extract information about any genetic risk conferred at that locus that may be important for these populations.

This study has a number of limitations. In the design, we allowed for both incident and prevalent diabetes cases as well as different case/control criteria depending on study; however, our sensitivity analysis of the different case groups (Supplementary Data) did not suggest systematic differences in effect sizes based on study design, case definition, or analytic approach. We also had no information about type 1 diabetes in some studies, although case subjects known to be diagnosed before age 30 years were excluded and most participants in these studies were middle-aged or older adults.

This is the largest effort to date to investigate the generalizability of T2D susceptibility variants in the major racial/ethnic groups of the U.S. The consistent patterns of association for these variants provide additional support for the importance of these loci in contributing to T2D risk in multiple populations. Identification of the underlying
biological functional allele(s) in each region, through fine-mapping, will be required to determine the extent to which these regions contribute to racial and ethnic disparities in T2D risk.

ACKNOWLEDGMENTS

The PAGE program is funded by the National Human Genome Research Institute, supported by U01HG004790 (CaliCo [Causal Variants Across the Life Course]), U01HG004798 (EAGLE [Epidemiologic Architecture of Genes Linked to Environment]), U01HG004802 (Mec [Multietnic Cohort]), U01HG004790 (WHI [Women’s Health Initiative]), and U01HG004801 (Coordinating Center).

No potential conflicts of interest relevant to this article were reported.

C.A.H. performed experiments, analyzed data, and wrote the manuscript. M.D.F., K.L.S., P.B., V.S.V., P.W., J.H., and N.F. performed experiments, analyzed data, and contributed to writing the manuscript. K.R.M., B.V.H., R.D.J., J.C.F., L.N.K., S.B., R.J.G., J.E.M., J.B.M., K.W., K.J.M., S.A.P., P.S., L.R.W., L.A.H., J.L.A., K.E.N., U.P., D.C.C., and L.L.M. contributed materials and to the study design, analysis tools, and interpretation of results and contributed to writing the manuscript. J.S.P. performed the experiments, analyzed data, and wrote the manuscript. C.A.H. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study-specific acknowledgments are listed in the Supplementary Data.

REFERENCES

2. Goyon AL, Weedon MN, Owen KR, et al. Large-scale association studies of variants in genes encoding the pancreatic beta-cell KATP channel subunits Kir6.2 (KCNJ11) and SUR1 (ABCC8) confirm that the KCNJ11 E23K variant is associated with type 2 diabetes. Diabetes 2003;52:568–572