American Journal of Epidemiology

Vol. 176, No. 11
DOI: 10.1093/aje/kws176
Advance Access publication:
November 9, 2012

Original Contribution

Evaluation of Genome-wide Association Study-identified Type 2 Diabetes Loci in African Americans

Jirong Long*, Todd Edwards, Lisa B. Signorello, Qiuyin Cai, Wei Zheng, Xiao-Ou Shu, and William J. Blot

* Correspondence to Dr. Jirong Long, Vanderbilt Epidemiology Center, Institute of Medicine and Public Health, Vanderbilt University Medical Center, 2525 West End Avenue, 8th Floor, Nashville, TN 37203-1738 (e-mail: jirong.long@vanderbilt.edu).

Initially submitted September 7, 2011; accepted for publication March 14, 2012.

Type 2 diabetes (T2D) is up to twice as prevalent among African Americans as Caucasians. Recent genome-wide association studies (GWAS) have identified multiple common genetic risk variants for T2D; however, none of these studies were conducted exclusively among subjects of African ancestry. Investigating these known loci in other populations would be an expedient way to evaluate the generalizability of the current findings. The authors evaluated 29 known T2D loci in a large southeastern US cohort study including 4,288 African Americans (1,554 cases and 2,734 controls) enrolled during 2002–2009. Seven of the 29 single nucleotide polymorphisms (SNPs) examined were found to be associated with T2D risk at $P \leq 0.05$, including rs6769511 (*IGF2BP2*), 2 SNPs in the *WFS1* gene (rs4689388 and rs1801214), rs7903146 (*TCF7L2*), and 3 SNPs in the *KCNQ1* gene (rs231362, rs2237892, and rs2237897). Notably, the association for rs7903146 reached the GWAS significance level ($P = 3.6 \times 10^{-8}$), with an odds ratio per T allele of 1.32 (95% confidence interval: 1.20, 1.46). Regional analyses using GWAS data from Vanderbilt University's BioVU DNA biobank showed significant associations ($P < 0.05$) with 9 loci, though no association was observed for the index SNPs reported in European- or Asian-ancestry populations. These results extend some of the recent GWAS findings to African Americans and may guide future efforts to identify causal variants for T2D.

African Americans; diabetes mellitus, type 2; genetics; genome-wide association study; molecular epidemiology; single nucleotide polymorphism

Abbreviations: CI, confidence interval; GWAS, genome-wide association studies; ICD-9, *International Classification of Diseases*, Ninth Revision; SCCS, Southern Community Cohort Study; SNP, single nucleotide polymorphism.

Type 2 diabetes is one of the leading health problems in the United States, affecting approximately 21 million persons or almost 10% of the US adult population (1). Type 2 diabetes is nearly twice as prevalent among African Americans as among Caucasians (1).

Although obesity and other environmental factors are major determinants of type 2 diabetes risk, genetic factors are believed to play an important role in its etiology (2). Several loci have been identified through traditional candidate gene and linkage approaches (3–7). Recent genome-wide association studies (GWAS) of type 2 diabetes among populations of European or Asian ancestry have discovered multiple novel, reproducible susceptibility loci (2, 8–14). However, little or no investigation of the associations between these loci and type 2 diabetes has been reported for African-derived populations. The genetic architecture of African populations differs considerably from that of other ethnic groups with regard to type 2 diabetes loci (15), as well as in general (16). In this study, we evaluated 29 type 2 diabetes loci in African Americans using data from a large ongoing cohort study, the Southern Community Cohort Study (SCCS), and from Vanderbilt University Medical Center's DNA biobank (BioVU).

MATERIALS AND METHODS

The SCCS is a prospective cohort study initiated in 2002 focusing on investigation of racial disparities in the risk of cancer and other chronic diseases. Men and women aged 40–79 years from 12 southeastern US states were enrolled between 2002 and 2009 (17). The SCCS includes approximately 86,000 participants; African Americans comprise two-thirds of the cohort. At baseline in the SCCS, participants completed a comprehensive in-person interview covering various aspects of health conditions and behavioral factors, including personal and family medical history, and other lifestyle factors. During this interview, participants were asked, "Has a doctor ever told you that you have had diabetes or high blood sugar?" Participants responding "yes" were asked follow-up questions regarding their age at first diagnosis and use (and names) of prescription medications taken to manage their diabetes. In the current study, we considered only those subjects who met the following additional criteria as cases: 1) use of type 2 diabetes medication, 2) age at diagnosis greater than 30 years, 3) no cancer at study enrollment, and 4) self-reported African-American race/ethnicity. Subjects who responded "no" to the question, "Has a doctor ever told you that you have had diabetes or high blood sugar?" and 1) had no cancer at study enrollment and 2) were self-reported African Americans were considered potentially eligible controls. Controls were frequency-matched to cases (2:1) by age at enrollment (within 5 years), age at type 2 diabetes onset, gender, and residence (same state). In total, 1,124 cases and 2,157 controls were successfully genotyped. Study participants provided written informed consent, and the study was approved by committees for the use of human subjects at both collaborating institutions.

Vanderbilt's DNA biobank project, BioVU, accrues DNA samples extracted from blood remaining from routine clinical testing. A full description of the BioVU resource and its ethical protections has been published elsewhere (18). Details on the definitions of type 2 diabetes cases and controls are given elsewhere (19). In brief, *International Classification of Diseases*, Ninth Revision (ICD-9) codes, laboratory data, and natural language processing techniques on unstructured patient records, such as records of medications, electrocardiograms, or medical history, were used to define cases and controls. The following 3 sets of criteria were used to define type 2 diabetes cases. Case definition I: 1) any of the following ICD-9 codes—250.3, 250.32, 250.2, 250.22, 250.9, 250.92, 250.8, 250.82, 250.7, 250.72, 250.6, 250.62, 250.5, 250.52, 250.4, 250.42, 250, and 250.02—and 2) use of non-insulin diabetes medication. Case definition II: 1) any of the above ICD-9 codes; 2) a glucose concentration greater than 200 mg/dL or a hemoglobin A_{1c} level greater than 6.5%; and 3) no use of insulin medication. Case definition III: 1) use of non-insulin diabetes medication and 2) a glucose concentration greater than 200 mg/dL or a hemoglobin A_{1c} level greater than 6.5%. The controls were defined as "record does not contain any of the following information": 1) ICD-9 codes from the case definition or any of following ICD-9 codes—790.21, 790.22, 790.29, 791.5, 648.8, 277.7, and 250; 2) use of insulin medication; 3) use of non-insulin diabetes medication; 4) history of diabetes; and 5) a glucose concentration ≥110 mg/dL or a hemoglobin A_{1c} level ≥6.0%.

SNP selection and genotyping

All single nucleotide polymorphisms (SNPs) associated with type 2 diabetes at a level of $P < 5 \times 10^{-8}$ in the GWAS catalog (20), plus SNPs identified through fine mapping linkage signal or candidate gene approaches (3–7), were included. SNPs with minor allele frequencies less than 0.05 in HapMap Yoruba data (http://hapmap.ncbi.nlm.nih.gov/) were excluded. For SNPs in linkage disequilibrium with each other with $r^2 \geq 0.4$ in the HapMap Yoruba data, the SNP with the lowest P value in reported studies was selected for each locus.

SCCS genotyping was completed using the iPLEX Sequenom MassArray platform (Sequenom, Inc., San Diego, California). A total of 30 SNPs were designed in 1 pool, and 29 SNPs were successfully genotyped. Included in each 96-well plate as quality control samples were 2 negative controls (water), 2 blinded duplicates, and 2 samples from the HapMap project. The mean concordance rate was 99.95% for the blind duplicates and 99.66% for HapMap samples compared with HapMap data.

BioVU samples were genotyped using the Illumina Human1M-Duov3 array (Illumina, Inc., San Diego, California) at the DNA Resources Core (Center for Human Genetics Research) at Vanderbilt University. SNPs were removed if they had a genotyping efficiency less than 0.95, a Hardy-Weinberg equilibrium P value less than 1×10^{-7}, or a minor allele frequency less than 0.01. Subjects' genotypes were imputed with IMPUTE software (Mathematical Genetics and Bioinformatics Groups, Department of Statistics, University of Oxford, Oxford, United Kingdom (https://mathgen.stats.ox.ac.uk/impute)) for all SNPs in the HapMap2 r22 CEU + YRI data, and only SNPs with imputation quality score information greater than 0.4 were extracted. Among the 29 SNPs successfully genotyped in SCCS, genotypes were directly available for 15 in the BioVU samples through the Illumina Human1M-Duov3 array, with genotypes being available for 11 SNPs through imputation. Data for 3 SNPs, including rs11634397, rs2237897, and rs4457053, were not available from the BioVU cohort. Because of the different genetic structures, SNPs associated with type 2 diabetes in African Americans may be different from those reported in other populations. Therefore, we also investigated all SNPs located within 1 millimorgan (0.1 cM) of the index SNP using the GWAS data for the BioVU cohort. However, the statistical power of this analysis was limited because of the small sample size in the BioVU GWAS.

Statistical analysis

Associations between the SNPs and type 2 diabetes risk were assessed using odds ratios and 95% confidence intervals derived from logistic regression models and adjusted for age, body mass index, sex, and study site, when appropriate.

Potential confounding by population structure was adjusted for by principal components analysis using EIGENSTRAT (21).

To evaluate the combined effect of the SNPs on type 2 diabetes risk, we created a genetic risk score by summing the number of risk alleles each subject carried at each SNP. In the SCCS cohort, persons with missing data on 5 or more of the 29 SNPs were excluded (13 cases and 26 controls). For persons with missing data on fewer than 5 SNPs, missing data for a particular SNP were assigned for the average of the number of risk alleles at that SNP for cases and controls separately. For the 3 SNPs for which data were not available in BioVU, we assigned the average of the number of risk alleles at each SNP from SCCS for cases and controls separately. We also carried out the genetic risk score analyses among the subjects with complete genotype data. Because data on 3 SNPs (rs11634397, rs2237897, and rs4457053) were not available in BioVU subjects, the genetic risk score analysis for the BioVU data set included 26 SNPs. All statistical analyses were conducted in SAS, version 9.2 (SAS Institute Inc., Cary, North Carolina), with the use of 2-tailed tests.

RESULTS

Characteristics of the study participants are shown in Table 1. In both the SCCS and the BioVU cohort, type 2 diabetes cases tended to be older and to have a higher body mass index than controls.

Among the 29 SNPs, 22 (75.9%) were associated with risk in the same direction as initial reports. This is higher than would be expected under the null hypothesis ($P = 0.008$, binomial sign test). Seven SNPs were nominally statistically significant ($P \leq 0.05$), including rs6769511 in the insulin-like growth factor 2 mRNA binding protein 2 gene (*IGF2BP2*), 2 SNPs in the Wolfram syndrome 1 gene (*WFS1*) (rs4689388 and rs1801214), rs7903146 in the transcription factor 7-like 2 (T-cell-specific, HMG-box) gene (*TCF7L2*), and 3 SNPs in the potassium voltage-gated channel, KQT-like subfamily, member 1, gene (*KCNQ1*) (rs231362, rs2237892, and rs2237897) (Table 2). The pattern of the association for these SNPs tended to be consistent in the SCCS and BioVU, with the SNPs rs6769511 (*IGF2BP2*) and rs7903146 (*TCF7L2*) showing significant associations in both studies. The association for rs7903146 (*TCF7L2*) reached GWAS-level significance, with an odds ratio per T allele of 1.32 (95% confidence interval (CI): 1.20, 1.46) and a P value of 3.6×10^{-8} (Table 2). After additional adjustment for population structure, these 7 SNPs still showed an association at $P < 0.05$, and the associations for the SNPs rs4689388, rs1801214, rs7903146, rs231362, and rs2237892 became slightly stronger (Table 2). Principal components analysis plotting showed that most subjects were clustered close to each other, except for 11 subjects from BioVU (see Web Figure 1, which appears on the *Journal*'s website (http://aje.oxfordjournals.org/)). After excluding these 11 outliers, we reran the association analyses, and the results did not change materially (Web Table 1). Considering that using age at diagnosis greater than 30 may not completely rule out the presence of type 1 diabetes, we also conducted analysis by excluding cases with ages of diagnosis under 35 or 40 years (Web Table 2). No material changes in results were noted. Additionally, we carried out analysis without adjustment for body mass index and found little change in the results (Web Table 2).

Table 3 presents the association of type 2 diabetes risk with the genetic risk score. A dose-response association was observed between the number of risk alleles and risk of type 2 diabetes; the odds ratio per unit of the genetic risk score was 1.05 (95% CI: 1.03, 1.07). The odds ratios for type 2 diabetes risk across increasing quartiles of the genetic risk score were 1.00 (reference), 1.28 (95% CI: 1.06, 1.54), 1.33 (95% CI: 1.12, 1.59), and 1.54 (95% CI: 1.27, 1.86), respectively ($P = 8.2 \times 10^{-6}$). After additional adjustment for the first 5 principal components for population structure, the results were very similar. Results were similar when the analysis was limited to persons with complete genotype data for all variants.

Regional analyses from the BioVU GWAS data showed that significant associations ($P < 0.05$) were observed for SNPs located in 9 other loci. At these loci, the index SNPs, which were reported in original GWAS in European- or Asian-ancestry populations, were not associated with type 2 diabetes. These 9 loci included rs1531343, rs243021,

Table 1. Characteristics of Type 2 Diabetes Cases and Controls in the Southern Community Cohort Study and the BioVU DNA Biobank, 2002–2012

Characteristic	SCCS				BioVU			
	Cases		Controls		Cases		Controls	
	No. of Persons	Mean (SD)	No. of Persons	Mean (SD)	No. of Persons	Mean (SD)	No. of Persons	Mean (SD)
Women	779		1,486		298		408	
Age, years		58.6 (8.8)		56.1 (9.2)		54.9 (12.8)		48.5 (12.4)
Body mass index[a]		34.0 (7.3)		31.2 (7.2)		36.3 (8.4)		31.8 (8.2)
Men	345		671		132		169	
Age, years		62.2 (8.3)		61.2 (8.7)		56.2 (11.6)		50.1 (11.4)
Body mass index		30.9 (5.4)		27.6 (5.3)		31.6 (6.3)		29.3 (5.7)

Abbreviations: SCCS, Southern Community Cohort Study; SD, standard deviation.
[a] Weight (kg)/height $(m)^2$.

Table 2. Association of Known Type 2 Diabetes Risk Alleles With Type 2 Diabetes Among African Americans, Southern Community Cohort Study and BioVU DNA Biobank, 2002–2012

Single Nucleotide Polymorphism	Chromosomal Region	Gene	Alleles[a]	No. of Cases	No. of Controls	Frequency[b]		Risk per Allele[c]		*P* Value[c]	Risk per Allele[d]		*P* Value[d]	Power[e]
						Cases	Controls	OR	95% CI		OR	95% CI		
rs10923931	1p12	*NOTCH2, ADAM30*	T/G	1,549	2,725	0.33	0.33	1.05	0.95, 1.15	3.6×10^{-1}	1.04	0.94, 1.15	4.7×10^{-1}	0.83
rs7578597	2p21	*THADA*	T/C	1,544	2,711	0.74	0.73	1.04	0.94, 1.15	4.6×10^{-1}	1.07	0.96, 1.19	2.3×10^{-1}	0.86
rs243021	2p16.1	*BCL11A*	A/G	1,545	2,714	0.37	0.37	1.01	0.92, 1.11	8.6×10^{-1}	1.02	0.92, 1.13	7.4×10^{-1}	0.51
rs7593730	2q24.2	*RBMS1/ITGB6*	C/T	1,545	2,713	0.64	0.63	1.06	0.96, 1.16	2.6×10^{-1}	1.05	0.95, 1.16	3.5×10^{-1}	0.72
rs4607103	3p14.1	*ADAMTS9*	C/T	1,419	2,513	0.72	0.71	1.06	0.96, 1.18	2.4×10^{-1}	1.05	0.95, 1.17	3.5×10^{-1}	0.53
rs4402960	3q27.2	*IGF2BP2*	T/G	1,546	2,715	0.53	0.51	1.07	0.97, 1.17	1.7×10^{-1}	1.06	0.96, 1.17	2.3×10^{-1}	0.90
rs6769511	3q27.2	*IGF2BP2*	C/T	1,541	2,715	0.79	0.76	1.23	1.10, 1.37	2.9×10^{-4}	1.23	1.10, 1.37	3.5×10^{-4}	0.99
rs4689388	4p16.1	*WFS1/PPP2R2C*	T/C	1,549	2,721	0.74	0.72	1.17	1.05, 1.29	3.9×10^{-3}	1.17	1.05, 1.30	3.7×10^{-3}	0.90
rs1801214	4p16.1	*WFS1*	T/C	1,192	2,246	0.69	0.67	1.11	1.00, 1.22	4.5×10^{-2}	1.11	1.00, 1.22	4.4×10^{-2}	0.81
rs4457053	5q13.3	*ZBED3*	G/A	1,105	2,126	0.19	0.19	1.00	0.88, 1.15	9.6×10^{-1}	1.01	0.87, 1.18	8.6×10^{-1}	0.32
rs4712523	6p22.3	*CDKAL1*	G/A	1,549	2,722	0.60	0.61	0.97	0.88, 1.07	5.3×10^{-1}	0.96	0.87, 1.07	5.0×10^{-1}	1.00
rs10440833	6p22.3	*CDKAL1*	A/T	1,453	2,565	0.20	0.20	1.02	0.91, 1.15	6.8×10^{-1}	1.03	0.91, 1.17	6.6×10^{-1}	0.99
rs864745	7p15.1	*JAZF1*	T/C	1,496	2,663	0.75	0.75	1.04	0.94, 1.16	4.3×10^{-1}	1.03	0.92, 1.16	5.6×10^{-1}	0.56
rs972283	7q32.3	*KLF14*	G/A	1,118	2,144	0.85	0.85	1.03	0.90, 1.17	6.9×10^{-1}	1.05	0.92, 1.20	4.8×10^{-1}	0.28
rs2383208	9p21.3	*CDKN2A/CDKN2B*	A/G	1,545	2,719	0.81	0.81	1.02	0.91, 1.14	7.5×10^{-1}	1.00	0.88, 1.13	9.6×10^{-1}	1.00
rs13292136	9q21.31	*CHCHD9*	C/T	1,532	2,701	0.92	0.92	1.05	0.88, 1.24	6.0×10^{-1}	1.05	0.88, 1.25	6.1×10^{-1}	0.33
rs1111875	10q23.33	*HHEX*	C/T	1,550	2,726	0.79	0.77	1.08	0.97, 1.21	1.6×10^{-1}	1.10	0.96, 1.25	1.8×10^{-1}	0.97
rs5015480	10q23.33	*HHEX,IDE*	C/T	1,497	2,629	0.62	0.62	1.03	0.94, 1.14	5.0×10^{-1}	1.02	0.91, 1.15	7.0×10^{-1}	0.97
rs7903146	10q25.2	*TCF7L2*	T/C	1,538	2,708	0.34	0.28	1.32	1.20, 1.46	3.6×10^{-8}	1.32	1.20, 1.46	3.2×10^{-8}	1.00
rs231362	11p15.5	*KCNQ1*	G/A	1,247	2,303	0.82	0.79	1.14	1.01, 1.28	2.9×10^{-2}	1.14	1.01, 1.28	2.9×10^{-2}	0.39
rs2237892	11p15.5	*KCNQ1*	C/T	1,551	2,725	0.91	0.89	1.19	1.02, 1.40	2.6×10^{-2}	1.20	1.02, 1.40	2.4×10^{-2}	1.00
rs2237897	11p15.4	*KCNQ1*	C/T	1,123	2,151	0.92	0.91	1.26	1.04, 1.53	1.7×10^{-2}	1.26	1.04, 1.53	1.7×10^{-2}	0.90
rs1531343	12q14.3	*HMGA2*	C/G	1,496	2,657	0.38	0.38	1.00	0.91, 1.10	9.6×10^{-1}	0.98	0.89, 1.08	6.9×10^{-1}	0.66
rs7961581	12q21.1	*TSPAN8/LGR5*	C/T	1,295	2,377	0.19	0.18	1.05	0.94, 1.18	3.7×10^{-1}	1.06	0.94, 1.19	3.8×10^{-1}	0.44
rs7957197	12q24.31	*HNF1A*	T/A	1,340	2,470	0.83	0.84	0.93	0.82, 1.05	2.3×10^{-1}	0.93	0.82, 1.05	2.3×10^{-1}	0.29
rs7172432	15q22.2	*C2CD4A/C2CD4B*	A/G	1,548	2,722	0.32	0.30	1.10	0.99, 1.21	7.0×10^{-2}	1.10	1.00, 1.22	5.9×10^{-2}	0.69
rs11634397	15q25.1	*ZFAND6*	G/A	1,116	2,146	0.42	0.43	1.00	0.90, 1.11	9.7×10^{-1}	1.00	0.89, 1.12	9.9×10^{-1}	0.30
rs8050136	16q12.2	*FTO*	A/C	1,550	2,721	0.44	0.44	0.98	0.90, 1.08	7.4×10^{-1}	0.98	0.89, 1.07	6.3×10^{-1}	1.00
rs391300	17p13.3	*SRR*	G/A	1,549	2,722	0.47	0.48	0.97	0.89, 1.06	5.2×10^{-1}	0.96	0.88, 1.06	4.3×10^{-1}	1.00

Abbreviations: CI, confidence interval; OR, odds ratio.

[a] Effect allele/reference allele.

[b] Frequency (proportion) of effect allele.

[c] Adjusted for age, gender, body mass index, and study site.

[d] Adjusted for age, gender, body mass index, study site, and first 5 principal components for population structure.

[e] Statistical power was estimated under the additive model. The reported effect size from the previous genome-wide association study and the effective allele frequency among controls in the current study were used to estimate the power ($\alpha = 0.05$, 1-sided).

Table 3. Association of Genetic Risk Score With Risk of Type 2 Diabetes in African Americans, Southern Community Cohort Study and BioVU DNA Biobank, 2002–2012

	No. of Cases	No. of Controls	OR[a]	95% CI	*P* Value	OR[b]	95% CI	*P* Value
GRS as continuous variable	1,541	2,708	1.05	1.03, 1.07	2.62×10^{-7}	1.06	1.03, 1.08	1.36×10^{-7}
Quartile of GRS					8.2×10^{-6}			6.7×10^{-6}
1 (lowest)	383	807	1.00	Reference		1.00	Reference	
2	361	636	1.28	1.06, 1.54		1.29	1.07, 1.56	
3	433	736	1.33	1.12, 1.59		1.36	1.13, 1.64	
4 (highest)	364	529	1.54	1.27, 1.86		1.59	1.30, 1.94	

Abbreviations: CI, confidence interval; GRS, genetic risk score; OR, odds ratio.
[a] Adjusted for age, gender, body mass index, and study site.
[b] Adjusted for age, gender, body mass index, study site, and first 5 principal components for population structure.

rs4402960, rs4457053, rs4607103, rs4712523, rs7172432, rs8050136, and rs864745. However, the associations for these SNPs need be validated in independent populations.

DISCUSSION

In the present study, we investigated associations of 29 diabetes susceptibility variants with risk of type 2 diabetes in 4,288 African-American subjects. Most of these variants were discovered in GWAS conducted among populations of European or Asian ancestry. We found 7 of these SNPs to be significantly associated with diabetes risk among African Americans. Notably, the SNP rs7903146 (*TCF7L2*) reached the GWAS significance level of $P = 3.6 \times 10^{-8}$. This finding suggests that rs7903146 is a key determinant of type 2 diabetes risk in the African-American population.

Among the 7 SNPs validated in this study, rs2237897 was investigated in the Multiethnic Cohort Study of Diet and Cancer; however, no association was observed among African Americans in that study (15). None of the following SNPs, including rs6769511, rs4689388, rs1801214, and rs2237892, were investigated in other African Americans. The association with rs7903146 has been replicated in multiple African-American studies (15, 22–24). However, it was not replicated in several small studies (22, 23, 25). Very recently, Palmer et al. (26) found that rs7903146 is the causal diabetes susceptibility variant in the *TCF7L2* gene via resequencing.

We did not validate the other 22 SNPs. For 12 of these SNPs, we had limited statistical power to detect evidence of association given the sample sizes, allele frequencies, and effect sizes. Especially, statistical power was less than 50% for 6 of these 22 SNPs, including rs4457053, rs972283, rs13292136, rs7961581, rs7957197, and rs11634397 (Table 2). Linkage disequilibrium structure often differs across populations, and GWAS hits are typically markers in linkage disequilibrium with causal alleles. The lack of confirmation was consistent with other studies (15, 22, 23). Recently, 40 known type 2 diabetes loci were investigated through meta-analysis in 8 African-American studies (1,986 cases and 7,695 controls), and only 3 showed an association, including rs9668162 (*HMGA2*), rs864745 (*JAZF1*), and rs7903146 (*TCF7L2*) (24). In the Diabetes Prevention Program study (23), 7 variants were investigated, and none of them were replicated. Lewis et al. (22) investigated 12 type 2 diabetes loci in more than 2,000 African Americans, and none of the other 11 loci showed an association, except for rs7903146. Recently, 19 GWAS variants were investigated in over 2,500 African Americans in the Multiethnic Cohort Study of Diet and Cancer, and only 4 of them, including rs7903146, were validated (15).

Even if different populations share common causal alleles, the marker SNPs which show strong association in Caucasians/Asians may show little or no association in African Americans. Additionally, the constellation of causal alleles may be unique for each geographic subpopulation of human subjects, where functional gene or regulatory regions are perturbed by independent sets of rare mutations which occurred after geographic or cultural barriers led to increased genetic distance (16). As a result, the same gene may be associated across populations but by different haplotype tag SNPs, due to either differences in linkage disequilibrium or underlying causal mutations. We investigated all SNPs flanking the index SNP using the BioVU GWAS data. There were 19 loci in which an association was observed for SNPs different from those reported previously (Web Table 3). However, this analysis was based on small sample sizes, and these associations need be further validated. On the other hand, absence of association in African Americans may suggest that different susceptibility genes exist between African Americans and subjects of European or Asian descent. The present study is one of the largest studies to investigate genetic susceptibility markers in African Americans, enhancing our ability to replicate 7 of the associations. In addition, the prevalence of both type 2 diabetes and obesity is particularly high among SCCS participants, and the genetic architecture of this population may differ from that in other studies.

This study had several limitations. First, we used a self-report of type 2 diabetes and use of diabetes medication to define cases and controls in the SCCS. However, in a review of medical records and assessment of hemoglobin A_{1c} levels for samples of the self-reported cases, confirmation of the diabetes diagnosis was achieved for over 96% (27). There are likely to have been some undiagnosed

diabetes cases among controls, although this would have led to a loss of power and increased type II error (false-negative findings). In addition, although we excluded all subjects under age 30 years, we cannot exclude the possibility that a few cases might have had type 1 diabetes. However, additional analyses excluding cases diagnosed before age 35 or 40 years showed similar results. We observed a strong association at the GWAS significance level for the *TCF7L2* gene, one of the strongest type 2 diabetes susceptibility genes identified to date, suggesting that the phenotypic characterization of our study was indeed consistent with type 2 diabetes. Another limitation is that we did not have genome-wide data to account for the potentially confounding effects of population stratification in the SCCS cohort. In the SCCS cohort, 20 SNPs that did not show an association with type 2 diabetes ($P > 0.2$) were used for principal components analysis. For the BioVU cohort, principal components analysis was conducted among a linkage disequilibrium-pruned set of 100,000 SNPs. The first 5 principal components were used to adjust for population structure, and we did not observe evidence of confounding by ancestry at the assayed SNPs. Lastly, although the present study is, to our knowledge, among the largest to have been conducted in African Americans, the statistical power to detect an association was limited for some loci, especially when we investigated all variants close to the index SNP based on the GWAS data from the BioVU cohort.

Results from this study demonstrate the challenges of directly replicating GWAS findings across ancestral groups for particular SNPs. Given the disproportionately higher rate of morbidity and mortality from type 2 diabetes and associated complications in African Americans, it is critical to understand the differential susceptibility to type 2 diabetes. Large-scale studies, such as GWAS, fine mapping, exome sequencing, or even whole-genome sequencing, are needed to identify genetic risk variants for type 2 diabetes in the understudied African-American population. Studies such as this will provide important motivation for carrying out those projects.

ACKNOWLEDGMENTS

Author affiliations: Division of Epidemiology, Department of Medicine, School of Medicine, Vanderbilt University, Nashville, Tennessee (Jirong Long, Todd Edwards, Lisa B. Signorello, Qiuyin Cai, Wei Zheng, Xiao-Ou Shu, William J. Blot); Vanderbilt Epidemiology Center, Institute of Medicine and Public Health, Vanderbilt University Medical Center, Nashville, Tennessee (Jirong Long, Todd Edwards, Lisa B. Signorello, Qiuyin Cai, Wei Zheng, Xiao-Ou Shu, William J. Blot); and International Epidemiology Institute, Rockville, Maryland (Lisa B. Signorello, William J. Blot).

Compilation of the Southern Community Cohort Study data set used for the present analyses was supported by National Institutes of Health grant R01CA92447. The BioVU data set was obtained from Vanderbilt University Medical Center, supported by institutional funding and by Vanderbilt Clinical and Translational Science Award 1UL1RR024975-01 from the National Center for Research Resources, National Institutes of Health. Genotyping assays using the Sequenom platform were conducted at Proactive Genomics, LLC (Clemmons, North Carolina). Sample preparation was conducted by the Survey Research Shared Resource Core and the Epidemiology Biospecimen Core, which are supported in part by the Vanderbilt-Ingram Cancer Center (grant P30 CA68485).

The authors thank the research staff for their contributions and commitment to this project, Regina Courtney for DNA preparation, and Mary Jo Daly for manuscript preparation.

The content of this article is solely the responsibility of the authors and does not necessarily represent the official views of the funding agents.

Conflict of interest: none declared.

REFERENCES

1. National Diabetes Information Clearinghouse, National Institute of Diabetes and Digestive and Kidney Diseases. *National Diabetes Statistics, 2011*. Bethesda, MD: National Institute of Diabetes and Digestive and Kidney Diseases; 2011. (http://diabetes.niddk.nih.gov/dm/pubs/statistics). (Accessed June 22, 2011).
2. Zeggini E, Scott LJ, Saxena R, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet*. 2008;40(5):638–645.
3. Winckler W, Weedon MN, Graham RR, et al. Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. *Diabetes*. 2007;56(3):685–693.
4. Sandhu MS, Weedon MN, Fawcett KA, et al. Common variants in WFS1 confer risk of type 2 diabetes. *Nat Genet*. 2007;39(8):951–953.
5. Grant SF, Thorleifsson G, Reynisdottir I, et al. Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet*. 2006;38(3):320–323.
6. Gloyn 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(2):568–572.
7. Altshuler D, Hirschhorn JN, Klannemark M, et al. The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet*. 2000;26(1):76–80.
8. Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*. 2007;445(7130):881–885.
9. Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nat Genet*. 2007;39(6):770–775.
10. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*. 2007;447(7145):661–678.

11. Yasuda K, Miyake K, Horikawa Y, et al. Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet*. 2008;40(9):1092–1097.
12. Scott LJ, Mohlke KL, Bonnycastle LL, et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science*. 2007;316(5829):1341–1345.
13. Saxena R, Voight BF, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*. 2007;316(5829):1331–1336.
14. Zeggini E, Weedon MN, Lindgren CM, et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science*. 2007;316(5829):1336–1341.
15. Waters KM, Stram DO, Hassanein MT, et al. Consistent association of type 2 diabetes risk variants found in Europeans in diverse racial and ethnic groups. *PLoS Genet*. 2010;6(8):pii:1001078. (doi:10.1371/journal.pgen.1001078).
16. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. *Science*. 2009;324(5930):1035–1044.
17. Signorello LB, Hargreaves MK, Steinwandel MD, et al. Southern Community Cohort Study: establishing a cohort to investigate health disparities. *J Natl Med Assoc*. 2005; 97(7):972–979.
18. Roden DM, Pulley JM, Basford MA, et al. Development of a large-scale de-identified DNA biobank to enable personalized medicine. *Clin Pharmacol Ther*. 2008;84(3):362–369.
19. Ritchie MD, Denny JC, Crawford DC, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet*. 2010;86(4):560–572.
20. National Human Genome Research Institute. *A Catalog of Published Genome-Wide Association Studies*. Bethesda, MD: National Human Genome Research Institute; 2011. (http://www.genome.gov/gwastudies). (Accessed October 11, 2010).
21. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet*. 2006;38(8): 904–909.
22. Lewis JP, Palmer ND, Hicks PJ, et al. Association analysis in African Americans of European-derived type 2 diabetes single nucleotide polymorphisms from whole-genome association studies. *Diabetes*. 2008;57(8): 2220–2225.
23. Moore AF, Jablonski KA, McAteer JB, et al. Extension of type 2 diabetes genome-wide association scan results in the Diabetes Prevention Program. *Diabetes*. 2008;57(9): 2503–2510.
24. Saxena R, Elbers CC, Guo Y, et al. Large-scale gene-centric meta-analysis across 39 studies identifies type 2 diabetes loci. *Am J Hum Genet*. 2012;90(3):410–425
25. Humphries SE, Gable D, Cooper JA, et al. Common variants in the TCF7L2 gene and predisposition to type 2 diabetes in UK European Whites, Indian Asians and Afro-Caribbean men and women. *J Mol Med (Berl)*. 2006; 84(12):1005–1014.
26. Palmer ND, Hester JM, An SS, et al. Resequencing and analysis of variation in the TCF7L2 gene in African Americans suggests that SNP rs7903146 is the causal diabetes susceptibility variant. *Diabetes*. 2011;60(2): 662–668.
27. Huizinga M, Elasy T, Villegas R, et al. Validation of diabetes self-report and characteristics of undiagnosed diabetes in the Southern Community Cohort Study [abstract]. *Diabetes*. 2009;58(suppl 1):A279.