

Transferability and Fine Mapping of Type 2 Diabetes Loci in African Americans

The Candidate Gene Association Resource Plus Study

Maggie C.Y. Ng,^{1,2} Richa Saxena,^{3,4} Jiang Li,² Nicholette D. Palmer,^{1,2,5} Latchezar Dimitrov,¹ Jianzhao Xu,² Laura J. Rasmussen-Torvik,⁶ Joseph M. Zmuda,⁷ David S. Siscovick,⁸ Sanjay R. Patel,⁹ Errol D. Crook,¹⁰ Mario Sims,¹¹ Yii-Der I. Chen,^{12,13} Alain G. Bertoni,¹⁴ Mingyao Li,¹⁵ Struan F.A. Grant,¹⁶ José Dupuis,¹⁷ James B. Meigs,^{18,19} Bruce M. Psaty,²⁰ James S. Pankow,²¹ Carl D. Langefeld,²² Barry I. Freedman,²³ Jerome I. Rotter,¹³ James G. Wilson,²⁴ and Donald W. Bowden^{1,2,5,23}

Type 2 diabetes (T2D) disproportionately affects African Americans (Afa) but, to date, genetic variants identified from genome-wide association studies (GWAS) are primarily from European and Asian populations. We examined the single nucleotide polymorphism (SNP) and locus transferability of 40 reported T2D loci in six Afa GWAS consisting of 2,806 T2D case subjects with or without end-stage renal disease and 4,265 control subjects from the Candidate Gene Association Resource Plus Study. Our results revealed that seven index SNPs at the *TCF7L2*, *KLF14*, *KCNQ1*, *ADCY5*, *CDKAL1*, *JAZF1*, and *GCKR* loci were significantly associated with T2D ($P < 0.05$). The strongest association was observed at *TCF7L2* rs7903146 (odds ratio [OR] 1.30;

$P = 6.86 \times 10^{-8}$). Locus-wide analysis demonstrated significant associations ($P_{emp} < 0.05$) at regional best SNPs in the *TCF7L2*, *KLF14*, and *HMGA2* loci as well as suggestive signals in *KCNQ1* after correction for the effective number of SNPs at each locus. Of these loci, the regional best SNPs were in differential linkage disequilibrium (LD) with the index and adjacent SNPs. Our findings suggest that some loci discovered in prior reports affect T2D susceptibility in Afa with similar effect sizes. The reduced and differential LD pattern in Afa compared with European and Asian populations may facilitate fine mapping of causal variants at loci shared across populations. *Diabetes* 62:965–976, 2013

From the ¹Center for Genomics and Personalized Medicine Research, the ²Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina; the ³Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts; the ⁴Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, Massachusetts; the ⁵Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina; the ⁶Department of Preventive Medicine, Northwestern University, Chicago, Illinois; the ⁷Department of Epidemiology, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, Pennsylvania; the ⁸Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington; the ⁹Division of Sleep Medicine, Brigham and Women's Hospital, Boston, Massachusetts; the ¹⁰Center for Healthy Communities, University of South Alabama College of Medicine, Mobile, Alabama; the ¹¹School of Medicine, University of Mississippi Medical Center, Jackson, Mississippi; the ¹²Department of Obstetrics and Gynecology, Cedars-Sinai Medical Center, Los Angeles, California; the ¹³Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California; the ¹⁴Department of Epidemiology and Prevention, Wake Forest School of Medicine, Winston-Salem, North Carolina; the ¹⁵Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania; the ¹⁶Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; the ¹⁷Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts; the ¹⁸General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts; the ¹⁹Department of Medicine, Harvard Medical School, Boston, Massachusetts; the ²⁰Cardiovascular Health Research Unit, Departments of Medicine, Epidemiology, and Health Services, University of Washington, Seattle, Washington; the ²¹Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, Minnesota; the ²²Department of Biostatistical Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina; the ²³Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina; and the ²⁴Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi.

Corresponding author: Maggie C.Y. Ng, mng@wakehealth.edu.

Received 2 March 2012 and accepted 21 August 2012.

DOI: 10.2337/db12-0266

This article contains Supplementary Data online at <http://diabetes.diabetesjournals.org/lookup/suppl/doi:10.2337/db12-0266/-/DC1>.

© 2013 by the American Diabetes Association. Readers may use this article as long as the work is properly cited, the use is educational and not for profit, and the work is not altered. See <http://creativecommons.org/licenses/by-nc-nd/3.0/> for details.

Type 2 diabetes (T2D) is a major public health problem affecting 25.8 million people in the U.S. (1). Marked racial differences in its prevalence have been observed, with African American (Afa) adults >40 years of age having nearly twofold higher prevalence than European Americans (27.1 and 15.5%, respectively) (2). In addition to socioeconomic and behavioral risk factors, genetic factors are likely contributors to T2D risk in Afa (3).

Genome-wide association studies (GWAS) for T2D and related traits have successfully identified >50 loci with common genetic variants associated with T2D risk in primarily European-descent populations (4–14) and more recently in East and South Asians (15–21). The reported index single nucleotide polymorphisms (SNPs) at these loci have been replicated in multiple populations (22–24) but less successfully in Afa (25–27). Although differences in environment and lack of study power may partly account for the lack of transferability across ethnicities, differences in linkage disequilibrium (LD) patterns, effect sizes, and risk allele frequency also likely impact the replication of index SNPs. Although the long-range LD in European populations allows for the identification of T2D loci using less dense markers, causal variants are not distinguishable from other nearby SNPs in high LD. This issue prompts the need to examine T2D loci in other populations with different allelic and LD architecture, which may help fine mapping of the underlying functional variants (28).

We performed a comprehensive evaluation of the LD region of T2D loci reported in European and Asian GWAS in a meta-analysis of six Afa GWAS. By testing the index and nearby SNPs, we evaluated the transferability of the previously reported loci for T2D association in Afa. We

demonstrated that the reduced and differential LD structure in AfA facilitated fine mapping of regions potentially harboring causal variants at some T2D loci.

RESEARCH DESIGN AND METHODS

Subjects. The study samples include AfA from the National Heart, Lung, and Blood Institute's (NHLBI's) Candidate Gene Association Resource (CARE) and the Wake Forest School of Medicine (WFSM) study. CARE is an NHLBI shared resource comprised of five cohorts with multiple phenotypes for GWAS in AfA. The study design of CARE has been described in detail elsewhere (26). The five CARE cohorts are as follows: Atherosclerosis Risk in Communities (ARIC), Coronary Artery Risk Development in Young Adults (CARDIA), Cleveland Family Study (CFS), Jackson Heart Study (JHS), and Multi-Ethnic Study of Atherosclerosis (MESA). Details of the study cohorts are described in the Supplementary Data. Written informed consent was obtained from all study participants. Recruitment and sample collection procedures were approved by the institutional review board from the respective institutions. The clinical characteristics of all cohorts are summarized in Table 1.

Clinical definitions. T2D was diagnosed according to the American Diabetes Association criteria (29) with at least one of the following: fasting glucose ≥ 126 mg/dL, 2-h oral glucose tolerance test glucose ≥ 200 mg/dL, random glucose ≥ 200 mg/dL, use of oral hypoglycemic agents and/or insulin, or physician-diagnosed diabetes. Subjects diagnosed before 25 years of age were excluded. Normal glucose tolerance was defined as fasting glucose < 100 mg/dL and 2-h oral glucose tolerance test glucose < 140 mg/dL (if available) without reported use of diabetes medications. Control subjects < 25 years of age were excluded.

Genotyping, imputation, and quality control. Genotyping was performed using the Affymetrix Genome-Wide Human SNP Array 6.0 in all samples. For the CARE study, genotyping, quality control, and data analyses were performed centrally by the CARE analytical group at the Broad Institute, and details are described elsewhere (26). For the WFSM study, genotyping was performed at the Center for Inherited Disease Research (CIDR), and analyses were performed at WFSM and described elsewhere (30,31). For all studies, imputation was performed using MACH with the function `-mle` (version 1.0.16, <http://www.sph.umich.edu/csg/abecasis/MaCH/>) to obtain missing genotypes and replace genotypes inconsistent with reference haplotypes. In general, SNPs with call rate $\geq 95\%$ and minor allele frequency (MAF) $\geq 1\%$ that passed study-specific quality control were used for imputation (26,32). A 1:1 HapMap II (NCBI Build 36) CEU:YRI (European:African) consensus haplotype was used as reference. Imputation was performed in two steps. The first step selected a random subset of unrelated samples to calculate recombination and error rate estimates. The second step used these rates to impute all samples across the SNPs in the entire reference panel. A total of 2,333,531–2,907,112 SNPs from each study with call rate $\geq 95\%$, MAF $\geq 1\%$, minor allele count (MAC) ≥ 10 , and Hardy-Weinberg P value ≥ 0.0001 for genotyped SNPs and MAF $\geq 1\%$, MAC ≥ 10 , and RSQ ≥ 0.5 for imputed SNPs were included in subsequent data analyses.

Principal component analysis. Principal component (PC) was computed on each study by using high-quality SNPs (26,32,33). The first PC was highly correlated ($r^2 > 0.87$) with global African-European ancestry, as measured by ANCESTRYMAP (34), STRUCTURE (35), or FRAPPE (36). The AfA samples had an average of 80% African ancestry. By analyzing unrelated samples from all studies using SMARTPCA (33), only the first PC appeared to account for substantial genetic variation in the screen plot (data not shown), whereas the subsequent PCs may reflect sampling noise and/or relatedness in samples (34). The first PC was used as a covariate in the association analyses to adjust for population substructure.

Statistical analyses

Single SNP association and meta-analysis. In each study, the association of genotyped and imputed (in dosage) SNPs with T2D was assessed under an additive model with adjustment for age, sex, study center (if applicable), and the first PC. Age at the last visit with other clinical parameters available for prospective studies (ARIC, CARDIA, CFS, and MESA), or at baseline for JHS and WFSM studies, was analyzed. Association tests were performed using logistic regression in PLINK (<http://pngu.mgh.harvard.edu/purcell/plink>) in unrelated samples and generalized estimating equation in GWAF in R (v2.9.0) (37) in related (CFS) samples. Association results with extreme values (absolute β coefficient or SE > 10), primarily due to low cell counts from small sample size and low MAF, were excluded. After genomic control correction within each study, association results were combined by fixed-effect inverse variance weighting implemented in METAL (<http://www.sph.umich.edu/csg/abecasis/metal/>). Results from SNPs with $< 50\%$ samples analyzed and those with allele frequency difference > 0.3 among studies were excluded. A total of 2,739,003 SNPs were analyzed in the meta-analysis. The mean SNP call rate was 99% in the locus-specific meta-analysis.

TABLE 1
Characteristics of study subjects

	ARIC		CARDIA		CFS		JHS		MESA		WFSM	
	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control	Case	Control
<i>n</i>	955	414	94	654	81	98	333	1450	411	793	932	856
Male (%)	35.8	31.2	19.2	38.2	50.6	32.7	33.9	38.7	47.2	42.4	38.8	43.7
Age (years)*	61.3 \pm 6.0	59.6 \pm 6.3	40.5 \pm 3.8	38.2 \pm 4.4	57.8 \pm 12.6	43.2 \pm 11.8	55.5 \pm 10.7	48.6 \pm 11.3	67.6 \pm 9.2	65.3 \pm 10.5	61.6 \pm 10.5	49.0 \pm 11.9
Age at diagnosis of T2D (years)	50.9 \pm 9.2	—	35.0 \pm 5.5	—	50.6 \pm 11.1	—	46.2 \pm 11.0	—	54.6 \pm 10.9	—	41.6 \pm 12.3	—
BMI (kg/m ²)*	32.0 \pm 6.7	27.6 \pm 5.7	33.8 \pm 8.1	29.6 \pm 6.8	38.9 \pm 9.4	31.8 \pm 7.7	35.2 \pm 7.5	31.4 \pm 7.5	31.7 \pm 6.2	28.6 \pm 5.8	29.7 \pm 7.1	30.0 \pm 7.1
African ancestry proportion	0.83 \pm 0.10	0.83 \pm 0.09	0.83 \pm 0.08	0.80 \pm 0.11	0.79 \pm 0.11	0.79 \pm 0.12	0.83 \pm 0.08	0.82 \pm 0.09	0.79 \pm 0.14	0.78 \pm 0.14	0.80 \pm 0.11	0.78 \pm 0.11

Data are shown as mean \pm SD or percentage. *Age and BMI are shown for the last available visit for the prospective studies including ARIC, CARDIA, CFS, and MESA; and the baseline visit for JHS and WFSM.

TABLE 2
Association of reported T2D index SNPs in AfA

Reported loci*	Index SNP	Chr	Position	RA/NRA†	RAF‡	OR (95% CI)	P§	P _{het}
<i>NOTCH2-ADAM30</i>	rs10923931	1	120319482	T/G	0.33	1.05 (0.96–1.14)	0.256	0.927
<i>PROX1</i>	rs340874	1	212225879	C/T	0.17	0.93 (0.83–1.04)	0.216	0.822
<i>GCKR</i>	rs780094	2	27594741	C/T	0.82	1.12 (1–1.24)	0.043	0.803
<i>THADA</i>	rs7578597	2	43586327	T/C	0.73	1.04 (0.95–1.14)	0.417	0.235
<i>BCL11A</i>	rs243021	2	60438323	A/G	0.39	0.93 (0.86–1.01)	0.091	0.780
<i>RBMS1-ITGB6</i>	rs7593730	2	160879700	C/T	0.62	1.01 (0.92–1.1)	0.875	0.173
<i>IRS1</i>	rs7578326	2	226728897	A/G	0.57	0.96 (0.88–1.05)	0.380	0.174
<i>PPARG</i>	rs1801282	3	12368125	C/G	0.98	0.82 (0.61–1.09)	0.175	0.491
<i>ADAMTS9</i>	rs4607103	3	64686944	C/T	0.71	0.94 (0.87–1.01)	0.103	0.054
<i>ADCY5</i>	rs11708067	3	124548468	A/G	0.85	1.18 (1.05–1.33)	4.74E-03	0.392
<i>IGF2BP2</i>	rs4402960	3	186994381	T/G	0.52	1.02 (0.94–1.10)	0.711	0.896
<i>WFS1-PPP2R2C</i>	rs4689388	4	6320957	A/G	0.71	0.96 (0.86–1.07)	0.485	0.103
<i>ZBED3</i>	rs4457053	5	76460705	G/A	—	—	—	—
<i>CDKAL1</i>	rs10440833	6	20796100	A/T	0.22	1.13 (1.02–1.24)	0.018	0.372
<i>DGKB-TMEM195</i>	rs2191349	7	15030834	T/G	0.60	1.03 (0.95–1.13)	0.444	0.134
<i>JAZF1</i>	rs864745	7	28147081	T/C	0.74	1.10 (1–1.21)	0.043	0.921
<i>GCK</i>	rs4607517	7	44202193	A/G	0.10	1.03 (0.88–1.20)	0.699	0.747
<i>KLF14</i>	rs972283	7	130117394	G/A	0.85	1.24 (1.09–1.41)	8.12E-04	0.225
<i>TP53INP1</i>	rs896854	8	96029687	T/C	0.69	1.09 (0.99–1.19)	0.067	0.258
<i>SLC30A8</i>	rs3802177	8	118254206	G/A	0.91	1.13 (0.97–1.31)	0.125	0.676
<i>PTPRD</i>	rs17584499	9	8869118	T/C	0.06	0.93 (0.70–1.23)	0.609	0.880
<i>CDKN2A-CDKN2B</i>	rs10811661	9	22124094	T/C	0.93	1.07 (0.91–1.27)	0.407	0.977
<i>CHCHD9</i>	rs13292136	9	81141948	C/T	0.91	1.08 (0.94–1.24)	0.284	0.797
<i>CDC123-CAMK1D</i>	rs10906115	10	12355003	A/G	0.69	1.03 (0.94–1.12)	0.534	0.472
<i>CDC123-CAMK1D</i>	rs12779790	10	12368016	G/A	0.13	1.07 (0.94–1.22)	0.291	0.572
<i>HHEX-IDE</i>	rs5015480	10	94455539	C/T	0.62	1.02 (0.94–1.11)	0.596	0.285
<i>TCF7L2</i>	rs7903146	10	114748339	T/C	0.31	1.30 (1.18–1.43)	6.86E-08	0.162
<i>KCNQ1</i>	rs231362	11	2648047	G/A	0.79	1.07 (0.95–1.20)	0.251	0.784
<i>KCNQ1</i>	rs2237892	11	2796327	C/T	0.89	1.25 (1.09–1.43)	1.80E-03	0.286
<i>KCNJ11</i>	rs5215	11	17365206	C/T	0.09	0.96 (0.83–1.11)	0.573	0.083
<i>CENTD2</i>	rs1552224	11	72110746	A/C	0.97	0.79 (0.61–1.03)	0.077	0.351
<i>MTNR1B</i>	rs1387153	11	92313476	T/C	0.38	0.99 (0.91–1.08)	0.827	0.048
<i>HMGA2</i>	rs1531343	12	64461161	C/G	0.38	0.91 (0.84–0.99)	0.022	0.836
<i>TSPAN8-LGR5</i>	rs7961581	12	69949369	C/T	0.20	0.98 (0.89–1.08)	0.667	0.948
<i>HNF1A</i>	rs7957197	12	119945069	T/A	0.86	1 (0.88–1.13)	0.986	0.854
<i>SPRY2</i>	rs1359790	13	79615157	G/A	0.89	1.13 (1–1.29)	0.057	0.783
<i>C2CD4A-C2CD4B</i>	rs7172432	15	60183681	A/G	0.31	0.99 (0.90–1.08)	0.767	0.749
<i>C2CD4B</i>	rs11071657	15	60221254	A/G	0.86	0.97 (0.84–1.11)	0.633	0.215
<i>ZFAND6</i>	rs11634397	15	78219277	G/A	0.44	0.9 (0.83–0.98)	0.011	0.051
<i>PRC1</i>	rs8042680	15	89322341	A/C	0.84	0.99 (0.88–1.11)	0.823	0.892
<i>FTO</i>	rs8050136	16	52373776	A/C	0.43	0.99 (0.92–1.08)	0.852	0.206
<i>SRR</i>	rs391300	17	2163008	C/T	0.49	0.96 (0.87–1.05)	0.336	0.718
<i>HNF1B</i>	rs4430796	17	33172153	G/A	—	—	—	—

Chr, chromosome. **ZBED3* rs4457053 and *HNF1B* rs4430796 failed quality control and results were not reported. †Risk allele (RA) and nonrisk allele (NRA) as reported previously in European or East Asian populations. Alleles were indexed to the forward strand of NCBI Build 36. ‡Risk allele frequency (RAF) in AfA from this study. §Significant associations ($P < 0.05$) are set in boldface type. ||Heterogeneity P values across studies for index SNPs.

Locus-specific analysis. We used two approaches to test for the transferability of 40 T2D loci in AfA. First, the most significant independent index SNPs ($P < 5 \times 10^{-8}$) from T2D loci identified through GWAS of T2D and related glucose homeostasis traits were selected from the catalog of published GWAS at the National Human Genome Research Institute until December 2010 (<http://www.genome.gov/gwastudies>) (38) (Supplementary Table 1). At the *CDC123-CAMK1D* and *KCNQ1* loci, independent T2D index SNPs had been identified in European and East Asian populations, respectively (10,12,16,19). At the *C2CD4A-C2CD4B* locus, rs11071657 was associated with T2D in East Asians (17). A nearby independent index SNP rs11071657 showing strong association with fasting glucose but modest association with T2D in Europeans (13) was also examined. Additionally, two index SNPs from *PPARG* and *HNF1B* that did not reach genome-wide significance were selected due to candidacy and consistent replication. Second, the region of interest for each locus was defined as the boundary of the farthest SNPs that show LD at $r^2 \geq 0.3$ with the index SNP in CEU or JPT+CHB (Asian [ASN]) populations and further extended by

100 kb. These regions will likely harbor causal variants that are in LD with the index SNPs reported by the original GWAS. This approach takes into account varying LD block size across the genome and absence of correlated SNPs for some index SNPs in HapMap. Regional pairwise LD was calculated in SNAP (<http://www.broadinstitute.org/mpg/snap/>) using the HapMap II release 22 CEU and ASN data for loci reported in Europeans and Asians, respectively. The regions of interest range from 200 to 807 kb (Supplementary Table 1), and the effective number of SNPs range from 45 to 156.

In the first approach, SNPs were examined for transferability by directly testing the reported index SNP for T2D association. SNP-specific significance was considered as $P < 0.05$ in the same direction of association in prior reports. In the second approach, locus transferability was assessed by testing all SNPs in the region of interest. In each locus, the most significant SNP was defined as the best SNP. The effective number of SNPs (independent SNPs) was estimated from the eigenvalues of the covariance matrix of the SNPs in each locus using the L_i and J_i method implemented in SOLAR (39). Empirical

TABLE 3
Transferability of reported T2D loci in AFA

Reported loci	Index SNP	Best regional SNP	Chr	Position	LD (r^2) with index SNP				OR (95% CI)	P	Eff	P_{het}	
					AFA	CEU (ASN)¶	YRI	RA/NRA*					RAFT†
NOTCH2-													
ADAM30	rs10923931	rs12075171	1	120161765	0.02	—	0.06	0.92	1.34 (1.15–1.55)	1.25E–04	85	0.011	0.911
PROX1	rs340874	rs4507982	1	212104884	0.15	0.12	0.01	0.77	1.13 (1.03–1.25)	0.013	89	1.000	0.644
GCKR	rs780094	rs1728918	2	27488967	0.26	0.55	0.01	0.93	1.32 (1.11–1.56)	1.43E–03	80	0.115	0.782
THADA	rs7578597	rs6544639	2	43363601	0.26	0.32	0.30	0.66	1.16 (1.06–1.26)	8.37E–04	156	0.130	0.814
BCL11A	rs243021	rs13034231	2	60531472	0.01	0.15	0.01	0.31	1.13 (1.03–1.23)	7.18E–03	101	0.726	0.206
RBMS1-													
ITGB6	rs7593730	rs1870102	2	160778828	0.04	0.05	0.01	0.10	1.25 (1.09–1.44)	2.01E–03	77	0.155	0.470
IRSI	rs7578326	rs6436586	2	226576239	0	0.09	0	0.20	1.17 (1.04–1.31)	6.32E–03	118	0.745	0.910
PPARG	rs1801282	rs9837686	3	12301825	0	—	—	0.91	1.22 (1.06–1.42)	7.12E–03	84	0.598	0.891
ADAMTS9	rs4607103	rs9875220	3	64626429	0	0.05	0	0.52	1.13 (1.04–1.22)	4.38E–03	102	0.447	0.530
ADCY5	rs11708067	rs11708067	3	124548468	Same	Same	Same	0.85	1.18 (1.05–1.33)	4.74E–03	97	0.460	0.392
IGFBP2	rs4402960	rs6780808	3	186924813	0	0.11	0.02	0.48	1.1 (1.02–1.19)	0.020	58	1.000	0.398
WFS1-													
PPP2R2C	rs4689388	rs4380588	4	6310789	0.01	0.03	0.02	0.47	1.16 (1.05–1.28)	4.09E–03	128	0.523	0.071
ZBED3	rs4457053	rs7708285	5	76461623	—	0.88	—	0.15	1.32 (1.09–1.6)	4.22E–03	74	0.312	0.600
CDKALI	rs10440833	rs2328549	6	20826219	0.07	0.30	0.06	0.23	1.14 (1.04–1.26)	6.41E–03	113	0.724	0.018
DGKB-													
TMEM195	rs2191349	rs7810554	7	15107715	0	0	—	0.79	1.2 (1.06–1.37)	4.97E–03	126	0.626	0.682
JAZF1	rs864745	rs6963763	7	28092602	0.17	0.04	0.23	0.59	1.17 (1.06–1.29)	1.09E–03	97	0.105	0.811
GCK	rs4607517	rs2971676	7	44161007	0.01	0.01	0.02	0.22	1.17 (1.05–1.3)	4.90E–03	45	0.220	0.605
KLF14	rs972283	rs13234269	7	130079726	0.36	1.00	0.01	0.80	1.26 (1.14–1.4)	1.55E–05	66	1.02E–03	0.469
TP53INP1	rs896854	rs3802191	8	96056760	0.04	0.23	—	0.09	1.24 (1.08–1.44)	3.14E–03	118	0.370	0.872
SLC30A8	rs3802177	rs2649102	8	118310401	0.02	0.02	0.04	0.40	1.11 (1–1.23)	0.045	92	1.000	0.067
PTRPD	rs17584499	rs16928642	9	8930693	0	(0)	0	0.03	1.39 (1.07–1.82)	0.015	124	1.000	0.829
CDKN2A-													
CDKN2B	rs10811661	rs10120722	9	22076840	0	—	—	0.73	1.12 (1.02–1.22)	0.015	72	1.000	0.556
CHCHD9	rs13292136	rs12551737	9	81098662	0.21	0.75	0.15	0.88	1.16 (1.03–1.32)	0.019	110	1.000	0.543
CDC123-													
CAMK1D	rs10906115	rs10906121	10	12390693	0	(0.01)	0.01	0.23	1.17 (1.04–1.32)	9.40E–03	73	0.686	0.346
CAMK1D	rs12779790	rs10906121	10	12390693	0	0.01	0	—	—	—	—	—	—
HHEX-IDE	rs5015480	rs7086285	10	94166068	0	0.03	0.01	0.31	1.2 (1.08–1.33)	6.44E–04	89	0.057	0.376
TCF7L2	rs7903146	rs7903146	10	114748339	Same	Same	Same	0.31	1.3 (1.18–1.43)	6.86E–08	65	4.46E–06	0.162
KCNQ1	rs231362	rs231361	11	2648076	0.13	0.40	0.09	0.36	1.17 (1.07–1.28)	6.64E–04	124	0.082	0.728
KCNQ1	rs2237892	rs2237892	11	2796327	Same	Same	Same	0.89	1.25 (1.09–1.43)	1.80E–03	124	0.223	0.811
KCNJ11	rs5215	rs11024251	11	17292987	0	0.14	0	0.94	1.24 (1.05–1.48)	0.014	102	1.000	0.811
CENTD2	rs1552224	rs7121935	11	72173796	0.01	0.15	—	0.26	1.19 (1.07–1.31)	8.54E–04	134	0.114	0.615
MTNR1B	rs1387153	rs750025	11	92284223	0.06	0.05	0.04	0.23	1.1 (0.99–1.23)	0.063	96	1.000	0.056
HMGCA2	rs1531343	rs12049974	12	64451469	0.43	0.01	0.67	0.44	1.24 (1.14–1.35)	1.73E–06	79	1.36E–04	0.821
TSPAN8-LGR5	rs7961581	rs41391144	12	69946900	0	—	—	0.99	2.32 (1.35–3.99)	2.34E–03	81	0.189	0.370

Continued on p. 969

TABLE 3
Continued

Reported loci	Index SNP	Best regional SNP	Chr	Position	LD (r^2) with index SNP				RA/NRA*	RAFT†	OR (95% CI)	P	Eff SNPs‡	P _{emp} §	P _{het}
					AFA	CEU (ASN)¶	YRI	RA/NRA*							
<i>HNFI1A</i>	rs7957197	rs7969196	12	119862039	0.01	0.06	—	T/G	0.12	1.2 (1.06–1.36)	4.99E–03	137	0.684	0.978	
<i>SPRY2</i>	rs1359790	rs3861148	13	79491295	0	(0)	0.01	T/C	0.70	1.14 (1.04–1.25)	3.82E–03	120	0.458	0.527	
<i>C2CD4A-C2CD4B</i>	rs172432	rs17238301	15	60131026	0.01	(0.03)	0	T/C	0.02	1.38 (1.01–1.89)	0.042	83	1.000	0.407	
<i>C2CD4B</i>	rs11071657	rs17238301	15	60131026	0	0.02	0	—	—	—	—	—	—	—	
<i>ZFAND6</i>	rs11634397	rs11634397	15	78219277	Same	Same	Same	A/G	0.56	1.11 (1.02–1.2)	0.011	79	0.888	0.051	
<i>PRCI</i>	rs8042680	rs1005060	15	89389551	0.02	0	0.01	T/C	0.76	1.14 (1.03–1.25)	8.35E–03	106	0.885	0.662	
<i>FTO</i>	rs8050136	rs7186637	16	52337603	0	0.01	0	C/T	0.61	1.13 (1.04–1.22)	5.10E–03	89	0.454	0.115	
<i>SRR</i>	rs391300	rs7225860	17	1948606	0.04	0.03	0.03	T/C	0.94	1.22 (1.03–1.46)	0.025	60	1.000	0.751	
<i>HNFI1B</i>	rs4430796	rs6607284	17	33100448	—	0	0.02	C/T	0.39	1.13 (1.04–1.22)	4.54E–03	101	0.459	0.808	

Chr, chromosome. *Risk allele (RA) and nonrisk allele (NRA) of best SNP in AFA from this study. Alleles were indexed to the forward strand of NCBI Build 36. †Risk allele frequency (RAF) of best SNP in AFA from this study. ‡Effective number of SNPs. §Significant associations after adjustment of effective number of SNPs ($P_{emp} < 0.05$) are in boldface type. ||Heterogeneity P values across studies for best SNPs. ¶LD of best SNP and index SNP was shown for HapMap CEU population in loci reported in Europeans, and for HapMap CHB+JPT (ASN) populations in loci reported in East Asians.

locus-specific P values, P_{emp} , were adjusted for multiple comparisons by Bonferroni correction for the effective number of SNPs.

Conditional and haplotype analyses. For loci showing significant regional associations, logistic regression was performed conditioned on both the index and best SNPs to reveal the presence of independent or residual associations. Haplotype analysis of the index and best SNPs was also performed by a haplotype-specific test using *-chap* in PLINK to compare differences of frequencies of each haplotype with a reference haplotype between cases and controls. Best-guess genotypes were analyzed for imputed SNPs. All analyses were performed separately in each study, adjusted for age, sex, study center, and the first PC. The conditional analyses also adjusted for inflation factor from GWAS in each study. Effect sizes were then combined by meta-analysis.

Population differentiation and natural selection. Four methods were applied to evaluate whether the differences in genetic architecture between the ancestries of AfA or between the discovery populations and AfA account for the differential association signals for AfA in this study. For the index SNPs, the absolute difference of risk allele frequency was assessed between AfA and CEU (or ASN) populations for loci identified in Europeans and East Asians, respectively. We also assessed two matrices using Haplotter, F_{ST} for measurement of population differentiation, and integrated haplotype score for the detection of recent positive selection in the CEU (or ASN) and YRI populations (40). To assess for interpopulation differences in LD patterns, the varLD method was used to assess genome-wide distribution of varLD scores between CEU (or ASN) and YRI (41). The varLD scores were standardized, and the 100-kb regions flanking the index T2D SNPs were examined. A standardized varLD score exceeding the 95th percentile of the distribution was considered a significant LD difference between the studied populations.

Power analysis. Posterior study power was calculated using the genetic power calculator (42) under an additive model, using the SNP-specific effective sample size (43) of this study and reported effect sizes from the replication phases (wherever available) or all phases in prior T2D reports to minimize winner's curse effect.

All statistical tests were performed by PLINK, GWAFF, or SAS v.9.1 (SAS Institute, Cary, NC), unless otherwise specified. A nominal P value < 0.05 for index SNPs was considered significant. A Bonferroni P value (P_{emp}) < 0.05 corrected for the effective number of SNPs was considered significant for regional SNPs.

RESULTS

Clinical characteristics of the study samples. Clinical characteristics of the six GWAS cohorts are shown in Table 1. A total of 2,806 T2D case subjects and 4,265 control subjects (6,701 effective sample size) were included. The mean age at diagnosis of T2D in case subjects varied from 35.0 to 54.6 years among studies.

GWAS and meta-analysis. A total of 2.3–2.9 million SNPs that passed quality control were tested for association with T2D in each cohort separately. Inflation factors for the associations were 1.022 for ARIC, 1.020 for CARDIA, 1.084 for CFS, 1.079 for JHS, 1.009 for MESA, and 1.054 for WFSM cohorts before genomic control. The inflation factor for the meta-analysis result was 1.027 after genomic control in 2,739,003 SNPs. Results from T2D candidate loci were selected for subsequent analyses. No correlation was observed between association results and F_{ST} with first PC adjustment. In addition, the inflation factor and association results with adjustment for the first 10 PCs are similar (data not shown), suggesting that adjustment for the first PC is sufficient to control for population substructure.

Association analyses of index SNPs. The association of 43 independent index SNPs from 40 T2D loci identified from GWAS of European and East Asian ancestries is shown in Table 2. No significant heterogeneity of associations was observed after Bonferroni correction of multiple comparisons despite heterogeneous study designs. Among 41 good-quality SNPs, 23 showed directionally consistent association, as in previous reports (binomial test, $P = 0.27$), and seven were significantly associated with T2D (Supplementary Fig. 1). The strongest association was observed at *TCF7L2* rs7903146 (odds ratio [OR] 1.30 [95% CI 1.18–1.43]; $P = 6.86 \times 10^{-8}$), followed by

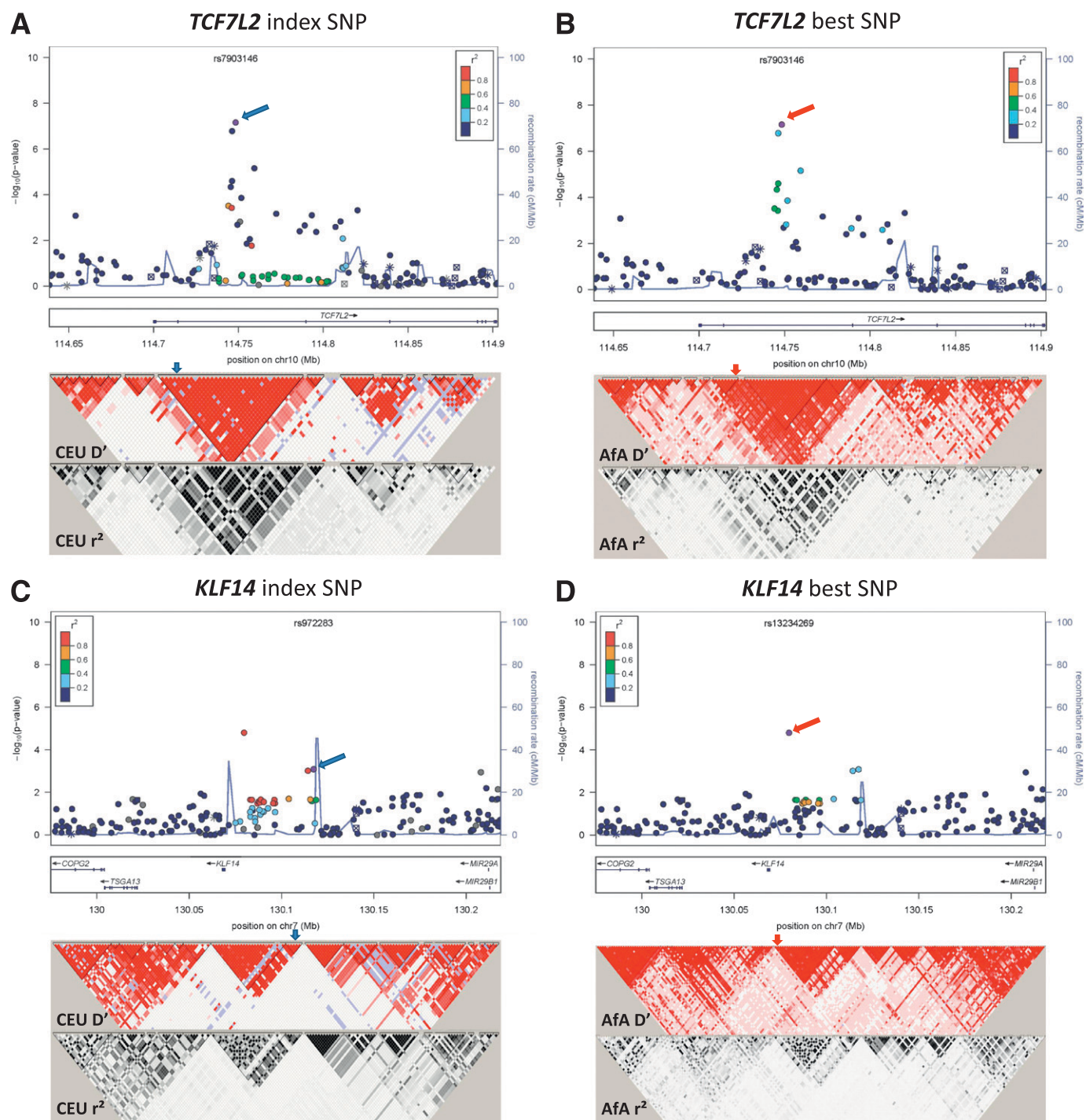


FIG. 1. Association plots and LD patterns at the regions flanking the index SNPs at *TCF7L2* (A and B), *KLF14* (C and D), *HMG2* (E and F), *NOTCH2-ADAM30* (G and H), and *KCNQ1* (I–L). At the top panel of each plot, the x-axis denotes genomic position and the y-axis denotes the $-\log(P$ value) for the association of each SNP in Afa. Each locus contains two plots. The plots on the left denote the location of the index SNPs (blue arrows) and the color of each data point represents its LD value (r^2) with the index SNPs in the HapMap II CEU or JPT+CHB (ASN) populations, for loci identified in Europeans and East Asians, respectively. The blue line represents the recombination rate in the respective HapMap populations. The LD plots (D' and r^2) in the respective HapMap populations are shown in the bottom panel. The plots on the right denote the location of the best SNPs (red arrows), and the color of each data point represents its LD value (r^2) with the best SNPs in our Afa samples. The blue line represents the recombination rate in the HapMap YRI population. The LD plot (D' and r^2) for our Afa samples is shown at the bottom panel.

KLF14, *KCNQ1*, *ADCY5*, *CDKAL1*, *JAZF1*, and *GCKR* (OR 1.10–1.25; $8.12 \times 10^{-4} < P < 0.05$). At *KCNQ1*, the association at the index SNP rs2237892 identified in East Asians (16) was significant and had stronger effect size (OR 1.25 [95% CI 1.09–1.43]; $P = 0.0018$) than the index

SNP rs231362 identified in European populations (12) ($1.07 [0.95–1.20]$; $P = 0.25$). Nominal associations were also observed at the index SNPs in *HMG2* and *ZFAND6* ($0.011 < P < 0.05$), but the reported risk alleles in European populations (12) were protective for T2D in Afa. The

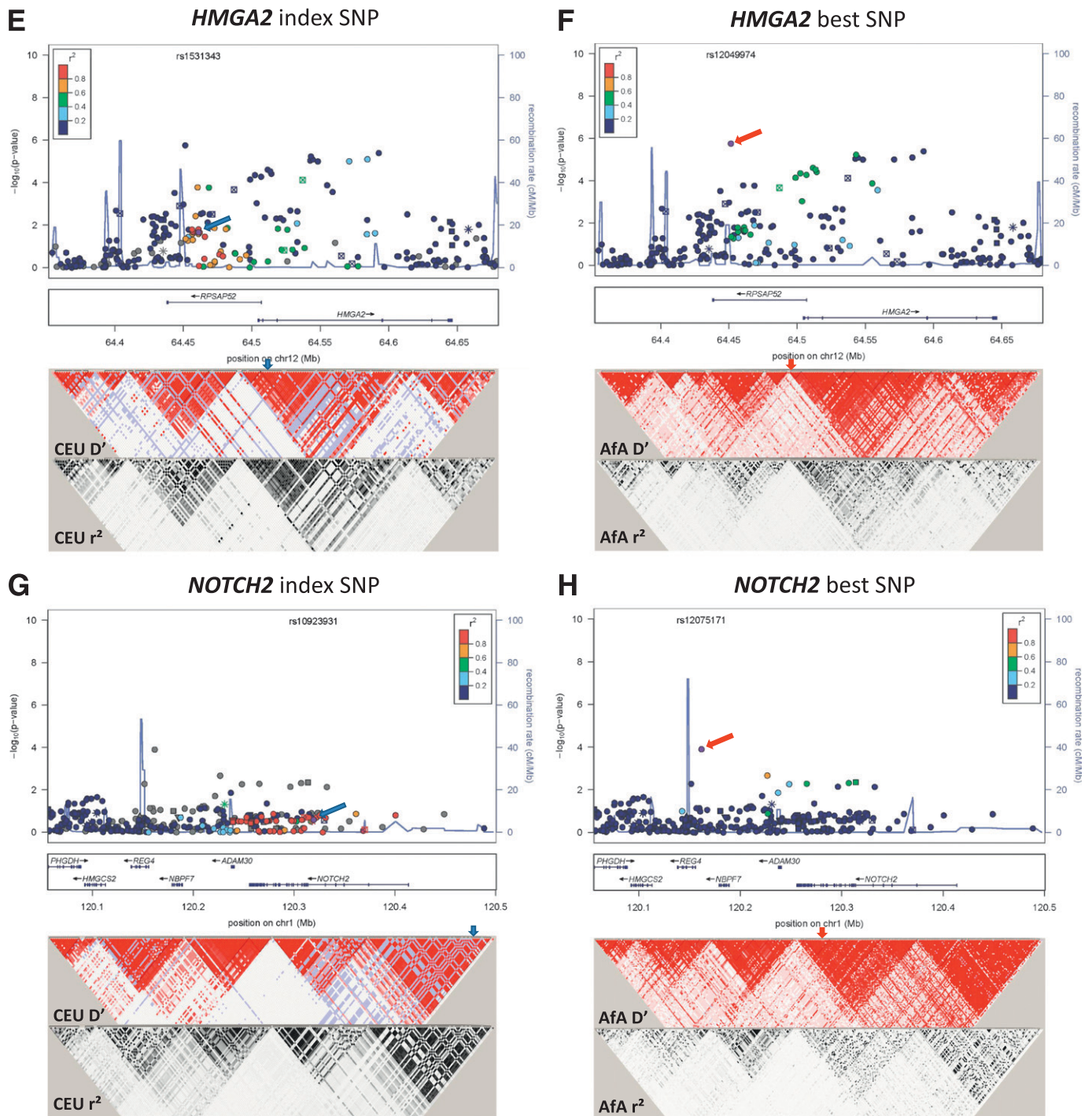


FIG. 1. Continued.

BMI-associated index SNP rs8050136 was not associated with T2D with ($P = 0.839$) or without ($P = 0.852$) BMI adjustment.

Best SNPs in regional association analyses. The regions defined by boundary SNPs in moderate LD ($r^2 \geq 0.3$) to the index SNPs were evaluated further. By defining the best SNP as the most significant SNP in each region, four of the significant index SNPs (rs7903146 at *TCF7L2*, rs11708067 at *ADCY5*, rs2237892 at *KCNQ1*, and rs11634397 at *ZFAND6*) were also the best SNPs in the respective regions (Tables 2 and 3). After correction for

multiple comparisons among the effective number of SNPs in each region, the association signal among the best SNPs at four loci, *TCF7L2*, *KLF14*, *HMG2*, and *NOTCH2-ADAM30*, remained significant ($4.46 \times 10^{-6} < P_{emp} < 0.05$) (Table 3).

***TCF7L2*.** The most significant best SNP was rs7903146 located at intron 3 of *TCF7L2* (OR 1.30; $P = 6.86 \times 10^{-8}$; $P_{emp} = 4.46 \times 10^{-6}$), which was also the index SNP reported in European and East Asian populations (12,44). Although rs7903146 was in strong LD ($r^2 > 0.8$) with several nearby SNPs in a 64-kb LD region in CEU (Fig. 1A), it

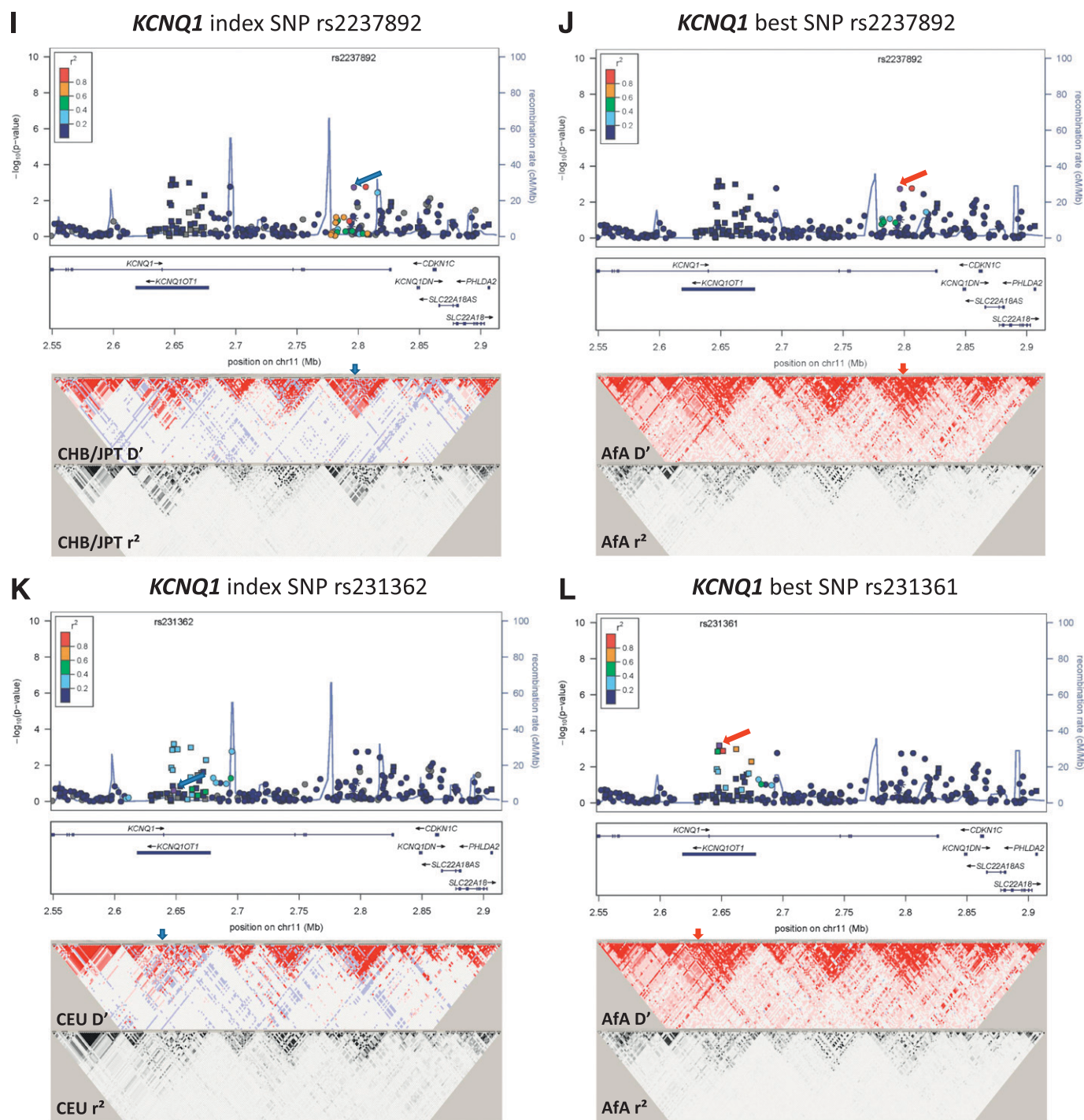


FIG. 1. Continued.

was located in a 9-kb LD block in Afa (Fig. 1B). No SNPs were in strong LD with rs7903146 in Afa ($r^2 \leq 0.6$) or YRI ($r^2 < 0.4$). The second strongest SNP was rs7069007 (OR 1.40 [95% CI 1.23–1.58]; $P = 1.63 \times 10^{-7}$; $P_{emp} = 1.1 \times 10^{-5}$). This SNP was no longer significant after adjustment for rs7903146 ($P = 0.21$), suggesting that rs7903146 represented the sole association signal in this region.

KLF14. At the *KLF14* locus, the best SNP rs13234269 (OR 1.26 [95% CI 1.14–1.40]; $P = 1.55 \times 10^{-5}$; $P_{emp} = 0.001$) was located at 5' of *KLF14*, 38 kb from the index SNP rs972283. The best and index SNPs resided in two adjacent

LD blocks and were perfectly correlated in CEU ($r^2 = 1$) (Fig. 1C) but showed weak or no correlations in Afa ($r^2 = 0.36$) (Fig. 1D) and YRI ($r^2 = 0.01$) (Table 3). The association was weaker for the index SNP rs972283 (OR 1.24; $P = 8.12 \times 10^{-4}$) (Table 2) as compared with the best SNP, and was no longer significant after conditioning on the best SNP rs13234269 ($P = 0.620$). Haplotype analysis of the best and index SNPs revealed three common haplotypes (AA, AG, and TG). As compared with the AA haplotype, the TG haplotype formed by the risk T allele at rs13234269 and the risk G allele at rs972283 was associated with increased risk

for T2D (OR 1.22; $P = 0.002$), but the AG haplotype did not show significant association (OR 0.99; $P = 0.898$). The TG haplotype was also associated with increased T2D risk when compared with the AG haplotype ($P = 0.002$) (Table 4).

HMGA2. At the *HMGA2* locus, the best SNP rs12049974 (T allele: OR 1.24 [95% CI 1.14–1.35]; $P = 1.73 \times 10^{-6}$; $P_{emp} = 1.36 \times 10^{-4}$) and the index SNP rs1531343, which were 100 kb apart, were located at a region of high LD near the 5' end of *HMGA2*. They were uncorrelated in CEU ($r^2 = 0.01$) (Fig. 1E) but moderately correlated in Afa ($r^2 = 0.43$) (Fig. 1F) and YRI ($r^2 = 0.67$) (Table 3). The association at the best SNP remained significant after adjustment for the effect of the index SNP ($P = 5.07 \times 10^{-6}$). The reported-risk C allele at index SNP showed trend of association with increased T2D risk after conditioning on the best SNP (OR 1.14; $P = 0.051$), in contrast to the protective effect in the unconditional analysis (OR 0.91; $P = 0.022$) (Table 2). The associations at other nearby SNPs were also substantially weaker ($P > 0.001$) after conditioning on the best SNP. Three common haplotypes (AC, TG, and AG) were observed using the best and the index SNPs. With reference to the AC haplotype, the TG haplotype was associated with increased T2D risk (OR 1.16 [1.06–1.27]; $P = 0.002$), whereas the AG haplotype was associated with decreased T2D risk (OR 0.86 [0.76–0.98]; $P = 0.024$) (Table 4).

NOTCH2-ADAM30. The best SNP at the *NOTCH2/ADAM30* locus was rs12075171 (OR 1.34 [95% CI 1.15–1.55]; $P = 1.25 \times 10^{-4}$; $P_{emp} = 0.011$) located at 5' end of the nearby gene, *REG4*. The best and index SNPs were located in discrete LD blocks and were uncorrelated with each other (Fig. 1G and H). None of the other SNPs in this region were significantly associated after correction for the effective number of SNPs. The best SNP did not appear to demonstrate transferability at this locus.

KCNQ1. Although the regional associations at the *KCNQ1* locus did not reveal significance after correction for the effective number of SNPs, several SNPs in two regions near the index SNPs showed nominal associations. At *KCNQ1* intron 15, the strongest associations were observed at the index SNP rs2237892 (OR 1.25 [95% CI 1.09–1.43]; $P = 0.0018$; $P_{emp} = 0.22$) and rs2283228 (1.23 [1.08–1.40]; $P = 0.0017$; $P_{emp} = 0.21$), which were highly correlated with each other ($r^2 = 0.91$ in Afa, 0.86 in ASN, 1 in CEU, and 0.92 in YRI) (Fig. 1I and J and Table 3). At *KCNQ1* intron 11, the best SNP was rs231361 (1.17 [1.07–1.28]; $P = 6.64 \times 10^{-4}$; $P_{emp} = 0.082$). This SNP was in weak LD ($r^2 = 0.13$ in Afa, 0.40 in CEU, and 0.09 in YRI) (Fig. 1K and L and Table 3)

to the insignificant index SNP rs231362. The effects of rs231361 were lower after conditioning on the index SNP alone (1.13 [1.02–1.25]; $P = 0.018$), and on both the index SNP and a surrogate SNP rs2283202 ($r^2 = 0.5$ to rs231362 in CEU) (1.15 [0.98–1.34]; $P = 0.086$) (Supplementary Table 2). Haplotype analysis of the index and best SNPs revealed three common haplotypes (AG, GA, and GG). As compared with the AG haplotype, the GA haplotype formed by the risk G allele at rs231362 and the risk A allele at rs231361 was associated with increased risk for T2D (OR 1.16; $P = 0.027$), but the GG haplotype did not reveal significant association (1.02; $P = 0.767$). The GA haplotype was also associated with increased T2D risk when compared with the GG haplotype ($P = 0.013$) (Table 4). The associations at the best SNPs in intron 11 (rs231361) and intron 15 (rs2237892) remained significant after conditioning on each other ($P = 0.003$ and 0.012, respectively), suggesting independent associations.

Population differentiation and natural selection at index SNPs. When comparing allele frequencies of the risk alleles at index SNPs in the respective European or East Asian populations, the absolute difference in risk allele frequency varied widely in our Afa samples from 0.01 at *HNF1A* to 0.58 at *PRC1*, regardless of whether the index SNPs were associated with T2D in Afa (Supplementary Table 3). Using the YRI population as a surrogate for Afa in this study, F_{ST} values at the index SNPs between CEU and YRI, or ASN and YRI populations, were highly significant at only one locus, *PRC1* (Supplementary Table 2), suggesting modest population differentiation. Three index SNPs at *NOTCH2-ADAM30*, *HMGA2*, and *FTO* showed significant integrated haplotype scores, suggesting recent positive selection (Supplementary Table 2). Finally, we performed varLD to assess for differential LD around the index SNPs. Four loci at *BCL11A*, *IRS1*, *DGKB/TMEM195*, and *PRC1*, and one locus at *PTPRD*, demonstrated significant differences in LD between YRI and CEU or ASN populations, respectively (data not shown).

DISCUSSION

We found that among 41 independent T2D-associated index SNPs, only seven in *TCF7L2*, *KLF14*, *KCNQ1*, *CDKALI*, *JAZF1*, *ADCY5*, and *GCKR* were significantly associated with T2D in Afa. The index SNPs in *ADCY5* and *GCKR* were initially identified for strong association with fasting glucose levels in European populations (13), suggesting that genes regulating glucose homeostasis may also affect

TABLE 4
Haplotype analyses of index and best SNPs for association with T2D in Afa

Reported loci	SNPs¶	Haplotype†	Frequency	OR (95% CI)	P ‡
<i>KLF14</i> §	rs13234269(b)	AA	0.10	Reference	—
	rs972283(i)	AG*	0.10	0.99 (0.83–1.18)	0.898
		T*G*	0.78	1.22 (1.08–1.39)	0.002
<i>HMGA2</i>	rs12049974(b)	AC*	0.39	Reference	—
	rs1531343(i)	T*G	0.42	1.16 (1.06–1.27)	0.002
		AG	0.18	0.86 (0.76–0.98)	0.024
<i>KCNQ1</i>	rs231362(i)	AG	0.18	Reference	—
	rs231361(b)	G*A*	0.38	1.16 (1.02–1.32)	0.027
		G*G	0.44	1.02 (0.90–1.15)	0.767

¶b, best SNP in this study; i, reported index SNP. †Risk allele is denoted by asterisk. ‡Significant associations ($P < 0.05$) are set in boldface type. §*KLF14*: AG (reference) vs. TG haplotypes; OR (95% CI) 1.44 (1.14–1.81); $P = 0.002$. ||*KCNQ1*: GG (reference) vs. GA haplotypes; OR (95% CI) 1.13 (1.03–1.25); $P = 0.013$.

T2D susceptibility in AfA. The lack of replication for most index SNPs may be partly due to insufficient power, as 16 and 28 SNPs have <50 and <80% power to detect association for the previously reported effect sizes at α level of 0.05, respectively (Supplementary Table 1). In addition, winner's curse may overestimate the true population effect size in the prior GWAS. Only seven index SNPs in AfA had an effect size greater than or equal to that reported in European and Asian populations (Supplementary Fig. 1). The lower effects in AfA further lower the power to detect the associations.

Several index sets of SNPs showed evidence of recent positive selection (e.g., *NOTCH2-ADAM30* rs10923931, *HMGA2* rs1531343, and *FTO* rs8050136), were rare (e.g., *PPARG* rs1801282 and *CENTD2* rs1552224), or revealed considerable differences in risk allele frequencies and population differentiation (e.g., *PRC1* rs8042680) in our AfA samples as compared with the European and Asian populations. These factors may also impact the direct replication of prior associations in AfA. Our findings of limited loci showing population differentiation or under-selection did not provide support of the thrifty genotype hypothesis, consistent with a study on 17 T2D loci (45). Interestingly, a locus-wide study of 16 T2D loci in worldwide populations revealed many moderately differentiated loci in sub-Saharan Africans (e.g., *TCF7L2*, *KCNJ11*, *IGF2BP2*, and *SLC30A8*) and several highly differentiated loci in East Asians (e.g., *HHEX*, *THADA*, and *FTO*) that contribute to the global differentiation pattern (46). Recently, Chen et al. (47) showed that risk alleles at 12 T2D loci showed high F_{ST} values, as well as a trend of decreasing frequencies from Africa through Europe to East Asia. Our study focused on the index SNP comparison between Africans and Europeans/Asians, whereas the latter two studies examined global differentiation either by locus or by index SNPs. The capture of causal rather than tagging variants in locus-wide analysis and multiethnic comparisons will likely reveal more loci undergoing differentiation and selection.

Differences in genetic architecture likely affect the pattern of associations, and lower degrees of LD in AfA may facilitate fine mapping of causal variants in loci shared by AfA and other populations (28,48). In the locus-specific analysis, our sample size had 80% power to detect an OR of at least 1.19 for risk allele frequency ≥ 0.2 at α level of 5×10^{-4} (corrected for average effective number of SNPs). Our results demonstrated that the best SNPs in *TCF7L2*, *KLF14*, and *HMGA2* were close to or the same as the index SNPs, and the associations remained significant after correction for multiple comparisons. In addition to *KCNQ1*, we were able to fine map these association signals by comparing LD patterns and analyzing haplotypes formed by the index and best SNPs.

The association at the *TCF7L2* index SNP rs7903146 approached genome-wide significance and was the best signal within the locus and among the reported T2D loci, consistent with prior GWAS showing that this SNP was one of the most significant signals in several populations (12,17,20). Indeed, rs7903146 was also the strongest SNP for the present GWAS, and in one of the CARE plus cohorts (WFSM) reported recently (31). Of note, rs7903146 was located in a 9-kb LD block in AfA and was weakly correlated with neighboring SNPs that were not significantly associated after adjustment for the effect at rs7903146. In contrast, rs7903146 resided in a large, 64-kb LD block in Europeans and was strongly correlated with a set of

different nearby SNPs (Fig. 1A and B). The differential LD pattern suggests that the risk T allele of rs7903146 is located on different haplotypes in AfA and Europeans. The differential association suggests that rs7903146 was the only SNP showing highly significant association in both populations (Fig. 1A and B). Taken together, T2D association at *TCF7L2* was transferable to multiple populations, including AfA, and rs7903146 is likely the causal variant, as suggested by a recent resequencing study (49), or it may share the same haplotype with the causal variant across different populations.

At *KCNQ1* intron 15, rs2237892 was the index SNP identified in East Asians and the best SNP in AfA. Similarly, the index and best SNPs at *KLF14* were highly correlated in Europeans. In both cases, the best SNPs in AfA were correlated with the same set of SNPs as the index SNPs in East Asians and Europeans, respectively, but at a reduced LD. This suggests that the best and index SNPs may capture a shared causal variant on the same haplotype in these populations. At *KCNQ1*, rs2237892 and rs2283228 were highly correlated in several populations, including Europeans, Asians, Africans, and AfA, so the two signals were indistinguishable. However, the reduced LD and the absence of association in other nearby SNPs in AfA suggest that those are not the causal variants. At *KLF14*, significant association was only observed for the TG haplotype carrying risk alleles of both the best and index SNPs, but not for the AG haplotype carrying only risk allele at the index SNP, suggesting that the causal variant may be located closer to the 5' end of *KLF14* as originally reported (12) and likely resides on the TG haplotype shared across the studied populations.

In contrast, the correlations of the best SNPs in AfA and the index SNPs in Europeans at *KCNQ1* intron 11 and *HMGA2* were relatively weak ($r^2 < 0.5$) in both populations. At *KCNQ1*, the best and index SNPs shared moderate correlations with some nearby SNPs (Fig. 1K and L). Haplotype analyses suggest that AfA and Europeans may share the same causal variant on the GA haplotype formed by risk alleles of both SNPs. The scenario at *HMGA2* is more complex. Haplotype analyses showed that the TG haplotype (frequency = 0.42) was at risk for T2D, whereas the AG haplotype (frequency = 0.18) was protective for T2D in AfA when compared with the AC haplotype (frequency = 0.39). The result of the TG haplotype being at risk was consistent with that of single SNP associations where the T allele of the best SNP and the G allele of the index SNP increased risk for T2D in AfA. Note that the AC haplotype was also at risk for T2D when compared with the AG haplotype. The stronger risk effect in the TG haplotype than in the AC haplotype explains the spurious opposite direction of association at the index SNP with or without conditioning on the best SNP. The respective haplotype frequencies are substantially different in CEU (0.04 for TG, 0.86 for AG, and 0.10 for AC). The TG haplotype is rare in Europeans. The AC haplotype is likely at risk for T2D as compared to the AG haplotype, leading to the observation of the index SNP C allele being at risk in Europeans. Recently, a multiethnic, gene-centric study revealed that rs9668162 at *HMGA2* was associated with T2D risk in AfA (27). Rs9668162 was moderately correlated ($r^2 = 0.41$) with the best SNP rs12049974 but weakly correlated ($r^2 = 0.18$) with the European index SNP rs1531343 in our AfA samples, supporting our finding of an independent signal at *HMGA2* in AfA. Together, this suggests either allelic heterogeneity with different causal

variants residing on different haplotypes in different populations, or a common causal variant residing on multiple haplotypes at different frequencies shared across populations.

To our knowledge, this is the first comprehensive fine-mapping study of reported T2D loci in AfA. We found that only 8 out of 40 loci at *TCF7L2*, *KLF14*, *KCNQ1*, *ADCY5*, *CDKAL1*, *JAZF1*, *GCKR*, and *HMGA2* were transferable to AfA with significant associations at the index or nearby SNPs. It should be noted that the magnitudes of association vary dramatically from strong association (*TCF7L2*) to very nominal evidence, e.g., *GCKR* and *JAZF1*. The lack of association is likely due to limitations in study power, population differentiation, and/or differential LD. Additional genetic variants, likely yet to be discovered, will unravel the high prevalence of T2D in AfA populations. Importantly, the reduced and differential LD patterns in AfA at the significant loci support the fine mapping of regions of association in prior reports. Subsequent studies, including higher-density imputation to the 1,000 genomes, trans-ethnic meta-analysis at loci demonstrating population variation in LD structure, and functional studies, will be valuable for localizing causal variants and confirming these findings.

ACKNOWLEDGMENTS

Genotyping services for the WFSM study were provided by the CIDR. CIDR is fully funded through a federal contract from the National Institutes of Health (NIH) to The Johns Hopkins University (Contract HHSC268200782096C). The work at Wake Forest was supported by NIH grants K99-DK-081350 (N.D.P.), R01-DK-066358 (D.W.B.), R01-DK-053591 (D.W.B.), R01-HL-56266 (B.I.F.), and R01-DK-070941 (B.I.F.), and in part by the General Clinical Research Center of the WFSM Grant M01-RR-07122. This work was also supported by the NHLBI. The following five parent studies have contributed parent study data, ancillary study data, and DNA samples through the Massachusetts Institute of Technology-Broad Institute (N01-HC-65226) to create this genotype/phenotype database for wide dissemination to the biomedical research community: ARIC, CARDIA, CFS, JHS, and MESA. ARIC is carried out as a collaborative study supported by NHLBI contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C) R01-HL-087641, R01-HL-59367, and R01-HL-086694, National Human Genome Research Institute contract U01-HG-004402, and NIH contract HHSN268200625226C. Infrastructure was partly supported by Grant UL1-RR-025005, a component of the NIH and NIH Roadmap for Medical Research. CARDIA is affiliated with University of Alabama at Birmingham (N01-HC-48047), University of Minnesota (N01-HC-48048), Northwestern University (N01-HC-48049), Kaiser Foundation Research Institute (N01-HC-48050), University of Alabama at Birmingham (N01-HC-95095), Tufts-New England Medical Center (N01-HC-45204), WFSM (N01-HC-45205), Harbor-UCLA Research and Education Institute (N01-HC-05187), and University of California Irvine (N01-HC-45134, N01-HC-95100). CFS is affiliated with Case Western Reserve University (NIH HL-46380 and M01-RR-00080). JHS is affiliated with Jackson State University (N01-HC-95170), University of Mississippi (N01-HC-95171), and Tougaloo College (N01-HC-95172). MESA

is conducted and supported by the NHLBI in collaboration with MESA investigators. Support is provided by grants and contracts N01-HC-95159, N01-HC-95160, N01-HC-95161, N01-HC-95162, N01-HC-95163, N01-HC-95164, N01-HC-95165, N01-HC-95166, N01-HC-95167, N01-HC-95168, N01-HC-95169, and RR-024156. Funding for SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. Funding for CARE genotyping was provided by NHLBI Contract N01-HC-65226. J.B.M. is supported by National Institute of Diabetes and Digestive and Kidney Diseases (R01-DK-078616 and K24-DK-080140).

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

M.C.Y.N. researched data and wrote and edited the manuscript. R.S. contributed to discussion and reviewed and edited the manuscript. J.L., L.D., and J.X. researched data. N.D.P. and D.W.B. researched data, contributed to discussion, and reviewed and edited the manuscript. L.J.R.-T., J.M.Z., D.S.S., S.R.P., E.D.C., M.S., Y.-D.I.C., A.G.B., M.L., S.F.A.G., B.M.P., J.S.P., C.D.L., B.I.F., J.I.R., and J.G.W. researched data and reviewed and edited the manuscript. J.D. and J.B.M. reviewed and edited the manuscript. M.C.Y.N. 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.

Part of this study was presented in abstract form at the 60th Annual Meeting of the American Society of Human Genetics, Washington, DC, 2–6 November 2010.

The authors would like to acknowledge the contributions of the involved research institutions, study investigators, field staff, and study participants of ARIC, CARDIA, CFS, JHS, and MESA in creating CARE for biomedical research (<http://public.nhlbi.nih.gov/GeneticsGenomics/home/care.aspx>).

REFERENCES

- Centers for Disease Control and Prevention. *National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States*. Atlanta, GA, U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2011
- Cowie CC, Rust KF, Ford ES, et al. Full accounting of diabetes and prediabetes in the U.S. population in 1988-1994 and 2005-2006. *Diabetes Care* 2009;32:287-294
- Rotimi C, Cooper R, Cao G, Sundarum C, McGee D. Familial aggregation of cardiovascular diseases in African-American pedigrees. *Genet Epidemiol* 1994;11:397-407
- Saxena R, Voight BF, Lyssenko V, et al.; Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* 2007;316:1331-1336
- 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;316:1341-1345
- Sladek R, Rocheleau G, Rung J, et al. A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* 2007;445:881-885
- Steinthorsdottir V, Thorleifsson G, Reynisdottir I, et al. A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat Genet* 2007;39:770-775
- Zeggini E, Weedon MN, Lindgren CM, et al.; Wellcome Trust Case Control Consortium (WTCCC). Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* 2007;316:1336-1341
- Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-678
- Zeggini E, Scott LJ, Saxena R, et al.; Wellcome Trust Case Control Consortium. Meta-analysis of genome-wide association data and large-scale

- replication identifies additional susceptibility loci for type 2 diabetes. *Nat Genet* 2008;40:638–645
11. Rung J, Cauchi S, Albrechtsen A, et al. Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat Genet* 2009;41:1110–1115
 12. Voight BF, Scott LJ, Steinthorsdottir V, et al.; MAGIC investigators; GIANT Consortium. Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat Genet* 2010;42:579–589
 13. Dupuis J, Langenberg C, Prokopenko I, et al.; DIAGRAM Consortium; GIANT Consortium; Global BPgen Consortium; Anders Hamsten on behalf of Procardis Consortium; MAGIC investigators. New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat Genet* 2010;42:105–116
 14. Qi L, Cornelis MC, Kraft P, et al.; Meta-Analysis of Glucose and Insulin-Related Traits Consortium (MAGIC); Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium. Genetic variants at 2q24 are associated with susceptibility to type 2 diabetes. *Hum Mol Genet* 2010;19:2706–2715
 15. Unoki H, Takahashi A, Kawaguchi T, et al. SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat Genet* 2008;40:1098–1102
 16. Yasuda K, Miyake K, Horikawa Y, et al. Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008;40:1092–1097
 17. Yamauchi T, Hara K, Maeda S, et al. A genome-wide association study in the Japanese population identifies susceptibility loci for type 2 diabetes at *UBE2E2* and *C2CD4A-C2CD4B*. *Nat Genet* 2010;42:864–868
 18. Tsai FJ, Yang CF, Chen CC, et al. A genome-wide association study identifies susceptibility variants for type 2 diabetes in Han Chinese. *PLoS Genet* 2010;6:e1000847
 19. Shu XO, Long J, Cai Q, et al. Identification of new genetic risk variants for type 2 diabetes. *PLoS Genet* 2010;6:pii: e1001127
 20. Kooner JS, Saleheen D, Sim X, et al.; DIAGRAM; MuTHER. Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat Genet* 2011;43:984–989
 21. Cho YS, Chen CH, Hu C, et al.; DIAGRAM Consortium; MuTHER Consortium. Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in east Asians. *Nat Genet* 2012;44:67–72
 22. Ng MC, Park KS, Oh B, et al. Implication of genetic variants near *TCF7L2*, *SLC30A8*, *HHEX*, *CDKAL1*, *CDKN2A/B*, *IGF2BP2*, and *FTO* in type 2 diabetes and obesity in 6,719 Asians. *Diabetes* 2008;57:2226–2233
 23. 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:pii: e1001078
 24. Sim X, Ong RT, Suo C, et al. Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet* 2011;7:e1001363
 25. 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;57:2220–2225
 26. Lettre G, Palmer CD, Young T, et al. Genome-wide association study of coronary heart disease and its risk factors in 8,090 African Americans: the NHLBI CARE Project. *PLoS Genet* 2011;7:e1001300
 27. 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;90:410–425
 28. Teo YY, Small KS, Kwiatkowski DP. Methodological challenges of genome-wide association analysis in Africa. *Nat Rev Genet* 2010;11:149–160
 29. Genuth S, Alberti KG, Bennett P, et al.; Expert Committee on the Diagnosis and Classification of Diabetes Mellitus. Follow-up report on the diagnosis of diabetes mellitus. *Diabetes Care* 2003;26:3160–3167
 30. McDonough CW, Palmer ND, Hicks PJ, et al. A genome-wide association study for diabetic nephropathy genes in African Americans. *Kidney Int* 2011;79:563–572
 31. Palmer ND, McDonough CW, Hicks PJ, et al.; DIAGRAM Consortium; MAGIC Investigators. A genome-wide association search for type 2 diabetes genes in African Americans. *PLoS ONE* 2012;7:e29202
 32. Hester JM, Wing MR, Li J, et al. Implication of European-derived adiposity loci in African Americans. *Int J Obes (Lond)* 2012;36:465–473
 33. Patterson N, Price AL, Reich D. Population structure and eigenanalysis. *PLoS Genet* 2006;2:e190
 34. Patterson N, Hattangadi N, Lane B, et al. Methods for high-density admixture mapping of disease genes. *Am J Hum Genet* 2004;74:979–1000
 35. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;155:945–959
 36. Keene KL, Mychaleckyj JC, Leak TS, et al. Exploration of the utility of ancestry informative markers for genetic association studies of African Americans with type 2 diabetes and end stage renal disease. *Hum Genet* 2008;124:147–154
 37. Chen MH, Yang Q. GWAF: an R package for genome-wide association analyses with family data. *Bioinformatics* 2010;26:580–581
 38. Hindorf LA, Sethupathy P, Junkins HA, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA* 2009;106:9362–9367
 39. Almasy L, Blangero J. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am J Hum Genet* 1998;62:1198–1211
 40. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol* 2006;4:e72
 41. Ong RT, Teo YY. varLD: a program for quantifying variation in linkage disequilibrium patterns between populations. *Bioinformatics* 2010;26:1269–1270
 42. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;19:149–150
 43. Yang Y, Remmers EF, Ogunwole CB, Kastner DL, Gregersen PK, Li W. Effective sample size: quick estimation of the effect of related samples in genetic case-control association analyses. *Comput Biol Chem* 2011;35:40–49
 44. Takeuchi F, Serizawa M, Yamamoto K, et al. Confirmation of multiple risk loci and genetic impacts by a genome-wide association study of type 2 diabetes in the Japanese population. *Diabetes* 2009;58:1690–1699
 45. Southam L, Soranzo N, Montgomery SB, et al. Is the thrifty genotype hypothesis supported by evidence based on confirmed type 2 diabetes and obesity-susceptibility variants? *Diabetologia* 2009;52:1846–1851
 46. Klimentidis YC, Abrams M, Wang J, Fernandez JR, Allison DB. Natural selection at genomic regions associated with obesity and type-2 diabetes: East Asians and sub-Saharan Africans exhibit high levels of differentiation at type-2 diabetes regions. *Hum Genet* 2011;129:407–418
 47. Chen R, Corona E, Sikora M, et al. Type 2 diabetes risk alleles demonstrate extreme directional differentiation among human populations, compared to other diseases. *PLoS Genet* 2012;8:e1002621
 48. Teo YY, Ong RT, Sim X, Tai ES, Chia KS. Identifying candidate causal variants via trans-population fine-mapping. *Genet Epidemiol* 2010;34:653–664
 49. 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;60:662–668