

# Genome-wide association study identifies novel loci association with fasting insulin and insulin resistance in African Americans

Guanjie Chen<sup>1,\*</sup>, Amy Bentley<sup>1</sup>, Adebowale Adeyemo<sup>1</sup>, Daniel Shriner<sup>1</sup>, Jie Zhou<sup>1</sup>, Ayo Doumatey<sup>1</sup>, Hanxia Huang<sup>1</sup>, Edward Ramos<sup>1</sup>, Michael Erdos<sup>2</sup>, Norman Gerry<sup>3</sup>, Alan Herbert<sup>4</sup>, Michael Christman<sup>3</sup> and Charles Rotimi<sup>1</sup>

<sup>1</sup>Center for Research on Genomics and Global Health and <sup>2</sup>Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892–8004, USA, <sup>3</sup>Coriell Institute for Medical Research, Camden, NJ 08103, USA and <sup>4</sup>Department of Genetics and Genomics, Boston University School of Medicine, Boston, MA 02118, USA

Received February 28, 2012; Revised June 28, 2012; Accepted July 9, 2012

Insulin resistance (IR) is a key determinant of type 2 diabetes (T2D) and other metabolic disorders. This genome-wide association study (GWAS) was designed to shed light on the genetic basis of fasting insulin (FI) and IR in 927 non-diabetic African Americans. 5 396 838 single-nucleotide polymorphisms (SNPs) were tested for associations with FI or IR with adjustments for age, sex, body mass index, hypertension status and first two principal components. Genotyped SNPs ( $n = 12$ ) with  $P < 5 \times 10^{-6}$  in African Americans were carried forward for *de novo* genotyping in 570 non-diabetic West Africans. We replicated SNPs in or near *SC4MOL* and *TCERG1L* in West Africans. The meta-analysis of 1497 African Americans and West Africans yielded genome-wide significant associations for SNPs in the *SC4MOL* gene: rs17046216 ( $P = 1.7 \times 10^{-8}$  and  $2.9 \times 10^{-8}$  for FI and IR, respectively); and near the *TCERG1L* gene with rs7077836 as the top scoring ( $P = 7.5 \times 10^{-9}$  and  $4.9 \times 10^{-10}$  for FI and IR, respectively). *In silico* replication in the MAGIC study ( $n = 37\,037$ ) showed weak but significant association (adjusted  $P$ -value of 0.0097) for rs34602777 in the *MYO5A* gene. In addition, we replicated previous GWAS findings for IR and FI in Europeans for *GCKR*, and for variants in four T2D loci (*FTO*, *IRS1*, *KLF14* and *PPARG*) which exert their action via IR. In summary, variants in/near *SC4MOL*, and *TCERG1L* were associated with FI and IR in this cohort of African Americans and were replicated in West Africans. *SC4MOL* is under-expressed in an animal model of T2D and plays a key role in lipid biosynthesis, with implications for the regulation of energy metabolism, obesity and dyslipidemia. *TCERG1L* is associated with plasma adiponectin, a key modulator of obesity, inflammation, IR and diabetes.

## INTRODUCTION

Elevated fasting insulin (FI) and insulin resistance (IR) are hallmarks of type 2 diabetes (T2D) (1,2). Both lifestyle characteristics (e.g. diet and physical inactivity) and genetic susceptibility are important etiologic risk factors for elevated FI, IR and T2D (3). Findings from single-gene defects in twin and family studies provided initial evidence supporting

the role of genes in the etiology of both IR and T2D (4). The recent success of genome-wide association study (GWAS) has significantly changed our understanding of the genetic basis of FI and IR with the identification of several well-replicated loci (5).

Relative to European Americans, African Americans have been reported to have a higher degree of IR (6) and display lower insulin sensitivity and C-peptide response to glucose

\*To whom correspondence should be addressed at: 12 South Drive, Building 12A, Room 4047, Bethesda, MD 20878, USA. Tel: +1 3015945065; Fax: +1 3014515426; Email: chengu@mail.nih.gov

challenge (7). Understanding the mechanisms behind these differences in IR will be important in the effort to reduce the well-documented ethnic disparities in morbidity and mortality associated with IR and T2D (7). Thus, this GWAS was designed to contribute to our understanding of the genetic basis of two important components (FI and IR) of the pathophysiology of T2D in African Americans and other African ancestry populations.

## RESULTS

Clinical characteristics of the 927 African-American participants included in this study and of the West Africans included in the replication sample are presented in Table 1. As expected, African Americans were heavier, had higher fasting glucose and FI concentrations and displayed a higher level of IR. Overall, the correlation between IR and FI is 0.89. Estimated heritability from pedigree data was 0.50 (SE = 0.007 and  $P$ -value =  $2.5 \times 10^{-15}$ ) and 0.46 (SE = 0.007 and  $P$ -value =  $1.4 \times 10^{-13}$ ) for log FI and log IR, respectively.

We tested 5 396 838 autosomal single-nucleotide polymorphisms (SNPs) for association with the log of FI and IR. The distributions of association  $P$ -values for FI and IR are displayed in Figure 1 (quantile–quantile plots are shown in Supplementary Material, Fig. S1). The values of the genomic control ( $\lambda_{GC}$ ) inflation factor were  $\sim 1$  for both outcomes, indicating negligible inflation of association test statistics due to population stratification. We used a  $P$ -value of  $2.5 \times 10^{-8}$  as the pre-determined genome-wide significance threshold (8).

Loci showing association with  $P$ -values of  $< 5.0 \times 10^{-6}$  for FI and IR are displayed in Supplementary Material, Tables S1 and S2 ( $n = 87$  FI only) and S3 ( $n = 91$  IR only). Forty-one SNPs within an  $\sim 76$  kb ( $\sim 100$  to  $\sim 170$  kb) region near *neurotrophic tyrosine kinase, receptor, type 2* (*NTRK2*) reached  $P < 5.0 \times 10^{-6}$ . The lowest  $P$ -value in this region was for the association of rs1077625 with IR ( $2.7 \times 10^{-8}$ ; Supplementary Material, Fig. S2); this SNP was also the most significantly associated with FI ( $5.2 \times 10^{-8}$ , Supplementary Material, Fig. S2). Other significant associations with FI or IR in this region were eliminated in conditional analyses adjusting for rs1077625 (Supplementary Material, Tables S2 and S3). Outside of this region, there were several other SNPs with  $P < 5 \times 10^{-6}$  that were associated with FI and/or IR (Supplementary Material, Table S1). SNP rs17046216 in *sterol-C4-methyl oxidase-like* (*SC4MOL*) was associated with FI and IR ( $P = 4.6 \times 10^{-7}$  and  $9.9 \times 10^{-7}$ , respectively; Supplementary Material, Fig. S3). Also, rs7077836, located 139 kb upstream of the *transcription elongation regulator 1-like* (*TCERGIL*) gene was associated with FI and IR ( $P = 3.3 \times 10^{-7}$  and  $1.0 \times 10^{-7}$ , respectively; Supplementary Material, Fig. S4). rs9524798, located in *glypican 6* (*GPC6*), was also associated with both traits ( $P = 7.9 \times 10^{-7}$  and  $8.3 \times 10^{-7}$  for FI and IR, respectively; Supplementary Material, Fig. S5).

We attempted to replicate the top-scoring associations observed in this study of African Americans in independent samples of West Africans and Europeans. The clinical and epidemiologic characteristics of the 570 non-diabetic unrelated West Africans are displayed in Table 1. We replicated the associations we observed in the primary analysis for SNPs

in or near *SC4MOL* and *TCERGIL* genes in West Africans (Table 2). Evidence of association for SNPs in *SC4MOL* and *TCERGIL* reached genome-wide significance in meta-analysis of the combined HUFs (The Howard University Family Study) and West African samples. Using a local replication approach, we observed significant, but weak, replication for rs34602777 (after adjustment for multiple tests,  $P = 0.0097$  for IR) in the *MYO5A* gene in the MAGIC study (Supplementary Material, Table S4) (5). Interestingly, the frequency of the risk allele is 0.289 in African Americans and 0.275 in HapMap CEU, but is absent in YRI. None of the other leading HUFs hits showed significant association in MAGIC (Supplementary Material, Table S4).

We also attempted to replicate previously reported GWAS findings for FI and IR in Europeans (5), using the direct (i.e. querying the reported SNP) and the local approach [i.e. querying SNPs in LD ( $r^2 \geq 0.3$ ) with the reported SNPs within a defined window of 250 kb]. Variants in or near the *GCKR* and *IGF1* genes were investigated, and we replicated the association for *GCKR* in this cohort of African Americans, using local replication approach (Supplementary Material, Table S5). Finally, we replicated ( $P < 0.05$ ) variants in four T2D loci (*FTO*, *IRS1*, *KLF14* and *PPARG*) (9) which exert their action via IR (Supplementary Material, Table S6).

## DISCUSSION

Compared with individuals of European ancestry, African Americans suffer disproportionately from the devastating health consequences of IR and related metabolic disorders including T2D (10). In this study, we used the genome-wide association approach to search for genes that influence the distribution of FI and IR in unrelated non-diabetic African Americans. Variants in or near two genes (*SC4MOL* and *TCERGIL*) were associated with FI and IR and were replicated in an independent sample. Available functional data on these genes support their plausibility as candidate genes for the physiologic modulation of insulin level and insulin-mediated metabolic processes.

First, the *sterol-C4-Methyl oxidase-like* (*SC4MOL*) gene has been shown to be under-expressed in an animal model of T2D and to play a key role in cholesterol biosynthesis with implications for the regulation of energy metabolism, obesity and dyslipidemia (11–15). rs17585739 in *SC4MOL* has been associated with HDL cholesterol (14). In Lu *et al.* (14), HDL level increased with increasing copies of the ‘A’ allele of rs17585739 (AA = 1.52 mmol/l, AG = 1.32 mmol/l, GG = 1.30 mmol/l). Notably, the association of SNPs within this locus is directionally consistent with what would be expected from a locus influencing IR (i.e. the alleles have inverse effects on IR and HDL levels, with alleles increasing FI and IR being negatively associated with HDL levels).

Second, a variant in the *TCERGIL* has been associated with plasma adiponectin (16), a key modulator of obesity, inflammation, IR and diabetes (17). *TCERGIL* was reported to be positively associated with T2D and negatively associated with adiponectin in a recent GWAS conducted among individuals of Indian ancestry. As expected, the association with adiponectin was negative and was most pronounced among

**Table 1.** Characteristics of the African-American participants included in this study and of the West Africans included in the replication sample

	Men			Women		
	African Americans	West Africans	<i>P</i> -values <sup>a</sup>	African Americans	West Africans	<i>P</i> -values <sup>a</sup>
Number of individuals	389	202		538	368	
Age (years)	46.05 ± 12.56 <sup>b</sup>	52.93 ± 11.65	<0.001	46.91 ± 13.45	48.00 ± 10.15	0.1986
BMI (kg/m <sup>2</sup> )	28.28 ± 6.91	24.00 ± 5.13	<0.001	31.35 ± 8.69	26.48 ± 6.04	<0.001
Fasting plasma glucose (mg/dl)	86.79 ± 13.39	82.61 ± 15.72	0.002	84.10 ± 12.63	80.54 ± 15.27	<0.001
FI (μU/ml)	9.41 ± 9.82	6.39 ± 7.78	<0.001	11.07 ± 9.86	9.39 ± 11.28	0.0061
HOMA-IR	2.35 ± 2.91	1.38 ± 1.86	<0.001	2.59 ± 4.19	2.02 ± 2.86	0.0167

<sup>a</sup>*P*-values for gender-stratified comparisons between African Americans (HUFS) and West Africans (AADM study).

<sup>b</sup>Means ± SD.

women (18). In our study, the risk allele ('T') of rs7077836 in the *TCERGIL* locus [minor allele frequency (MAF) = 0.12] was also associated with increased risk of T2D (odds ratio = 1.14, 95% CI: 1.00–1.60) and was positively associated with FI and IR. The association with adiponectin was directionally consistent with what one might expect from a locus influencing risk of T2D and IR.

We observed a significant, but weak, replication of the *MYO5A* locus in the MAGIC study. Inhibition of the *MYO5A* gene, present in the pancreatic islet cells, diminishes insulin-dependent *GLUT4* translocation and glucose uptake in animal models (19,20). The lack of replication of our other top signals in the MAGIC study that included over 37 000 subjects suggests that these findings are still tentative and need to be confirmed in other studies. Although this observation may be due to true population differences (for example, in allele frequencies, LD patterns and effect sizes), the possibility exists that some may be false-positive findings.

Finally, although it was not replicated in the available West African and European samples, the observed genome-wide significant association for SNP rs1077625 near the *NTRK2* gene for FI and IR (*P*-values of  $5.2 \times 10^{-8}$  and  $2.7 \times 10^{-8}$ , respectively) deserves further investigation. Mutation of *NTRK2*, which encodes *TrkB*, was reported to result in a unique human syndrome of hyperphagic obesity and this syndrome involves a signaling molecule that plays important roles in both energy balance and other aspects of central nervous system function (21). Furthermore, we observed a significant association between rs715243 (~65 kb from *NTRK2*) and body mass index (BMI) (*P* = 0.0005) in this study.

In summary, variants in/near *SC4MOL* and *TCERGIL* are associated with FI and IR in African Americans. Despite the small size, this study was successful in identifying promising candidate loci for FI and IR. We postulate that this success in African ancestry population and the lack of replication of these loci in MAGIC are likely due to multiple factors, including the possibility that the genetic architecture of these traits may differ in European and African ancestry populations. Differences in allele frequency, as observed in this study, and currently uncharacterized evolutionary forces manifesting in ancestry-specific patterns in terms of the genetic and environmental determinants of metabolic disorders such as IR offer plausible explanations. For example, the key role of *SC4MOL* in lipid biosynthesis with implications for the regulation of energy metabolism, obesity and dyslipidemia may represent a significant insight into the higher degree of IR,

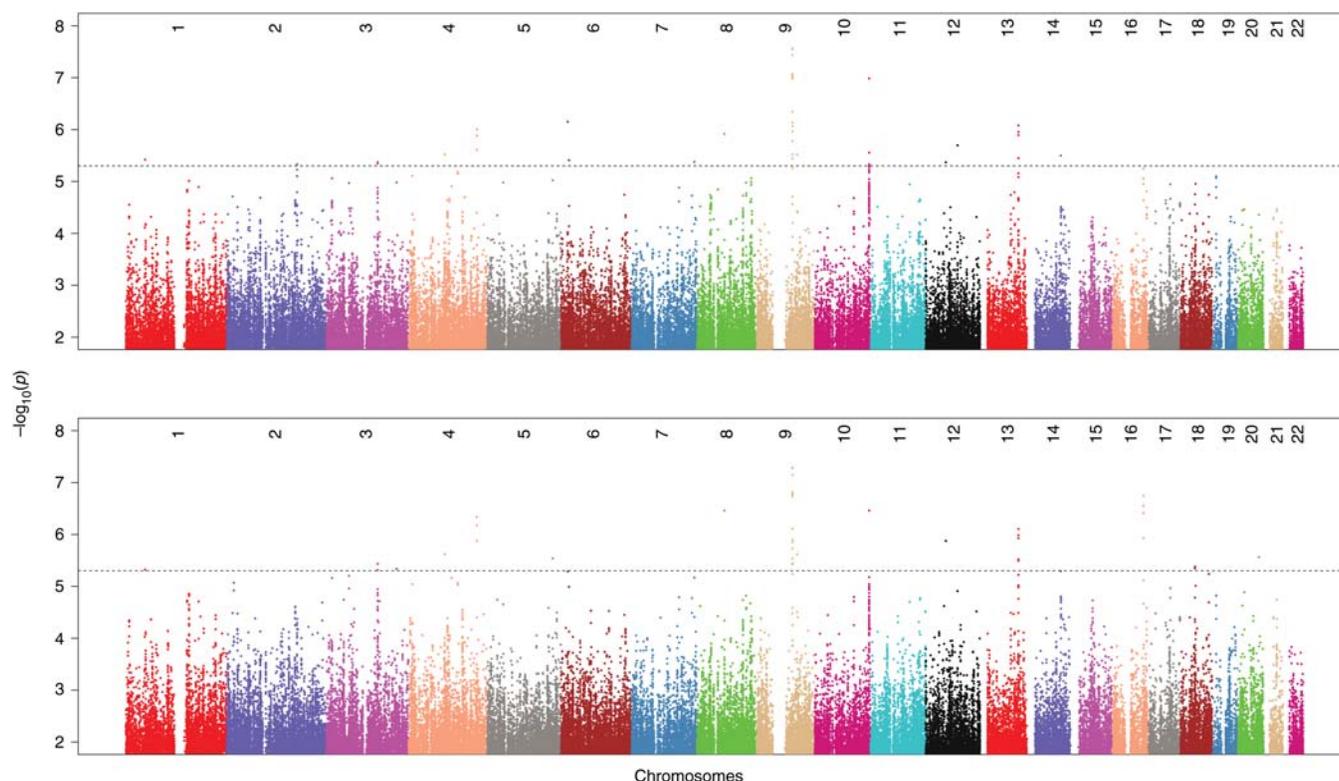
lower insulin sensitivity and C-peptide responses to glucose challenge (6,7) and the well-documented differences in the distribution of lipid profiles (especially triglycerides and HDL) in African ancestry populations compared with populations of European ancestry. Generally, African Americans have a healthier lipid profile than other US ethnic groups, and similar healthier lipid distributions are observed among other African ancestry groups with distinct lifestyle characteristics, suggesting the importance of inherited factors. However, the observed healthier serum lipid profile does not seem to provide corresponding protection against several metabolic disorders including IR, T2D and cardiovascular disease in African Americans. Finally, we note that it is important to be careful in the choice of replication sample for admixed African Americans due to the possibility of observing different allele frequencies between the parental populations, as evident for the *MYO5A* loci (frequency of risk allele is 0.289 in African Americans, 0.275 in HapMap CEU and absent in HapMap YRI; Supplementary Material, Table S1). More studies are needed to confirm the validity of these associations as well as the potential role of these loci in multiple populations.

## MATERIALS AND METHODS

### Genome-wide scan study subjects and clinical characteristics

HUFS is a population-based genetic epidemiology study of African Americans from the Washington, DC, metropolitan area (22). Ethical approval for the study was obtained from the Howard University Institutional Review Board. The major objective of the HUFS was to enroll and examine a randomly ascertained sample of African-American families, along with a set of unrelated individuals, from the Washington, DC, metropolitan area to study the genetic and environmental bases of common complex diseases, including hypertension, obesity, diabetes and associated phenotypes. In order to maximize the utility of this sample for the study of multiple common traits, families and unrelated individuals were not ascertained based on any phenotypes.

During a clinical examination, demographic information was collected by interview. Weight was measured in light clothes on an electronic scale to the nearest 0.1 kg, and height was measured with a stadiometer to the nearest 0.1 cm. BMI was computed as weight (kg) divided by the



**Figure 1.** Manhattan plots for genome-wide association for FI (top) and HOMA-IR (bottom). The Y-axis starts from 1 ( $P$  value = 0.1). The block dotted lines indicate  $P$  value =  $5 \times 10^{-6}$ .

square of height ( $m^2$ ). For this study, we first selected non-diabetic persons with fasting glucose, insulin measurements and genotype data. Second, we selected eligible unrelated individuals and randomly selected one unrelated person per pedigree for a total of 927 unrelated non-diabetic subjects. We studied 1314 individuals within 328 pedigrees in the larger data set (HUFs) to estimate heritability using SOLAR (23).

### Phenotype definitions

Fasting blood samples were obtained from all participants after an overnight fast. Individuals were classified as non-diabetic if they meet the following criteria: a fasting plasma glucose concentration of  $<100$  mg/dl and no history of diabetes and not on treatment for diabetes. Fasting plasma glucose was measured on integra 400 *plus* (Roche Diagnostics) using an enzymatic method with hexokinase, and the unit was mg/dl. FI was measured on an Eleceys1010 immunoassay analyzer (Roche Diagnostics) using an electrochemiluminescence technique, and the unit of measurement was  $\mu U/ml$ . IR was calculated using the formula:  $IR = FI (\mu U/ml) \times \text{fasting glucose (FG)} (mmol/l) / 22.5$  (24). For analysis of the continuous traits (FI and IR), phenotypic values were first natural log-transformed to reduce skew and then winsorized at  $\pm 3$  standard deviations to reduce kurtosis (25). Hypertension was defined as systolic BP  $\geq 140$  mmHg or diastolic BP  $\geq 90$  mmHg or taking medications for high blood pressure.

### DNA sample preparation, genotyping and quality control

DNA was extracted from buffy coat samples following the manufacturer's instructions using a Gentra PUREGENE DNA Isolation Kit (QIAGEN, Valencia, CA, USA). After sample processing, genome-wide genotyping was performed using the Affymetrix Genome Wide Human SNP Array 6.0 (26). Genotypes were called using birdseed, version 2. SNPs were excluded if they had a success rate of  $<95\%$  (41 885 SNPs excluded), an MAF of  $\leq 0.01$  (19 154 SNPs excluded) or had a  $P$ -value for the Hardy-Weinberg test of equilibrium (HWE)  $\leq 10^{-3}$  (6317 SNPs excluded). The current analysis focuses on the 809 465 autosomal SNPs that passed these filters. The sample genotyping rate for this set of SNPs in these individuals was 99.5%. The concordance of blind duplicates was 99.74%. Lower throughput genotyping for replication was carried out using Sequenom Homogenous MassEXTEND or iPLEX Gold SBE assays at the National Human Genome Research Institute at NIH.

### Evaluation of sample population stratification

EIGENSTRAT (27) detects and corrects for population stratification in genome-wide association studies. The method is based on principal components analysis. The resulting correction is specific to a candidate marker's variation in frequency across ancestral populations, minimizing spurious associations while maximizing power to detect true associations. The significant principal components were included as covariates in the regression analysis.

**Table 2.** Replication and meta-analysis for FI and HOMA-IR in West Africans and African Americans

Chr	SNPs	Genes	Trait <sup>a</sup>	African Americans (HUFS)			West Africans (AADM)			Meta (joint) analysis			
				Eff. (freq.)	<i>P</i> -value <sup>b</sup>	Beta (SE)	Eff. (freq.)	<i>P</i> -value <sup>b</sup>	Beta (SE)	<i>P</i> -value	Beta (SE)	Dir <sup>c</sup>	<i>P</i> -value (Het.) <sup>d</sup>
4	rs10050311	<i>SLC10A6</i>	FI	T (0.09)	2.39 × 10 <sup>-6</sup>	-0.32 (0.07)	T (0.07)	0.4023	-0.08 (0.08)	8.21 × 10 <sup>-6</sup>	-0.25 (0.06)	-	0.0671
			IR		3.01 × 10 <sup>-6</sup>	-0.34 (0.07)		0.3268	-0.10 (0.10)	1.06 × 10 <sup>-5</sup>	-0.26 (0.06)	-	0.0561
4	rs17046216	<i>SC4MOL</i>	FI	A (0.48)	4.61 × 10 <sup>-7</sup>	0.20 (0.04)	A (0.48)	0.0095	0.14 (0.04)	1.65 × 10 <sup>-8</sup>	0.18 (0.03)	++	0.4040
			IR		9.91 × 10 <sup>-7</sup>	0.21 (0.04)		0.0086	0.16 (0.05)	2.88 × 10 <sup>-8</sup>	0.19 (0.03)	++	0.5051
	rs17488534	<i>SC4MOL</i>	FI	A (0.48)	6.68 × 10 <sup>-7</sup>	0.20 (0.04)	A (0.47)	0.0147	0.14 (0.04)	3.98 × 10 <sup>-8</sup>	0.18 (0.03)	++	0.3414
			IR		1.31 × 10 <sup>-6</sup>	0.21 (0.04)		0.0125	0.15 (0.05)	5.83 × 10 <sup>-8</sup>	0.19 (0.03)	++	0.4366
5	rs1895320	<i>PTTG1</i>	FI	C (0.12)	2.88 × 10 <sup>-6</sup>	0.28 (0.06)	C (0.11)	0.2534	0.10 (0.06)	5.64 × 10 <sup>-6</sup>	0.22 (0.05)	++	0.0897
			IR		9.51 × 10 <sup>-6</sup>	0.28 (0.06)		0.2335	0.11 (0.08)	1.26 × 10 <sup>-5</sup>	0.23 (0.05)	++	0.1393
10	rs10829848	<i>TCERGIL</i>	FI	T (0.13)	1.35 × 10 <sup>-5</sup>	0.26 (0.06)	T (0.09)	0.0025	0.25 (0.06)	1.02 × 10 <sup>-7</sup>	0.25 (0.05)	++	0.9102
			IR		6.48 × 10 <sup>-6</sup>	0.28 (0.06)		0.0012	0.34 (0.08)	2.61 × 10 <sup>-8</sup>	0.30 (0.05)	++	0.6322
	rs7083202	<i>TCERGIL</i>	FI	C (0.13)	2.52 × 10 <sup>-5</sup>	0.25 (0.06)	C (0.09)	0.0024	0.24 (0.06)	1.80 × 10 <sup>-7</sup>	0.25 (0.05)	-	0.9316
			IR		1.15 × 10 <sup>-5</sup>	0.28 (0.06)		0.0010	0.34 (0.08)	3.95 × 10 <sup>-8</sup>	0.29 (0.05)	-	0.5980
	rs7069620	<i>TCERGIL</i>	FI	T (0.13)	1.06 × 10 <sup>-5</sup>	0.26 (0.06)	T (0.09)	0.0126	0.20 (0.06)	4.47 × 10 <sup>-7</sup>	0.24 (0.05)	++	0.5341
			IR		4.89 × 10 <sup>-6</sup>	0.29 (0.06)		0.0074	0.28 (0.08)	1.02 × 10 <sup>-7</sup>	0.28 (0.05)	++	0.9663
	rs7077836	<i>TCERGIL</i>	FI	T (0.12)	3.46 × 10 <sup>-7</sup>	0.31 (0.06)	T (0.09)	0.0005	0.22 (0.06)	7.50 × 10 <sup>-9</sup>	0.28 (0.05)	++	0.3579
			IR		1.03 × 10 <sup>-7</sup>	0.35 (0.06)		0.0016	0.32 (0.08)	4.86 × 10 <sup>-10</sup>	0.34 (0.05)	++	0.8355
	rs10829854	<i>TCERGIL</i>	FI	A (0.13)	9.97 × 10 <sup>-6</sup>	0.26 (0.06)	A (0.09)	0.0056	0.22 (0.06)	1.80 × 10 <sup>-7</sup>	0.24 (0.05)	++	0.6435
			IR		4.66 × 10 <sup>-6</sup>	0.29 (0.06)		0.0026	0.31 (0.08)	3.61 × 10 <sup>-8</sup>	0.29 (0.05)	++	0.8541
	rs3097970	<i>TCERGIL</i>	FI	C (0.17)	9.05 × 10 <sup>-6</sup>	0.23 (0.05)	C (0.14)	0.0368	0.14 (0.05)	1.48 × 10 <sup>-6</sup>	0.20 (0.04)	-	0.2878
			IR		6.72 × 10 <sup>-6</sup>	0.25 (0.05)		0.0222	0.20 (0.07)	4.35 × 10 <sup>-7</sup>	0.23 (0.05)	-	0.6297
	rs12773697	<i>TCERGIL</i>	FI	G (0.13)	2.62 × 10 <sup>-5</sup>	0.25 (0.06)	G (0.09)	0.0057	0.22 (0.06)	4.53 × 10 <sup>-7</sup>	0.24 (0.05)	-	0.7359
			IR		1.52 × 10 <sup>-5</sup>	0.28 (0.06)		0.0019	0.32 (0.08)	9.23 × 10 <sup>-8</sup>	0.29 (0.05)	-	0.7264
13	rs9524298	<i>GPC6</i>	FI	T (0.16)	7.85 × 10 <sup>-7</sup>	-0.28 (0.05)	T (0.15)	0.4852	-0.05 (0.05)	1.20 × 10 <sup>-5</sup>	-0.19 (0.05)	-	0.0137
			IR		8.30 × 10 <sup>-7</sup>	-0.29 (0.06)		0.5340	-0.05 (0.07)	2.01 × 10 <sup>-4</sup>	-0.18 (0.05)	-	0.0080

<sup>a</sup>FI, fasting insulin; IR, HOMA-insulin resistance.<sup>b</sup>*P*-values under additive genetic model.<sup>c</sup>All association results are expressed relative to the forward strand of the reference genome.<sup>d</sup>Test of heterogeneity between the African-American and the West African samples.

## Imputation

Imputation was performed with MACH, version 1.16 (28), using the 1000 Genome CEU and YRI references (<http://www.sph.umich.edu/csg/abecasis/MACH/download>) 2010–06 releases. Our imputation strategy for this admixed population has been published elsewhere (29). Briefly, we conducted two separate imputations for HUFs using CEU and YRI references and then merge the resulting data sets. If a reference SNP yielded an imputed genotype for both the CEU and YRI reference panels or in YRI reference panel only, we preferentially accepted the genotype, using the YRI reference panel. Otherwise, the CEU reference-imputed genotypes were used. Based on this imputation strategy, 4 587 373 autosomal SNPs [after QC of call  $r^2 \geq 0.3$  (between true allele counts and estimated allele counts), imputation call rate  $>90\%$ ,  $MAF > 0.01$ ,  $HWE > 10^{-3}$ ] were successfully imputed, bringing the total number of genotype and imputed SNPs to 5 396 838. The means (SD) of quality and  $r^2$  scores (28) are 0.99 (0.03) and 0.97 (0.07) for SNPs from Affymetrix 6.0, and 0.95 (0.05) and 0.83 (0.16) for imputed SNPs.

## Association testing

All association testing was performed using PLINK (30). Association between an autosomal marker and a continuous trait was assessed using linear regression assuming an additive genetic effect model. The two traits were analyzed with adjustments for age, sex, BMI, hypertension status and the significant principal components (first two principal components) from EIGENSTRAT. We estimated the genomic inflation factor ( $\lambda_{GC}$ ) based on the median  $\chi^2$  test of all study participants (31).

## Follow-up study in West Africans

We carried forward genotyped SNPs (Table 2) with  $P$ -values  $< 5 \times 10^{-6}$  for IR and/or FI and SNPs in LD ( $r^2 \geq 0.8$ ) with the genotype SNPs in African Americans for replication in a sample of 570 non-diabetic West Africans enrolled as part of the African America Diabetes Mellitus (AADM) study (32). However, SNPs that had nominal association of  $P < 5 \times 10^{-6}$  in African Americans and  $MAF = 0$  or were absent from HapMap YRI were not genotyped in our West African sample. After quality filter ( $HWE > 10^{-3}$ ,  $MAF > 0.01$  and SNP call rate  $>0.95$ ), a total of 12 SNPs within 5 regions were analyzed. We also attempted *in silico* replication of our major findings ( $P$ -value  $< 5 \times 10^{-6}$  for genotyped and imputed SNPs) in European participants in the MAGIC study (5). All association analyses and trait transformation were the same as in the discovery samples of African Americans with adjustment for age, sex, BMI and hypertension status.  $P$ -values for the discovery (African Americans) samples and the replication (West Africans) samples were combined using METAL (<http://www.sph.umich.edu/csg/abecasis/Metal>). All association results were expressed relative to the forward strand of the reference genome. The METAL calculates a  $z$ -statistic for each marker, summarizing the magnitude and direction of the effect relative to the reference allele in each sample and then calculates an overall

$z$ -statistic and  $P$ -value from the weighted average of the statistics. Weights are proportional to the square root of the sample size of each study. Cochran's  $Q$  was estimated to test for heterogeneity across the two samples.

## Transferability of previous GWAS findings

We attempted to replicate published GWAs of FI and/or IR (5) in our African-American samples in two stages: direct replication (i.e. same SNPs as reported) and local replication (i.e. SNPs in reference ancestry LD with reported SNPs) as previously reported (33). For the identification of SNPs in LD with published variants, we used the appropriate reference data set (e.g. HapMap CEU reference data for European ancestry population). To adequately account for multiple testing, we estimated the effective degrees of freedom (d.f.) for the spectrally decomposed covariance matrix for the block of SNPs, using the HUFs genotype data as previously described (34). Power calculations were performed using QUANTO (<http://hydra.usc.edu/gxe>). To estimate power for replication of an SNP in West Africans, we used effect size from the discovery sample (HUFs) and the allele frequency estimates in our West African sample (or HpaMap YRI), assuming an additive genetic model with  $\alpha = 2.5 \times 10^{-8}$ . For power estimation of a variant discovered in HUFs to be replicated in MAGIC (i.e. transferability), we used the effect size from the discovery sample (HUFs) and the allele frequency from MAGIC (or HapMap CEU). For variants discovered in MAGIC, we did the reverse.

## SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

## ACKNOWLEDGEMENTS

C.R. is the guarantor of this work and, as such, had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Conflict of Interest statement.* None declared.

## FUNDING

The study was supported by grants S06GM008016-320107 to C.R. and S06GM008016-380111 to A.A., both from the NIGMS/MBRS/SCORE Program. Participant enrollment was carried out at the Howard University General Clinical Research Center (GCRC), which is supported by grant number 2M01RR010284 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official view of NIH. Additional support was provided by the Coriell Institute for Medical Research. This research was supported in part by the Intramural Research Program of the Center for Research in Genomics and Global Health, NHGRI/NIH. *In silico* replication in European populations

was performed in the MAGIC study ([www.magicinvestigators.org](http://www.magicinvestigators.org)).

## REFERENCES

- Moller, D.E. and Flier, J.S. (1991) Insulin resistance—mechanisms, syndromes, and implications. *N. Engl. J. Med.*, **325**, 938–948.
- Falkner, B. (2003) Insulin resistance in African Americans. *Kidney Int. Suppl.*, **83**, S27–S30.
- Uusitupa, M. (2002) Lifestyles matter in the prevention of type 2 diabetes. *Diabetes Care*, **25**, 1650–1651.
- Carmelli, D., Cardon, L.R. and Fabsitz, R. (1994) Clustering of hypertension, diabetes, and obesity in adult male twins: same genes or same environments? *Am. J. Hum. Genet.*, **55**, 566–573.
- Dupuis, J., Langenberg, C., Prokopenko, I., Saxena, R., Soranzo, N., Jackson, A.U., Wheeler, E., Glazer, N.L., Bouatia-Naji, N., Gloyn, A.L. *et al.* (2010) New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.*, **42**, 105–116.
- Elliott Proctor, J. (2005) *Joslin's Diabetes Mellitus*. Lippincott Williams & Wilkins, c2003.
- Kasim-Karakas, S.E. (2000) Ethnic differences in the insulin resistance syndrome. *Am. J. Clin. Nutr.*, **71**, 670–671.
- Pe'er, I., Yelensky, R., Altshuler, D. and Daly, M.J. (2008) Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. *Genet. Epidemiol.*, **32**, 381–385.
- Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
- Raymond, N.R. and D'Eramo-Melkus, G. (1993) Non-insulin-dependent diabetes and obesity in the black and Hispanic population: culturally sensitive management. *Diabetes Educ.*, **19**, 313–317.
- Rodriguez, C., Raposo, B., Martinez-Gonzalez, J., Llorente-Cortes, V., Vilahur, G. and Badimon, L. (2003) Modulation of ERG25 expression by LDL in vascular cells. *Cardiovasc. Res.*, **58**, 178–185.
- Hailleselle Sene, K., Porter, C.J., Palidwor, G., Perez-Iratxeta, C., Muro, E.M., Campbell, P.A., Rudnicki, M.A. and Andrade-Navarro, M.A. (2007) Gene function in early mouse embryonic stem cell differentiation. *BMC Genomics*, **8**, 85.
- Sato, Y., Koshioka, S., Kirino, Y., Kamimoto, T., Kawazoe, K., Abe, S., Minakuchi, K. and Nakahori, Y. (2011) Role of dipeptidyl peptidase IV (DPP4) in the development of dyslipidemia: DPP4 contributes to the steroid metabolism pathway. *Life Sci*, **88**, 43–49.
- Lu, Y., Dolle, M.E., Imholz, S., van 't Slot, R., Verschuren, W.M., Wijmenga, C., Feskens, E.J. and Boer, J.M. (2008) Multiple genetic variants along candidate pathways influence plasma high-density lipoprotein cholesterol concentrations. *J. Lipid Res.*, **49**, 2582–2589.
- Homo-Delarche, F., Calderari, S., Irminger, J.C., Gangnerau, M.N., Coulaud, J., Rickenbach, K., Dolz, M., Halban, P., Portha, B. and Serradas, P. (2006) Islet inflammation and fibrosis in a spontaneous model of type 2 diabetes, the GK rat. *Diabetes*, **55**, 1625–1633.
- Heid, I.M., Henneman, P., Hicks, A., Coassin, S., Winkler, T., Aulchenko, Y.S., Fuchsberger, C., Song, K., Hivert, M.F., Waterworth, D.M. *et al.* (2010) Clear detection of ADIPOQ locus as the major gene for plasma adiponectin: results of genome-wide association analyses including 4659 European individuals. *Atherosclerosis*, **208**, 412–420.
- Tilg, H. and Moschen, A.R. (2008) Role of adiponectin and PBEF/visfatin as regulators of inflammation: involvement in obesity-associated diseases. *Clin. Sci. (Lond.)*, **114**, 275–288.
- Sim, X., Ong, R.T., Suo, C., Tay, W.T., Liu, J., Ng, D.P., Boehnke, M., Chia, K.S., Wong, T.Y., Seielstad, M. *et al.* (2011) Transferability of type 2 diabetes implicated loci in multi-ethnic cohorts from Southeast Asia. *PLoS Genet.*, **7**, e1001363.
- Varadi, A., Tsuboi, T. and Rutter, G.A. (2005) Myosin Va transports dense core secretory vesicles in pancreatic MIN6 beta-cells. *Mol. Biol. Cell*, **16**, 2670–2680.
- Yoshizaki, T., Imamura, T., Babendure, J.L., Lu, J.C., Sonoda, N. and Olefsky, J.M. (2007) Myosin 5a is an insulin-stimulated Akt2 (protein kinase Bbeta) substrate modulating GLUT4 vesicle translocation. *Mol. Cell. Biol.*, **27**, 5172–5183.
- Yeo, G.S., Connie Hung, C.C., Rochford, J., Keogh, J., Gray, J., Sivaramakrishnan, S., O'Rahilly, S. and Farooqi, I.S. (2004) A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat. Neurosci.*, **7**, 1187–1189.
- Adeyemo, A., Gerry, N., Chen, G., Herbert, A., Doumatey, A., Huang, H., Zhou, J., Lashley, K., Chen, Y., Christman, M. *et al.* (2009) A genome-wide association study of hypertension and blood pressure in African Americans. *PLoS Genet.*, **5**, e1000564.
- Almasy, L. and Blangero, J. (1998) Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.*, **62**, 1198–1211.
- Matthews, D.R., Hosker, J.P., Rudenski, A.S., Naylor, B.A., Treacher, D.F. and Turner, R.C. (1985) Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia*, **28**, 412–419.
- Fernandez, J.R., Etzel, C., Beasley, T.M., Shete, S., Amos, C.I. and Allison, D.B. (2002) Improving the power of sib pair quantitative trait loci detection by phenotype winsorization. *Hum. Hered.*, **53**, 59–67.
- McCarroll, S.A., Kuruvilla, F.G., Korn, J.M., Cawley, S., Nemesh, J., Wysoker, A., Shapero, M.H., de Bakker, P.I., Maller, J.B., Kirby, A. *et al.* (2008) Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*, **40**, 1166–1174.
- Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
- Li, Y., Willer, C.J., Ding, J., Scheet, P. and Abecasis, G.R. (2010) MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.*, **34**, 816–834.
- Shriner, D., Adeyemo, A., Chen, G. and Rotimi, C.N. (2009) Practical considerations for imputation of untyped markers in admixed populations. *Genet. Epidemiol.*, **34**, 258–265.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
- Devlin, B., Roeder, K. and Wasserman, L. (2000) Genomic control for association studies: a semiparametric test to detect excess-haplotype sharing. *Biostatistics*, **1**, 369–387.
- Rotimi, C.N., Dunston, G.M., Berg, K., Akinsete, O., Amoah, A., Owusu, S., Acheampong, J., Boateng, K., Oli, J., Okafor, G. *et al.* (2001) In search of susceptibility genes for type 2 diabetes in West Africa: the design and results of the first phase of the AADM study. *Ann. Epidemiol.*, **11**, 51–58.
- Ramos, E., Chen, G., Shriner, D., Doumatey, A., Gerry, N.P., Herbert, A., Huang, H., Zhou, J., Christman, M.F., Adeyemo, A. *et al.* (2011) Replication of genome-wide association studies (GWAS) loci for fasting plasma glucose in African-Americans. *Diabetologia*, **54**, 783–788.
- Bretherton, S.C., Widmann, M., Dymnikov, P.V., Wallace, M.J. and Blade, I. (1999) The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, **12**, 1990–2009.