G; Zelenika, Diana; Zethelius, Björn; Zhai, Guangju; Zhao, Jing Hua; Zillikens, M Carola; Borecki, Ingrid B; Loos, Ruth J F; Meneton, Pierre; Magnusson, Patrik K E; Nathan, David M; Williams, Gordon H; Silander, Kaisa; Salomaa, Veikko; Smith, George Davey; Bornstein, Stefan R; Schwarz, Peter; Spranger, Joachim; Karpe, Fredrik; Shuldiner, Alan R; Cooper, Cyrus; Dedoussis, George V; Serrano-Ros, Manuel; Lind, Lars; Palmer, Lyle J; Franks, Paul W; Ebrahim, Shah; Marmot, Michael; Kao, W H Linda; Pramstaller, Peter Paul; Wright, Alan F; Stumvoll, Michael; Hamsten, Anders; Buchanan, Thomas A; Valle, Timo T; Rotter, Jerome I; Siscovick, David S; Penninx, Brenda W J H; Boomsma, Dorret I; Deloukas, Panos; Spector, Timothy D; Ferrucci, Luigi; Cao, Antonio; Scuteri, Angelo; Schlessinger, David; Uda, Manuela; Ruokonen, Aimo; Jarvelin, Marjo-Riitta; Waterworth, Dawn M; Vollenweider, Peter; Peltonen, Leena; Mooser, Vincent; Sladek, Robert (2012) A genome-wide association search for type 2 diabetes genes in African Americans. PloS one, 7 (1). e29202. ISSN 1932-6203

Downloaded from: http://researchonline.lshtm.ac.uk/20693/

DOI:

Usage Guidelines

Please refer to usage guidelines at http://researchonline.lshtm.ac.uk/policies.html or alternatively contact researchonline@lshtm.ac.uk.

Available under license: http://creativecommons.org/licenses/by/2.5/
A Genome-Wide Association Search for Type 2 Diabetes Genes in African Americans

Nicholette D. Palmer1,2,3, Caitrin W. McDonough2,3,5, Pamela J. Hicks1,2,3, Bong H. Roh1,2,3, Maria R. Wing2,3,7, S. Sandy An1,2,3, Jessica M. Hester2,3,7, Jessica N. Cooke2,3,5, Meredith A. Bostrom1,2,3, Megan E. Rudock1,2, Matthew E. Talbert2,3,5, Joshua P. Lewis2,3,7, DIAGRAM Consortium4, MAGIC Investigators5, Assiamira Ferrara11, Lingyi Lu8, Julie T. Ziegler8, Michele M. Sale9, Jasmin Divers10, Daniel Shriner10, Adebowale Adeyemo10, Charles N. Rotimi10, Maggie C. Y. Ng2,3,6, Carl D. Langefeld8, Barry I. Freedman4, Donald W. Bowden1,2,3,4

1 Department of Biochemistry, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 2 Center Genomics and Personalized Medicine Research, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 3 Center for Diabetes Research, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 4 Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 5 Program in Molecular Medicine and Translational Science, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 6 Department of Pediatrics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 7 Program in Molecular Genetics and Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 8 Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, North Carolina, United States of America, 9 Center for Public Health Genomics, University of Virginia School of Medicine, Charlottesville, Virginia, United States of America, 10 Center for Research on Genomics and Global Health, National Human Genome Center, Howard University, Bethesda, Maryland, United States of America, 11 Division of Research, Kaiser Permanente, Oakland, California, United States of America

Abstract

African Americans are disproportionately affected by type 2 diabetes (T2DM) yet few studies have examined T2DM using genome-wide association approaches in this ethnicity. The aim of this study was to identify genes associated with T2DM in the African American population. We performed a Genome Wide Association Study (GWAS) using the Affymetrix 6.0 array in 965 African-American cases with T2DM and end-stage renal disease (T2DM-ESRD) and 1029 population-based controls. The most significant SNPs (n = 550 independent loci) were genotyped in a replication cohort and 122 SNPs (n = 98 independent loci) were further tested through genotyping three additional validation cohorts followed by meta-analysis in all five cohorts totaling 3,132 cases and 3,317 controls. Twelve SNPs had evidence of association in the GWAS (P<0.0071), were directionally consistent in the Replication cohort and were associated with T2DM in subjects without nephropathy (P<0.05). Meta-analysis in all cases and controls revealed a single SNP reaching genome-wide significance (P<2.5×10⁻⁸), SNP rs7560163 (P=7.0×10⁻⁹, OR (95% CI) = 0.75 (0.67–0.84)) is located intergenically between RND3 and RBM43. Four additional loci (rs7542900, rs4659485, rs2722769 and rs7107217) were associated with T2DM (P<0.05) and reached more nominal levels of significance (P<2.5×10⁻⁶) in the overall analysis and may represent novel loci that contribute to T2DM. We have identified novel T2DM-susceptibility variants in the African-American population. Notably, T2DM risk was associated with the major allele and implies an interesting genetic architecture in this population. These results suggest that multiple loci underlie T2DM susceptibility in the African-American population and that these loci are distinct from those identified in other ethnic populations.

Introduction

African Americans have a disproportionately high risk for developing type 2 diabetes (T2DM) with an estimated prevalence twice that observed for their European-American counterparts [1]. In addition to socioeconomic and behavioral risk factors, genetic factors are likely contributors to the disproportionate risk observed in this population. Genome-Wide Association Studies (GWAS) have been used extensively with great success to identify common genetic variants associated with T2DM in primarily
European-derived populations [2,3,4]. Until recently, comparable studies have been difficult to perform in African Americans due to the greater complexity of their African-derived genome compounded by recent admixture of European-derived genes. With the development of high density SNP arrays that give reasonable coverage of the African-American genome and methods to account for admixture in this population, it has become possible to perform informative GWAS in the African-American population. The aim of this study was to identify loci that contribute to T2DM by GWAS and replication in multiple African-American samples.

**Results**

**Clinical characteristics of the study samples**

The clinical characteristics of the study samples used in the GWAS, Replication and Validation phases are shown in Table 1. The GWAS and Replication populations were similar. In both groups, the age at enrollment for the T2DM-ESRD subjects was older than for the control groups. However, the mean age at enrollment for the control groups in the GWAS and Replication phases was older than the mean age of T2DM diagnosis in the T2DM-ESRD and T2DM subjects. Notably, the use of population-based controls has not precluded the identification of bona fide associations in other efforts (e.g., [2]). All of the case groups with T2DM (T2DM-ESRD and T2DM) had a higher proportion of females; possibly reflecting the increased prevalence of T2DM among African-American women [5], participation bias and/or survival. On average, all of the groups were overweight or obese at the time of enrollment. Among case subjects, those with T2DM-ESRD had the lowest average body mass index (BMI; 29.7 kg/m², Table 1), and the T2DM subjects without nephropathy (T2DM) had the highest average BMI (35.5 kg/m², Table 1).

**GWAS**

After the application of SNP and sample quality control metrics, 832,357 directly-genotyped, autosomal SNPs were analyzed in 965 African-American T2DM-ESRD case subjects and 1,029 African-American controls lacking T2DM and ESRD. Given the modest increase of the inflation factor with inclusion of related individuals (1.04 versus 1.06) cryptic first degree relatives were retained in the analysis. A summary of the association results is shown in Figure 1 and Figure S1. The top hit was rs5750250 located on chromosome 22 in the MYH9 (non-muscle myosin heavy chain 9) gene (P-value = 3.0 × 10⁻⁷, Figure 1). This gene has been previously associated with non-diabetic and diabetic forms of ESRD [6,7,8,9]. In total, there were 126 SNPs with P-values<1.0×10⁻⁴ (Figure 1). In addition, we also evaluated previously identified T2DM index variants and their corresponding CEU LD blocks for association with T2DM in the African-American population (Table S1). Among the 37 T2DM index variants [3,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27] identified to date from candidate gene studies, large scale association studies and GWAS, 33 were directly-genotyped or imputed. Among these, 20 SNPs showed consistency with the Caucasian-defined risk allele, although most were non-significant. Only rs1634397 and rs903146 were nominally associated (P = 0.016 and 4.9E-05, respectively) although the direction of effect was inconsistent for rs1634397 with previous studies (OR = 0.86 with respect to the Caucasian risk allele G). Notably, additional signals of association were observed in CEU LD blocks containing the index SNP. After correction for multiple comparisons, only SNP rs4506565 in TCF7L2 remained significant (Bonferroni-corrected P = 0.027; n = 18 SNPs (10 effective tests) contained in the CEU LD block and genotyped in the African-American GWAS). The flow of the study through the GWAS, Replication and Validation phases is outlined in Table 2.

**Replication and GWAS + Replication Analysis of T2DM-ESRD cases and controls lacking both T2DM and ESRD**

In an effort to replicate the GWAS results, the most significant 712 SNPs (n = 550 independent loci) were successfully genotyped in an additional sample of 709 African-American T2DM-ESRD cases and 690 African-American controls lacking both T2DM and ESRD (Table 2). In this replication analysis, 70 of the 712 SNPs (9.8%) showed nominal evidence of replication; a P-value<0.05 under an additive genetic model with association in the same direction. Although no SNP reached genome-wide significance (P-value ≤ 2.5×10⁻⁸), P-values ranged from 7.6×10⁻⁴ to 6.5×10⁻⁷ (GWAS + Replication). The top hit from the GWAS, rs5750250, did not reach nominal significance in the replication cohort (P-value = 0.054).

**Validation of T2DM loci**

A total of 122 SNPs were genotyped in three independent cohorts comprising a total of 1,458 African-American T2DM cases and 1,598 controls lacking both T2DM and ESRD (Table 2). These included 56 of the 70 SNPs with evidence of

---

**Table 1. Clinical Characteristics of Study Samples.**

<table>
<thead>
<tr>
<th></th>
<th>GWAS T2DM Case-Control</th>
<th>GWAS IRAS</th>
<th>GWAS IRASFS</th>
<th>Replication T2DM Case-Control</th>
<th>Replication T2DM Control</th>
<th>Validation T2DM Case-Control</th>
<th>Validation T2DM Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>965</td>
<td>1029</td>
<td>709</td>
<td>690</td>
<td>1246</td>
<td>927</td>
<td>115</td>
</tr>
<tr>
<td>Female (%)</td>
<td>61.2%</td>
<td>57.3%</td>
<td>55.7%</td>
<td>51.3%</td>
<td>64.0%</td>
<td>58.0%</td>
<td>53.9%</td>
</tr>
<tr>
<td>Age at Enrollment (years)</td>
<td>61.6±10.5</td>
<td>49.0±11.9</td>
<td>60.2±10.4</td>
<td>48.5±12.8</td>
<td>57.2±11.7</td>
<td>46.6±13.1</td>
<td>56.8±8.0</td>
</tr>
<tr>
<td>Age at T2DM diagnosis (years)</td>
<td>41.6±12.4</td>
<td>39.4±12.5</td>
<td>46.1±12.6</td>
<td>51.1±10.7</td>
<td>51.2±11.9</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Age at ESRD diagnosis (years)</td>
<td>58.0±10.9</td>
<td>56.7±10.9</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>T2DM to ESRD duration (years)</td>
<td>16.2±10.9</td>
<td>20.4±10.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>BMI (kg/m²)</td>
<td>29.7±7.0</td>
<td>30.0±7.0</td>
<td>29.8±6.9</td>
<td>29.4±7.6</td>
<td>33.5±7.6</td>
<td>30.0±7.7</td>
<td>32.1±6.0</td>
</tr>
</tbody>
</table>

Values are presented as trait mean and standard deviation.

doi:10.1371/journal.pone.0029202.t001
replication and 66 SNPs with more nominal evidence of significance in the combined analysis (Table S2). These samples allowed differentiation between association with T2DM or T2DM-ESRD while increasing power of detection for suspected T2DM loci through meta-analysis. Meta-analysis of the five putative T2DM SNPs in Validation samples, revealed association signals with \(P\)-values ranging from 0.011–1.86\(10^{-2}\) (Table S3, Table 3). The most significant SNP was rs7560163 (\(P\)-value = 1.86\(10^{-2}\), odds ratio (OR) (95% confidence interval (95%CI) = 0.74 (0.63–0.87)) located intergenically between \(RND3\) (Ras homolog gene family, member E) and \(RBM43\) (RNA binding motif protein 43).

Meta-analysis of all African-American study samples

The association results of all 122 SNPs successfully genotyped in all five cohorts (GWAS, Replication, T2DM, IRAS and IRASFS) were used in a meta-analysis to compute an overall test of association (Table 3). This analysis combined results from cases (T2DM-ESRD and T2DM; \(n = 3,132\)) and controls (lacking both T2DM and ESRD; \(n = 3,317\)) for a sample size of 6,449 individuals. As a result of this analysis, one SNP reached genome-wide significance (\(P\)-value\(\leq 2.5\times10^{-8}\), Table 3 and Figure S2). SNP rs7560163 (\(P\)-value = 7.06\(10^{-9}\), OR (95% CI) = 0.75 (0.67–0.84)) is located intergenically between \(RND3\) and \(RBM43\). This SNP was tested for association with T2DM, in silico, by the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium [3] however failed quality control filters and was not included in analysis likely due to being monomorphic as seen in a representative Caucasian population from the HapMap project (Table S4).

Quantitative Trait Analysis

Exploration of putative T2DM variants with quantitative glycemic traits in a subset of African-American samples (\(n = 671\) from the IRAS and IRASFS control samples, Table S5) revealed

Table 2. Study Design.

<table>
<thead>
<tr>
<th>Stage</th>
<th>SNPs (Independent Loci)</th>
<th>Cases</th>
<th>Control</th>
<th>Admixture Adjustment</th>
</tr>
</thead>
<tbody>
<tr>
<td>GWAS</td>
<td>832,357</td>
<td>T2DM-ESRD Cases ((n = 965))</td>
<td>Population-based Controls ((n = 1029))</td>
<td>Principal Component 1 (PC1)</td>
</tr>
<tr>
<td>Replication</td>
<td>712 (550)</td>
<td>T2DM-ESRD Cases ((n = 709))</td>
<td>Population-based Controls ((n = 690))</td>
<td>FRAPPE (70 AIMs)</td>
</tr>
<tr>
<td>GWAS + Replication</td>
<td>712 (550)</td>
<td>T2DM-ESRD Cases ((n = 1,674))</td>
<td>Population-based Controls ((n = 1,719))</td>
<td></td>
</tr>
<tr>
<td>Validation</td>
<td>122 (98)</td>
<td>T2DM Cases ((n = 1,246))</td>
<td>Controls ((n = 927))</td>
<td>FRAPPE (76 AIMs)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IRAS T2DM Cases ((n = 115))</td>
<td>IRAS Controls ((n = 164))</td>
<td>FRAPPE (70 AIMs)</td>
</tr>
<tr>
<td>Validation Meta-analysis</td>
<td>122 (98)</td>
<td>IRASFS T2DM Cases ((n = 97))</td>
<td>IRASFS Controls ((n = 507))</td>
<td>FRAPPE (70 AIMs)</td>
</tr>
<tr>
<td>Overall Meta-analysis (All 5 cohorts)</td>
<td>122 (98)</td>
<td>T2DM Cases ((n = 3,132))</td>
<td>Controls ((n = 3,317))</td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pone.0029202.t002

Figure 1. Genome-Wide Association Study Results. Results are adjusted for admixture using PC1 as a covariate in the analysis. \(P\)-values are shown under the additive model. The blue line at \(-\log_{10}(P\text{-value}) = 3\) represents an additive \(P\)-value = 0.001 and the red line at \(-\log_{10}(P\text{-value}) = 5\) represents a \(P\)-value = 1.0\(\times10^{-5}\). doi:10.1371/journal.pone.0029202.g001
### Table 3. GWAS + Replication, Validation and Overall P-values for susceptibility loci identified from the Overall meta-analysis (P<2.5×10^{-5}).

<table>
<thead>
<tr>
<th>Locus</th>
<th>GWAS</th>
<th>GWAS + Replication</th>
<th>Validation</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T2DM-ESRD (n = 965)</td>
<td>T2DM-ESRD (n = 1,674)</td>
<td>T2DM (n = 1,458)</td>
<td>T2DM-ESRD + T2DM (n = 3,132)</td>
</tr>
<tr>
<td>SNP</td>
<td>Position</td>
<td>Alleles</td>
<td>Nearest Gene(s)</td>
<td>MAF</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>----------------</td>
<td>-----</td>
</tr>
<tr>
<td>rs7542900</td>
<td>Chr1:94842629</td>
<td>C/T</td>
<td>SLC44A3 F3</td>
<td>0.46</td>
</tr>
<tr>
<td>rs4659485</td>
<td>Chr1:235212541</td>
<td>T/C</td>
<td>RYR2 MTR</td>
<td>0.10</td>
</tr>
<tr>
<td>rs7560163</td>
<td>Chr2:151346182</td>
<td>C/G</td>
<td>RBM43 RND3</td>
<td>0.13</td>
</tr>
<tr>
<td>rs3775045</td>
<td>Chr4:96345907</td>
<td>C/T</td>
<td>UNC5C</td>
<td>0.28</td>
</tr>
<tr>
<td>rs6451146</td>
<td>Chr5:34599780</td>
<td>T/C</td>
<td>RAI1 -</td>
<td>0.18</td>
</tr>
<tr>
<td>rs7103576</td>
<td>Chr6:148746647</td>
<td>G/A</td>
<td>SASH1 -</td>
<td>0.28</td>
</tr>
<tr>
<td>rs17103805</td>
<td>Chr10:86418457</td>
<td>A/G</td>
<td>- FAM190B</td>
<td>0.10</td>
</tr>
<tr>
<td>rs2722769</td>
<td>Chr11:11184950</td>
<td>C/G</td>
<td>GALNT4 LOC729013</td>
<td>0.10</td>
</tr>
<tr>
<td>rs7107217</td>
<td>Chr11:128978900</td>
<td>C/A</td>
<td>TMEM45B BARX2</td>
<td>0.48</td>
</tr>
<tr>
<td>rs1271784</td>
<td>Chr18:30972955</td>
<td>A/T</td>
<td>MAPRE2</td>
<td>0.22</td>
</tr>
</tbody>
</table>

SNPs are ordered by chromosome and position (NCBI Build 36.1, hg18) with the major/minor alleles (positive strand) and corresponding gene (underlined) or nearest annotated genes (+/- 500 kb). For each phase of the study, GWAS + Replication, Validation and Overall (GWAS+Replication+T2DM+IRRAS+IRRASFS) analyses, the minor allele frequency (MAF) in controls, additive P-value and odds ratio (OR) with associated 95% confidence interval (CI) with respect to the minor allele is listed.

doi:10.1371/journal.pone.0029202.t003
limited insight into the biological mechanism associated with T2DM risk. In addition, the five putative African-American T2DM susceptibility loci were tested for association with quantitative measures of glucose homeostasis in the European Caucasian population, in *silico*, by the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC; [16]). These results did not provide further insight into the probable role these variants may have in disease susceptibility (Table S6). The most significantly associated SNP in African Americans, rs7560163, failed quality controls filters and was not included in analysis likely due to being monomorphic as seen in a representative Caucasian population from the HapMap project (Table S4).

**Exploration of eQTLs for T2DM loci**

Evaluation of three of the five putative African-American T2DM susceptibility loci for association with altered expression levels of neighboring genes revealed no strong evidence of association. However, SNP rs7542900 trended toward association with CNN3 (*β* = 0.20 vs. −0.12, *P* = 0.095). Lack of association could be due to the small sample size (n = 90), ethnic differences (African vs. African American) or lack of identification of the causal variant. SNPs rs4539465 and rs2722769 were not evaluated as they are monomorphic in the YRI population.

**Discussion**

We performed a high-density genome-wide association study to investigate the genetic determinants of T2DM in the African-American population. Meta-analysis of five study cohorts revealed a single SNP, rs7560163, near RND3 that contributes to T2DM in the African-American population. It is noteworthy that this locus and more nominally associated loci are distinct from those implicated in previous GWAS of T2DM in primarily European-derived populations. These results are consistent with our prior observations [28,29] that “European” genes appear to make only modest contributions to inter-individual risk of T2DM in the African-American population.

Although the associations observed reside intergenically, several neighboring genes could be implicated and have characteristics relevant to the pathophysiology of T2DM. The nearest annotated gene to SNP rs7560163, the only SNP identified to reach stringent levels of genome-wide significance in the Overall analysis (*P* = 7.0 × 10⁻⁸, OR = 0.75 (0.67–0.84); Table 3), is *RND3*. This gene encodes the Rho family GTPase 3 which is ubiquitously expressed and has been implicated as a regulator of actin cytoskeleton organization in response to extracellular growth factors [30,31]. Additional SNPs that reached nominal significance in the Validation samples but failed to reach stringent criteria for genome-wide significance in the Overall analysis included SNP rs7542900 located upstream of coagulation factor III precursor (*F3*). Higher expression levels of *F3* have been measured in monocytes from patients with T2DM [32] although this association could be related to unmeasured vascular complications [33,34]. In addition, SNP rs46539485 is located intergenically between the cardiac muscle ryanodine receptor (*RYR2*) and 5-methyltetrahydrofolate-homocysteine (*MTR*) genes however, a biological relationship with T2DM is not clearly evident. SNP rs2722769 resides ~64 kb upstream of UDP-N-acetyl-alpha-D-galactosamine polypolymer (*GALNTL4*); *GALNTL4* is a member of the large subfamily of glycosyltransferases and although little is known about its biological function, *GALNTL2* has been implicated in cholesterol metabolism in a large GWAS meta-analysis [35]. Among other top hits, rs7107217 is located downstream of BarH-like homeobox 2 (*BHR2*), a transcription factor expressed in smooth and skeletal muscle and involved in muscle differentiation [36,37,38].

Exploration of putative T2DM variants with quantitative glycemic traits in a subset of the samples (n = 671, Table S5) revealed limited insight into the biological mechanism associated with T2DM risk. Notably among the SNPs and traits examined, only SNP rs7107217 was nominally associated with fasting insulin (*P* = 0.011). Exploration of these variants in European Caucasian populations represented by the DIAGRAM Consortium [3] and MAGIC [16] revealed only nominal evidence of association with T2DM (rs7107217 *P* = 0.086, located intergenically between *BHR2* and *NFER1*; Table S4) and did not provide further insight into the probable role of these variants in disease susceptibility through examination of quantitative measures of glucose homeostasis (Table S6), respectively.

To put these findings into context, the association of TCFL2 with T2DM has been widely replicated across multiple ethnicities (reviewed in [39] including prior analysis of African-American samples included in this study [28,40]). SNP rs7903146 has been the most strongly associated variant within this locus with one of the largest allelic odds ratio (OR) for a common variant, i.e. OR ~1.35 [3]. Although rs7903146 is not typed on the Affymetrix 6.0 array and given that the genomic interval is not tagged well (max *r²* = 0.45), only nominal evidence of association was observed in our African-American GWAS (*P* = 0.0015, rs4506565; Table S1). Direct genotyping of rs7903146 in the GWAS + Replication (n = 1,674 T2DM-ESRD cases and 1,719 controls lacking both T2DM and ESRD) resulted in the most strongly associated signal observed (*P* = 2.46 × 10⁻⁸) with an odds ratio (OR = 1.33, 95% CI = 1.19–1.48). This odds ratio is in the range of other signals which were observed (Table 3).

A notable observation common to all putative T2DM loci (Table 3) is the association of “protection”, i.e. and odds ratio less than 1.0, with the minor allele. Comparison with data from the International HapMap Consortium [41] confirms that the major allele in all instances is more common in the representative African samples (YRI) from Ibadan, Nigeria. This could suggest that selection for diabetogenic traits is occurring and that the more common, African-derived allele is deleterious in a more westernized environment. This is consistent with a trend we observed in prior tests of “European” T2DM associated variants in African Americans [20].

Since obesity is known to be a significant risk factor for the development of T2DM we explored the potential influence using a surrogate measure of adiposity, body mass index (BMI). As seen in Table 1, BMI differs significantly in the validation cohorts (*P* < 0.0001). Given this significant difference, association analyses were repeated with inclusion of BMI as a covariate in the analysis. Adjustment for BMI did not substantially affect the strength of the associations observed. For example, the most significant hit from the validation analysis, rs7560163, was significantly associated with T2DM in the Validation cohorts (n = 1,149 T2DM cases and 919 controls with BMI data; *P* = 3.59 × 10⁻⁸; data not shown). Additionally, all the case groups with T2DM (T2DM-ESRD and T2DM) had a higher proportion of females (Table 1); possibly reflecting the increased prevalence of disease in women [5]. Gender stratified analyses revealed seven of the ten most strongly associated loci were more significant in women (*P* = 0.0010-3.2E-07; Table S7) although none of the loci remained significantly associated in men (*P* = 0.028-3.9E-06; Table S7). Notably, the power to detect association is diminished when the sample size is reduced (n = 2648 men and 3781 women).

This study has similar limitations to other GWAS conducted in minority populations. Although the current study has a modest sample size for the GWAS discovery phase compared to large-scale
meta-analyses in European-derived populations, power calculations (Tables S8 and S9) show that this study has greater than 80% statistical power to detect effects for common variants (MAF = 0.20) consistent with published effect sizes (OR = 1.20) for T2DM (e.g., transcription factor 7-like 2 (TCF7L2) and potassium voltage-gated channel, Kv11.1-like subfamily, member 1 (KCNJ11) with ORs 1.3–1.4; reviewed by [4]) and more modest power (<70%) to detect effects for less common variants (MAF = 0.10). The power to detect and replicate moderate level contributions to T2DM susceptibility should increase with meta-analysis of this GWAS data and other GWAS currently being conducted in African-American populations. In addition this study reports results from only directly genotyped SNPs. Effective imputation of additional SNPs would undoubtedly improve coverage of the African-American genome.

While recent imputation methods development [42] show encouraging progress, rigorous empirical testing continues. A potential bias of the current study design may be that the GWAS was conducted in an African-American population of individuals with type 2 diabetes with nephropathy however; there is no specific reason why this African-American population should differ substantially from African Americans with T2DM without ESRD. For example, TCF7L2 is strongly associated in our studies of African-American T2DM-ESRD subjects [28,40]. In addition it should be noted that although every precaution was taken to account for population structure, as with any GWAS or candidate gene study, there may be residual population substructure. The major strength of this study is the genotyping and replication in four additional populations, thus providing support for the evidence of association observed. In addition, the study design which includes individuals with T2DM and ESRD allows for the identification of ESRD loci which are distinct from those presented herein (Table S10; [43]).

In conclusion, we have performed a GWAS for T2DM-ESRD in an African-American population from the southeastern United States. These results were then replicated in an additional sample recruited under identical ascertainment criteria. As a second stage of replication, a Validation study was carried out in three independent cohorts to confirm the association of suspected loci with T2DM. As a result, we have identified SNP rs7560163 that reached genome-wide significance and four additional loci (rs7650163; KCNQ1) with ORs 1.3–1.4; validated by [4]) and more modest power (<70%) to detect effects for less common variants (MAF = 0.10). The power to detect and replicate moderate level contributions to T2DM susceptibility should increase with meta-analysis of this GWAS data and other GWAS currently being conducted in African-American populations. In addition to these findings require stringent levels of genome-wide significance and four additional loci with more nominal evidence of association. These findings require additional replication in independent African-American populations as well as in additional ethnicities to confirm these findings and aid in the identification of the causal variant(s).

Materials and Methods

Ethics Statement

Recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University (GWAS, Replication, T2DM, IRAS and IRASFS samples) and Howard University (HUFs samples). Written informed consent was obtained from all study participants.

Subjects

Genome-Wide Association Study (GWAS) samples and clinical characteristics. Recruitment and sample collection procedures were approved by the Institutional Review Board at Wake Forest University and informed consent was obtained from all study participants. Patients with T2DM-ESRD were recruited from dialysis facilities. T2DM was diagnosed in African Americans who reported developing T2DM after the age of 25 and who did not receive only insulin therapy since diagnosis. In addition, cases had to have at least one of the following three criteria for inclusion: a) T2DM diagnosed at least 5 years before initiating renal replacement therapy, b) background or greater diabetic retinopathy and/or c) ≥100 mg/dl proteinuria on urinalysis in the absence of other causes of nephropathy (T2DM-ESRD cases). Unrelated African-American controls without a current diagnosis of diabetes or renal disease were recruited from the community and internal medicine clinics (controls). All T2DM-ESRD cases and controls lacking T2DM and ESRD were born in North Carolina, South Carolina, Georgia, Tennessee or Virginia. DNA extraction was performed using the PureGene system (Genta Systems; Minneapolis, MN).

Replication study samples and clinical characteristics. African-American T2DM-ESRD cases and controls lacking T2DM and ESRD were recruited using the same criteria as the case and control subjects that were used in the GWAS.

Validation study samples and clinical characteristics. T2DM Cases. Subjects with T2DM without evidence of nephropathy were recruited from medical clinics, churches, health fairs and community resources. Individuals were unrelated and self-described African Americans. All subjects were born in North Carolina, South Carolina, Georgia, Virginia or Tennessee. The PureGene system (Genta Systems; Minneapolis, MN) was used for DNA extraction. Controls. The Howard University Family Study (HUFs) is a population-based study of African-American families enrolled from the Washington, D.C. metropolitan area. Families were not ascertained based on a given phenotype. In a second phase of recruitment, additional unrelated individuals from the same geographic area were enrolled to facilitate a nested case-control study design. A total of 1,976 samples remained after data cleaning. Diagnosis of T2DM was based on the criteria established by the American Diabetes Association Expert Committee: a fasting plasma glucose concentration ≥126 mg/dL (7.0 mmol/l) or a 2h-postload value in the oral glucose tolerance test ≥200 mg/dL (11.1 mmol/l) on more than one occasion or receiving medication for T2DM. From this sample, a subset of 927 unrelated control individuals was used for analysis.

IRAS. The Insulin Resistance Atherosclerosis Study (IRAS) is a multicenter population-based cohort study that recruited men and women from 40 to 69 years of age living in four U.S. communities from 1992 to 1993 [44]. The study recruited approximately equal numbers of persons with normal glucose tolerance, impaired glucose tolerance and T2DM. Diabetes was defined by self-report or a fasting glucose measure >126 mg/dL at baseline or follow-up visits. The IRAS protocol was approved by local institutional review committees and all participants gave informed consent. IRASFS. Study design, recruitment and phenotyping for the IRAS Family Study (IRASFS) have been described [45]. Briefly, the IRASFS is a multicenter study designed to identify the genetic determinants of quantitative measures of glucose homeostasis. Members of large families of self-reported African Americans (n = 581 individuals in 42 pedigrees from Los Angeles, California) were recruited. Diabetes was defined by self-report, use of diabetes medications or fasting glucose measures >126 mg/dL at baseline or follow-up visits. The IRASFS protocol was approved by local institutional review committees and all participants gave informed consent.

Genotyping and Quality Control

GWAS. Genotyping was performed at the Center for Inherited Disease Research (CIDR) using 1 μg of genomic DNA (diluted in 1 x TLE buffer and at 30 ng/μl) on the Affymetrix Genome-wide Human SNP array 6.0. DNA from cases and controls were equally interleaved on 96-well master plates to ensure technical uniformity during sample processing. To confirm sample identity, a SNP barcode (96 SNPs) was generated prior to
genotyping on the Affymetrix arrays and confirmed on downstream released genotyping data. Genotypes were called using Birdseed version 2; APT 1.10.0 by grouping samples by DNA plate to determine the genotype cluster boundaries. All autosomal SNPs (n = 688,157) were included in analysis but classified on data quality with primary inference drawn from SNPs (n = 632,357) which had less than 5% missing data, Hardy-Weinberg $P$-values in cases greater than 0.0001 and in controls greater than 0.01, no significant difference in missing data rate between cases and controls and were polymorphic. The average sample call rate was 99.16% for all autosomal SNPs. Forty-six blind duplicates were included in genotyping and had a concordance rate of 99.59%. In addition, individuals whose gender call from X chromosome genotype data was discordant with the gender obtained from patient interviews were excluded from the analysis (n = 1). Cryptic relatedness was estimated by pairwise identity-by-descent (IBD) analysis implemented in the PLINK analysis software package (http://pngu.mgh.harvard.edu/~purcell/plink/). Two duplicate samples were identified, and one sample in each duplicate pair was removed. In addition, 104 individuals were identified as cryptic first degree relatives. We also assessed heterozygosity by estimating the inbreeding coefficient using PLINK. One subject had an F value >4 standard deviations; this excess of homozygosity would suggest population substructure and this subject was removed. Our final dataset consisted of 1994 individuals in which we performed the association analysis.

**Replication.** The replication sample consisted of a population recruited under identical ascertainment criteria to that of the GWAS. A total of 749 SNPs (including 272 SNPs captured in 104 linkage disequilibrium (LD) blocks defined by an $r^2$ > 0.50 at consecutive loci as assessed in 988 unrelated GWAS control subjects; 581 independent loci) were selected for genotyping on the Sequenom MassArray platform (Sequenom; San Diego, CA). Case and control samples were genotyped and analyzed together to avoid sample-dependent SNP calling bias. SNPs were included in analysis if genotyping was greater than 90% efficient, had a Hardy-Weinberg $P$-value $\geq 0.001$ in the replication cohort and were polymorphic (n = 712 SNPs, including 264 SNPs captured in 102 LD blocks defined by an $r^2$ > 0.50 at consecutive loci as assessed in 988 unrelated GWAS control subjects; 550 independent loci). Forty five blind duplicate samples included in genotyping had a concordance rate of >99.9%.

**Validation.** Among the 712 SNPs genotyped during the replication phase, 122 (including 41 SNPs captured in 17 linkage disequilibrium (LD) blocks defined by an $r^2$ > 0.50 at consecutive loci as assessed in 690 unrelated Replication control subjects; 98 independent loci) were genotyped using the iPLEX HD Sequenom MassARRAY platform (T2DM, IRAS and IRASFS) or on the Affymetrix Genome-wide Human SNP array 6.0 (Controls) for the validation phase. Genotyping was greater than 90% efficient and the 20 blind duplicate samples included in genotyping had a concordance rate of 100%.

**Analysis**

**GWAS.** To address the effect of admixture in this African-American dataset we performed a Principal Components Analysis (PCA) which utilized all high quality data from the GWAS excluding regions of high LD and inversions. This approach was an iterative process whereby all high quality autosomal SNPs were used to calculate the top 50 principal components. Once calculated, the principal components were examined to determine if they were tied to regions of the genome. If so, those SNPs were excluded and the analysis repeated. The first principal component (PC1) explained the largest proportion of variation at 22% and was used as a covariate in all analyses. A direct comparison of the PCA with FRAPPE [46] analysis of 70 ancestry informative markers (AIMs; [47]) resulted in a high correlation between PC1 and the AIMs ancestry estimates, $r^2 = 0.07$. The mean (SD) African ancestry proportion in 965 T2DM-ESRD cases and 1,029 controls was 0.50 ± 0.11 and 0.70 ± 0.13, respectively, as estimated by FRAPPE analysis. Other principal components were associated with regions of the genome, representing another unclassified source of variance. To test for association with T2DM-ESRD, genotypic tests of association were performed on each SNP individually using SNPGWA (http://www.sph.washington.edu/ [48]), an analytic package which includes the capability to perform association calculations adjusting for covariates. The primary inference was based on the additive genetic model; with note when there is strong evidence of a departure from additivity. The inflation factor was calculated as the observed mean of the chi squared statistic and compared to its theoretical expectation of 1 under the null hypothesis.

Imputation was performed for autosome using MACH (version 1.0.16, http://www.sph.umich.edu/csg/abecasis/MaCH/) to obtain missing genotypes for previously identified T2DM index variants and to provide support for regions associated with T2DM in the African American dataset. SNPs with minor allele frequency $\geq 1$, call rate $\geq 95%$ and Hardy–Weinberg $P$-value $\leq 10^{-4}$ were used for imputation. A 1:1 mixture of the HapMap II release 22 (NCBI build 36) CEU:YRI consensus haplotypes (http://math-gen.stats.ox.ac.uk/impute/) were used as a reference panel. Imputation was performed in two steps. For the first step, 484 unrelated African-American samples were randomly selected to calculate recombination and error rate estimates. In the second step, these rates were used to impute all samples across the SNPs in the entire reference panel. Imputation results were filtered at an rqs threshold of $\geq 0.3$ and a minor allele frequency $\geq 0.05$.

We examined previously identified T2DM loci for association with T2DM in the African American GWAS dataset. For SNPs not available on the Affymetrix 6.0 array or from direct genotyping (n = 10), genotypes were determined from imputation. In addition to the index variant, we identified the corresponding LD block using the HapMap phase II CEU data as defined by Gabriel et al. [49] and implemented in Haplovew. These intervals were then extracted from the African-American GWAS and the most significant SNP identified. These results were corrected for the effective number of SNPs (independent SNPs) in each locus counted using the Li and Ji method implemented in SOLAR [50]. The empirical locus-specific P-values were adjusted for multiple comparisons by Bonferroni correction for the effective number of SNPs (Table S1).

**Replcation in T2DM-ESRD cases and controls lacking T2DM and ESRD.** To account for admixture in the replication cohort, ancestral allele proportions were estimated by comparing allele frequencies to 70 AIMs [47] genotyped in 44 Yoruba Nigerians and 39 European Americans. Individual ancestral proportions were generated for each subject using FRAPPE [46], an EM algorithm-based approach, under a two-population model and used as covariates in all analyses. The mean (SD) African ancestry proportion in 709 T2DM-ESRD cases and 690 controls was 0.80 ± 0.12 and 0.76 ± 0.13, respectively. Association analysis was performed as described for the GWAS.

**Validation in T2DM, non-nephropathy cases and controls lacking T2DM and ESRD.** In order to discriminate between association with T2DM and T2DM-ESRD, meta-analysis of three additional association analyses was performed. For the T2DM population, individual admixture proportions were estimated by comparing allele frequencies from 76 AIMs genotyped on the
Sequenom MassArray (T2DM cases) or Affymetrix 6.0 array (controls) to frequencies reported in the HapMap CEU and YRI populations (unrelated samples only). Individual ancestral proportions were generated for each subject using FRAPPE [46] under a two-population model and used as covariates in all analyses. The mean (SD) African ancestry proportion in T2DM cases and controls was 0.78 ± 0.11 and 0.76 ± 0.12, respectively. Association analysis was performed as described for the GWAS. For the IRAS and IRASFS cohorts, ancestral allele frequencies were estimated using 70 AIMs genotyped in 44 Yoruba Nigerians and 39 European Americans. Individual ancestral proportions were generated for each subject using FRAPPE [46] under a two-population model and used as covariates in all analyses. For the IRASFS cohort, each SNP was examined for Mendelian inconsistencies using PedCheck [51]. Genotypes inconsistent with Mendelian inheritance were converted to missing. Maximum likelihood estimates of allele frequencies were computed using the largest set of unrelated African-American individuals (n = 58), and then genotypes were tested for departures from Hardy-Weinberg proportions. For the IRAS (unrelated individuals) and IRASFS (related individuals) cohorts, data was analyzed using a variance component measured genotype model [50]. To model T2DM as the outcome, a threshold model of the variance component measured genotype model was used. Likelihood ratio tests were computed for the tests of association with the individual SNP, modeling the correlation structure suggested by the familial relationships as appropriate, i.e. IRASFS. The family data has already been examined in detail and familial relationships corrected based on a linkage panel. P-values were calculated from the threshold model while the odds ratios were calculated from a logistic regression model.

**Meta-Analyses.** In order to perform GWAS + Replication, Validation (T2DM, IRAS and IRASFS) and Overall (GWAS + Replication + Validation) analyses a meta-analysis approach was taken. Meta-analysis was performed using the weighted Z-method implemented in METAL (www.sph.umich.edu/csg/abecasis/software/), using B-statistic was derived from the sample-specific P-value and direction of effect which were then summed with weights [52]. The Z-statistic was derived from Fisher’s test incompatibility between phenotype units as in the Fisher method [53], but with improved power and precision over Fisher’s test [54]. The Z-statistic was derived from the sample-specific P-values and directionality of effect which were then summed with weights proportional to the square root of the sample size for each sub-study.

**Quantitative Trait Analysis.** To test for association between individual SNPs and quantitative measures of glucose homeostasis in the IRAS and IRASFS cohorts, differences in trait values by genotype were tested using the variance components model that explicitly models the correlation among related individuals as implemented in SOLAR [12]. For statistical testing, trait values were transformed to best approximate the distributional assumptions of the test and to minimize heterogeneity of the variance. The primary statistical inference was the additive genetic model. All tests were computed after adjustment for age, gender, BMI and admixture adjustment.

**Exploration of eQTLs for T2DM loci**

To identify potential T2DM-susceptibility genes we explored association of the putative African-American T2DM loci with transcript levels for flanking genes using gene expression profiles from the publically available HapMap Yoruba (YRI) dataset [54]. Coupling the YRI expression dataset with genotypes from the most associated loci we explored the association of SNPs with flanking genes using the variance components model and accounting for correlation among related individuals as implemented in SOLAR [12].

**Supporting Information**

**Figure S1** Quantile-Quantile plot of the genome-wide association study results. (DOC)

**Figure S2** African-American T2DM candidate regions. A) rs7542900 region. B) rs4659485 region. C) rs7560163 region. D) rs2722769 region. E) rs7107217 region. −log_{10} additive P-value from the GWAS are plotted versus position (NCBI Build 36.1, hg18). The large red diamond indicates the additive P-value from the GWAS of the marker(s) displayed. The large blue diamond and corresponding P-value indicates the additive P-value from the Overall analysis of the marker(s) displayed. r² based on the control samples is color-coded with respect to the most significant SNP: red (0.8–1.0), orange (0.5–0.8), yellow (0.2–0.5) and white (<0.2). Gene annotations were obtained from UCSC Genome Browser (RefSeq Genes, b36). Arrows represent direction of transcription. (DOC)

**Table S1** GWAS P-values for previously associated T2DM loci. Loci are ordered by chromosome and position (NCBI Build 36.1, hg18) and referenced (Ref) by the initial publication. The African American major/minor alleles are presented on the positive strand with the Caucasian risk allele underlined. For each T2DM Index SNP, results from the African American GWAS (the minor allele frequency (MAF) for the T2DM-ESRD and control populations or combined for imputed SNPs with the corresponding additive P-value and odds ratio (OR) with associated 95% confidence interval (CI)) are presented with respect to the published risk allele (underlined). In addition, association results (additive P-value and odds ratio (OR) with associated 95% confidence interval (CI)) from recent Caucasian large-scale meta-analyses with associated references (Ref) are listed for comparison. For each index SNP, the corresponding LD block was identified using the HapMap phase II CEU data as defined by Gabriel et al. and implemented in Haploview. These intervals were then extracted from the African-American GWAS and the most significant SNP listed. From the GWAS, the minor allele frequency (MAF) for the T2DM-ESRD and control populations are listed with the corresponding additive P-value (nominal and corrected for the effective number of tests at the locus (number of SNPs genotyped in the GWAS and effective number of SNPs determined from the Li and Ji method and implemented in SOLAR)) and odds ratio (OR) with associated 95% confidence interval (CI) with respect to the African-American minor allele. (DOC)

**Table S2** GWAS, Replication, T2DM, IRAS and IRASFS P-values for 122 GWAS SNPs genotyped on replication and validation samples. SNPs are ordered by chromosome and position (NCBI Build 36.1) with the major/minor alleles (positive strand). For each cohort, minor allele frequency (MAF) for case and control populations are listed with the reference allele (minor allele) and corresponding additive P-value and odds ratio (OR) with associated 95% confidence interval (CI) with respect to the minor allele. Note: For IRAS-FS MAFs are derived from the overall sample including relatives. In addition, allele frequencies has been extracted from HapMap Yoruba (YRI) and CEPH (CEU) samples for comparison. Rows in red type represent the five loci which are the focus of the manuscript. (XLS)
Table S5  Quantitative trait meta-analysis for African-American T2DM loci across the genome. SNPs are ordered by chromosome and position (NCBI Build 36.1, hg18) with the major/minor alleles (positive strand) and corresponding gene (underlined) or nearest annotated gene. For each phase of the study, GWAS + Replication, Validation and Overall analyses, the additive $P$-value and odds ratio (OR) with associated 95% confidence interval (CI) are listed.

Table S6  Association results for African-American T2DM loci in the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC). SNPs are ordered by chromosome and position (NCBI Build 36.1, hg18) with the major/minor alleles (positive strand) and corresponding gene (underlined) or nearest annotated gene. For each phase of the study, GWAS + Replication, Validation and Overall analyses, the additive $P$-value and odds ratio (OR) with associated 95% confidence interval (CI) are listed.

Table S7  Gender stratified association analysis with T2DM. SNPs are ordered by chromosome and position (NCBI Build 36.1, hg18) with the major/minor alleles (positive strand) and corresponding gene (underlined) or nearest annotated genes (+/-500 kb). For males and females, the additive $P$-value and odds ratio (OR) with associated 95% confidence interval (CI) with respect to the minor allele and heterozygosity $P$-value are listed.

Table S8  Power Calculations. Table S8a. Genome-wide association study power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 965 cases and 1029 controls. Table S8b, Replication power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 709 cases and 690 controls. Table S8c. GWAS + Replication sample power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 1719 controls. Table S8d. T2DM power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 1246 cases and 927 controls. Table S8e. IRAS power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 115 cases and 164 controls. Table S8f. IRASFS power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 97 cases and 507 controls. Table S8g. Validation meta-analysis power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 1458 cases and 1598 controls. Table S8h. Overall power analysis for causal variant in complete and incomplete linkage disequilibrium with a typed variant given minor allele frequency ($p$) in 3132 cases and 3317 controls.

Acknowledgments

We wish to thank the patients, their relatives and staff of the Southeastern Kidney Council, Inc./ESRD Network 6 for their participation. We wish to thank the patients, their relatives and staff of the Southeastern Kidney Council, Inc./ESRD Network 6 for their participation.
References


Author Contributions

Conceived and designed the experiments: NDP CWM DWB MMS. Performed the experiments: NDP CWM DWB MMS. Analyzed the data: NDP DWB MMS. Wrote the paper: NDP CWM DWB MMS. Provided consortium data for replication: DIAGRAM Consortium, MAGIC Investigators. Reviewed the manuscript: AF.