Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes

Maggie C. Y. Ng, Wake Forest School of Medicine
Daniel Shriner, National Human Genome Research Institute
Brian H. Chen, University of California Los Angeles
Jiang Li, Wake Forest School of Medicine
Wei-Min Chen, University of Virginia
Xiuqing Guo, Harbor-UCLA Medical Center
Jiankang Liu, University of Mississippi
Suzette J. Bielinski, Mayo Clinic
Lisa R. Yanek, Johns Hopkins University
Michael A. Nalls, National Institute on Aging

Only first 10 authors above; see publication for full author list.

Journal Title: PLoS Genetics
Volume: Volume 10, Number 8
Publisher: Public Library of Science | 2014-08-01, Pages e1004517-e1004517
Type of Work: Article | Final Publisher PDF
Publisher DOI: 10.1371/journal.pgen.1004517
Permanent URL: https://pid.emory.edu/ark:/25593/s2d05

Final published version: http://dx.doi.org/10.1371/journal.pgen.1004517

Copyright information:
This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration, which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. This is an Open Access work distributed under the terms of the Creative Commons Universal : Public Domain Dedication License (http://creativecommons.org/publicdomain/zero/1.0/).

Accessed September 11, 2020 12:31 PM EDT
Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes


1 Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 2 Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 3 Center for Research on Genomics and Global Health, National Human Genome Research Institute, Bethesda, Maryland, United States of America, 4 Program on Genomics and Nutrition, School of Public Health, University of California Los Angeles, Los Angeles, California, United States of America, 5 Center for Metabolic Disease Prevention, School of Public Health, University of California Los Angeles, Los Angeles, California, United States of America, 6 Center for Public Health Genomics, University of Virginia, Charlottesville, Virginia, United States of America, 7 Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia, United States of America, 8 Institute for Translational Genomics and Population Sciences, Los Angeles BioMedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, United States of America, 9 Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, United States of America, 10 Division of Epidemiology, Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, United States of America, 11 The GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 12 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, United States of America, 13 Center for Public Health Genomics, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 14 Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 15 Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America, 16 Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, United States of America, 17 Department of Medicine, University of Washington, Seattle, Washington, United States of America, 18 San Francisco Coordinating Center, California Pacific Medical Center Research Institute, San Francisco, California, United States of America, 19 Department of Epidemiology and Biomedical Informatics, Emory University, Atlanta, Georgia, United States of America, 20 Division of Statistical Genomics, Washington University School of Medicine, St. Louis, Missouri, United States of America, 21 Division of Sleep Medicine, Brigham and Women’s Hospital, Boston, Massachusetts, United States of America, 22 The Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, 23 The Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, United States of America, 24 Division of Epidemiology, Department of Medicine, Vanderbilt Epidemiology Center, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, 25 Division of Endocrinology, Metabolism and Molecular Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, United States of America, 26 Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 27 Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 28 Department of Biology, Center for Health Disparities, East Carolina University, Greenville, North Carolina, United States of America, 29 Department of Internal Medicine, Wake Forest School of Medicine, Winston-Salem, North Carolina, United States of America, 30 Medical Genetics Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, United States of America, 31 Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi, United States of America, 32 Jackson State University, Tougaloo College, Jackson, Mississippi, United States of America, 33 Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States of America, 34 Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, United States of America, 35 Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, United States of America, 36 Division of Allergy and Clinical Immunology, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, United States of America, 37 Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, United States of America, 38 Laboratory of Personality and Cognition, National Institute on Aging, National Institutes of Health, Baltimore, Maryland, United States of America, 39 Department of Epidemiology and Biostatistics, Case
Abstract

Type 2 diabetes (T2D) is more prevalent in African Americans than in Europeans. However, little is known about the genetic risk in African Americans despite the recent identification of more than 70 T2D loci primarily by genome-wide association studies (GWAS) in individuals of European ancestry. In order to investigate the genetic architecture of T2D in African Americans, the META-analysis of type 2 Diabetes in African Americans (MEDIA) Consortium examined 17 GWAS on T2D comprising 8,284 cases and 15,543 controls in African Americans in stage 1 analysis. Single nucleotide polymorphisms (SNPs) association analysis was conducted in each study under the additive model after adjustment for age, sex, study site, and principal components. Meta-analysis of approximately 2.6 million genotyped and imputed SNPs in all studies was conducted using an inverse variance-weighted fixed effect model. Replications were performed to follow up 21 loci in up to 6,061 cases and 5,483 controls in African Americans, and 8,130 cases and 38,987 controls of European ancestry. We identified three known loci (TCF7L2, HMG2A and KCNQ1) and two novel loci (HLA-B and INS-IGF2) at genome-wide significance (4.15 × 10^{-8} < P < 5.0 × 10^{-8}, odds ratio (OR) = 1.09 to 1.36). Fine-mapping revealed that 88 of 158 previously identified T2D or glucose homeostasis loci demonstrated nominal to highly significant association (2.2 × 10^{-3} < \text{locus-wide } P < 0.05). These novel and previously identified loci yielded a sibling relative risk of 1.19, explaining 7.5% of the phenotypic variance of T2D on the liability scale in African Americans. Overall, this study identified two novel susceptibility loci for T2D in African Americans. A substantial number of previously reported loci are transferable to African Americans after accounting for linkage disequilibrium, enabling fine mapping of causal variants in trans-ethnic meta-analysis studies.


Editor: Eleftheria Zeggini, Wellcome Trust Sanger Institute, United Kingdom

Received January 21, 2014; Accepted June 5, 2014; Published August 7, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

Funding: Atherosclerosis Risk in Communities Study (ARIC) is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute contracts (HHSN26820100005C, HHSN26820100006C, HHSN26820100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL06694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN26820062526C. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research. Funding for Coronary Risk and Disease (CARDIA) include support to University of Alabama at Birmingham (U01-HC-48047), University of Minnesota (U01-HC-48048), Northwestern University (U01-HC-48049), Kaiser Foundation Research Institute (U01-HC-48050), University of Alabama at Birmingham (U01-HC-95095), Tufts-New England Medical Center (U01-HC-45224), Wake Forest School of Medicine (U01-HC-45205), Harbor-UCLA Research and Education Institute (U01-HC-45187), University of California Irvine (U01-HC-45134, U01-HC-95100). Funding to Candidate-gene Association Resource (CAR) include support to University of Technology - Broad Institute (U01-HC-65226). Cleveland Family Study (CFS) was supported by a grant to Case Western Reserve University (NIH HL 46380, N01RR00808); Cardiovascular Health Study (CHS) was supported by NHLBI contracts HL085351, HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL105756, HL103612, and HL120393 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023692 from the National Institute on Aging (NIA); A full list of principal CHS investigators and institutions can be found at CHS- NINDS. The provision of genotyping data was supported in part by National Human Genome Research Institute contract U01HG003747 and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. DNA handling and genotyping was supported in part by National Center of Advancing Translational Technologies CTSA grant UL1TR000124, the National Institute of Diabetes and Digestive and Kidney Diseases grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The Electronic Medical Records and Genomics Network (eMERGE) Network was initiated and funded by NHLBI, with additional funding from NIGMS through the following grants: U01-HG-004610 (Group Health Cooperative); U01-HG-004608 (Marshfield Clinic); U01-HG-004599 (Mayo Clinic); U01HG004609 (Northwestern University); U01-HG-00463 (Vanderbilt University, also serving as the Coordinating Center), and the State of Washington Life Sciences Discovery Fund award to the Northwest Institute of Medical Genetics. The Northwestern University Enterprise Data Warehouse was funded in part by a grant from the National Research Resources, UL1TR000124. The data analyses described were obtained from Vanderbilt University Medical Center’s BioVU which is supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. Family Heart Study (FamHS) was supported by NHLBI grants R01-HL-087700 and R01-HL-088215 (MAP, Principal Investigator) from NHLBI; and R01-DK-8925601 and R01-DK-075881 (IBB, Principal Investigator) from NIDDK. Family Investigation of Nephropathy in Diabetes (FIND) was supported by FIND grants U01DK57292, U01DK57329, U01DK057306, U01DK057298, U01DK057249, U01DK057295, U01DK070657, U01DK057303, and U01DK070657, U01DK057304. This project has been funded in whole or in part with federal funds.
from the NIH National Cancer Institute (NCI) under contract N01-CO-12400 and the Intramural Research Program of the NIH-NCI Center for Cancer Research. This work also was sponsored by the National Center for Research Resources for the General Clinical Research Center grants: Case Western Reserve University, M01 RR00089; Wake Forest University, M01 RR-07122; Harbor-University of California, Los Angeles Medical Center, M01 RR-04245; College of Medicine, University of California, Irvine, M01 RR-00827-29; University of New Mexico, HSC, M01 RR-00997; and Frederic C. Bartter, M01 RR-01346. The CHOICE Study was supported in part by HS05365 from the Agency for Healthcare Research and Quality and Pool, MD, and HL02985 from the National Heart, Lung, and Blood Institute, Bethesda, MD. Genetic Study of Atherosclerosis Risk (GenSTAR) was supported by grants from the National Heart, Lung, and Blood Institute (HL58625-01A1, HL59684, HL071025-01A1, U01HL72518, and HL07698) and the National Institute of Nursing Research (NR0224103) and by M01-RR000052 to the Johns Hopkins General Clinical Research Center. Genetic Epidemiology Network of Atherosclerotic Youth (GENOA) study is supported by the National Heart, Lung, and Blood Institute (HL087660, HL100245, and HL100185 from the National Heart, Lung, and Blood Institute. Healthy Aging in Neighborhoods of Diverse Communities (HANDLS) was supported by the Intramural Research Program of the NIH, National Institute on Aging and the National Center on Minority Health and Health Disparities (project # C01 AG000053 and intramural protocol # 2009-149). Health ABC Study was supported by NIA contracts N01-AG1-2103, N01-AG1-2105, and N01AG62106. The genome-wide association study was funded by NIA grant RO1AG032098-01A1 to Wake Forest University Health Sciences and genotyping services were provided by the Center for Inherited Disease Research (CIDR). CIDR is fully funded through a federal contract from the National Institutes of Health to The Johns Hopkins University, contract number HS026692.01A (Phase II release 21–24 CEU S2). Missing genotypes in individual studies were imputed to one of the HapMap reference panels (Phase II release 21–24 CEU+YRI, Phase II release 22 all populations, Phase II+III release 27 CEU+YRI, Phase II+III release 27 CEU+YRI+ASW or Phase II+III release 27 all populations) using MACH, IMPUTE2 or BEAGLE (Table S3). Genomic control corrections [11] were applied to each study (\(\lambda = 1.01–1.08\)) and after meta-analysis (\(\lambda = 1.06\)) due to modest inflated association results (Table S3) [12]. Association results for \(\sim 2.6\) SNPs were subsequently examined.

From stage 1 meta-analysis, 49 SNPs moderately associated with T2D (\(P < 10^{-5}\)) and two candidate SNPs near the p value threshold \((rs231356 at KCNQ1, P = 2.84 \times 10^{-5}\) and rs2244020 at HLA-B, \(P = 1.02 \times 10^{-5}\)) totaling 51 SNPs in 21 loci were followed up for replication. rs231356 is 14 kb downstream of the reported T2D index SNP, rs231362, in Europeans [3]. Moderate associations have also been observed across the HLA region in Europeans [3]. The stage 2 replication included \(m\) silico and \(de\) novo replication in up to 11,544 African American T2D cases and controls, as well as in \(m\) silico replication in 47,117 individuals of European ancestry from DIAGRAM v2 [3] (Table S4). Meta-analyses were performed to combine results from African Americans (stage 1+2a, n=35,371, Table S4) and both African Americans and Europeans (stage 1+2a+2b, n=82,468, Table S4).

**T2D loci reaching genome-wide significance**

Five independent loci reached genome-wide significance (\(P<5 \times 10^{-8}\)). Stage 1 meta-analysis identified the established TCF7L2 locus. Stage 1+2a meta-analysis identified the established KCNQ1 and HMGA2 loci. Stage 1+2a+2b meta-analysis identified a second signal at KCNQ1 and a novel HLA-B locus. Secondary analysis including body mass index (BMI) adjustment in stage 1+2a meta-analysis identified the second novel locus at INS-IGF2 (Table 1 and Figure 1). None of the most strongly associated SNPs at these loci demonstrated significant heterogeneity of effect sizes among studies within each stage, between African Americans in

---

**Introduction**

The prevalence of type 2 diabetes (T2D) among adults in the USA is currently 11.3%, with substantially higher prevalence in African Americans (18.7%) than in European Americans (10.2%) [1]. To date, genome-wide association studies (GWAS) have identified >70 susceptibility loci for T2D [2–8]. While it is known that T2D is heritable in African Americans [9], it is unclear how much heritability is explained by the known genetic associations identified. While it is known that T2D is heritable in African Americans [9], it is unclear how much heritability is explained by the known genetic associations identified. While it is known that T2D is heritable in African Americans [9], it is unclear how much heritability is explained by the known genetic associations identified.

---

**Results**

**Study overview**

We conducted a meta-analysis of 17 African American GWAS on T2D comprising 8,284 cases and 15,543 controls (Tables S1 and S2). Missing genotypes in individual studies were imputed to one of the HapMap reference phases (Phase II release 21–24 CEU+YRI, Phase II release 22 all populations, Phase II+III release 27 CEU+YRI, Phase II+III release 27 CEU+YRI+ASW or Phase II+III release 27 all populations) using MACH, IMPUTE2 or BEAGLE (Table S3). Genomic control corrections [11] were applied to each study (\(\lambda = 1.01–1.08\)) and after meta-analysis (\(\lambda = 1.06\)) due to modest inflated association results (Table S3) [12]. Association results for \(\sim 2.6\) SNPs were subsequently examined.

From stage 1 meta-analysis, 49 SNPs moderately associated with T2D (\(P < 10^{-5}\)) and two candidate SNPs near the p value threshold \((rs231356 at KCNQ1, P = 2.84 \times 10^{-5}\) and rs2244020 at HLA-B, \(P = 1.02 \times 10^{-5}\)) totaling 51 SNPs in 21 loci were followed up for replication. rs231356 is 14 kb downstream of the reported T2D index SNP, rs231362, in Europeans [3]. Moderate associations have also been observed across the HLA region in Europeans [3]. The stage 2 replication included \(m\) silico and \(de\) novo replication in up to 11,544 African American T2D cases and controls, as well as in \(m\) silico replication in 47,117 individuals of European ancestry from DIAGRAM v2 [3] (Table S4). Meta-analyses were performed to combine results from African Americans (stage 1+2a, n=35,371, Table S4) and both African Americans and Europeans (stage 1+2a+2b, n=82,468, Table S4).

---

**Type 2 Diabetes GWAS in African Americans**

**T2D loci reaching genome-wide significance**

Five independent loci reached genome-wide significance (\(P<5 \times 10^{-8}\)). Stage 1 meta-analysis identified the established TCF7L2 locus. Stage 1+2a meta-analysis identified the established KCNQ1 and HMGA2 loci. Stage 1+2a+2b meta-analysis identified a second signal at KCNQ1 and a novel HLA-B locus. Secondary analysis including body mass index (BMI) adjustment in stage 1+2a meta-analysis identified the second novel locus at INS-IGF2 (Table 1 and Figure 1). None of the most strongly associated SNPs at these loci demonstrated significant heterogeneity of effect sizes among studies within each stage, between African Americans in
Author Summary

Despite the higher prevalence of type 2 diabetes (T2D) in African Americans than in Europeans, recent genome-wide association studies (GWAS) were examined primarily in individuals of European ancestry. In this study, we performed meta-analysis of 17 GWAS in 8,284 cases and 15,543 controls to explore the genetic architecture of T2D in African Americans. Following replication in additional 6,061 cases and 5,483 controls in African Americans, and 8,130 cases and 38,987 controls of European ancestry, we identified two novel and three previous reported T2D loci reaching genome-wide significance. We also examined 158 loci previously reported to be associated with T2D or regulating glucose homeostasis. While 56% of these loci were shared between African Americans and the other populations, the strongest associations in African Americans are often found in nearby single nucleotide polymorphisms (SNPs) instead of the original SNPs reported in other populations due to differential genetic architecture across populations. Our results highlight the importance of performing genetic studies in non-European populations to fine map the causal genetic variants.

stages 1 and 2a, or between African Americans in stage 1+2a and Europeans in stage 2b after Bonferroni correction of multiple comparisons (Pcorr > 0.001) (Figure S1).

At the TCF7L2 locus, the most strongly associated SNP in stage 1+2a African Americans samples was rs7903146 (OR = 1.33, P = 4.78 × 10⁻⁴⁴, Table 1 and Figure 2), rs7903146 is also the index SNP (most significantly associated with T2D in prior studies) in Europeans (OR = 1.40, P = 2.21 × 10⁻⁴⁵) [3], South Asians (OR = 1.25, P = 3.4 × 10⁻¹⁵) [4] and East Asians (OR = 1.48, P = 2.44 × 10⁻¹⁵) [13].

Two association signals were observed at KCNQ1 (Table 1 and Figure 2). The first association signal was represented by rs2283228 located at the 3' end of KCNQ1 (stage 1+2a OR = 1.20, P = 9.90 × 10⁻¹¹; stage 1+2a+2b OR = 1.19, P = 4.87 × 10⁻¹³). Using data from individuals of African ancestry in Southwest USA (ASW) from the 1000 Genomes Project (1KG) [14], rs2283228 mapped to the same linkage disequilibrium (LD)-based interval as index SNPs from other populations (rs2283228 [15] and rs2237892 [16-17] in Japanese, rs2237892 in Hispanics [18], rs163182 [19]and rs2237895 [20] in Han Chinese). The second association signal was represented by rs231356 (rs² = 0 with rs2283228 in both ASW and CEU) (stage 1+2a OR = 1.11, P = 1.94 × 10⁻⁰⁷; stage 1+2a+2b OR = 1.09, P = 3.93 × 10⁻⁰⁸), located 144 kb upstream of the first signal, rs231356 is located at the same LD interval as the index SNPs rs231362 in Europeans [3] and rs231359 in Chinese [20].

At the HMGAI2 locus, the most strongly associated SNP was rs343092 (stage 1+2a OR = 1.16, P = 8.79 × 10⁻⁵; stage 1+2a+2b OR = 1.14, P = 2.75 × 10⁻¹², Table 1 and Figure 2), rs343092 is located 76 kb downstream and at the same LD interval as of the index SNP rs1531343 reported in Europeans [3].

Two novel T2D loci were identified. The effect sizes of rs2244020 located near HLA-B were similar in African Americans and Europeans (OR = 1.11 vs. 1.07, Pcorr = 0.26; stage 1+2a+2b P = 6.57 × 10⁻⁸) (Table 1 and Figure 2). HLA-B encodes the class I major histocompatibility complex involved in antigen presentation in immune responses.

The most strongly associated SNP near INS-IGF2 was rs3142770 in African Americans (OR = 1.14, P = 2.73 × 10⁻⁸, stage 1+2a BMI adjusted, Table 1 and Figure 2) but the risk A allele was absent in the CEU population. Insulin plays a key role in glucose homeostasis. Mutations at INS lead to neonatal diabetes, type 1 diabetes, and hyperinsulinemia [21]. Insulin-like growth factor 2 (IGF2) is involved in growth and development. IGF2 overexpression in transgenic mice leads to insulin resistance [22] and IGF2 deficiency in the Goto-Kakizaki rat leads to beta cell mass anomaly [23].

Associations at previously reported T2D and glucose homeostasis loci

We investigated index SNPs from 158 independent loci associated with T2D and/or glucose homeostasis from prior genome-wide and candidate gene studies in individuals of European, East Asian, South Asian, or African American ancestry (Table S5). Among the 104 T2D-associated index SNPs, 19 were associated with T2D in stage 1 African American samples (P < 0.05). Most of the 17 T2D-associated SNPs that showed consistent direction of effects had similar effect sizes between this study and prior reports, despite that rs10440833 at CDKAL1 had substantially stronger effect size in Europeans (OR = 1.25) than in African Americans (OR = 1.06, Pcorr = 5.86 × 10⁻⁶). Additionally, 3 out of 54 trait-increasing alleles from glucose homeostasis-associated index SNPs were associated with increased T2D risk in African Americans (P < 0.05).

We also performed a locus-wide analysis to test for associations of all SNPs within the LD region at r² ≥ 0.3 with the previously reported index SNPs and results were corrected for the effective number of SNPs [24]. Since the causal variant(s) at each locus may be different or reside on different haplotypes across populations with different LD structures, this approach allows the identification of the most strongly associated SNPs in African Americans that may or may not be in LD with the index SNPs reported in other populations. A total of 55 T2D- and 29 glucose-associated loci were associated with T2D in African Americans (Pcorr < 0.05, corrected for LD in ASW for SNPs within a locus; Table S6). We compared the genetic architecture between the previously reported index SNPs and our fine-mapped SNPs for these 84 loci. The respective average risk allele frequencies were 0.51 and 0.46, and the distributions or pairwise differences of risk allele frequencies were not significantly different (P = 0.255, Wilcoxon rank sum test; and P = 0.295, Wilcoxon signed-rank test, respectively, Figure S2). In contrast, the average odds ratios for the risk alleles were higher for the fine-mapped SNPs as compared to the index SNPs (1.14 vs. 1.05). The distributions and pairwise differences of risk allele odds ratios were significantly different (P = 1.18 × 10⁻¹⁰ and 5.55 × 10⁻¹⁴, respectively, Figure S2). Thus, the locus-wide analysis identified variants with larger effect sizes and similar allele frequencies.

We leveraged differences in LD between African Americans and Europeans to fine-map and re-annotate several established loci. The association signal spanning ~100 kb at INTS8 in African Americans overlapped the ~200 kb TPF3INP1 T2D locus in Europeans [3]. The most strongly associated SNP in MEDIA tended to have larger effect size in African Americans than in Europeans (rs17359493, OR = 1.13 vs. 1.06, P = 1.39 × 10⁻⁷ vs. 3.20 × 10⁻⁸, respectively, Pcorr = 0.06) (Table S4). However, rs17359493 at intron 10 of INTS8 was only in weak LD with the reported index SNP rs896854 in Europeans (r² = 0.21 in CEU, 0.10 in ASW). Neither the reported index SNP rs896854 nor its proxies from the CEU data demonstrated significant association to T2D in African Americans (Table S6 and Figure S3a, b), suggesting that rs17359493 may be an independent novel signal. INTS8 encodes a subunit of the integrator complex which is involved in the cleavage of small nuclear RNAs. At KCNQ1, the most strongly associated SNP rs231356 was in weak LD with the
Table 1. Novel and previously identified loci associated with T2D at $P<5\times10^{-8}$.

<table>
<thead>
<tr>
<th>Loci</th>
<th>Chr</th>
<th>Position (Build 36)</th>
<th>SNP</th>
<th>Allelesa</th>
<th>RAFb</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>$P_{\text{het}}$</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>$P_{\text{het}}$</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>$P_{\text{het}}$</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>$P_{\text{het}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Previously identified T2D loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCF7L2c</td>
<td>10</td>
<td>114748339</td>
<td>rs7903146</td>
<td>T/C</td>
<td>0.30</td>
<td>1.32</td>
<td>(1.25–1.4)</td>
<td>6.62E-24</td>
<td>1.81E-01</td>
<td>1.34</td>
<td>(1.26–1.43)</td>
<td>8.38E-20</td>
<td>6.01E-03</td>
<td>1.33</td>
<td>(1.28–1.39)</td>
<td>4.78E-44</td>
<td>7.34E-01</td>
</tr>
<tr>
<td>KCNQ1c</td>
<td>11</td>
<td>2661919</td>
<td>rs231356</td>
<td>T/A</td>
<td>0.27</td>
<td>1.14</td>
<td>(1.07–1.21)</td>
<td>2.84E-05</td>
<td>9.11E-01</td>
<td>1.05</td>
<td>(0.98–1.14)</td>
<td>1.68E-01</td>
<td>3.26E-01</td>
<td>1.11</td>
<td>(1.06–1.16)</td>
<td>1.94E-05</td>
<td>1.08E-01</td>
</tr>
<tr>
<td>KCNQ1c</td>
<td>11</td>
<td>2806106</td>
<td>rs2283228</td>
<td>A/C</td>
<td>0.89</td>
<td>1.22</td>
<td>(1.14–1.31)</td>
<td>6.10E-08</td>
<td>9.48E-02</td>
<td>1.17</td>
<td>(1.06–1.28)</td>
<td>1.04E-03</td>
<td>7.10E-01</td>
<td>1.2</td>
<td>(1.14–1.27)</td>
<td>9.90E-11</td>
<td>4.34E-01</td>
</tr>
<tr>
<td>HMGA2c</td>
<td>12</td>
<td>64537207</td>
<td>rs343092</td>
<td>T/G</td>
<td>0.81</td>
<td>1.16</td>
<td>(1.09–1.24)</td>
<td>1.91E-06</td>
<td>9.48E-01</td>
<td>1.15</td>
<td>(1.04–1.26)</td>
<td>3.99E-03</td>
<td>3.37E-01</td>
<td>1.16</td>
<td>(1.1–1.22)</td>
<td>8.79E-09</td>
<td>7.93E-01</td>
</tr>
<tr>
<td><strong>Newly identified T2D loci</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HLA-Bd</td>
<td>6</td>
<td>31455430</td>
<td>rs2244020</td>
<td>G/A</td>
<td>0.69</td>
<td>1.12</td>
<td>(1.06–1.17)</td>
<td>1.02E-05</td>
<td>2.11E-02</td>
<td>1.1</td>
<td>(0.98–1.22)</td>
<td>1.01E-01</td>
<td>1.11</td>
<td>(1.07–1.16)</td>
<td>1.14E-06</td>
<td>7.57E-01</td>
<td>1.07</td>
</tr>
<tr>
<td>INS-IGFBP3d</td>
<td>11</td>
<td>2135246</td>
<td>rs3842770</td>
<td>A/G</td>
<td>0.23</td>
<td>1.18</td>
<td>(1.11–1.25)</td>
<td>8.18E-08</td>
<td>7.16E-01</td>
<td>1.07</td>
<td>(0.99–1.16)</td>
<td>8.09E-02</td>
<td>7.16E-01</td>
<td>1.14</td>
<td>(1.09–1.19)</td>
<td>2.78E-08</td>
<td>7.37E-02</td>
</tr>
</tbody>
</table>

Abbreviations: Chr, chromosome; SNP, single nucleotide polymorphism; RAF, risk allele frequency; OR, odds ratio for risk allele; CI, confidence interval; $P_{\text{het}}$, heterogeneity $P$ value.
aAlleles are ordered as risk allele/other allele aligned to the forward strand of NCBI Build 36.
bRisk allele frequency in Stage 1 samples.
cAssociations were performed with adjustment for age, sex, study sites, and study-specific principal components.
dAssociations were performed with adjustment for age, sex, study sites, study-specific principal components and body mass index.
P<5×10$^{-8}$ are in bold.
doi:10.1371/journal.pgen.1004517.t001
Effect of obesity on T2D susceptibility loci

We investigated the influence of obesity by comparing the stage 1 meta-analysis results with or without adjustment for BMI at the 51 most significantly associated SNPs from the GWAS for follow up (Tables S4 and S7) and 158 established T2D or glucose homeostasis index SNPs (Table S5). Association results were highly similar with and without BMI adjustment (correlation coefficients were 0.99 for both effect sizes and $-\log P$ values). Of particular note, FTO is suggested to influence T2D primarily through modulation of adiposity in Europeans [3,25], but evidence is contradictory across multiple ethnic groups [26–28]. The index SNP rs116142841 was not significantly associated with T2D in African Americans without and with BMI adjustment ($P=0.06$ and 0.23, respectively) (Table S5). The frequency of the risk A allele was 0.13 in this study. It had 100% power to detect association at the reported OR of 1.13 at type 1 error rate of 0.05, suggesting that FTO is unlikely a key T2D susceptibility gene in African Americans.

Gene expression and bioinformatics analyses

Among the six genome-wide significant loci (Table 1), we found no coding variants in the most significantly associated SNPs or their proxies. These SNPs demonstrated only weak associations with expression quantitative trait loci (eQTLs) ($P>0.001$, Table S8). Examination of the ENCODE data [29] revealed that several SNPs at TCF7L2, KCNQ1, and HMGA2 were located at protein binding sites or were predicted to alter motif affinity for transcription factors implicated in energy homeostasis (Table S9). The most strongly associated SNP rs13791346 in TCF7L2 is predicted to alter the binding affinity for a POU3F2 regulatory motif [30], POU3F2 is a neural transcription factor that enhances the activation of genes regulated by corticotropin-releasing hormone which stimulates adrenocorticotropic hormone (ACTH). ACTH is synthesized from pre-pro-opiomelanocortin (pre-POMC) which regulates energy homeostasis. For the 3’ signal at KCNQ1, several tag SNPs are predicted to alter the binding affinity for regulatory motifs, including SREBP, CTCF and HNF4A. SREBP is a transcription factor involved in sterol biosynthesis. CTCF regulates the expression of IGFI2 [31], HNF4A is a master regulator of hepatocyte and islet transcription. The tag SNP rs2257883 at HMGA2 is predicted to alter the binding affinity of ME2, which regulates GLUT4 transcription in insulin responsive tissues [32].

Discussion

We have performed the largest genetic association analysis to date for T2D in African Americans. Our data support the hypothesis that risk for T2D is partly attributable to a large number of common variants with small effects [7]. We identified HLA-B and INS-IGF2 as novel T2D loci, the latter specific to African Americans. We found evidence supporting association for 88 previously identified T2D and glucose homeostasis loci. Taken together, these 90 loci yielded a sibling relative risk of 1.19. The phenotypic variance measured on the liability scale is substantially larger in African Americans than in European Americans (17.5% vs. 5.7%) [7] due to larger effect sizes upon fine-mapping as well as higher disease prevalence in African Americans.

The two novel T2D loci, HLA-B and INS-IGF2, have been implicated in type 1 diabetes (T1D) risk in Europeans [33–35]. One limitation of our study is the lack of autoantibody measurement. However, our results are unlikely to be confounded by the presence of misclassified patients. Among diabetic youth
aged <20 years, T2D characterized by insulin resistance without autoimmunity is more prevalent in African Americans (40.1%) than in European Americans (6.2%), while African Americans less often present with autoimmunity and insulin deficiency resembling T1D compared to European Americans (32.5% vs. 62.9%, respectively) [36]. Autoimmunity is also uncommon in African American diabetic adults [37]. Furthermore, associations for T1D are stronger at HLA class II (HLA-DRB1, -DQA1, and -DQB1) than HLA class I regions in Europeans [33–34,38–41] (http://www.t1dbase.org). In African Americans, T1D individuals showed both shared and unique risk and protective HLA class II haplotypes as compared to European T1D individuals [42–43].

Figure 2. Regional plots of five previously and newly identified T2D loci in African Americans. Association P values (on a $-\log_{10}$ scale) of genotyped and imputed SNPs from stage 1 meta-analysis are plotted as a function of genomic position (NCBI Build 36). Plots for HLA-B, TCF7L2, KCNQ1, and HMGA2 used the model without BMI adjustment whereas plots for INS-IGF2 used the model with BMI adjustment. In each panel, the most strongly associated SNP from stage 1 and stage 1+2a+2b meta-analysis is denoted by a purple circle and a purple diamond, respectively. The color of all other SNPs indicates LD with the most strongly associated SNP based on the HapMap 2 YRI data. At KCNQ1, two independent signals are shown. doi:10.1371/journal.pgen.1004517.g002
More importantly, these individuals also showed substantially stronger associations at HLA class II (P<1×10−25) than class I regions (P<1×10−18) [42], which is in contradiction with our finding of stronger associations at HLA class I than class II regions in T2D individuals (HLA-B, Figure S4). The observed HLA-B association may be due to LD with nearby causal gene(s) since there is long range LD in this region. Recently, rs3130501 near POU5F1 and TCF719 was reported for association with T2D in a trans-ancestry meta-analysis [8]. rs3130501 was located 211 kb upstream of rs2244029 and mapped to the same LD interval. However, the two SNPs were not correlated in both CEU (D’ = 0.57, r² = 0.05) and ASW (D’ = 0.68, r² = 0.16) from 1KGP nor strongly associated with T2D in the stage 1 meta-analysis (P = 0.04). Other potential non-HLA candidate genes may include TNFA which regulates immune and inflammatory response. It has been hypothesized that activated innate and adaptive immune cells stimulate release of cytokines such as TNFα and IL-1β, which promote both systemic insulin resistance and β-cell damage [44]. On the other hand, evidence has implicated T1D loci to promote both systemic insulin resistance and participate in the pathogenesis of autoimmune diabetes in adults (LADA) and/or T2D [7,34,45–46], while T2D loci such as PPARγ and TCF712 were associated with T1D [47] and LADA [46,48], respectively. More comprehensive studies are needed to understand the shared and distinct genetic risks in different forms of diabetes which will facilitate diagnosis and personalized treatment.

Our results have several implications regarding the genetic architecture of T2D. First, fine-mapping suggests that currently known loci explain more of the risk than previously estimated. Second, the loci conferring the largest risk for T2D appear to act through regulatory rather than protein-coding changes. Third, many, but not all, of the previously identified T2D loci are shared across ancestries. The differential LD structure of African-ancestry populations at shared loci provides an opportunity for fine mapping in trans-ethnic meta-analysis. Fourth, the ~2.6M MEDIA SNPs achieved only 43.3% coverage of the 1KGP ASW common SNPs, suggesting that risk loci that are specific to African-ancestry individuals are difficult to discover with the genotyping arrays being used. Large-scale sequencing studies, such as those focusing on whole genomes, exomes, and targeted resequencing for personalized treatment, will facilitate diagnosis and treatment of different forms of diabetes which will facilitate diagnosis and personalized treatment.

Materials and Methods

Samples and clinical characterization

Stage 1 discovery and clinical characterization included 17 T2D GWAS studies (ARIC, CARDIA, CFS, CHS, FamHis, GENOA, HANDLS, Health ABC, HUFS, JHS, MESA, MESA Family, SIGNET-REGARDS, WFSM, FIND, and WHI) with up to 23,827 African American subjects (8,284 cases and 15,543 controls). Stage 2 replication sample included up to 11,544 African American subjects (6,061 cases and 5,483 controls), using mendiceto replication of GWAS data from eMERGE and IPM Biobank and de novo genotyping in IRAS, IRASFS, SCGS, and WFSM. In general, T2D cases were defined as having at least one diagnosis of fasting plasma glucose ≥126 mg/dl, 2 hour glucose during oral glucose tolerance test (OGTT) ≥200 mg/dl, random glucose ≥200 mg/dl, oral hypoglycemic agent or insulin treatment, or physician-diagnosed diabetes. All cases were diagnosed at ≥25 years (or age at study ≥25 years if age at diagnosis was not available). For cohort studies, individuals who met the criteria at any of the visits were defined as cases. Controls with normal glucose tolerance (NGT) were defined by satisfying all the following criteria: fasting plasma glucose <100 mg/dl, 2 hour OGTT<140 mg/dl (if available), no treatment of diabetes, and age ≥25 years. For cohort studies, individuals who met the criteria at all visits were defined as controls. All study participants provided written informed consent, except for eMERGE that use an opt out program, and approval was obtained from the institutional review board (IRB) from the respective local institutions. Detailed descriptions of the participating studies are provided in Text S1.

Genotyping, imputation and quality control

For stage 1 and 2 GWAS studies, genotyping was performed with Affymetrix or Illumina genome-wide SNP arrays. Imputation of missing genotypes was performed using MACH [49], IMPUTE2 [50] or BEAGLE [51] using HapMap reference haplotypes. For each study, samples reflecting duplicates, low call rate, gender mismatch, or population outliers were excluded. In general, SNPs were excluded by the following criteria: call rate < 0.95, minor allele frequency (MAF)<0.01, minor allele count < 10, Hardy-Weinberg P-value <1×10−4, or imputation quality score <0.5 (Table S3). For de novo replication studies, genotyping was performed using the Sequenom MassArray platform (Sequenom; San Diego, CA). Sample and SNP quality controls were performed as with GWAS data.

Statistical analysis

Single SNP association was performed for each study by regressing T2D case/control status on genotypes. To account for uncertainty of genotype calls during imputation, genotype probabilities or dosage were used for association tests in imputed SNPs. The association tests assumed an additive genetic model and adjusted for age, sex, study centers, and principal components. Principal components were included to control for confounding effects of admixture proportion and population structure. Secondary analysis with additional adjustment for BMI was performed for SNPs with P<1×10−7 in stage 1 meta-analysis and index SNPs previously reported to be associated with T2D or glucose homeostasis traits. BMI adjustment allows increasing power to detect T2D loci independent of BMI effect and diminish associations at T2D loci with effects modulated through BMI. Logistic regression was used for samples of unrelated individuals.

Generalized estimating equations [52] or SOLAR [53] were used for samples of related individuals. Association results with extreme values (absolute beta coefficient or standard error >10), primarily due to low cell counts resulting from small sample sizes and/or low minor allele frequencies, were excluded (Table S3).

Meta-analysis

In stage 1, association results were combined by a fixed effect model with inverse variance weighted method using the METAL software [12]. Genomic control correction [11] was applied to each study before meta-analysis, and to the overall results after meta-analysis. Results from SNPs genotyped in <10,000 samples and those with allele frequency difference >0.3 among studies were excluded. A total of 2,579,389 SNPs were analyzed in the meta-analysis (Table S3). In stage 2a, association results from African American replication studies were also combined using a fixed effect inverse variance weighted method. To assess the overall effects in African Americans (stage 1+2a) and both African Americans and Europeans (stage 1+2a+2b), association results from studies in the respective stages were combined using a fixed effect inverse variance weighted method. Genome-wide significance is declared at P<5×10−8 from the meta-analysis result of all stages, which has better power than the replication-based strategy [54].
Among the 51 SNPs carried forward for replication, heterogeneity of effect sizes across studies within each was assessed using Cochran’s Q statistic implemented in METAL. Meta-analysis results from stages 1 and 2a, stage 1a+2a and 2b were used to assess heterogeneity of effect sizes between discovery and replication stages in African Americans, and between African Americans and Europeans, respectively. For SNPs with significant heterogeneous effect size after multiple comparison corrections (P_{het}<0.001), meta-analysis results including studies of all stages assessed by the random effect model implemented in GWAMA [55] were reported.

Heterogeneous associations may partly due to differences in ascertainment scheme across studies. For index SNPs reported in prior studies, assessment of heterogeneity using Cochran’s Q statistic was also reported.

Transferability analysis

Index SNPs associated with T2D or glucose homeostasis traits from prior GWAS and candidate gene studies were examined for association with T2D in African Americans (Table S3). For the index SNP association tests, a per-SNP P value <0.05 was defined as significant. In the locus-wide analysis, the boundaries of a locus were defined by the most distant markers (within ±500 kb) using the 1KGP CEU data with r²≥0.3 with the index SNP. All MEDIA SNPs within these bounds were examined for association analysis. All pairwise LD values within each locus were estimated using the 1KGP CEU and ASW data. To estimate the effective number of SNPs at a locus, we retrieved genotypes from the 1KGP ASW data for markers present in MEDIA, estimated the sample covariance matrix from those genotypes, and spectrally decomposed the covariance matrix [24]. The effective number of SNPs was estimated using the relationship $N_{eff} = \left(\sum_{k=1}^{K} \lambda_k^2 \right) / \left(\sum_{k=1}^{K} \lambda_k \right)^2$, in which $\lambda_k$ is the kth eigenvalue of the K×K covariance matrix for the K SNPs in the locus [24]. The per-locus significance level was defined as 0.05/effective number of SNPs (Table S6). By accounting for all SNPs within the bounds of LD, the per-locus significance level is corrected to account for markers in LD with the index SNP as well as markers not in LD with the index SNP, thereby potentially allowing for discovery of new associations at markers not tagged by the index SNP.

Liability-scale variance explained

For each independent locus, we estimated the sibling relative risk using the most strongly associated SNP within that locus. Let p_i and y_i be the risk allele frequency and the corresponding odds ratio at the ith SNP, respectively. Assuming the additive genetic model and independence between SNPs, the contribution to the sibling relative risk $\lambda_s$ for a set of N SNPs is given by

$$\lambda_s = \frac{N}{\Pi_{i=1}^{N} \left(1 + \frac{p_i(1-p_i)(y_i-1)^2}{2(1-(1-p_i)+p_iy_i)}\right)^2} \left[2 \left(\frac{1}{\omega + T_1 (\omega - T)} \right)^{-1} \right]^{\frac{K}{2}}$$

[36]. Let K be the disease prevalence. The liability-scale variance $h_L^2$ explained by the set of N SNPs is given by

$$h_L^2 = \frac{2}{\omega + T_1 (\omega - T)} \left[2 \left(\frac{1}{\omega + T_1 (\omega - T)} \right)^{-1} \right]$$

in which $T = \phi^{-1}(1-K), T_1 = \phi^{-1}(1-\lambda_s K), and \omega = \frac{z}{K}$ with $\phi^{-1}$ representing the standard normal quantile function and z representing the standard normal density at T [57].

Coverage

The coverage of MEDIA SNPs to the human genome was estimated using HaploView [58] via pairwise tagging at the r²=0.8 threshold. We used all SNPs with minor allele frequencies ≥1% in both MEDIA and the 1KGP ASW sequence data. Coverage was estimated using non-overlapping bins of 1,000 SNPs.

Power analysis

Study power was calculated using the genetic power calculator [59]. For SNPs with MAF≥0.3, our study had >80% power to detect odds ratios for T2D at OR≥1.06 and ≥1.13 at P<0.05 and P<5×10⁻⁶ respectively, in stage 1 samples under an additive model. The observed odds ratios among our stage 1 most significantly associated SNPs with P<1×10⁻⁵ ranged from 1.11 to 1.56 (Table S4). Given our African American sample size in stage 1a+2a, our study had >80% power to detect OR≥1.1 at P<5×10⁻⁶ at MAF≥0.3, thus provided good power to detect genome-wide significance among the most significantly associated SNPs using all African American samples. For T2D SNPs reported from the literature, power was also calculated from the reported effect size using the risk allele frequency from this study for stage 1 samples at P<0.05 and P<5×10⁻⁶, respectively (Table S5).

Gene expression analysis

The MuTHER resource (www.muther.ac.uk) includes lymphoblastoid cell lines (LCLs), skin, and adipose tissue derived simultaneously from a subset of well-phenotyped healthy female twins from the TwinsUK adult registry [60]. Whole-genome expression profiling of the samples, each with either two or three technical replicates, was performed using the Illumina Human HT-12 V3 BeadChips (Illumina Inc.) according to the protocol supplied by the manufacturer. Log₂-transformed expression signals were normalized separately per tissue as follows: quantile normalization was performed across technical replicates of each individual followed by quantile normalization across all individuals. Genotyping was performed with a combination of Illumina arrays (HumanHap300, HumanHap610Q, 1M-Duo, and 1.2M-Duo 1M). Untyped HapMap2 SNPs were imputed using the IMPUTE2 software package. In total, 776 adipose and 777 LCL samples had both expression profiles and imputed genotypes. Association between all SNPs (MAF≥5%), IMPUTE info ≥0.8) within a gene or within 1 MB of the gene transcription start or end site and normalized expression values were performed with the GenABEL/ProbaBEL packages [61–62] using the polygenic linear model incorporating a kinship matrix in GenABEL followed by the ProbaBEL mmscore score test with imputed genotypes. Age and experimental batch were included as cofactors.

Genotype and gene expression in LCL in HapMap samples were also available [63]. Association of genotypes and gene expression of transcripts within 1 MB of tested SNPs were analyzed separately for CEU and YRI populations. The variance components model implemented in SOLAR was used for association analysis which accounts for correlation among related individuals [53].

In this study, we examined the association of the most significantly associated SNPs from the six genome-wide significant loci and their proxies (r²≥0.8 in ASW) within 1 MB of the associated SNPs with cis-expression quantitative trait loci (eQTLs) in peripheral blood leukocytes (LCL) and adipose tissue (Table S8).

**ENCODE data analysis**

We examined putative function of non-coding genome-wide significant SNPs and their proxies within 1 MB (r²≥0.8 in 1KGP ASW) using HaploReg [30] and RegulomeDB [64]. These databases interrogated multiple chromatin features from the
Encyclopedia of DNA Elements (ENCODE) project [29]. High priority was given to variants annotated as protein-binding via ChIP-seq, and motif-changing via position weight matrices, with the respective transcription factors implicated in diabetes pathogenesis and related biological processes.

Supporting Information

Figure S1  Forest plots of the most strongly associated SNPs at five previously and newly identified T2D loci in African Americans. Odds ratio and 95% CIs are presented for individual studies (black circle and line) and meta-analysis results (red diamond and line). At KCNQ1, two independent associated SNPs are shown. (PDF)

Figure S2  (A) Distributions of risk allele frequencies for the previously reported index SNPs (in black) vs. the MEDIA most strongly associated SNPs (in red) in African Americans from stage 1 meta-analysis. (B) Distributions of odds ratios for risk alleles of the index SNPs (in black) vs. the most strongly associated MEDIA SNPs (in red) in African Americans from stage 1 meta-analysis. (PDF)

Figure S3  Regional plots of stage 1 meta-analysis association results in African Americans for the most strongly associated SNPs from this study and the index SNPs from previous studies. (A-B) INTS8-TP53INPI region; (C-D) KCNQ1 region; (E-F) HMG42 region. (A, C, E) The most strongly associated SNP in MEDIA is denoted by a purple circle and a red arrow with LD colored based on the HapMap 2 YRI data. (B, D, F) The index SNP is denoted by a purple circle and a red arrow with LD colored based on the HapMap 2 CEU data. (PDF)

Figure S4  Regional plots of HLA-B and HLA-DQ/DR regions for (A, C) stage 1 meta-analysis association results in African Americans and HapMap 2 YRI LD data, and (B, D) stage 3 DIAGRAMv2 results in Europeans using HapMap 2 CEU LD data. (A, B) The most strongly associated SNP rs2244020 at HLA-B region from this study is denoted by a purple circle and a red arrow. (C, D) The index SNP rs99727346 from Burton PR et al (2007) [65] is denoted by a purple circle and a blue arrow. (PDF)

Table S1  Design of studies in stage 1 GWAS and stage 2a replication in African Americans. (PDF)

Table S2  Clinical characteristics of study samples in stage 1 GWAS and stage 2a replication studies in African Americans. (PDF)

Table S3  Genotyping methods, quality controls, imputation and statistical analysis in stage 1 GWAS and stage 2a replication studies in African Americans. (PDF)

Table S4  SNPs with $P$ value $\leq 1 \times 10^{-5}$ from stage 1 GWAS meta-analysis (BMI unadjusted) selected for stage 2 in silico and de novo replication in African Americans and in silico replication in individuals of European ancestry from DIAGRAMv2. (PDF)

Table S5  Stage 1 GWAS meta-analysis results for index SNPs at established T2D or glucose homeostasis loci in African Americans. (PDF)

Table S6  Locus-wide association at established T2D or glucose homeostasis loci in stage 1 GWAS meta-analysis in African Americans. (PDF)

Table S7  BMI-adjusted association for SNPs from stage 1 GWAS meta-analysis selected for replication. (PDF)

Table S8  Expression Quantitative Trait Loci (eQTL) analysis for the genome-wide significant SNPs for T2D. Results are shown for suggestive evidence of cis-association ($P<0.05$) between the genome-wide significant SNPs and their proxies within 1 Mb of the associated SNPs. (PDF)

Table S9  Putative regulatory SNPs predicted from the ENCODE project for the genome-wide significant SNPs and their proxies at TCF7L2, INS-IGF2, KCNQ1 and HMG42. (PDF)

Text S1  Description of GWAS and replication studies. (PDF)

Acknowledgments

We thank all the study participants for their valuable contributions to the parent studies (ARIC, CARDIA, CFS, CHS, eMERGE, FamHS, FIND, Genestar, GenoVA, HANDLS, Health ABC, HUFS, IPM Biobank, IRAS, IRASFS, JHS, MESA, MESA Family, SCCS, SIGNET-REGARDS, WFSM and WHI) of the MEDIA consortium. We thank the contributions of investigators and staff of the parent studies for data collection, genotyping, data analysis and sharing of association results to the MEDIA consortium at the discovery and replication stages. We also thank the DIAGRAM Consortium for sharing the European T2D GWAS meta-analysis results for replication. We thank the MuTHER Consortium to share the expression quantitative trait loci data. The Candidate-gene Association Resource (CARE) acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the involved research institutions, study investigators, field staff, and study participants of ARIC, CARDIA, CFS, CHS, and MESA in creating the Candidate-gene Association Resource for biomedical research [http://public.nlm.nih.gov/GeneticsGenomics/home/care.aspx]. The parent studies have contributed parent study data, ancillary study data, and DNA samples through the Massachusetts Institute of Technology - Broad Institute to create this genotype/phenotype database for wide dissemination to the biomedical research community. Women’s Health Initiative (WHI) investigators have reviewed and approved this manuscript. WHI investigators are listed at http://www.whiscience.org/publications/WHI_investigators_shortlist_2010-2015.pdf. The datasets used for the analyses described in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap through dbGaP accession phs000200.v1.p1.

Members of the FIND Consortium are:


Clinical centers:

Case Western Reserve University, Cleveland, OH: JR Sedor, J Schelling, A Seghal, A Pickens, L Humbert, L Gutz-Drayle.


University of California, Davis, CA.

University of California, Irvine, CA.

Johns Hopkins University, Baltimore, MD: M Klaj, J Coresh, L Kao, L McManus, R Parekh, N Find, P Bayton, Y Long, L Wei, T Whitehead.


University of California, Los Angeles, CA: MF Saad, N Gui, J Rotter, K Taylor, M Budgrett.

University of New Mexico, Albuquerque, NM: P Zager, V Shah, M Scavini, A Bobelou.

University of Texas Health Science Center at San Antonio, San Antonio, TX: H Abdoul, N Arat, R Duggirala, B Kasimath, R Plaetke, M Stem, C Goyes, V Sartorio.
Members of the MuTHER Consortium are:
Kourosh R. Ahmadi1, Chrysanthi Ainali2, Amy Barrett3, Veronique Bataille1, Jordana T. Bell1,4, Alfonso Buil5, Panos Deloukas6, Emmanouil Kerrin S. Small1,6, Nicole Soranzo1,6, Tim D. Spector1, Gabriel Surdulescu1, Mary E. Travers3, Loukia Tsaprouni6, Sophia Tsoka2, Alicja Wilk6, Tsz-Po Yang6, Krina T. Zondervan4

1 Department of Internal Medicine, Catharina Hospital, Eindhoven, The Netherlands.
2 Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes, France.
3 Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.
4 Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK.
5 Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK.
6 Institute of Biometrics and Epidemiology, German Diabetes Center, Duesseldorf, Germany.
7 Department of Internal Medicine, University Medical Center Utrecht, Utrecht, The Netherlands.
8 Molecular Genetics, Medical Biology Section, Department of Pathology and Medical Biology, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands.
9 Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands.
10 Department of Physiology and Biophysics, University of Southern California School of Medicine, Los Angeles, California, USA.
11 National Institute of Health, Bethesda, Maryland, USA.
12 Genomic Medicine, Imperial College London, Hammersmith Hospital, London, UK.
13 Type 2 Diabetes GWAS in African Americans
1. Department of Twin Research and Genetic Epidemiology, King’s College London, London, UK
2. Department of Informatics, School of Natural and Mathematical Sciences, King’s College London, Strand, London, UK
3. Oxford Centre for Diabetes, Endocrinology & Metabolism, University of Oxford, Churchill Hospital, Oxford, UK
4. Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK
5. Department of Genetic Medicine and Development, University of Geneva Medical School, Geneva, Switzerland
6. Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
7. University of Cambridge, Cambridge, UK
8. European Bioinformatics Institute, Hinxton, UK
9. University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke’s Hospital, Cambridge, UK
10. Cambridge NIHR Biomedical Research Centre, Addenbrooke’s Hospital, Cambridge, UK
11. Oxford NIHR Biomedical Research Centre, Churchill Hospital, Oxford, UK
12. St. John’s Institute of Dermatology, King’s College London, London, UK

Author Contributions
Conceived and designed the experiments: MCYN DS XG JL Liu LRY MAN WLb EP NG Mh MP Cc JD AG FL Rz Wz Pz Mj. Performed the experiments: MAN NDP AA AD ABS VU Lf ZI Bw Wb Dc. Analyzed the data: MCYN Ds Bh Jl Wm Xg Jl Lr MAN Mc RJ Ds Vb Sp Rb Yl Fa Lo La Ly Bm Rd Pm Cd Lr Dv Ab BM Kr Wc Wzh Ak Og Jp EG Mg Rh. Contributed reagents/materials/analysis tools: Sb LW Bf Jc Mn Lj Mc Rh Jl Jl. Wrote the paper: MCYN Ds.

References


42. Howson JM, Roy MS, Zeitels L, Stevens H, Todd JA (2013) HLA class II gene associations in African American Type 1 diabetes reveal a protective HLA-DRB1*03 haploype. Diabetes 62: 3292–3299.


