Large-Scale Gene-Centric Meta-Analysis across 39 Studies Identifies Type 2 Diabetes Loci


1Center for Human Genetic Research, Massachusetts General Hospital, Boston, MA 02114, USA; 2Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA; 3Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA 02114, USA; 4Department of Genetics, University of Pennsylvania, School of Medicine, Philadelphia, PA 19104, USA; 5Complex Genetics Section, Department of Medical Genetics, University Medical Center Utrecht, The Netherlands; 6Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, The Netherlands; 7Center for Applied Genomics, Abramson Research Center, The Children’s Hospital of Philadelphia, Philadelphia, PA 19104, USA; 8BGI Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China; 9Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York, NY 10029 USA; 10Medical Research Council Centre for Causal Analyses in Translational Epidemiology, Department of Social Medicine, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK; 11Thrombosis in Myocardial Infarction Study Group, Cardiovascular Division, Brigham and Women’s Hospital, Boston, MA 02115 USA; 12Department of Biochemistry, University of Western Ontario, London, ON N6A 5C1, Canada; 13Servicio de Genética Instituto de Investigación Sanitaria-Fundación Jiménez Díaz, Avda. Reyes Católicos 228040, Madrid, Spain; 14Clinical Pharmacology, Barts and the London Genome Centre, Queen Mary University of London, London EC1M 6BQ, UK; 15William Harvey Research Institute, Barts and the London School of Medicine, Queen Mary University of London, London EC1M 6BQ, UK; 16LifeLines Cohort Study and Biobank, University Medical Center Groningen, University of Groningen, the Netherlands; 17Children’s Environmental Health and Disease Prevention Center at Dartmouth, Hanover, NH 03755, USA; 18Section of Biostatistics and Epidemiology, Department of Community and Family Medicine, Dartmouth Medical School, Hanover, NH 03756, USA; 19Department of Medicine (Medical Genetics), University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA; 20Department of Cardiovascular Medicine, Barts and the London Genome Centre, Queen Mary University of London, London EC1M 6BQ, UK; 21Center for Human Genetic Research, Institute of Medical and Population Genetics, Cambridge, MA 02142, USA; 22Department of Physiology and Biophysics, Massachusetts General Hospital, Boston, MA 02114, USA; 23Department of Nutrition, Institute for Cancer Research, University of Delaware, Newark, DE 19716, USA; 24Division of Endocrinology, Diabetes and Nutrition, University of Maryland School of Medicine, Baltimore, MD 21201, USA; 25Department of Epidemiology, Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany; 26Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, OX3 7BN, UK; 27British Heart Foundation Glasgow Cardiovascular Research Centre, Division of Cardiovascular and Medical Sciences, Western Infirmary, University of Glasgow, Glasgow G12 8TA, UK; 28Department of Epidemiology and Public Health, University College London, London WC1E 6BT, UK; 29LURC Study, Freiburg im Breisgau 79098, Germany; 30Synlab Center of Laboratory Diagnostics Heidelberg, Heidelberg 69037, Germany; 31Division of Cardiology,
To identify genetic factors contributing to type 2 diabetes (T2D), we performed large-scale meta-analyses by using a custom ~50,000 SNP genotyping array (the ITMAT-Broad-CARe array) with ~2000 candidate genes in 39 multiethnic population-based studies, case-control studies, and clinical trials totaling 17,418 cases and 70,298 controls. First, meta-analysis of 25 studies comprising 14,073 cases and 57,489 controls of European descent confirmed eight established T2D loci at genome-wide significance. In silico follow-up analysis of putative association signals found in independent genome-wide association studies (including 8,130 cases and 38,987 controls) performed by the DIAGRAM consortium identified a T2D locus at genome-wide significance (GATAD2A/CILP2/PBX4; \( p = 5.7 \times 10^{-9} \)) and two loci exceeding study-wide significance (SREBF1, and TH/INS; \( p < 2.4 \times 10^{-6} \)). Second, meta-analyses of 1,986 cases and 7,695 controls from eight African-American studies identified study-wide-significant (\( p = 2.4 \times 10^{-7} \)) variants in HMGA2 and replicated variants in TCF7L2 (\( p = 5.1 \times 10^{-15} \)). Third, conditional analysis revealed multiple known and novel independent signals within five T2D-associated genes in samples of European ancestry and within HMGA2 in African-American samples. Fourth, a multiethnic meta-analysis of all 39 studies identified T2D-associated variants in BCL2 (\( p = 2.1 \times 10^{-3} \)). Finally, a composite genetic score of SNPs from new and established T2D signals was significantly associated with increased risk of diabetes in African-American, Hispanic, and Asian populations. In summary, large-scale meta-analysis involving a dense gene-centric approach has uncovered additional loci and variants that contribute to T2D risk and suggests substantial overlap of T2D association signals across multiple ethnic groups.

Introduction

Type 2 Diabetes (T2D [MIM 125853]) is a complex disease caused by multiple genetic and environmental factors; heritability is estimated at 22%–73% from twin and family studies.\(^1\)–\(^5\) The age-adjusted prevalence of T2D in adults has recently been estimated at 7.6% in European Americans, 14.9% in non-Hispanic African Americans, 4.3%–8.2% in Asian Americans, and 10.9%–15.6% in Hispanics.\(^6\)–\(^9\) Researchers have identified more than 40 T2D-associated genetic loci, but these loci have been revealed primarily on the basis of studies of individuals of European ancestry. Candidate-gene association studies discovered association between T2D and missense variants in PPARG (MIM 601487) and KCNJ11 (MIM 600937), which are targets for anti-diabetic medications, and implicated common genetic variants responsible for Mendelian forms of diabetes in T2D (e.g., such variants include those in the Wolfram-syndrome-associated locus WFS1 [MIM 606201], those in HNF1A [MIM 142410] and HNF4A [MIM 600281], and those in genes involved in maturity-onset diabetes of the young [MODY (MIM 125851)]\(^10\)–\(^15\)). Association testing near a linkage peak identified common variants in TCF7L2 (MIM 602228), which remains the strongest signal for T2D and replicates robustly across many,\(^16\) but not all, ethnic groups.\(^15\) Early genome-wide association studies (GWASs) for T2D\(^18\)–\(^22\) and fasting glucose\(^23\) successfully identified multiple loci. Recent meta-analyses of GWASs of T2D\(^24\) and glycemic quantitative traits\(^25\) have dramatically increased the number of genome-wide-significant T2D-associated loci in European populations; most of these variants act through defects in beta-cell function rather than insulin action. Together, variants known to be associated with T2D explain ~10% of the genetic variance,\(^24\)\(^26\) indicating that additional loci and independent signals in established loci are likely to contribute to disease risk.

Genetic contributors to T2D are less well understood in non-European populations. One novel locus (KCNQ1 [MIM 607542]) was identified on the basis of a GWAS in a Japanese population\(^27\)\(^28\) and has subsequently been shown to harbor independent alleles in individuals of European descent.\(^24\) More recently, GWASs in Chinese,\(^29\)\(^30\) Japanese,\(^31\) and South Asian populations\(^32\) describe additional T2D loci surpassing genome-wide significance. To date, T2D GWASs in African-Americans have been underpowered to detect novel loci.\(^33\)

An important first step toward understanding genetic risk across populations is to establish whether known T2D association signals span ethnicities or are population specific. Consistent association of T2D risk variants discovered in Europeans was reported in a multiethnic case-control study of five US populations,\(^34\) in studies of Chinese,\(^30\) Japanese,\(^31\) Hispanic\(^35\) and South Asian\(^36\) populations, and in a study focusing on fasting glucose in an African-American\(^36\) population, despite possible differences in linkage disequilibrium (LD) between marker and causal variants in each population. Indeed, multietnic differences in regional LD help with refinement of association signals and can distinguish causal variants from correlated markers.\(^33\)\(^37\) Furthermore, independent association signals in the same gene (for example, KCNQ1) in different ethnicities could be useful for pinpointing genes that harbor causal mutations. Recent power analyses suggest that large-scale multiethnic association studies might have greater statistical power to detect causal alleles because random genetic drift can elevate global risk variants to a higher allele frequency in different populations.\(^38\)

The 50K SNP Human CVD beadchip, or ITMAT-Broad-CARe (IBC) array, captures genetic diversity across more than 2,000 candidate gene regions related to cardiovascular, inflammatory, and metabolic phenotypes, and a large portion of loci are captured with marker density equal to or greater than that found by GWAS.\(^39\) SNPs were selected on the basis of the International HapMap Consortium and publically available resequencing data such as those from the SeattleSNPs and National Institute of Environmental Health Sciences (NIEHS) SNPs consortia. The focus was on the inclusion of rare variants and variants with a high likelihood of functionality. More than 5,000 SNPs from 49 candidate genes (Table S1, available online) were specifically selected on the basis of prior evidence of a role in Mendelian forms of diabetes,\(^40\) diabetes pathophysiology (including insulin signaling, endocrine pathways, and energy metabolism),\(^41\) linkage studies,\(^42\) and meta-analyses of T2D GWASs.\(^18\)\(^22\) GWAS efforts have
identified additional loci since the design of this array; therefore, of the currently documented T2D loci, 21 candidate genes from eighteen loci and index SNPs from nine additional T2D loci were included on the array. SNP association for this array has been reported for a range of phenotypes, including those involving coronary artery disease, lipids, and blood traits; the TH/INS region (MIMs 191290 and 176730)) from the DIAGRAM consortium, and 22 association results for each component study of the DIAGRAM GWAS meta-analysis were obtained for SNPs rs9273363 (HLA-DQB1) and rs10770141 (TH/INS region (MIMs 191290 and 176730)). The genomic control inflation factor, \( \lambda \), was calculated in each case-control study and used for within-study correction before meta-analysis. \( \lambda \) ranged from 1.0 to 1.04. ARIC individuals used for the Look AHEAD meta-analysis reported here (1,278 ARIC cases and 2,600 independent ARIC controls) sample were not included in the ARIC IBC association analysis reported here (1,278 ARIC cases and 2,600 independent ARIC controls). Notably, Look AHEAD participants were more obese (with an average BMI of 36.0 \pm 5.9 kg/m\(^2\)), and ARIC controls selected for matching were not (average BMI of 27.7 \pm 5.2 kg/m\(^2\)). Despite BMI adjustment, this discrepancy could lead to T2D association signals arising from SNP associations to BMI.

Genotyping and Quality Control
Genotyping in each component study of the IBC meta-analysis was performed with the IBC array. SNPs were clustered into genotypes with the Illumina Beadstudio software and subjected to quality-control filters at the sample and SNP levels separately within each cohort. Samples were excluded for individual call rates <90%, gender mismatch, and duplicate discordance. SNPs were removed for call rates <95% or Hardy-Weinberg equilibrium \( p < 10^{-7} \) in controls from each cohort (regardless of ethnicity). Because of the low-frequency SNPs included in the design and the aim to capture low-frequency variants of large effect across the large dataset, we filtered only on minor allele frequency (MAF) < 0.005.

Statistical Analyses
Evaluation of Population Stratification
For the primary meta-analysis, only individuals of European ancestry were included. Self-reported ethnicity was verified by multidimensional scaling (MDS) analysis of identity-by-state distances as implemented in PLINK; HapMap panels were included as reference standards. After SNPs in LD (\( r^2 > 0.3 \)) were pruned out, Eigenstrat was used for computing principal components on the subset of nonexcluded individuals for use as covariates in the regression analyses.

Association Testing
We performed T2D association analysis in each study by using an additive genetic model. Our primary association analysis included adjustment for age, sex, body-mass index (BMI), study site(s), and principal components (PCs) if population structure was evident. The genomic control inflation factor, \( \lambda \), was calculated in each case-control study and used for within-study correction before meta-analysis. \( \lambda \) ranged from 1.0 to 1.077. As the Look AHEAD study was a case-only study, we merged this dataset with 4,124 randomly selected ARIC controls of European ancestry for association analysis. After performing stringent quality control, we found limited population structure between the two datasets and a \( \lambda \) of 1.04. ARIC individuals used for the Look AHEAD (cases)/ARIC (controls) sample were not included in the ARIC IBC association analysis reported here (1,278 ARIC cases and 2,600 independent ARIC controls). Notably, Look AHEAD participants were obese (with an average BMI of 36.0 \pm 5.9 kg/m\(^2\)), and ARIC controls selected for matching were not (average BMI of 27.7 \pm 5.2 kg/m\(^2\)). Despite BMI adjustment, this discrepancy could lead to T2D association signals arising from SNP associations to BMI.

Meta-analyses within each ethnic group were performed by two independent analysts who used a fixed-effect inverse-variance approach in two different software packages: MANTEL and METAL. After association tests in each ethnicity, a multietnic meta-analysis that included all available participants was performed. Additionally, the direction of effect of lead SNPs from previously identified loci was evaluated for consistency in African-Americans, Hispanics, and Asians.

Previous studies using the IBC array have used different significance thresholds, from \( p < 1 \times 10^{-5} \) to \( p < 1 \times 10^{-6} \). To calculate an appropriate significance threshold, we used data from the Candidate Gene Association Resource (CARe) IBC array studies and determined that after LD was accounted for, the effective number of independent tests was ~26,500 for African Americans and ~20,500 for Europeans. This produces an experimental or study-wide statistical threshold of \( p = 1.9 \times 10^{-6} \) and \( p = 2.4 \times 10^{-6} \), respectively, if a false-positive rate of 5% is to be
maintained. Consistent with the hypothesis-driven candidate gene approach taken here, we have adopted these “array-wide” or study-wide statistical thresholds for this study, but we also highlight loci significantly associated at a more conventional genome-wide-significant threshold of $p < 5.0 \times 10^{-8}$.

For confirmation of signals previously associated with T2D, we selected the exact SNP from the most recent discovery study or meta-analysis when possible. If the SNP was not present on the array, we selected from the IBC array a SNP that was in highest linkage disequilibrium with the previously identified SNP in the HapMap CEU (Utah residents with northern and Western European ancestry from the CEPH collection) population (this SNP was identified with the online tool SNAP). Selected SNPs and $r^2$ to previously associated SNPs are shown in Table S1.

**Conditional Analyses**

Forty-six loci harboring nominally significant evidence for association ($p < 1.0 \times 10^{-4}$) were examined for additional signals via conditional analyses in PLINK.88 A term was added to the regression model so that the lead SNP was included as a covariate, and SNPs within the same candidate gene, or $\pm 200$ kb if the candidate gene region was $< 200$ kb, were evaluated for significance. Conditional analysis was performed in fifteen European and eight African-American cohorts with individual-level genotype data.

We applied a locus-specific Bonferroni correction (i.e., correction for SNPs tested within the same candidate gene or within 200 kb if the candidate gene region was $< 200$ kb) to determine the significance of independent signals within candidate genes genotyped at each locus. Table S1 shows the number of SNPs (and therefore tests) per locus on which the locus-wide significance level was based.

**Genetic-Risk-Score Analyses**

In eight African-American, Asian, and Hispanic cohorts from the CARe study, we generated a genetic-risk score of 27 T2D-associated SNPs (26 previously established genome-wide-significant T2D-associated SNPs and the index SNP from the GATA2A signal from this study) weighted by the log of the OR from meta-analyses of GWAS24,25 as described previously.59 We evaluated the contribution of the weighted genetic risk score to T2D in logistic regression models adjusting for age, gender, BMI, and ten principal components and compared the relative ORs across quartiles of risk.

**Results**

**Meta-Analysis of Samples of European Ancestry**

We performed association testing for T2D status (tests included 14,073 cases and 57,489 controls) and adjusted for age, gender, BMI, and three or more PCs in 25 European IBC studies. After a fixed-effects, inverse-variance meta-analysis, independent SNPs at HLA-DQB1, SREBF1 (MIM 184756), GATA2A/CILP2 (MIM 612419)/PBX4 (MIM 608127), BCL2 (MIM 151430), and 16 previously described loci were significantly associated with T2D at study-wide significance ($p < 2.4 \times 10^{-5}$), and SNPs at eight known loci surpassed the traditional genome-wide-significance threshold ($p < 5.0 \times 10^{-8}$; Tables 1 and 2; Figures S1A and S1B). Of the study-wide significant loci, SNPs correlated to the most significant SREBF1 polymorphism ($r^2 > 0.85$ in HapMap CEU) have been reported previously in candidate-gene analyses60,61 but not in large-scale genomic studies with robust replication. Our sample-selection and study-design strategies are validated by study-wide-significant independent replication of 14 out of 29 previously reported T2D-associated loci present on the IBC chip (Table 2) and nominal association of 23/29 loci with a consistent direction of effect for all signals. For five T2D association signals tagged by variants on the IBC array (these signals were in or near RBMS1 [MIM 602310], CENTD2 [MIM 606646], ZFAND6 [MIM 610183], HMGAA2 [MIM 600968], or HNF1A), overlapping samples from four component studies (ARIC, CCGS, FHS, and KORA) were included in the initial report,24 and therefore our results are not independent.

In order to confirm putative novel signals, we carried out in silico follow-up analyses of 25 SNPs from previously undescribed T2D-associated signals ($p < 1.0 \times 10^{-4}$) in a meta-analysis of eight GWASs ($n = 8,130$ cases and $38,987$ controls of European ancestry) from the DIAGRAM consortium.24 Combined meta-analyses of the discovery and replication studies led to genome-wide-significant signals at the GATA2A/CILP2/PBX4 ($p = 5.7 \times 10^{-9}$) and HLA-DQB1 ($p = 1.1 \times 10^{-8}$) loci; the signal at SREBF1 remained study-wide significant, and a signal at TH/INS became study-wide significant (Table 1, Figures 1A–1D). Furthermore, 12/18 additional SNPs displayed a direction of effect consistent with our discovery dataset, suggesting that additional loci might have weaker effects that are undetectable as a result of limited statistical power (Table S3). Lead or correlated SNPs for three signals with $p < 10^{-5}$ in the IBC meta-analysis at the BCL2, CDKNI1B (MIM 600778), and SLC39A4 (MIM 607059) loci were not included in the DIAGRAM meta-analysis (Table S3), and these signals, as well as our study-wide-significant findings, will need independent follow-up in future studies of European ancestry.

The HLA-DQB1 allele has previously been demonstrated to exhibit a strong association with type 1 diabetes (odds ratio [OR] > 5.4). To determine whether the association signal observed at this locus was driven by cases with LADA, we examined association of this SNP in DIAGRAM component studies with and without exclusion of GAD-antibody-positive cases. We observed a trend toward more significant association of the SNP in meta-analysis of studies including LADA than in the study set that excluded LADA (Table S4), suggesting this signal might indeed stem from a subset of cases with autoimmune diabetes misclassified as T2D; however, this difference was not statistically significant ($p_{het} = 0.51$). To explore whether rs9273363 in HLA-DQB1 in fact represents a T1D signal, we imputed classical HLA alleles in 10,636 cases and 38,063 controls from our IBC dataset by using a reference set of 2,767 European individuals with four-digit genotype data for HLA-A (MIM 142800), HLA-B (MIM 142830), HLA-C (MIM 142840), HLA-DQA1 (MIM 146880), HLA-DQB1, HLA-DRB1 (MIM 142857), HLA-DPA1 (MIM 142880), and HLA-DPB1 (MIM 142858). After imputation, we tested imputed classical HLA alleles for association with T2D by using logistic regression including age, sex, BMI, study site, and three PCs as
Table 1. Loci Associated with T2D in Individuals of European Ancestry

| Chr | NCBI 36 Position | Candidate Gene | SNP | RA | RAF | OR (95% CI)  | p     | % I² | OR (95% CI)  | p     | OR (95% CI)  | p     | Percent I² |
|-----|------------------|---------------|-----|----|-----|-------|-------|-----|-------|-------|-------|-------|-------|-----------|
| 19  | 19741596          | GATAD2A       | rs3794991 T | 0.08 | 1.14 (1.08–1.20) | 8.95 × 10⁻⁷ | 25 | 1.11 (1.04–1.17) | 1.55 × 10⁻³ | 1.12 (1.08–1.15) | 5.70 × 10⁻⁹ | 22 |
| 17  | 17662182          | SREBF1        | rs4925115 A | 0.38 | 1.09 (1.05–1.12) | 2.04 × 10⁻⁷ | 0  | 1.04 (0.99–1.08) | 1.01 × 10⁻¹ | 1.07 (1.04–1.1)  | 2.62 × 10⁻⁷ | 0  |
| 11  | 2150416           | TH/INS        | rs10770141 A | 0.39 | 1.07 (1.04–1.11) | 6.44 × 10⁻⁶ | 0  | 1.05 (1.00–1.10) | 7.08 × 10⁻² | 1.07 (1.04–1.10) | 1.57 × 10⁻⁶ | 0  |
| 18  | 58996864          | BCL2          | rs12454712 T | 0.63 | 1.08 (1.04–1.11) | 2.29 × 10⁻⁶ | 21 | -     | -     | -     | -     | -     | -     |
| 11  | 49127350          | FOLH1         | rs16906158 C | 0.09 | 1.13 (1.07–1.19) | 8.75 × 10⁻⁶ | 32 | 1.07 (1.00–1.14) | 5.17 × 10⁻² | 1.10 (1.06–1.15) | 2.57 × 10⁻⁶ | 32 |
| 3   | 172217793         | SLC2A2        | rs11924032 G | 0.74 | 1.08 (1.04–1.12) | 1.87 × 10⁻⁵ | 0  | 1.05 (1.00–1.10) | 3.31 × 10⁻² | 1.06 (1.04–1.10) | 2.60 × 10⁻⁶ | 0  |
| 19  | 50864118          | GIPR          | rs11671664 A | 0.11 | 1.12 (1.07–1.18) | 2.86 × 10⁻⁶ | 0  | 1.06 (0.98–1.18) | 1.71 × 10⁻¹ | 1.10 (1.06–1.15) | 2.61 × 10⁻⁶ | 0  |
| 6   | 32734250          | HLA-DQB1      | rs9273363 A | 0.27 | 1.10 (1.06–1.14) | 7.99 × 10⁻⁸ | 17 | 1.06 (1.01–1.10) | 1.64 × 10⁻² | 1.08 (1.05–1.11) | 1.10 × 10⁻⁸ | 19 |

Abbreviations are as follows: Chr, chromosome; RA, risk allele; RAF, risk-allele frequency; OR, odds ratio; and CI confidence interval.

- **a**14,073 cases and 57,489 controls.
- **b**8,130 cases and 38,987 controls.
- **c**22,203 cases and 94,476 controls.
- **d**Candidate gene selected for testing on the array.
- **e**IBC Study-wide significance = p < 2.4 × 10⁻⁵.
- **f**See Tables S4 and S5.
Table 2. Association Results for Known T2D Loci in Meta-Analysis of IBC Studies of European Ancestry

<table>
<thead>
<tr>
<th>Chr</th>
<th>Prev. Assoc. SNP</th>
<th>Pos (NCBI 36)</th>
<th>Candidate Gene</th>
<th>Lead SNP on IBC array</th>
<th>r2 to Prev. Assoc. SNP</th>
<th>RA</th>
<th>RAF IBC</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>I² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>rs10923931</td>
<td>120,319,482</td>
<td>NOTCH2</td>
<td></td>
<td></td>
<td>T</td>
<td>0.10</td>
<td>1.06 (1.01–1.12)</td>
<td>3.23 × 10⁻²</td>
<td>21.0</td>
</tr>
<tr>
<td>1</td>
<td>rs120,239,407</td>
<td></td>
<td>NOTCH2</td>
<td>rs2641348</td>
<td>0.85 (rs10923931)</td>
<td>G</td>
<td>0.11</td>
<td>1.06 (1.01–1.12)</td>
<td>1.95 × 10⁻²</td>
<td>16.0</td>
</tr>
<tr>
<td>2</td>
<td>rs780094</td>
<td>27,594,741</td>
<td>GCKR</td>
<td>rs780094</td>
<td>0.93 (rs1260326)</td>
<td>C</td>
<td>0.59</td>
<td>1.09 (1.05–1.12)</td>
<td>2.12 × 10⁻⁷</td>
<td>43.0</td>
</tr>
<tr>
<td>2</td>
<td>rs7578597</td>
<td>43,586,327</td>
<td>THADA</td>
<td>rs7578597</td>
<td></td>
<td>T</td>
<td>0.90</td>
<td>1.16 (1.10–1.22)</td>
<td>2.91 × 10⁻⁷</td>
<td>34.4</td>
</tr>
<tr>
<td>2</td>
<td>rs7593730</td>
<td>160,879,700</td>
<td>RBMS1</td>
<td>rs6718526</td>
<td>0.83 (rs7593730)</td>
<td>C</td>
<td>0.79</td>
<td>1.03 (0.99–1.07)</td>
<td>1.20 × 10⁻¹</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>rs1801282</td>
<td>12,368,125</td>
<td>PPARG</td>
<td></td>
<td></td>
<td>C</td>
<td>0.88</td>
<td>1.11 (1.10–1.16)</td>
<td>2.15 × 10⁻⁴</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>rs1801282</td>
<td>12,367,272</td>
<td>PPARG</td>
<td>rs7649970</td>
<td>1 (rs1801282)</td>
<td>C</td>
<td>0.88</td>
<td>1.11 (1.06–1.16)</td>
<td>7.87 × 10⁻⁶</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>rs4607103</td>
<td>64,686,944</td>
<td>ADAMTS9</td>
<td></td>
<td></td>
<td>C</td>
<td>0.74</td>
<td>1.05 (1.02–1.10)</td>
<td>4.11 × 10⁻³</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>rs4607103</td>
<td>64,676,186</td>
<td>ADAMTS9</td>
<td>rs9860730</td>
<td>0.69 (rs4607103)</td>
<td>A</td>
<td>0.68</td>
<td>1.05 (1.02–1.09)</td>
<td>1.24 × 10⁻³</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>rs1470579</td>
<td>187,011,774</td>
<td>IGF2BP2</td>
<td>rs1470579</td>
<td></td>
<td>C</td>
<td>0.32</td>
<td>1.14 (1.10–1.18)</td>
<td>9.19 × 10⁻¹⁶</td>
<td>8.8</td>
</tr>
<tr>
<td>4</td>
<td>rs10010131</td>
<td>6,343,816</td>
<td>WFS1</td>
<td></td>
<td></td>
<td>G</td>
<td>0.60</td>
<td>1.10 (1.01–1.20)</td>
<td>3.37 × 10⁻²</td>
<td>76.9</td>
</tr>
<tr>
<td>4</td>
<td>rs10010131</td>
<td>6,336,616</td>
<td>WFS1</td>
<td>rs4688985</td>
<td>0.58 (rs10010131)</td>
<td>G</td>
<td>0.73</td>
<td>1.10 (1.05–1.14)</td>
<td>2.88 × 10⁻⁷</td>
<td>34.8</td>
</tr>
<tr>
<td>6</td>
<td>rs7754840</td>
<td>20,769,229</td>
<td>CDKAL1</td>
<td></td>
<td></td>
<td>C</td>
<td>0.32</td>
<td>1.16 (1.12–1.20)</td>
<td>4.58 × 10⁻¹⁹</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>rs7754840</td>
<td>20,794,975</td>
<td>CDKAL1</td>
<td>rs9368222</td>
<td>0.63 (rs7754840)</td>
<td>A</td>
<td>0.27</td>
<td>1.18 (1.14–1.22)</td>
<td>5.37 × 10⁻²¹</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>rs10244051</td>
<td>15,030,358</td>
<td>DGKB-TMEM195</td>
<td>rs10244051</td>
<td>1 (rs2191349)</td>
<td>G</td>
<td>0.54</td>
<td>1.04 (0.95–1.14)</td>
<td>4.02 × 10⁻¹</td>
<td>55.3</td>
</tr>
<tr>
<td>7</td>
<td>rs864745</td>
<td>28,147,081</td>
<td>JAZF1</td>
<td>rs864745</td>
<td>0.97 (rs849134)</td>
<td>T</td>
<td>0.50</td>
<td>1.11 (1.08–1.14)</td>
<td>2.22 × 10⁻⁹</td>
<td>31.3</td>
</tr>
<tr>
<td>7</td>
<td>rs4607517</td>
<td>44,202,193</td>
<td>GCK</td>
<td></td>
<td></td>
<td>A</td>
<td>0.18</td>
<td>1.10 (0.98–1.23)</td>
<td>1.07 × 10⁻¹</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>rs4607517</td>
<td>44,189,327</td>
<td>GCK</td>
<td>rs1990458</td>
<td>0.18 (rs4607517)</td>
<td>C</td>
<td>0.59</td>
<td>1.08 (1.04–1.11)</td>
<td>1.67 × 10⁻⁶</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>rs13266634</td>
<td>118,253,964</td>
<td>SLC30A8</td>
<td>rs13266634</td>
<td></td>
<td>C</td>
<td>0.70</td>
<td>1.11 (1.08–1.15)</td>
<td>2.49 × 10⁻⁹</td>
<td>29.5</td>
</tr>
<tr>
<td>9</td>
<td>rs10811661</td>
<td>22,124,094</td>
<td>CDKN2A/B</td>
<td>rs10811661</td>
<td></td>
<td>T</td>
<td>0.82</td>
<td>1.19 (1.14–1.23)</td>
<td>4.83 × 10⁻¹⁶</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>rs12779790</td>
<td>12,368,016</td>
<td>CDC123/CAMK1D</td>
<td>rs12779790</td>
<td></td>
<td>G</td>
<td>0.18</td>
<td>1.07 (0.99–1.15)</td>
<td>7.28 × 10⁻²</td>
<td>4.7</td>
</tr>
</tbody>
</table>
Table 2.  

<table>
<thead>
<tr>
<th>Chr</th>
<th>Prev. Assoc. SNP</th>
<th>Pos (NCBI 36)</th>
<th>Candidate Gene</th>
<th>Lead SNP on IBC array</th>
<th>r² to Prev. Assoc. SNP</th>
<th>RA</th>
<th>RAF IBC</th>
<th>OR (95% CI)</th>
<th>p</th>
<th>I² (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>rs5015480</td>
<td>94,455,539</td>
<td>HHEX/IDE</td>
<td>rs5015480</td>
<td>C</td>
<td>0.60</td>
<td>1.12</td>
<td>(1.10–1.16)</td>
<td>2.68 × 10⁻¹⁴</td>
<td>12.3</td>
</tr>
<tr>
<td>10</td>
<td>rs7903146</td>
<td>114,748,339</td>
<td>TCF7L2</td>
<td>rs7903146</td>
<td>T</td>
<td>0.30</td>
<td>1.44</td>
<td>(1.40–1.49)</td>
<td>1.21 × 10⁻¹⁰⁹</td>
<td>27.6</td>
</tr>
<tr>
<td>11</td>
<td>rs163184</td>
<td>2,803,645</td>
<td>KCNQ1</td>
<td>rs163184</td>
<td>G</td>
<td>0.48</td>
<td>1.08</td>
<td>(1.05–1.11)</td>
<td>6.08 × 10⁻⁷</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>rs2237892</td>
<td>2,796,327</td>
<td>KCNQ1</td>
<td>rs2237892</td>
<td>C</td>
<td>0.94</td>
<td>1.11</td>
<td>(0.92–1.37)</td>
<td>2.61 × 10⁻¹</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>rs231362</td>
<td>2,648,047</td>
<td>KCNQ1</td>
<td>rs231362</td>
<td>G</td>
<td>0.52</td>
<td>1.08</td>
<td>(1.04–1.11)</td>
<td>5.90 × 10⁻⁶</td>
<td>3.8</td>
</tr>
<tr>
<td>11</td>
<td>rs5215</td>
<td>17,365,206</td>
<td>KCNJ11</td>
<td>rs5215</td>
<td>0.93 (rs5219)</td>
<td>C</td>
<td>0.37</td>
<td>1.09 (1.06–1.13)</td>
<td>1.65 × 10⁻⁸</td>
<td>15.2</td>
</tr>
<tr>
<td>11</td>
<td>rs1552224</td>
<td>72,110,746</td>
<td>CENTD2 region</td>
<td>rs613937</td>
<td>A</td>
<td>0.81</td>
<td>1.08</td>
<td>(1.04–1.13)</td>
<td>1.17 × 10⁻⁴</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>rs10830963</td>
<td>92,348,358</td>
<td>MTNR1B</td>
<td>rs10830963</td>
<td>G</td>
<td>0.30</td>
<td>1.05</td>
<td>(1.01–1.08)</td>
<td>8.07 × 10⁻⁵</td>
<td>21.0</td>
</tr>
<tr>
<td>11</td>
<td>rs2943634</td>
<td>226,776,324</td>
<td>IRS1 region</td>
<td>rs2943634</td>
<td>C</td>
<td>0.67</td>
<td>1.09</td>
<td>(1.05–1.12)</td>
<td>4.32 × 10⁻⁷</td>
<td>16.4</td>
</tr>
<tr>
<td>12</td>
<td>rs7961581</td>
<td>69,949,369</td>
<td>TSPAN8/LGR5</td>
<td>rs7961581</td>
<td>G</td>
<td>0.71</td>
<td>1.06</td>
<td>(1.01–1.10)</td>
<td>1.57 × 10⁻²</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>rs7957197</td>
<td>119,945,069</td>
<td>HNF1A</td>
<td>rs7957197</td>
<td>G</td>
<td>0.77</td>
<td>1.07</td>
<td>(1.03–1.12)</td>
<td>4.76 × 10⁻⁴</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>78,200,439</td>
<td>17179453</td>
<td>HMGA2</td>
<td>rs17179453</td>
<td>C</td>
<td>0.09</td>
<td>1.10</td>
<td>(1.05–1.16)</td>
<td>3.73 × 10⁻⁴</td>
<td>11.9</td>
</tr>
<tr>
<td>12</td>
<td>rs975197</td>
<td>64,569,867</td>
<td>ZFAND6</td>
<td>rs975197</td>
<td>G</td>
<td>0.72</td>
<td>1.01</td>
<td>(0.98–1.05)</td>
<td>4.62 × 10⁻¹</td>
<td>4.2</td>
</tr>
<tr>
<td>16</td>
<td>rs2943634</td>
<td>52,402,988</td>
<td>FTO</td>
<td>rs2943634</td>
<td>A</td>
<td>0.41</td>
<td>1.06</td>
<td>(1.03–1.09)</td>
<td>1.78 × 10⁻⁴</td>
<td>14.8</td>
</tr>
<tr>
<td>17</td>
<td>rs4430796</td>
<td>33,172,153</td>
<td>HNF1B</td>
<td>rs4430796</td>
<td>A</td>
<td>0.51</td>
<td>1.08</td>
<td>(1.05–1.12)</td>
<td>2.48 × 10⁻⁷</td>
<td>0</td>
</tr>
<tr>
<td>X</td>
<td>rs945326</td>
<td>152,553,116</td>
<td>DUSP9</td>
<td>rs945326</td>
<td>A</td>
<td>0.77</td>
<td>1.09</td>
<td>(1.03–1.15)</td>
<td>2.53 × 10⁻³</td>
<td>0</td>
</tr>
</tbody>
</table>

The study included 14,073 cases and 57,489 controls. Abbreviations are as follows: Chr, chromosome; Prev. assoc., previously reported T2D-associated SNP; Pos, position; RA, risk allele; RAF, risk-allele frequency; OR, odds ratio; and CI, confidence interval.
covariates. \textit{HLA-DQA1} was identified as the signal that was the most significantly associated with T2D (\textit{HLA-DQA1*03}; \(p = 2.8 \times 10^{-7}\)) and was closely followed by \textit{HLA-DRB1} (\textit{HLA-DRB1*04}; \(p = 3.6 \times 10^{-7}\); Table S5). \textit{HLA-DQB1}, \textit{HLA-DQA1}, and \textit{HLA-DRB1} are major genetic determinants of T1D,\textsuperscript{63–65} suggesting that the observed association at HLA SNPs is an artifact caused by LADA pollution of our cases. We repeated association analysis conditioning on lead SNPs to identify other HLA-region loci that were associated with T2D (both rs9273363 and HLA-DQA1*03), but we found no additional association signals.

**Independent Signals at European T2D-Associated Loci**

An advantage of our study over a GWAS is the inclusion of rare missense SNPs at candidate loci. Such SNPs can help identify independent T2D signals within genes at known loci. We performed regional conditionalassociation analyses by adjusting for lead SNPs at 46 signals with \(p < 10^{-4}\) in 10,636 cases and 38,063 controls of European

---

**Figure 1.** Regional Plots for T2D Loci with Study-Wide-Significant Regions in IBC Meta-Analysis of Data from Individuals with European Ancestry

Loci are shown as the lead SNP with a flanking region depicting the candidate gene and nearby genes included on the array. The purple diamond represents the lead SNP in the IBC meta-analysis, and the dots represent the surrounding SNPs; colors show the LD relationship with the lead SNP on the basis of CEU HapMap II information. \(-\log_{10} p\) values for association with T2D are shown for each SNP (left-hand axis). Recombination rates in CEU HapMap II are shown in blue traces (right-hand axis).
ancestry. After Bonferroni correction for SNPs at the candidate gene locus, we found independent novel T2D signals at the PPARG, THADA (MIM 611800), and JAZF1 (MIM 606246) loci and confirmed known secondary signals at CDKN2A (MIM 600160)/CDKN2B (MIM 600431) and KCNQ1 (rs9668162 p = 2.4×10^{-7}; Table S7). HMG2 encodes a transcriptional regulator of IGF2BP2 (MIM 608289), a gene known to be associated with T2D, and a T2D-associated signal ~43 kb upstream of HMG2 was recently described in a European GWAS meta-analysis performed by the DIAGRAM consortium (r^2 = 0.005 in CEU, r^2 = 0.35 in YRI [Yoruba in Ibadan, Nigeria] to best European SNP, 24). Because our array coverage was limited to the HMG2 region, we cannot confirm whether the African-American signal observed is independent from the previously reported HMG2 signal in Europeans; however, conditional analysis in African-American studies revealed two locus-wide-significant associations (Table 3).

T2D Meta-Analysis in Multiethnic Populations
We next performed meta-analysis across IBC-array T2D-association results independently in eight African-American studies (1,986 cases, 7,695 controls), three Hispanic-Latino studies (592 cases, 1,410 controls), and three Asian studies (767 cases, 3704 controls). Although these analyses were underpowered to detect novel loci at genome-wide significance in these populations, we could evaluate the contribution of known and identified T2D loci to disease risk across ethnicities (Table S7).

In African Americans, study-wide significant association was observed for SNPs within TCF7L2 (rs7903146 p = 5.1×10^{-15}) and within HMG2 (rs9668162 p = 2.4×10^{-7}; Table S7). HMG2 encodes a transcriptional regulator of IGF2BP2 (MIM 608289), a gene known to be associated with T2D, and a T2D-associated signal ~43 kb upstream of HMG2 was recently described in a European GWAS meta-analysis performed by the DIAGRAM consortium (r^2 = 0.005 in CEU, r^2 = 0.35 in YRI [Yoruba in Ibadan, Nigeria] to best European SNP, 24). Because our array coverage was limited to the HMG2 region, we cannot confirm whether the African-American signal observed is independent from the previously reported HMG2 signal in Europeans; however, conditional analysis in African-American studies revealed two locus-wide-significant associations (Table 3).
diabetes-associated loci, verified known T2D-associated loci, and uncovered multiple independent T2D association signals, demonstrating the power of large collaborative approaches to uncover genetic insights in T2D. Furthermore, we have evaluated the impact of established T2D risk variants on risk of T2D in multiethnic populations.

Samples were assembled primarily through population-based studies and clinical trials; only one study (Look AHEAD) solely recruited T2D cases. As a validation of samples included in this study, 22 of 25 studies of European ancestry showed the expected association of TCF7L2 SNPs at \( p < 0.05 \) in the expected direction (Table S8), and 25 of 27 known T2D index SNPs or close proxies were replicated with \( p < 0.05 \) as described. A particular strength of our study as compared to other large-scale meta-analyses of T2D is that all participating studies performed association analyses adjusted for sex, age, and BMI, reducing confounding effects.

We identified a novel diabetes-associated locus at GATAD2A/CILP2/PBX4, included on the IBC array because it was previously described as being associated with lipid levels.\(^{39} \) GATAD2A rs3794991 is in strong LD \(( r^2 > 0.90 \) in HapMap CEU) with rs16996148 (near CILP2/PBX4), previously associated with low-density lipoprotein (LDL) cholesterol and triglycerides (TGs) in meta-analysis of GWAS\(^{67} \) and in previous IBC studies of TG levels.\(^{45} \) Interestingly, the LDL and TG-lowering allele is associated with increased risk of T2D, reminiscent of the contrasting effects on T2D and triglyceride levels of the pleiotropic missense P446L SNP in GCKR.\(^{68} \) GATAD2A encodes the GATA zinc finger domain containing 2A, a transcriptional repressor that interacts with the methyl-CpG-binding domain proteins MBD2 (MIM 603547) and MBD3 (MIM 603573). Methyl-CpG-binding domain proteins mediate functional responses of methylated DNA. PBX4 encodes cartilage intermediate layer protein 2. Further studies are needed to establish the genetic variants that contribute to diabetes and lipid traits and to establish the causal mechanisms at this locus.

T2D-associated SNPs at the second genome-wide significant locus, the HLA-DQB1 region, have previously been strongly associated with type 1 diabetes (T1D) as one of four independent signals from the HLA region.\(^{65} \) This signal most likely represents association with LADA in both the IBC array and DIAGRAM T2D datasets; among most T2D studies, approximately 10% of “cases” are actually misdiagnosed individuals with LADA.\(^{66} \) An alternative hypothesis is that T1D risk alleles confer a much weaker risk to T2D.\(^{70} \) In this study, we show that the association of HLA SNPs with T2D is most likely an artifact caused by the inclusion of misdiagnosed individuals with LADA in our study, as indicated both by the comparison of DIAGRAM component studies with and without LADA cases and fine mapping of the major signal in the HLA SNPs rs13266634, which alters codon R324 (MAF = 0.10; \( p = 0.017 \)) and was rare in Europeans (MAF = 0.02, \( p = 0.28; \) Figures S2A–S2C).

In a multiethnic meta-analysis of all available IBC case-control samples, 14 loci retained study-wide significance (Table S7), and a variant in BCL2 attained genome-wide significance (rs12454712T OR = 1.09 [95% confidence interval (CI): 1.05–1.11], \( p = 2.1 \times 10^{-9} \)). Concordance in direction of effect was observed for 27 of 40 T2D association signals between Europeans and African Americans (\( p = 0.011 \)). Thirty-two SNPs were concordant in direction of effect between Europeans and Hispanics (\( p = 7.0 \times 10^{-5} \)), and 27 SNPs were concordant between Europeans and Asians (\( p = 0.011 \)). A total of 17 out of 40 SNPs were concordant across all four ethnic groups. A combined genetic risk score, comprising 26 previously described, robustly T2D-associated variants (listed in Table S1) and the genome-wide-significant GATAD2A-region signal discovered in this study, demonstrated a significant per-allele risk effect in African-American (OR = 1.06 [95% CI: 1.04–1.08]; \( p < 10^{-10} \)), Hispanic (OR = 1.07 [95% CI: 1.03–1.12]; \( p < 10^{-3} \)), and Asian cohorts (OR = 1.12 [95% CI: 1.04–1.22]; \( p < 10^{-3} \)) from the CARe study (Table 4), suggesting overlap of causal T2D risk alleles across multiple ethnic groups.

**Discussion**

This study reports a large meta-analysis of T2D-candidate-gene association studies and has identified three additional

### Table 4. Prediction of Risk of T2D in CARe African-American, Hispanic, and Asian Populations via a Weighted Genetic Score of 27 T2D Variants

<table>
<thead>
<tr>
<th>Additive Risk per Allele</th>
<th>African-American</th>
<th>Hispanics</th>
<th>Asians</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (cases/controls)</td>
<td>1801/7253</td>
<td>297/874</td>
<td>109/529</td>
</tr>
<tr>
<td>OR (95% CI)</td>
<td>1.06 (1.04–1.08)</td>
<td>1.07 (1.03–1.12)</td>
<td>1.12 (1.04–1.22)</td>
</tr>
<tr>
<td>p</td>
<td>( 1.07 \times 10^{-10} )</td>
<td>0.0021</td>
<td>0.0047</td>
</tr>
</tbody>
</table>

**Quartiles of Risk Alleles**

<table>
<thead>
<tr>
<th>Q1 OR (95% CI)</th>
<th>1.00 (ref)</th>
<th>1.00 (ref)</th>
<th>1.00 (ref)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2 OR (95% CI)</td>
<td>1.13 (0.96–1.33)</td>
<td>1.06 (0.69–1.61)</td>
<td>1.77 (0.88–3.57)</td>
</tr>
<tr>
<td>p</td>
<td>0.14</td>
<td>0.802</td>
<td>0.114</td>
</tr>
<tr>
<td>Q3 OR (95% CI)</td>
<td>1.45 (1.24–1.70)</td>
<td>1.28 (0.85–1.94)</td>
<td>1.64 (0.81–3.33)</td>
</tr>
<tr>
<td>p</td>
<td>5.01 \times 10^{-6}</td>
<td>0.24</td>
<td>0.17</td>
</tr>
<tr>
<td>Q4 OR (95% CI)</td>
<td>1.53 (1.31–1.80)</td>
<td>1.69 (1.12–2.57)</td>
<td>2.63 (1.37–5.05)</td>
</tr>
<tr>
<td>p</td>
<td>1.90 \times 10^{-7}</td>
<td>0.013</td>
<td>0.0035</td>
</tr>
</tbody>
</table>

The 27 T2D variants on the IBC chip include 26 T2D-associated variants in Table S1 and the lead SNP in the GATAD2A region. p = 0.34 for heterogeneity of allelic effects across ethnic groups.

SNP rs13266634, which alters codon R324 (MAF = 0.10; \( p = 0.017 \)) and was rare in Europeans (MAF = 0.02, \( p = 0.28; \) Figures S2A–S2C).
region. If we assume an OR of 5.49 for the HLA signal in the LADA cases,\textsuperscript{71} we estimate the percentage of LADA cases in our total cases set to be 7.6%, consistent with previous estimates.\textsuperscript{69}

A second T1D risk allele (rs10770141, located in the promoter region of tyrosine hydroxylase [TH] and 11 kb upstream of the insulin (INS) gene) was study-wide significant, but interestingly, this T1D risk allele was protective for T2D. Although located close to the insulin gene, this SNP regulates expression of TH; the T2D risk allele increases expression, and the T1D risk allele lowers expression.\textsuperscript{72,73} The signal is not correlated with a minisatellite upstream of the insulin gene; this minisatellite has previously been inconsistent associated with risk of T2D.\textsuperscript{74,75} Replication in independent cohorts will be important for the validation of this finding, and functional studies will be required to establish whether causal variants at this locus indeed act antagonistically to contribute to type 1 and type 2 diabetes. No significant association with T2D was found for other SNPs that had genome-wide-significant associations with T1D (Table S9).

The fourth signal resides in 

SREBF1,

which encodes the transcription factors sterol-regulatory-element-binding protein (SREBP)-1a and -1c. Whereas SREBP1a is ubiquitously expressed, SREBP1c is particularly expressed in insulin-sensitive tissues such as liver and adipose. SREBP1a and SREBP1c control lipid synthesis and glucose metabolism by regulating the expression of key genes involved in glucose, fatty acid, and triglyceride metabolism.\textsuperscript{76,77} Variants in SREBF1 have previously been shown to be associated with T2D in several candidate-gene studies.\textsuperscript{50,61,78–80} All reported associated variants are in substantial LD with our most significantly associated SNP, rs4925115 (1000 Genomes Pilot 1: $r^2 = 0.81$–0.89). However, thus far GWASs have not detected SREBF1 as a T2D locus, perhaps because most GWASs did not systematically adjust for the confounding factors of sex, age and BMI, and these adjustments were demonstrated to be essential for detection of this association in a previous study.\textsuperscript{60} Specifically, in the DIAGRAM meta-analysis, most component studies did not adjust for sex, age, and BMI, providing a likely explanation as to why the SREBF1 signal was not strongly replicated (Table 1).

T2D association signals that were close to study-wide significance include compelling candidates previously confirmed as associated with glycemic diabetes-related traits; such candidates include MADD (MIM 603584) near FOLH1 (MIM 600934),\textsuperscript{25} the glucose transporter 2 isoform SLC2A2 (MIM 138160),\textsuperscript{25} and gastric inhibitory polypeptide receptor GIPR (MIM 137241).\textsuperscript{66} These suggestive findings are consistent with the hypothesis that a large number of common variants and genes with modest effects contribute to the risk of T2D and that current studies are underpowered to detect these effects. Indeed, despite the large sample sizes assembled here, a limitation of our study is that significant and borderline-significant associations based on combined analysis of the European IBC and DIAGRAM datasets will need additional independent replication and follow-up.

Multiethnic meta-analysis across all 39 studies identified association of a common variant in BCL2 with T2D, an anti-apoptotic protein that has not previously been implicated in this disease. Meta-analysis of IBC array studies in African Americans also identified study-wide-significant association of independent alleles at HMGA2, suggesting that this gene might be causal across different ethnicities. Our observations that a genetic risk score of 27 variants associates with a risk of T2D in African Americans, Asians, and Hispanic populations is also consistent with the idea that at least a subset of T2D causal risk alleles spans ethnicities. The genetic risk score analysis confirms and extends findings in a previous study by Waters et al.,\textsuperscript{34} who report consistent association of 19 European T2D risk variants in several racial and ethnic groups from the US. Notably, our study included additional correction for population structure, which may contribute to the slightly reduced effect sizes observed.

In addition to common variants, the IBC array was selected to capture rare missense variants at select loci that are not directly captured or imputable by conventional GWASs. Robust association of an established missense SNP in HNF1a (rs1800574) with MAF = 0.029 ($p = 1.4 \times 10^{-7}$ in samples of European ancestry; Table S7)\textsuperscript{13} validates the platform’s ability to detect rare variation that contributes to the risk of T2D. Although no rare variants reached study-wide significance, several follow-up candidate variants in known T2D loci were identified in African Americans.

In conclusion, this large-scale gene-centric meta-analysis of 39 multiethnic T2D association studies identified three European T2D risk loci (GATAD2A/CILP2/PBX4, previously known to have protective effects on lipids; TH/INS, previously known to have protective effects on T1D, and SREBF1), one African-American T2D risk locus (HMGA2), and one multiethnic risk locus (BCL2) and confirmed that a genetic score of T2D risk alleles influence risk of T2D in multiethnic populations including African-Americans, Hispanics, and Asians. Thus, well-powered, multiethnic GWASs of T2D should lead to the discovery of additional diabetes-associated genes relevant to multiple ethnic groups.

Supplemental Data

The Supplemental Data include acknowledgments and funding information for each cohort, a list of members of the DIAGRAM consortium and Look AHEAD research group, disclosures, nine tables, and two figures.

Acknowledgments

The CARe Consortium wishes to acknowledge the support of the National Heart, Lung, and Blood Institute and the contributions of the research institutions, study investigators, field staff, and study participants in creating this resource for biomedical
research. The following nine parent studies have contributed parent study data, ancillary study data, and DNA samples through the Broad Institute of Harvard University and the Massachusetts Institute of Technology (N01-HC-65226) to create this genotype/phenotype database for wide dissemination to the biomedical research community: the Atherosclerosis Risk in Communities (ARIC) study, the Cardiovascular Health Study (CHS), the Cleveland Family Study (CFS), the Cooperative Study of Sickle Cell Disease (CSSCD), the Coronary Artery Risk Development in Young Adults (CARDIA) study, the Framingham Heart Study (FHS), the Jackson Heart Study (JHS), the Multi-Ethnic Study of Atherosclerosis (MESA), and the Sleep Heart Health Study (SHHS). R.S. is partially supported by NIDDK NIH R21 (DK089378). C.C.E is supported by a Rubicon grant from the Netherlands Organization for Scientific Research (NWO). Additional acknowledgements for each cohort are listed in the Supplemental Data. Also see the Supplemental Data for conflict-of-interest disclosures for some of the authors.

Received: August 22, 2011
Revised: December 6, 2011
Accepted: December 31, 2011
Published online: February 9, 2012

Web Resources
The URLs for data presented herein are as follows:
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org

References


