A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants

Laura J. Scott,1 Karen L. Mohlke,2 Lori L. Bonnycastle,3 Cristen J. Willer,4 Yun Li,5 William L. Duren,6 Michael R. Erdos,7 Heather M. Stringham,8 Peter S. Chines,9 Anne U. Jackson,10 Ludmila Prokunina-Olsson,11 Chia-Jen Ding,12 Amy J. Swift,13 Narisu Narisu,14 Tienle Hu,15 Randall Pruim,16 Rui Xiao,17 Xiao-Yi Li,18 Karen N. Connelly,19 Nancy L. Riebow,20 Andrew G. Sprau,21 Maurine Tong,22 Peggy P. White,23 Kurt N. Hetrick,24 Michael W. Barnhart,25 Craig W. Bark,26 Janet L. Goldstein,5 Lee Watkins,5 Fang Xiang,7 Jouko Saramies,6 Andrew G. Sprau,3 Maurine Tong,3 Peggy P. White,1 Kurt N. Hetrick,5 Michael W. Barnhart,5 William L. Duren,1 Michael R. Erdos,3 Heather M. Stringham,1 Peter S. Chines,3

Identifying the genetic variants that increase the risk of type 2 diabetes (T2D) in humans has been a formidable challenge. Adopting a genome-wide association strategy, we genotyped 1161 Finnish T2D cases and 1174 Finnish normal glucose-tolerant (NGT) controls with >315,000 single-nucleotide polymorphisms (SNPs) and imputed genotypes for an additional >2 million autosomal SNPs. We carried out association analysis with these SNPs to identify genetic variants that predispose to T2D, compared our T2D association results with the results of two similar studies, and genotyped 80 SNPs in an additional 1215 Finnish T2D cases and 1258 Finnish NGT controls. We identify T2D-associated variants in an intergenic region of chromosome 11p12, contribute to the identification of T2D-associated variants near the genes IGF2BP2 and CDKAL1 and the region of CDKN2A and CDKN2B, and confirm that variants near TCF7L2, SLCOA8, HHEX, FTO, PPARG, and KCNJ11 are associated with T2D risk. This brings the number of T2D loci now confidently identified to at least 10.

Type 2 diabetes (T2D) is a disease characterized by insulin resistance and impaired pancreatic beta-cell function that affects >170 million people worldwide (1). With first-degree relatives having ~3.5 times as much risk as compared to individuals in the general middle-aged population (2), hereditary factors, together with lifestyle and behavioral factors, play an important role in determining T2D risk (3). To date, intense efforts to identify genetic risk factors in T2D have met with only limited success. This study, reports from our collaborators (4–6), and the recently published work of Sladek et al. (7) describe results of genome-wide association (GWA) studies that further define the genetic architecture of T2D and identify biological pathways involved in T2D pathogenesis.

We genotyped 1161 Finnish T2D cases and 1174 Finnish NGT controls on 317,503 SNPs on the Illumina HumanHap300 BeadChip in stage 1 of a two-stage GWA study of T2D (8). These samples are from the Finland–United States Investigation of Non–Insulin-Dependent Diabetes Mellitus Genetics (FUSION) (9, 10) and Finrisk 2002 (11) studies (tables S1 and S2A). Among the 317,503 GWA SNPs, 315,635 had >10 copies of the less common allele [minor allele frequency (MAF) > 0.002] and passed quality-control criteria (8). We tested these 315,635 SNPs for association with T2D using a model that is additive on the log-odds scale (Table 1 and tables S3 and S4) (8). We observed a modest excess (41 observed versus 31.6 expected; P = 0.19) of SNPs with P values < 10−4 (fig. S1). These results argue against the existence of multiple common SNPs with a large impact on T2D disease risk but are consistent with the presence of multiple common SNPs that each confer modest risk. The results also suggest that the matching of cases and controls by birth province, sex, and age (8) has been successful; in support of this conclusion, the genomic control (12) correction value is 1.026.

Analysis of our Illumina HumanHap300 data allowed us to query much of the known SNP variation in the genome. To increase this proportion, we developed an imputation method (6, 13) that uses genotype data and linkage disequilibrium (LD) information from the HapMap Centre d’Etude du Polymorphisme Humain (Utah residents with ancestry from northern and
Table 1. Confirmed T2D susceptibility loci based on all available data from the FUSION, DGI, and WTCCC/UKT2D samples.

<table>
<thead>
<tr>
<th>New T2D Loci</th>
<th>FUSION Chr (bp)</th>
<th>Genes</th>
<th>Risk allele / nonrisk allele freq.</th>
<th>FUSION stage 1</th>
<th>FUSION stage 2</th>
<th>FUSION stage 1 + 2</th>
<th>DGI All Data</th>
<th>WTCCC/UKT2D All Data</th>
<th>Total sample size for 80% power**</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4402960</td>
<td>3</td>
<td>IGF2BP2</td>
<td>T/G</td>
<td>0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7754840†</td>
<td>6</td>
<td>CDKAL1</td>
<td>C/G</td>
<td>0.36</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10811661†</td>
<td>9</td>
<td>CDKN2A/B</td>
<td>T/C</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs9300039††</td>
<td>11</td>
<td>CDKN2A/B</td>
<td>C/A</td>
<td>0.89</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs8050136</td>
<td>16</td>
<td>FTO</td>
<td>A/C</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1801282‡‡</td>
<td>3</td>
<td>PPARG</td>
<td>C/G</td>
<td>0.82</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13266634‡‡</td>
<td>8</td>
<td>SLC30A8</td>
<td>C/T</td>
<td>0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1111875††</td>
<td>10</td>
<td>HHEX</td>
<td>C/T</td>
<td>0.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs7903146‡‡</td>
<td>10</td>
<td>TCF7L2</td>
<td>T/C</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs5219††</td>
<td>11</td>
<td>KCNJ11</td>
<td>T/C</td>
<td>0.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Total sample size

Number of cases/controls

| FUSION-DGI-WTCCC/UKT2D | 2,335 | 2,473 | 4,808 | 13,781 | 13,965 | 32,544 |

**Approximate total sample size for 80% power to detect T2D SNP association at significance level 0.05 is based on the FUSION control risk allele frequency and the risk ratio calculated from FUSION-DGI-WTCCC/UKT2D all-data analyses, assuming 0.10 T2D prevalence. The sample sizes vary slightly from those of (4) because study-specific allele frequencies were used in the calculations.

*rs10946398 WTCCC/UKT2D (r² = 1). †Multimarker tag for rs9300039 DGI and rs1514823 WTCCC/UKT2D (r² = 0.965). ‡rs5015480 WTCCC GWA only (r² = 1). §rs7901695 WTCCC/UKT2D (r² = 0.849). ¶rs5215 WTCCC/UKT2D (r² = 0.995). ††DGI GWA samples. ‡‡WTCCC GWA samples.
western Europe) (CEU) samples to predict genotypes of autosomal SNPs not genotyped in our subjects. A total of 2.09 million HapMap CEU SNPs (14) had imputed MAF >1% in FUSION and passed our imputation quality-control criteria. In the HapMap CEU sample, imputed SNPs passing these criteria increased coverage of SNPs with MAF >1% from 71.9 to 89.1% at an \( r^2 \) threshold of 0.8.

To increase the statistical power to detect T2D predisposing variants, we compared our stage 1 results to GWA results from the Diabetes Genetics Initiative (DGI) and the Wellcome Trust Case Control Consortium (WTCCC). We selected 82 SNPs for FUSION stage 2 follow-up genotyping based on evidence from: (i) FUSION-genotyped and FUSION-imputed SNPs; (ii) a combined analysis of GWA results from FUSION, DGI, and WTCCC; and (iii) previous T2D association results. For (i) and (ii), we used a prioritization algorithm that advantaged SNPs based on genome annotation (8) (table S7) and gave preference to genotyped SNPs over nearby imputed SNPs. We successfully genotyped 80 of the 82 SNPs in our stage 2 sample of 1215 Finnish T2D cases and 1258 Finnish NGT controls (8) (table S2B) and carried out joint analysis of the combined FUSION stage 1 + 2 sample (table S5). DGI (4) and United Kingdom T2D Genetics Consortium (UKT2D) (5) investigators also followed up DGI and WTCCC GWAs by genotyping replication samples.

We confirmed well-established T2D associations with TCF7L2, PPARG, and KCNJ11 (Table 1) (15–18). SNPs in TCF7L2 reached genome-wide significance in the FUSION stage 1 + 2 sample [odds ratio (OR) = 1.34, \( P = 1.3 \times 10^{-8} \)] and in the FUSION-DGI-WTCCC/UKT2D “all-data” (i.e., all GWA and follow-up samples) meta-analysis (OR = 1.37, \( P = 1.0 \times 10^{-8} \)) (Table 1 and table S5). PPARG Pro12→Ala12 (rs1801282) and KCNJ11 Glu23→Lys23 (rs5219) were not genotyped in the FUSION GWA, but nearby SNPs showed some evidence for T2D association, as did the imputed genotypes for the coding variants. All-data meta-analysis resulted in genome-wide significant T2D association with KCNJ11 Glu23→Lys23 (OR = 1.14, \( P = 6.7 \times 10^{-11} \)) and strong evidence for PPARG Pro12→Ala12 (OR = 1.14, \( P = 1.7 \times 10^{-9} \)). The PPARG and KCNJ11 results emphasize the value of combining data across studies and suggest that other T2D-associated loci remain to be found.

The combined samples from the three studies provide evidence for seven additional T2D loci. For the first three of these loci, we had strong evidence in the FUSION stage 1 GWA data and, for the latter four, our FUSION stage 1 evidence was more modest.

A cluster of variants in the IGF2BP2 (insulin-like growth factor 2 mRNA binding protein 2) region was associated with T2D in our stage 1 sample \( (e.g., \text{rs}1470579 \text{ with } \text{OR} = 1.27, \text{ } P = 1.6 \times 10^{-8}) \) (Fig. 1A). The all-data meta-analysis for \( \text{rs}4402960 \) resulted in genome-wide significance \( (\text{OR} = 1.14, \text{ } P = 8.9 \times 10^{-10}) \). Including the \( \text{rs}4402960 \) genotype as a covariate essentially eliminates evidence for T2D association for other variants in the cluster (Fig. 1A), which is consistent with all SNPs representing the same T2D-predisposing variant(s). IGF2BP2 is a paralog of IGF2BP1, which binds to the 5′ untranslated region of the insulin-like growth factor 2 (IGF2) mRNA and regulates IGF2 translation (19). IGF2 is a member of the insulin family of polypeptide growth factors involved in the development, growth, and stimulation of}

![Fig. 1. Plots of T2D association and LD in FUSION stage 1 samples for regions surrounding IGF2BP2 (A) and rs9300039 (B). (A) and (B) each contain six panels. The top panels display RefSeq genes; there are none in the rs9300039 region. The second panels (i.e., directly below the top panels) show the T2D association \( -\log_{10} P \) values in FUSION stage 1 samples for SNPs genotyped in the GWA panel (closed blue circles) or imputed (open blue circles). The third panels show T2D association \( -\log_{10} P \) values for each SNP in a logistic regression model correcting for the reference SNP [indicated by the red circle for rs4402960 in (A) and for rs9300039 in (B)]. SNP rs7480010, reported by Sladek et al. (7), is also labeled in the rs9300039 plot (B) (green circle). A decrease in the \( -\log_{10} P \) value from the second to the third panels indicates that the association signal of the tested SNPs can be explained, at least in part, by the reference SNP. In both regions, the reference SNP was chosen for convenience; the choice of another strongly associated SNP nearby would have resulted in a similar picture. The fourth panels show recombination rate in centimorgans per megabase for the HapMap CEU sample (14). The fifth and sixth panels show LD \( r^2 \) and \( D' \) based on FUSION stage 1–genotyped and FUSION stage 1–imputed data.](www.sciencemag.org)
insulin action. The most strongly associated \textit{IGF2BP2} SNPs are located in a 50-kb region within intron 2 (Fig. 1A); diabetes-predisposing variants may therefore affect regulation of \textit{IGF2BP2} expression.

SNP rs13266634, a nonsynonymous Arg\textsuperscript{235→Trp} variant in the pancreatic beta-cell-specific zinc transporter \textit{SLC30A8} (20), showed (through our annotation-based algorithm) evidence for T2D association in stage 1 (Table 1 and fig. S2). Modest evidence in stage 2 resulted in stronger evidence in our stage 1 + 2 sample (OR = 1.18, \(P = 7.0 \times 10^{-5}\)) (Table 1 and table S5). Subsequent DGI and UKT2D genotyping resulted in strong evidence in the combined samples (OR = 1.12, \(P = 5.3 \times 10^{-8}\)). Sladek et al. (7) recently reported independent T2D association evidence with the same allele in two French samples (\(P = 1.8 \times 10^{-5}\) and \(P = 5.0 \times 10^{-7}\)). SLC30A8 transports zinc from the cytoplasm into insulin secretory vesicles (20, 21), where insulin is stored as a hexamer bound with two \(\text{Zn}^{2+}\) ions before secretion (22). Variation in \textit{SLC30A8} may affect zinc accumulation in insulin granules, affecting insulin stability, storage, or secretion. In high-glucose conditions, overexpression of \textit{SLC30A8} in insulinoma (INS-1E) cells enhanced glucose-induced insulin secretion (21).

SNP rs9300039 in an intergenic region on chromosome 11 showed evidence for T2D association in stage 1 (Table 1 and Fig. 1B); genotyping our stage 2 sample resulted in stronger evidence in our stage 1 + 2 sample (OR = 1.48, \(P = 5.7 \times 10^{-8}\)) (Table 1 and tables S3 and S5). In the WTCCC and DGI scans, the nearby SNP rs1514823 (\(r^2 = 0.97\) with rs9300039) provided weak evidence for T2D association with the appropriate allele; combining results across all three studies gave OR = 1.25 and \(P = 4.3 \times 10^{-7}\). Fifty-six imputed SNPs and two more genotyped SNPs spanning 219 kb are in LD with rs9300039 and show substantial evidence for T2D association (\(P < 10^{-8}\)) in our stage 1 sample (table S3 and Fig. 1B). Including the genotype for rs9300039 as a covariate essentially eliminates evidence for T2D association with the remaining SNPs (Fig. 1B). This region includes three sets of spliced Expressed Sequence Tags but no annotated genes. The identification of a T2D-associated variant \(>1\) Mb from the nearest annotated gene highlights the value of a genome-wide approach. Sladek et al. (7) reported strongly associated SNPs in two nearby regions on chromosome 11. SNP rs7480010 near hypothetical gene \textit{LOC387761} is 331 kb centromeric to rs9300039. LD between rs9300039 and rs7480010 is essentially zero (\(r^2 = 0.00063\) and \(D' = 0.936\)), and rs7480010 showed little evidence for association in our stage 1 + 2 sample (OR = 1.03, \(P = 0.54\)). Sladek et al. (7) also reported T2D association with three intronic variants of \textit{EXT2}, located \(2.4\) Mb centromeric of rs9300039; we found no evidence for association with \textit{EXT2} SNPs.

SNP rs4712523, located within intron 5 of \textit{CDKAL1}, showed modest evidence for T2D association in our FUSION stage 1 sample, which strengthened slightly in our combined stage 1 + 2 sample (OR = 1.12, \(P = 0.0073\)) (table S5). Nearby SNPs in strong LD with rs4712523 including rs7754840 showed modest evidence for T2D association in the DGI scan and considerably stronger evidence in the WTCCC scan. Including strong DGI and UKT2D replication data resulted in genome-wide significance (OR = 1.13, \(P = 5.7 \times 10^{-10}\)) in the all-data meta-analysis. These SNPs are in LD (\(r^2 = 0.70\)) in a region that includes \textit{HHEX} (hematopoietically expressed homeobox), which is critical for development of the ventral pancreas (30), the insulin-degrading enzyme gene \textit{IDE}, and the kinesin-interacting factor 11 gene \textit{KIF11}. Sladek et al. (7) recently reported independent genome-wide significant evidence for T2D association with these SNPs.

The WTCCC/UKT2D groups identified evidence for T2D and body mass index (BMI) associations with a set of SNPs including rs8050136 in the \textit{FTO} region; the T2D association appears to be mediated through a primary effect on adiposity (5, 6, 31). We observed modest evidence for association with T2D in the combined FUSION stage 1 + 2 sample (OR = 1.11, \(P = 0.016\)) (Table 1 and table S5).

T2D can be a component of a larger syndrome of metabolic abnormalities, and we were interested to assess the effects of T2D-related traits on our association results. We repeated our T2D association analysis for the 10 SNPs in Table 1 with one of several variables included as an additional covariate. Adjustment for BMI strengthened T2D association with \textit{TCF7L2} and \textit{SLC30A8}, weakened association with rs9300039 and \textit{FTO}, and had little effect on the other loci. The effect of waist circumference was similar to that of BMI; blood pressure variables had essentially no effect.
We previously carried out T2D linkage analysis in the families of many of our stage 1 cases (10). None of the 10 loci in Table 1 had large T2D logarithm of the odds (LOD) scores, although those for FTO and TCFTL2 were 0.63 and 0.60 and so were nominally significant. LOD scores for six of the 10 loci were greater than 0.2, as compared to 2.2 that would be expected for random genome locations. This suggests enrichment for T2D-associated loci in regions with modest evidence of T2D linkage (P = 0.01) but that the power of the linkage approach was insufficient to distinguish these signals from background noise.

The ability to construct a list of ten robust and replicated T2D-associated loci (Table 1) represents a landmark in efforts to identify genetic variants that predispose to complex human diseases, although the specific predisposing variants and even the relevant genes remain to be defined. We examined the combined risk of T2D based on these 10 loci in our stage 1 + 2 sample by constructing a logistic regression model and predicting T2D risk for each person (8). We found a fourfold variation in T2D risk from the lowest to highest predicted risk groups, which is of potential interest for a personalized preventive-medicine program (Fig. 2). However, these predictions from our data may be biased as compared to predictions based on the general population, likely owing to the overestimation of ORs due to the “winner’s curse,” enrichment for familial T2D cases, and exclusion of individuals with impaired glucose tolerance or impaired fasting glucose.

Thirty years ago, James V. Neel labeled T2D as “the geneticist’s nightmare” (32), predicting that the discovery of genetic factors in T2D would be thoroughly challenging. Until recently, his prediction has proven true. Although large samples and collaboration among three groups were required, we can confidently state that new samples and collaboration among three groups would be thoroughly challenging. Until recently, 1Department of Medicine, University of Cambridge, Cambridge, U.K. 2Section for Infection and Immunity, College of Medicine, University of Wales, Heath Park, Cardiff, CF14 4XX, UK.

**Complex I Binding by a Virally Encoded RNA Regulates Mitochondria-Induced Cell Death**

Matthew B. Reeves,1* Andrew A. Davies,1 Brian P. McSharry,2 Gavin W. Wilkinson,2 John H. Sinclair†

Human cytomegalovirus infection perturbs multiple cellular processes that could promote the release of proapoptotic stimuli. Consequently, it encodes mechanisms to prevent cell death during infection. Using rotenone, a potent inhibitor of the mitochondrial enzyme complex I (reduced nicotinamide adenine dinucleotide—ubiquinone oxido-reductase), we found that human cytomegalovirus infection protected cells from rotenone-induced apoptosis, a protection mediated by a 2.7-kilobase virally encoded RNA (β2.7). During infection, β2.7 RNA interacted with complex I and prevented the relocation of the essential subunit genes associated with retinoid/interferon—induced mortality—19, in response to apoptotic stimuli. This interaction, which is important for stabilizing the mitochondrial membrane potential, resulted in continued adenosine triphosphate production, which is critical for the successful completion of the virus’ life cycle. Complex I targeting by a viral RNA represents a refined strategy to modulate the metabolic viability of the infected host cell.

**References and Notes**

8. Materials and methods are available as supporting material on Science Online.

DURING primary infection or reactivation of human cytomegalovirus (HCMV), especially in the immunocompromised, the virus is able to replicate in a number of cell types, often resulting in life-threatening disease (1). HCMV exhibits a relatively protracted life cycle (upwards of 5 days) and at early times of infection (12 to 24 hours) encodes a highly abundant 2.7-kb RNA transcript (β2.7), accounting for >20% of total viral gene transcription (2, 3) of unknown function. The RNA may be associated with mitochondria (4), and no protein product of this RNA has ever been detected in infected cells (5), suggesting that it functions as a noncoding RNA (5).

We investigated the possibility that β2.7 could function as a noncoding RNA. A

*Department of Medicine, University of Cambridge, Addenbrooke’s Hospital, Hills Road, Cambridge, CB2 2QO, UK. 2Section for Infection and Immunity, College of Medicine, University of Wales, Heath Park, Cardiff, CF14 4XX, UK.

†Present address: Novartis Institutes for Biomedical Research, 500 Technology Square, Cambridge, MA 02139, USA.

To whom correspondence should be addressed. E-mail: js@ mole.bio.cam.ac.uk