New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk

Yarnell, J. (2010). New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. Nature Genetics, 42(2), 105-119. https://doi.org/10.1038/ng.520

Published in:
Nature Genetics

Queen's University Belfast - Research Portal:
Link to publication record in Queen's University Belfast Research Portal

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.
New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk

Levels of circulating glucose are tightly regulated. To identify new loci influencing glycemic traits, we performed meta-analyses of 21 genome-wide association studies informative for fasting glucose, fasting insulin and indices of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) in up to 46,186 nondiabetic participants. Follow-up of 25 loci in up to 76,558 additional subjects identified 16 loci associated with fasting glucose and HOMA-B and two loci associated with fasting insulin and HOMA-IR. These include nine loci newly associated with fasting glucose (in or near ADCY5, MAD2, ADRA2A, CRY2, FADS1, GLI53, SLC2A2, PROX1 and C2CD4B) and one influencing fasting insulin and HOMA-IR (near IGF1). We also demonstrated association of ADCY5, PROX1, GCK, GCKR and DGKB-TMEM195 with type 2 diabetes. Within these loci, likely biological candidate genes influence signal transduction, cell proliferation, development, glucose-sensing and circadian regulation. Our results demonstrate that genetic studies of glycemic traits can identify type 2 diabetes risk loci, as well as loci containing gene variants that are associated with a modest elevation in glucose levels but are not associated with overt diabetes.

Impaired beta-cell function and insulin resistance are key determinants of type 2 diabetes (T2D). Hyperglycemia in the fasting state is one of the criteria that defines T2D\(^1\), it can predict definitive clinical endpoints in nondiabetic individuals\(^2,3\) and, when corrected in subjects with T2D, may help prevent microvascular\(^4,5\) and long-term macrovascular\(^6,7\) complications. To date, there are nearly 20 published loci reproducibly associated with T2D\(^8\); most of these are also associated with decreased insulin secretion\(^9\) due to defective beta-cell function or beta-cell mass. Association studies for diabetes-related quantitative traits in participants without diabetes have also identified loci influencing fasting glucose levels, whose effects appear to be mediated by impairment of the glucose-sensing machinery in beta cells\(^10\)-\(^17\).

We recently formed the Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) to conduct large-scale meta-analyses of genome-wide data for continuous diabetes-related traits in participants without diabetes\(^13\). We aimed to identify additional loci that influence glycemic traits in individuals free of diabetes and investigate their impact on related metabolic phenotypes. We were also interested in understanding variation in the physiological range of glycemia and evaluating the extent to which the same variants influence pathological fasting glucose variation and T2D risk. The initial MAGIC collaboration identified the fasting glucose- and T2D-associated locus in MTNR1B\(^13\), which was also reported by others\(^16,17\); this finding demonstrated that studies of continuous glycemic phenotypes in nondiabetic individuals can complement the genetic analyses of diabetes as a dichotomous trait and can improve our understanding of the mechanisms involved in beta-cell function and glucose homeostasis. Here, we extend our previous approach by performing meta-analyses of \(-2.5\) million directly genotyped or imputed autosomal SNPs from 21 genome-wide association studies (GWAS). These 21 cohorts include up to 46,186 nondiabetic participants of European descent informative for fasting glucose and 20 GWAS including up to 38,238 nondiabetic individuals informative for fasting insulin, as well as the surrogate estimates of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) derived from fasting variables by homeostasis model assessment\(^18\). Follow-up of 25 lead SNPs in up to 76,558 additional individuals of European ancestry identified nine new genome-wide significant associations (empirically determined as \(P < 5 \times 10^{-8}\))\(^9\) with fasting glucose and one with fasting insulin and HOMA-IR. Five of these loci also demonstrated genome-wide significant evidence for association between the glucose-raising allele and T2D risk in up to 40,655 cases and 87,022 nondiabetic controls.

The wealth of loci newly discovered to be associated with fasting glucose and HOMA-B contrasts with the single new locus identified for fasting insulin and HOMA-IR and suggests that there is a different genetic architecture for beta-cell function and insulin resistance. Furthermore, our data support the hypothesis that not all loci that influence glycemia within the physiological range are also associated with pathological levels of glucose and T2D risk.

RESULTS

Genome-wide association meta-analysis of glycemic traits

We conducted a two-stage association study in individuals of European descent (Online Methods, Supplementary Fig. 1 and Supplementary Table 1a,b). Because we sought to identify variants that influence fasting glucose in the unaffected population, hyperglycemia in the diabetic range exerts deleterious effects on beta-cell function\(^20,21\) and treatment can confound glucose and insulin measurements, we excluded individuals with known diabetes, those on anti-diabetic treatment, and those with fasting glucose \(\geq 7\) mmol/l. We combined data from 21 stage 1 discovery GWAS for fasting glucose (\(n = 46,186\)) and 20 GWAS for fasting insulin (\(n = 38,238\)), HOMA-B (\(n = 36,466\))

\(^{1}\)A full list of authors and affiliations appears at the end of the paper.

Received 13 August 2009; accepted 14 October 2009; published online 17 January 2010; doi:10.1038/ng.520
Table 1 SNPs associated with fasting glucose-related or insulin-related traits at genome-wide significance levels

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene(s)</th>
<th>Alleles (effect/other)</th>
<th>Freq</th>
<th>Discovery P</th>
<th>Global P</th>
<th>Joint analysis n</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs560887</td>
<td>G6PC2</td>
<td>C/T</td>
<td>0.70</td>
<td>4.4 × 10⁻⁷</td>
<td>0.31 (0.18)</td>
<td>8.7 × 10⁻²³</td>
</tr>
<tr>
<td>rs10830963</td>
<td>MTRNR1B</td>
<td>G/C</td>
<td>0.30</td>
<td>1.2 × 10⁻⁸</td>
<td>0.00 (1.00)</td>
<td>5.8 × 10⁻₁⁷</td>
</tr>
<tr>
<td>rs4407517</td>
<td>GCKR</td>
<td>A/G</td>
<td>0.16</td>
<td>5.3 × 10⁻⁶</td>
<td>0.19 (0.46)</td>
<td>6.5 × 10⁻⁹</td>
</tr>
<tr>
<td>rs2191349</td>
<td>DGKB-TMEM195</td>
<td>T/G</td>
<td>0.52</td>
<td>7.8 × 10⁻¹⁷</td>
<td>0.10 (0.68)</td>
<td>3.0 × 10⁻⁴⁴</td>
</tr>
<tr>
<td>rs780994</td>
<td>GCKR</td>
<td>C/T</td>
<td>0.62</td>
<td>2.5 × 10⁻¹²</td>
<td>0.00 (1.00)</td>
<td>5.6 × 10⁻³⁸</td>
</tr>
<tr>
<td>rs11709067</td>
<td>ADCYS</td>
<td>A/G</td>
<td>0.78</td>
<td>8.7 × 10⁻⁹</td>
<td>0.04 (0.89)</td>
<td>7.1 × 10⁻²²</td>
</tr>
<tr>
<td>rs9744584</td>
<td>MAD2</td>
<td>A/T</td>
<td>0.75</td>
<td>1.5 × 10⁻⁹</td>
<td>0.00 (1.00)</td>
<td>2.0 × 10⁻¹⁸</td>
</tr>
<tr>
<td>rs10885122</td>
<td>ADRA2A</td>
<td>G/T</td>
<td>0.87</td>
<td>8.4 × 10⁻¹¹</td>
<td>0.00 (1.00)</td>
<td>2.9 × 10⁻¹⁶</td>
</tr>
<tr>
<td>rs174550</td>
<td>FADS1</td>
<td>T/C</td>
<td>0.64</td>
<td>1.5 × 10⁻⁸</td>
<td>0.00 (1.00)</td>
<td>1.7 × 10⁻¹⁵</td>
</tr>
<tr>
<td>rs11605924</td>
<td>CYR2</td>
<td>A/C</td>
<td>0.49</td>
<td>1.5 × 10⁻⁹</td>
<td>0.00 (1.00)</td>
<td>1.0 × 10⁻¹⁴</td>
</tr>
<tr>
<td>rs11920090</td>
<td>SLCA2</td>
<td>T/A</td>
<td>0.87</td>
<td>1.9 × 10⁻⁸</td>
<td>0.00 (1.00)</td>
<td>8.1 × 10⁻¹³</td>
</tr>
<tr>
<td>rs7034200</td>
<td>GLIS3</td>
<td>A/C</td>
<td>0.49</td>
<td>1.2 × 10⁻⁸</td>
<td>0.00 (1.00)</td>
<td>1.0 × 10⁻¹²</td>
</tr>
<tr>
<td>rs440874</td>
<td>PROX1</td>
<td>C/T</td>
<td>0.52</td>
<td>7.1 × 10⁻⁸</td>
<td>0.00 (1.00)</td>
<td>6.6 × 10⁻¹²</td>
</tr>
<tr>
<td>rs11071657</td>
<td>C2CD4B</td>
<td>A/G</td>
<td>0.63</td>
<td>2.8 × 10⁻⁷</td>
<td>0.00 (1.00)</td>
<td>3.6 × 10⁻⁸</td>
</tr>
<tr>
<td>rs11558471</td>
<td>SLCA2A</td>
<td>A/G</td>
<td>0.68</td>
<td>2.6 × 10⁻¹¹</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>rs45056565</td>
<td>TCP2L2</td>
<td>T/C</td>
<td>0.31</td>
<td>1.2 × 10⁻⁸</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

**In Table 1, HOMA-B-IR and HOMA-IR:**

- rs780094 and rs35767 were chosen as HOMA-IR loci associated with fasting glucose and/or HOMA-B at genome-wide significance levels.
- rs780094 was chosen as a HOMA-IR locus associated with fasting glucose.
- rs35767 was chosen as a HOMA-IR locus associated with fasting insulin.

**Replication studies and global meta-analysis for 25 loci**

We carried forward to stage 2 all independent loci with association to any of the four traits at P < 2 × 10⁻⁵; we did not include SNPs in the known T2D genes TCF7L2 and SLC30A8, for which no further validation was sought (Table 1 and Supplementary Table 2). We also included the nominally associated top SNP from a likely biological candidate (IRS1, P = 10⁻⁴ for HOMA-IR) and a locus with P values that approached genome-wide significance in several stage 1 discovery cohorts (PLXDC2-NEBL), even though their overall stage 1 P values were > 2 × 10⁻⁵ (Table 1 and Supplementary Table 2). In total, 25 loci were chosen for replication.

We directly genotyped 25 variants in 26 additional stage 2 studies with up to 63,850 nondiabetic participants of European ancestry for fasting glucose and 25 studies and up to 52,892 participants for fasting insulin, HOMA-IR and HOMA-B (Supplementary Table 1b and Online Methods). We also obtained in silico replication data for 12,708 additional individuals from seven studies for fasting glucose (9,372 participants and five studies for fasting insulin, HOMA-IR and HOMA-B), for a total of up to 76,558 individuals for fasting glucose and 62,264 for fasting insulin, HOMA-IR and HOMA-B in stage 2 association analyses.

Our combined stage 1 and 2 meta-analysis, including a total of up to 122,743 participants for fasting glucose (98,372 for fasting insulin, HOMA-IR and HOMA-B), established genome-wide significant associations for nine new loci for fasting glucose and/or HOMA-B (in or near ADCYS5, MADD, ADRA2A, CRY2 and FADS1 (Table 1 and Fig. 1a–j), four previously reported fasting glucose-associated loci in or near GCK, GCKR, G6PC2 and MTRNR1B, the recently reported 24 locus in DGKB-TMEM195, and two loci in the T2D susceptibility genes TCF7L2 (rs45056565, r² = 0.92 with the previously reported SNP rs7903146) and SLC30A8 (rs11558471, r² = 0.96 with the previously reported SNP rs13266634). Seven additional loci had replicable evidence for association with fasting glucose and/or HOMA-B across studies at the arbitrary summary threshold of P < 2 × 10⁻⁵, chosen to prioritize SNPs for follow-up (Table 1 and Supplementary Table 2). After excluding SNPs within the four previously discovered genome-wide significant fasting glucose loci in GCK, GCKR, G6PC2 and MTRNR1B, we still observed an excess of small P values compared to the distribution expected under the null hypothesis (Fig. 2a,b), suggesting that some of these additional loci are likely to represent new fasting glucose– and/or HOMA-B–associated loci that merit additional investigation.

Stage 1 analyses of fasting insulin and HOMA-IR revealed no loci that reached genome-wide significance, but there were six loci with consistent evidence for association across study samples at P < 2 × 10⁻⁵ (Table 1, Supplementary Table 2 and Supplementary Fig. 2c,d). Comparison of the observed P values with the distribution expected under the null hypothesis showed an excess of small P values that warrant further investigation (Fig. 2c,d).
We further conducted a global meta-analysis of cohort results adjusted for body mass index (BMI) to test whether these diabetes-related quantitative trait associations may be mediated by associations with adiposity. The adjustment for BMI did not materially affect the strength of the associations with any of the traits (data not shown).

**Effect size estimates for genome-wide significant loci**

We restricted our effect size estimates (Table 2 and Supplementary Table 2) to the stage 2 replication samples (up to n = 76,558) to avoid inflation introduced by the discovery cohorts (the so-called 'winner’s curse'). The previously identified loci in G6PC2, MTNR1B and GCK showed the largest effects on fasting glucose (0.075, 0.067 and 0.062 mmol/l per allele, respectively), with the remaining loci examined showing smaller effects (0.008 to 0.030 mmol/l per allele; Table 2). The proportion of variance in fasting glucose explained by the 14 fasting glucose–associated loci with replication data (that is, all fasting glucose loci except for those on TCF7L2 and SLC30A8) ranged from 3.2%–4.4% in the six replication studies providing this information. Because results from our largest unselected community-based cohort (Framingham) were on the lower bound of these estimates (3.2%), we felt reassured that the winner’s curse was not a major concern in this instance and selected the Framingham cohort to estimate the proportion of heritability explained and the genotype score. With a heritability estimate of 30.4% in the Framingham cohort, these 14 loci explain a substantial proportion (~10%) of the inherited variation in fasting glucose. Given the possibility that these same loci harbor additional independent variants (for example, those due to low-frequency alleles not captured by this analysis) that also influence fasting glucose, this estimate of the heritability attributable to these loci is likely to be conservative.

We estimated the combined impact of the 16 loci associated with fasting glucose (the 14 loci included in the effect size estimates plus those on TCF7L2 and SLC30A8) in some of the largest cohorts (Framingham, the Northern Finland Birth Cohort (NFBC) of 1966 and the Atherosclerosis Risk in Communities (ARIC) study) by constructing a genotype score equal to the sum of the expected number of risk alleles at each SNP weighted by their effect sizes (see Online Methods). Fasting glucose levels were higher in individuals with higher genotype scores (Fig. 3), with mean differences of −0.4 mmol/l (5.93 versus 5.51 mmol/l in NFBC 1966; 5.36 versus 5.03 mmol/l in

---

**Figure 1** Regional plots of ten newly discovered genome-wide significant associations. (a) ADCY5. (b) MAD. (c) ADRA2A. (d) FADS1. (e) CRY2. (f) SLC2A2. (g) GLIS3. (h) PROX1. (i) C2CD4B. (j) IGF1. For each region, directly genotyped and imputed SNPs are plotted with their meta-analysis P values (−log10 values) as a function of genomic position (NCBI Build 35). In each panel, the stage 1 discovery SNP taken forward to stage 2 replication is represented by a blue diamond (with global meta-analysis (i.e., a red diamond. Estimated recombination rates (taken from HapMap) are plotted to reflect the local LD structure around the associated SNPs and their correlated proxies (according to a white-to-red scale from r² = 0 to 1, based on pairwise r² values from HapMap CEU). Gene annotations were taken from the UCSC genome browser.
Impairment of glucose homeostasis may be characterized by elevated fasting glucose or fasting insulin, elevated glucose or insulin at 2 h after oral glucose tolerance test (OGTT), or elevated glycated hemoglobin (HbA1c). We tested associations of each of the 17 loci of interest in a subset of MAGIC cohorts with GWAS data informative for these traits. Because HbA1c is a measure of average glycaemia over the preceding 2–3 months, we hypothesized that if an association of these loci with additional traits was present, it should be directionally consistent. The three loci with the largest effect sizes on fasting glucose—G6PC2, MTNR1B and GCK—all showed genome-wide significant and directionally consistent associations with HbA1c in DGKB-TMEM195, ADCYS, SLC2A2, PROX1, SLC30A8 and TCF7L2 showed nominal (P < 0.05) evidence of directionally consistent association (Table 2). The fasting glucose–raising alleles at TCF7L2, SLC30A8, GCK and ADCYS were associated (P < 0.0002) with increased 2-h glucose (Table 2); a parallel MAGIC project reports the genome-wide significant association with 2-h glucose of another ADCYS SNP in strong linkage disequilibrium (LD) with our lead SNP (r2 = 0.82)39. In contrast, and consistent with previous reports that the fasting glucose–raising allele of GCKR is associated with greater insulin release during OGTT31,12,30, this allele was associated with lower 2-h glucose.

Testing of these loci for association with T2D as a dichotomous trait in up to 40,655 cases and 87,022 nondiabetic controls demonstrated that the fasting glucose–raising alleles at seven loci (in or near ADCYS, PROX1, GCK, GCKR and DGKB-TMEM195 and the known T2D genes TCF7L2 and SLC30A8) are robustly associated (P < 5 × 10−8) with increased risk of T2D (Table 2). The association of a highly correlated SNP in ADCYS with T2D in partially overlapping samples is reported by our companion manuscript29. We found less significant T2D associations (P < 5 × 10−3) for variants in or near CRY2, FADS1, GLI3 and C2CD4B (Table 2). These data clearly show that loci with similar fasting glucose effect sizes may have very different T2D risk effects (see, for example, ADCYS and MADD in Table 2).

Given that several alleles associated with higher fasting glucose levels were also associated with increased T2D risk and that the T2D–related genes TCF7L2 and SLC30A8 showed association with fasting glucose, we systematically investigated association of all established T2D loci with the same four fasting diabetes–related quantitative traits. We found directionally consistent nominal associations (P < 0.05) of T2D risk alleles with higher fasting glucose for 11 of 18 established T2D loci, including MTNR1B (Supplementary Table 3). These data demonstrate that a large T2D effect size does not always translate to an equivalently large fasting glucose effect in nondiabetic persons, as clearly highlighted when contrasting the remarkably small effects of TCF7L2 on fasting glucose compared to MTNR1B (Table 2).

### Impact of newly discovered loci on other metabolic traits

Next, we used available GWAS results for additional metabolic phenotypes (BMI from GIANT31, blood pressure from Global BPgen32 and lipids from ENGAGE33) to assess the impact of the newly discovered glycemic loci on these traits. None of the newly discovered loci had significant (P < 0.01) associations with BMI or blood pressure (Table 3). Notably, the FADS1 glucose-raising allele was associated with increased total cholesterol (P = 2.5 × 10−6),...
low-density lipoprotein cholesterol ($P = 8.5 \times 10^{-6}$) and high-density lipoprotein cholesterol ($P = 2.9 \times 10^{-5}$), but was associated with lower triglyceride levels ($P = 1.9 \times 10^{-6}$) (Table 3); a consistent association of this locus with lipid levels has been previously reported14. The fasting glucose–associated variant in MADD was not associated with lipid levels and is not in LD ($r^2 < 0.1$) with a previously reported high-density lipoprotein cholesterol SNP (rs7395662)15, suggesting two independent signals within the same locus, one affecting lipid levels and the other affecting fasting glucose levels (Table 3).

### Potential functional roles of newly discovered loci

We investigated the likely functional role of genes mapping closest to the lead SNPs using several sources of data, including human disease

---

**Table 2** Association of newly discovered SNPs with glycemic traits in MAGIC and type 2 diabetes replication meta-analyses

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene(s)</th>
<th>Alleles (effect/other)</th>
<th>Fasting glucose (mmol/l)</th>
<th>HOMA-B</th>
<th>Fasting insulin (mmol/l)</th>
<th>HOMA-IR</th>
<th>HbaA1c (%)</th>
<th>2-h glucose (mmol/l)</th>
<th>2-h insulin (mmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>n560887</td>
<td>6GPD C/T</td>
<td>Effect*</td>
<td>0.075 (0.003)</td>
<td>0.007 (0.004)</td>
<td>0.006 (0.004)</td>
<td>0.017 (0.002)</td>
<td>0.017 (0.002)</td>
<td>-0.031 (0.013)</td>
<td>0.97 (0.95–0.99)</td>
</tr>
<tr>
<td>n10830963</td>
<td>MTNR1 B/G</td>
<td>Effect*</td>
<td>0.067 (0.003)</td>
<td>0.030 (0.004)</td>
<td>0.004 (0.004)</td>
<td>0.024 (0.004)</td>
<td>0.056 (0.022)</td>
<td>0.34 (0.015)</td>
<td>1.09 (1.06–1.12)</td>
</tr>
<tr>
<td>n4607517</td>
<td>GCK C/T</td>
<td>Effect*</td>
<td>0.062 (0.004)</td>
<td>0.006 (0.006)</td>
<td>0.015 (0.006)</td>
<td>0.041 (0.005)</td>
<td>0.097 (0.026)</td>
<td>-0.012 (0.015)</td>
<td>1.07 (1.05–1.10)</td>
</tr>
</tbody>
</table>

---

**Figure 3** Variation in levels of fasting glucose depending on the number of risk alleles at newly identified loci, weighted by effect size in an aggregate genotype score for the Framingham Heart Study. The bar plots show the average and standard error of fasting glucose in mmol/l for each value of the genotype score based on the regression coefficient (right y-axis), and the histogram denotes the number of individuals in each genotype score category (left y-axis). Comparable results were obtained for the NFBC 1966 and ARIC cohorts. On average, the range spans ~0.4 mmol/l (~7.2 mg/dl) from low to high genotype score.
Expression analyses

We measured expression of the genes mapping closest to our lead SNPs (in DGKB–TMEM195, ADC5Y, MADD, and PROX1), cell proliferation and development (GLIS3, MADD, and PROX1), glucose transport and sensing (SLC2A2, GCK, GCKR, and G6PC2), and circadian rhythm regulation (MTNR1B and CRY2). All of these pathways represent further avenues for physiological characterization and possible therapeutic intervention for T2D. However, we note that other genes could be causal (Box 1 and Supplementary Table 4), and further experimental evidence will be needed to unequivocally link specific genes with phenotypes.

Potential causal variants, eQTLs and copy number variants

Our results interrogate only a fraction of the common variants in any given genomic region; we therefore expect that for the majority of the loci here described, the underlying causal variant has yet to be identified. Nevertheless, for some loci there are possible SNP...
proximity (130 bp) to rs174548, a SNP highly associated with mRNA expression levels of nearby genes (Online Methods). The lead SNP participates in beta-cell ontogeny and predicted by PolyPhen to be “tolerated” by SIFT (r59 = 0.92). A missense mutation in this gene (leading to a S59N amino acid substitution) segregates with diabetes in one family affected with a Mendelian form of the disease (data not shown). These results are substantiated by previous work showing strong associations (P = 1.0 with rs174550), whose impact of copy number variation on glycemic traits, we took advantage of the hypothesis that not only the abundance of fatty acids, but also their precise composition and degree of desaturation, may influence glucose homeostasis.

The DGKB-TMEM195 locus was recently reported to be associated with fasting glucose24, here we report genome-wide significant replication of that finding and evaluate the genes mapping closest to the lead SNP in further detail. DGKB encodes the β (1 of 10) isotype of the catalytic domain of diacylglycerol kinase, which regulates the intracellular concentration of the second messenger diacylglycerol. In rat pancreatic islets, glucose increases diacylglycerol46, which activates protein kinase C (PKC) and thus potentiates insulin secretion50. TMEM195 encodes transmembrane protein 195, an integral membrane phosphoprotein highly expressed in liver.

**Box 1: Genes nearest to loci associated with fasting diabetes-related quantitative traits**

The DGKB-TMEM195 locus was recently reported to be associated with fasting glucose24, here we report genome-wide significant replication of that finding and evaluate the genes mapping closest to the lead SNP in further detail. DGKB encodes the β (1 of 10) isotype of the catalytic domain of diacylglycerol kinase, which regulates the intracellular concentration of the second messenger diacylglycerol. In rat pancreatic islets, glucose increases diacylglycerol46, which activates protein kinase C (PKC) and thus potentiates insulin secretion50. TMEM195 encodes transmembrane protein 195, an integral membrane phosphoprotein highly expressed in liver.

DGKB encodes adenylate cyclase 5, which catalyzes the generation of cAMP. Upon binding to its receptor in pancreatic beta cells, glucagon-like peptide 1 (GLP-1) induces cAMP-mediated activation of protein kinase A, transcription of the proinsulin gene and stimulation of insulin secretory processes51.

**MADD** encodes mitogen-activated protein kinase (MAPK) activating death domain, an adaptor protein that interacts with the tumor necrosis factor α receptor to activate MAPK. Both PKC and MAPK have been implicated in the proliferation of beta cells induced by GLP-1 (ref. 51), suggesting that DGKB and MADD may contribute to beta-cell mass and insulin secretion in this manner as well. Also in this region, SLC39A13 encodes a putative zinc transporter required for connective tissue development and BMP/TGF-β signaling52. NRIH3 encodes the liver X receptor alpha (LXRA) protein, which contains the retinoid response element. Glucose stimulates the transcriptional activity of LXR, which acts as a molecular switch that integrates hepatic glucose metabolism and fatty acid synthesis53.

**ADRA2A** encodes the α2a adrenergic receptor, which is expressed in beta cells and whose activation leads to an outward potassium current independent of the islet potassium-sensitive ATP (KATP) channel, thus possibly modifying insulin release54. Mice with null mutations display abnormal glucose homeostasis in addition to cardiac hypertrophy and abnormal heart rate and blood pressure.

**FADS1** encodes fatty acid desaturase 1, which catalyzes the biosynthesis of highly unsaturated fatty acids from precursor essential polysaturated fatty acids. One such product is arachidonic acid, in rodent beta cells, arachidonic acid liberated by phospholipase A2 augments glucose-mediated insulin release55. Two other members of the same family, FADS2 and FADS3, also reside in this region. By directing fatty acids down this metabolic pathway, increased activity of these enzymes may lower circulating triglyceride concentrations.

**CRY2** encodes cryptochrome 2, an integral component of the mammalian circadian pacemaker56. Mice with null mutations in this gene present with abnormal circadian rhythmicity and several metabolic abnormalities including impaired glucose tolerance, increased insulin sensitivity, decreased body weight and adipose tissue, and abnormal heart rate. Together with MTNR1B57–59, this is the second circadian gene associated with fasting glucose in humans, contributing further evidence to the emerging idea that this pathway regulates glucose homeostasis57. In the same region, MAPK8IP2 encodes the scaffolding protein JIP1. Cross-talk between JIP1 and JIP3 has been implicated in the regulation of ASK1-SEK1-JNK signaling during glucose deprivation59. A missense mutation in this gene (leading to a S55N amino acid substitution) segregates with diabetes in one family affected with a Mendelian form of the disease59.

**SLC2A2** encodes the GLUT2 transporter responsible for transporting glucose into beta cells and triggering the glucose-mediated insulin secretion cascade. In humans, recessive mutations in this gene lead to Fanconi-Bickel syndrome, a rare disorder characterized by hepatoenral glycogen accumulation, proximal renal tubular dysfunction and impaired utilization of glucose and galactose60; mouse mutants also show hyperglycemia and abnormal glucose homeostasis61.

**GLIS3** encodes the transcription factor GLIS3 family zinc finger 3 isoform, a Krüppel-like zinc finger protein that both activates and represses transcription and participates in beta-cell ontogeny62,63. Functional mutations in this gene cause a syndrome of neonatal diabetes and congenital hypothyroidism63. Polymorphisms within this gene have recently been associated with type 1 diabetes risk (t1dgc.org).

**PROX1** encodes the prospero homeobox protein 1, a novel co-repressor of hepatocyte nuclear factor 4e64 that plays a crucial role in beta-cell development; mutations in its target gene HNF4A cause maturity-onset diabetes of the young, type 1 (ref. 65).

**C2CD4B** (formerly FAM148B) encodes the nuclear localized factor 2 (NLF2). It is expressed in endothelial cells and upregulated by proinflammatory cytokines65. As shown here, it has a high level of expression in the pancreas, although its putative molecular connection with glucose homeostasis is presently unclear.

**IGF1** encodes the insulin-like growth factor 1 and is the sole genome-wide significant locus associated with HOMA-IR in our study. Humans and mice null for IGF1 display abnormal glucose homeostasis, with insulin resistance, increased circulating insulin and sensitivity to growth hormone67.

candidates; in SLC2A2, the lead SNP (rs11920090) is in perfect LD (r2 = 1.0) with rs5400 (stage 1 discovery association P = 5.9 × 10−6), which codes for the amino acid substitution T110I, predicted to be “possibly damaging” by PolyPhen55 and PANTHER (Pdel = 0.92)56. In GCKR, the lead SNP is in strong LD (r2 = 0.93) with rs1260326, encoding P446L, a nonsynonymous variant previously associated with fasting glucose and HOMA-IR11,12,30 and predicted by PolyPhen to be “probably damaging.” A recent functional study has demonstrated that this variant indirectly leads to increased GCK activity, resulting in the observed effects on fasting glucose and triglyceride levels57. Both the SLC2A2 T110I and GCKR P446L substitutions were predicted to be “tolerated” by SIFT58, highlighting the difficulties in obtaining consensus functional predictions from different informatic approaches.

We used publicly available expression quantitative trait locus (eQTL) datasets for liver39, cortex40 and Epstein-Barr virus–transformed lymphoblastoid cell lines41 to explore additional possible causal mechanisms by testing for association between replicated loci and mRNA expression levels of nearby genes (Online Methods). The lead SNP in FADS1, rs174550, is in strong LD with (r2 = 0.80) and is in close proximity (130 bp) to rs174548, a SNP highly associated with FADS1 mRNA expression levels in liver (P = 1.7 × 10−5) and with FADS2 mRNA expression levels in lymphoblastoid cells (P = 3.1 × 10−4). The SNP rs174548 has also been associated (up to P = 4.5 × 10−8) with a number of serum glycerophospholipid concentrations in a GWAS investigating metabolomic profiles42, and rs174550 also showed strong associations (P < 5.2 × 10−7) with the same metabolites (data not shown). These results are substantiated by previous work associating SNPs in this region with the fatty acid composition of phospholipids43. The latter data suggest that the minor allele variant of rs174550 results in a reduced efficiency of the fatty acid delta-5 desaturase reaction42. Finally, bioinformatic analysis identifies a perfect proxy, rs174545 (r2 = 1.0 with rs174550), whose glucose-raising allele abolishes a predicted target site for the mir-124 microRNA (see Online Methods). Taken together, these data support the hypothesis that not only the abundance of fatty acids, but also their precise composition and degree of desaturation, may influence glucose homeostasis.

Although our study was not designed to explicitly investigate the impact of copy number variation on glycemic traits, we took advantage of existing data44 to investigate whether any of our lead SNPs are...
in LD with common, diallelic copy number polymorphisms (CNPs) mapping within a 1-Mb window. Of the fasting glucose loci, only DGKB-TMEM195 has a validated, common CNP affecting sequence within 1 Mb of the index SNP\(^4\). Despite the proximity of this CNP to the associated SNP (\(r^2 = 0.25\)), the CNP is essentially uncorrelated with the index SNP (\(r^2 = 0.01\) in HapMap CEU) and is therefore unlikely to explain the observed association with fasting glucose level.

**DISCUSSION**

In this meta-analysis of 21 stage 1 discovery GWAS cohorts followed by targeted stage 2 replication of 25 loci in 33 additional cohorts (totaling up to 122,743 nondiabetic participants), we report new genome-wide significant associations of SNPs in or near ADCYS5, MADD, ADRA2A, CRY2, FADS1, GLIS3, SLC2A2, PROX1 and C2CD4B with fasting glucose and one SNP near IGF1 associated with fasting insulin and HOMA-IR. We have also confirmed associations of variants in GCK, GCKR, G6PC2 and MTNR1B with fasting glucose and achieved genome-wide significance for the recently reported DGKB-TMEM195 locus\(^4\) and for variants in the known T2D-associated genes TCF7L2 and SLC30A8. All of the fasting glucose–associated SNPs showed consistent nominal associations with HOMA-B, and those in GCK, G6PC2, MTNR1B, DGKB-TMEM195, ADCY5, FADS1 and GLIS3 did so at genome-wide significant levels. As previously reported\(^{11,12,30}\), GCKR is also associated with fasting insulin and HOMA-IR.

Notably, in addition to the established T2D-associated loci in TCF7L2, SLC30A8 and MTNR1B, five of the loci that are associated with elevated fasting glucose levels in nondiabetic individuals (in ADCYS5, GCK, GCKR, PROX1 and DGKB-TMEM195) also increase the risk of T2D in separate T2D case-control studies. However, this overlap is incomplete and highlights the fact that the magnitude of the effect on fasting glucose is not predictive of the effect on T2D risk, as shown when comparing fasting glucose and T2D effect sizes for MTNR1B and TCF7L2, or for ADCY5 and MADD (Table 2). Loci on the latter two genes have similar effect sizes on fasting glucose and have similar allele frequencies, and yet the former is robustly associated with T2D risk (OR 1.12, \(P = 5.5 \times 10^{-21}\)) whereas the latter is not (OR 1.01, \(P = 0.3\)) in the same samples. This suggests that not all loci associated with fasting glucose within the ‘physiological’ range are also associated with ‘pathological’ fasting glucose levels and T2D risk. Thus, variation in fasting glucose in healthy individuals is not necessarily an endophenotype for T2D, which posits the hypothesis that the mechanism by which glucose is raised, rather than a mere elevation in fasting glucose, is a key contributor to disease progression. On the other hand, we cannot rule out the existence of separate T2D-protective variants within loci for which elevated fasting glucose does not progress to manifest T2D; we also cannot rule out the effect of cohort selection in the detection of the loci with variable effects on fasting glucose and T2D risk. Nevertheless, this work shows that targeting quantitative traits in GWAS searches can help identify genetic determinants of overt disease.

With regard to insulin resistance, our analyses resulted in only one novel genome-wide significant locus associated with fasting insulin and HOMA-IR. The associated SNP rs35767 is 1.2 kb upstream of IGF1, raising the possibility that it may influence IGF1 expression levels (we have found no direct support for this notion in the limited eQTL data available). Although not reaching genome-wide significance, we note that SNP rs4675095 in IRS1 (the insulin receptor substrate-1 gene) was also associated with HOMA-IR (\(P = 4.6 \times 10^{-3}\)), which, given IRS1’s excellent biological credentials, will warrant further investigation. This SNP is not in LD with the widely studied missense SNP substitution G972R (rs1801278), nor is it in LD with the newly discovered T2D SNP rs2943641 (ref. 45), whose C risk allele was only nominally associated with increased fasting insulin (\(P = 0.02\)) and HOMA-IR (\(P = 0.04\)) in our discovery dataset. The previously reported associations of SNPs in PANK1 with fasting insulin\(^3\) did not receive strong support in our discovery cohorts (\(P = 0.04\) and \(P = 0.17\) for rs11185790 and rs1075374, respectively).

Notably, our large-scale meta-analyses produced more than a dozen robust associations with fasting glucose and only two with fasting insulin and HOMA-IR (GCKR and IGF1). Although the somewhat smaller sample size for the insulin analysis may have contributed to this discrepancy, a comparison of the similarly powered HOMA-B and HOMA-IR analyses reveals associations with HOMA-B several orders of magnitude more significant than those seen with HOMA-IR (Fig. 2). Because insulin itself is a component of the numerator in both measures, one cannot attribute this discrepancy to technical differences in insulin measurements across cohorts. Similarly, because the quantile-quantile plots are very similar for fasting insulin and HOMA-IR, we do not believe that the use of a mathematical formula (as was used with HOMA-IR) rather than a direct measurement (as was used with fasting insulin) has affected our analyses substantially. HOMA-B and HOMA-IR have comparable heritability estimates (0.26 and 0.27 in the Framingham Heart Study, respectively), and their correlation is substantial (\(r = 0.55\) in the Framingham Heart Study). Thus, not only may there be a difference in the identity of specific genetic determinants for each trait\(^46\), but the genetic architecture may be distinct for each trait, with more modest effects, fewer loci, rarer variants, or a stronger environmental modification underlying HOMA-IR. In addition, HOMA-IR (which is composed of fasting values) is an imperfect estimate of global insulin resistance, as it addresses mostly hepatic sensitivity to insulin and is partially affected by beta-cell function. The heritability of HOMA-IR is lower than the heritability for insulin sensitivity derived from the minimal model\(^47\). Exploration of gene × environment interactions and analysis of datasets that include 2-h glucose and insulin values may reveal other genetic factors that increase insulin resistance in humans\(^29\).

In conclusion, our large-scale meta-analysis of GWAS has identified ten new loci associated with glycemic traits whose in-depth physiological investigation should further our understanding of glucose homeostasis in humans and may reveal new pathways for diabetes therapeutics.

**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Note: Supplementary information is available on the Nature Genetics website.

**ACKNOWLEDGMENTS**

A full list of Acknowledgments appears in the Supplementary Note.

**AUTHOR CONTRIBUTIONS**

A full list of Author Contributions appears in the Supplementary Note.

**COMPETING INTERESTS STATEMENT**

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/naturegenetics/.

Published online at http://www.nature.com/naturegenetics/. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

**Supplementary information**

The Applications for protein sequence-function evolution data: advance online publication, Early diabetes and abnormal postnatal pancreatic islet responses to an oral glucose challenge.

Pe'er, I., Yelensky, R., Altshuler, D. & Daly, M.J. Estimation of the multiple testing burden in genome-wide association studies by imputation of genotypes.

Matthews, D.R. Mutations in the LKB1 gene cause hepatic phosphorylase b kinase deficiency.

Bouatia-Naji, N. et al. Mutations in the hepatocyte nuclear factor-4 alpha (HNF4A) gene cluster and non-alcoholic fatty liver disease.

Song, K.-H., Li, T. & Chiang, J.Y.L. A prospero-related homeodomain protein is a prospero homolog that mediates cell-cell junctions and promotes posterior development.

Dixon, A.L. et al. The nuclear receptor LXR is a glucose sensor.

Santer, R. et al. The phosphorylation of SEK1–JNK2 and Akt1 act as mediators of glucose-stimulated insulin secretion in pancreatic islets.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.


Matthews, D.R. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

Prentki, M. & Matschinsky, F.M. Ca-dependent phospholipases.

Kume, K. et al. The P446L variant in the GCKR gene is associated with increased plasma triglyceride and C-reactive protein but not fasting glucose and insulin resistance.

ONLINE METHODS

Cohort description. The consortia participating in MAGIC contributed a maximum total of 122,743 individuals. The stage 1 discovery set included 36,466–46,186 individuals (depending on trait) from 17 population-based cohort studies and four case-control studies. The stage 2 replication set included up to 76,558 individuals from 33 sample collections, including 28 population-based and 5 case-control collections. Detailed information on all studies is provided in Supplementary Table 1a (stage 1 discovery) and 1b (stage 2 replication). All participants were adults of white European ancestry from the United States or Europe. Individuals were excluded from the analysis if they had a physician diagnosis of diabetes, were on diabetes treatment (oral or insulin) or had a fasting plasma glucose ≥7 mmol/l. Some individuals with fasting glucose <7 mmol/l but who would have tested abnormally after an oral glucose challenge could have been included; we estimated this number to be as low as <1% in the Framingham Heart Study and 1.6% in Inter99, two population cohorts in which all relevant data were available. Individual studies applied further sample exclusions, including pregnancy, non-fasting individuals, type 1 diabetes, or outliers ±3 s.d. of distribution for either fasting glucose or fasting insulin, as detailed in Supplementary Table 1a and 1b. Individual stage 1 discovery cohort sizes ranged between 458 and 6,479 samples; stage 2 replication cohorts ranged between 554 and 8,010 samples. All studies were approved by local research ethic committees, and all participants gave informed consent.

Type 2 diabetes association. The association analysis of lead SNPs with T2D as a dichotomous trait was carried out under the additive genetic model in 27 case-control cohorts totaling 40,655 cases and 87,022 controls of European descent. These included 8,130 cases and 38,987 controls from eight DIAGRAM+ Consortium studies and 32,525 additional T2D cases and 48,035 additional controls from 19 cohorts genotyped de novo, listed as cohort (n cases/n controls): FUSION stage1 (1,203/1,261), METSIM_CC (854/3,469), Addition/Ely (892/1,612), Cambridgeshire Case Control Study (541/527), Norfolk Diabetes Case Control Study (6,056/6,428), deCODE (1,465/23,194), DGDG (690/730), DGI (1,022/1,075), ERGO (1,178/4,761), EUROSPAN (268/3,710), FUSION (1,161/1,174), KORA S3 (433/438), T2D Wellcome Trust Case Control Consortium (1,924/2,938), HPFS (1,146/1,241), Nurses’ Health Study (1,532/1,754), Danish (3,652/4,992), KORA replication consisting of cases from KORAS1-S4 and the Augsburg Diabetes Family Study (ADFS) and controls from KORA S4 (1,047/1,491), OxnGen_58BCC (UKR52) (612/1,596), UKT2DGC (4,979/6,454), Framingham Heart Study_CC (674/7,664), NHANES (289/1,219), Partners/Roche (534/622, Umeå (1,327/1,424), French_CC (2,155/1,682), GCI Poland_DGI_Stage2 (969/969), GCL_US_DGI_Stage2 (1,191/1,171) and MDC_MDR_DGI_Stage2 (2,814/3,234). According to the best sample-specific model, in some cohorts, age and BMI were used as covariates for adjustment of the case-control association. The meta-analysis of the cohort-specific summary statistics (odds ratios and 95% confidence intervals) was performed using a fixed effects inverse-variance approach with GWAMA (see URLs).

Quantitative trait measurements. Fasting glucose (in mmol/l) was measured from fasting whole blood, plasma or serum or a combination of these. Whole-blood fasting glucose levels were corrected to plasma fasting glucose using a correction factor of 1.13. Fasting insulin was measured as described in Supplementary Table 1a and 1b for each of the cohorts. Indices of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) were derived from paired fasting glucose and insulin measures using the homeostasis model assessment14.

Genotyping, imputation and quality control. Genotyping of individual cohorts was carried out using commercial genome-wide arrays as detailed in Supplementary Table 1a and 1b. For genome-wide SNP sets, different criteria were used to filter out poor-quality SNPs and samples before imputation. Criteria generally applied for exclusion of samples were (i) call-rate <0.95, (ii) individuals with heterozygosity outside the population-specific bounds and (iii) ethnic outliers. Criteria generally applied for exclusion of SNPs were (i) minor allele frequency (MAF) <0.01, (ii) Hardy-Weinberg equilibrium P < 10^-4 or 10^-6 and (iii) call-rate <0.95. Imputation of additional autosomal SNPs from the HapMap CEU reference panel was performed using the software MACH23, IMPUTE24 or BIMBAM25 with parameters and pre-imputation filters as specified in Supplementary Table 1a and 1b. SNPs were also excluded if the cohort-specific imputation quality as assessed by r2 hat was <0.3 (MACH) or proper-info was <0.4 (IMPUTE) or observed/expected dosage variance was <0.3 (BIMBAM), or if their mapping and/or strand annotation was ambiguous. In total, up to 2.5 million genotyped or imputed autosomal SNPs were considered for meta-analysis. SNPs were considered for meta-analysis if they were available for at least 20% of maximum available sample size or if ≥10,000 individuals were informative for each SNP.

Statistical analyses. We excluded from analysis people with diabetes (those on diabetes treatment or with fasting glucose ≥7 mmol/l), non-fasting participants and pregnant women. In each cohort, we used log-transformed trait values for fasting insulin, HOMA-IR and HOMA-B and untransformed fasting glucose as the dependent variable in linear regression models that included terms for sex, age (except NFBC 1966, where all subjects were 31-years-old), study site (if applicable), geographical covariates (if applicable) and age squared (Framingham only) to assess the association of additively coded genotypes with trait values. Association testing was performed using software that takes genotype and imputation uncertainty into account, using a missing-data likelihood test implemented in SNPTEST26 or by using allele dosages in the linear regression model in MACH2QTL27, GenABEL28 or Lmkin from the R kinship package29. Regression estimates for the effect of the additively coded SNPs were pooled across studies in a meta-analysis using a fixed effect inverse-variance approach31. The individual cohort results, but not the final meta-analysis results, were corrected for residual inflation of the test statistic using the genomic control method27. Final GC values were 1.05 for fasting glucose, 1.046 for HOMA-B, 1.04 for HOMA-IR and 1.041 for fasting insulin.

Replication SNP selection and analysis. Twenty-five lead SNPs from among the most significant association results in the stage 1 discovery meta-analyses were selected for replication. To account for the correlation between traits and to ensure independent signals, highly significant associations detected in two or more traits were selected only once. All selected loci had an r^2 < 0.5 with the nearest other selected loci. From each unique locus, the SNP with the smallest P value was chosen. All SNPs had a minimum sample size of at least 80% of the overall discovery sample. Variants known to be associated with T2D (in SLC30A8 and TCP7L2) and reaching the genome-wide significance threshold (P ≤ 5 × 10^-8) were not included in the replication list. SNPs were also selected on the basis of low heterogeneity between studies, although loci with biologic plausibility were selected even if there was some evidence of heterogeneity. Seventeen SNPs from the glucose and HOMA-B analyses and eight SNPs from the insulin and HOMA-IR analyses were taken for stage 2 replication. Although previously described variants in G6PC2, GCK, GCKR and MTNR1B were selected for replication to serve as ‘positive controls’ in all study samples. Up to four alternate proxy SNPs (maximizing LD with the index SNP) were selected for each locus to accommodate the capacities of different platforms. In the cases where index SNPs failed in the initial stage of genotyping, replication results were obtained for proxy SNPs in strong LD with the original index SNP whenever possible. SNPs with Hardy-Weinberg equilibrium P values ≤0.001 were excluded. In cases where more than one proxy SNP was genotyped but the index SNP was unavailable, the proxy SNP’s LD with the index SNP and its call rate was used to select the SNP with the best-quality genotyping to be included in the meta-analysis.

Genotype data for 25 signals or proxies were obtained from 33 independent replication cohorts, including both in silico data from pre-existing GWAS (8) and de novo genotyping (25). Phenotype definition and association testing between fasting traits and these 25 SNPs was performed in the same manner in each cohort. The inverse variance method was then applied to derive pooled effect estimates from the stage 2 replication samples using METAL (see URLs) and GWAMA software. We then carried out a pooled analysis of the stage 1 discovery cohorts and stage 2 replication samples to determine which SNPs reached genome-wide significance, as determined by a P < 5 × 10^-8. Heterogeneity was assessed using the P index28.

Notes on replication genotyping. Amish. The Amish trait data is reported for the Heredity and Phenotype Interaction Heart Study (HAPI), Amish Family Longevity Study (LS), Amish Family Diabetes Study (AFDS), Amish
Family Calcification Study (AFCS) and Pharmacogenomics of Anti-Platelet Intervention (PAPI) Study. All studies genotyped 15 SNPs (rs10830963, rs4607517, rs11605924, rs1700807, rs1416002, rs588263, rs6475095, rs947696, rs912494, rs1192090, rs174550, rs703420, rs4234291, rs4537420, rs18814133). Other SNPs were typed on different sample subsets: AFDS only (rs2191349, rs10498346); HAPI only (rs560887); HAPI, LS and Pharmacogenomics of Anti-Platelet Interaction study (PAPI) (rs780904, rs6479526, rs340835, rs11167682); HAPI, LS, AFDS, PAPI (rs9418635, rs855228); HAPI, LS and AFCS (rs1039149). The genotyping statistics in Supplementary Table 1b are reported for the AFDS + HAPI + LS cohorts.

FUSION stage 2. The FUSION stage 2 cohort includes some Health 2000 samples, none of which overlap with the Health 2000 cohort.

SNP score. For the 16 SNPs reaching genome-wide significance of association (either in the discovery stage alone or in the combined replication and discovery meta-analysis), we defined a risk score as the weighted sum of the number of expected risk alleles, where the sum of the weights was set to the number of SNPs (16) and the weights were proportional to the estimate of the effect size for each SNP. Mean fasting glucose levels according to the number of weighted risk alleles were computed in some of the largest cohorts (Framingham, ARIC, NFBC 1966) with all 16 SNPs available (genotyped or imputed).

Bioinformatic analysis and functional annotation. To perform a preliminary assessment of the underlying functionality at the associated loci, we first expanded the set of SNPs to include those in strong LD with the index SNP (defined as pairwise r^2 > 0.8 according to HapMap Phase II CEU data). We then mapped the genomic locations of all the SNPs in this expanded set to several non-mutually-exclusive genomic annotation sets: nonsynonymous sites, splice sites, intergenic regions, 5' UTR, 3' UTR and introns from dbSNP version 129 (see URLs section for URLs of this and other software mentioned in this paragraph); 1-kb and 5-kb regions upstream of transcription start sites from Ensembl version 49; intergenic predicted transcription factor binding sites, CpG islands, ORegAnno elements, Encode region ancestral repeats, EviFold elements, multispecies conserved sequences and positively selected gene regions from the University of California Santa Cruz human table browser; predicted microRNA target sites from TargetScan 4.2; validated enhancers from the Vista Enhancer Browser; predicted cis-regulatory modules from the PreMod database; and validated noncoding RNAs from RNAdb. The potential functional effect of nonsynonymous substitutions were evaluated using three prediction programs: SIFT, PolyPhen and PANTHER.

GRAIL. We used GRAIL (see URLs) to examine the putative relationship between candidate genes at validated loci based on concomitant appearance in published scientific text. GRAIL is a bioinformatic annotation tool that, given a list of genes and specific regions or SNPs associated with a particular phenotype or disease, searches for similarities in the published scientific text among the associated genes. It scores regions for functional relatedness by defining associated regions based on the interval between recombination hotspots flanking furthest neighboring SNPs with r^2 > 0.5 to the index SNP, and identifies overlapping genes in that region. Based on textual relationships between genes (as determined from a download of PubMed abstracts on 16 December 2006), GRAIL assigns a P value to each region suggesting its degree of functional connectivity, and picks the best candidate gene after taking into account multiple comparisons.

We considered the following SNPs and candidate genes: rs10830963 (MTN1R1B), rs2191349 (DGKB), rs4607517 (GCK), rs1192090 (SLCA2A), rs11708067 (ADCT5), rs560887 (GPPC2), rs780904 (GCKR), rs11605924 (CR2Y), rs7034200 (GL33), rs3408374 (PROX1), rs10885122 (ADRA2A), rs10498346 (NR1H3), rs174550 (FEN1, FADS1, C11orf9, C11orf10, FADS2) and rs11071657 (C2CD4B). In addition, the following keywords describing functional connections were used: “glucose,” “diabetes,” “islet,” “diacylglycerol,” “circuitad,” “insulin,” “drosophila,” “liver,” “clock,” “cyclase,” “pancreatic,” “adenyl,” “memory,” “beta,” “mice,” “islets,” “phosphatase,” “camp,” “light,” “activity.” A total of 7 genes (MTN1R1B, DGKB, GCK, SLCA2A, ADCT5, GPPC2 and GCKR) out of 14 had a significant association with functional connectivity (at P < 0.1) compared to 1.4 expected under the null, demonstrating that this gene set is enriched in relationships with each other.

eQTL analysis. The validated association signals were searched for previous evidence of expression quantitative trait loci (eQTLs) using several data sources. Liver eQTL association results were obtained from Schadt et al.25. Cortex eQTL association results were obtained from Myers et al.26. Epstein-Barr virus-transformed lymphoblastoid cell eQTLs from ref. 41 were retrieved using the mRNA by SNP browser (see URLs). For each region, we limited our analysis to cis eQTLs given the difficulty of reliably interpreting trans effects. Genes or SNPs within 1 Mb from the lead SNP were considered. The r^2 values between the lead SNPs and eQTL SNPs were retrieved from the HapMap Phase 2 data (CEU Panel), and only SNPs with r^2 > 0.6 were considered.

Of the 12 SNPs showing association with liver and located at < 1 Mb from the lead SNP, five had no r^2 data in HapMap and were located at large distances from the MAGIC lead SNP (mean 320 kb, range 48-725 kb). Of the remaining seven, rs174548 at the FADS1 (fatty acid desaturase 1, P_eQTL = 1.7 × 10^-5) locus was located 130 bp away from the lead SNP rs174550 and in strong LD (pairwise r^2 = 0.8). All the remaining SNPs did not fit our criteria for selection, although we note that a second lead SNP (rs780094 at GCKR) was also moderately associated (r^2 = 0.49, distance = 74 kb) with a strong effect eQTL (rs4665969 at IFT172, P_eQTL = 3.97 × 10^-22). For circulating lymphoblastoid cells, the only cis effect fitting our criteria was observed for the MAGIC SNP rs174550 (FADS1), which was located 24 kb from a known eQTL centered on the FADS2 gene (P_eQTL = 3.1 × 10^-5). Finally, for cortex, the only eQTL was found at four SNPs within LOC131076 (rs6769837, rs7648255, rs12636058, rs6438726), all located >870 kb from the MAGIC lead SNP rs11708067 on ADCYS, LD metrics not available.

Gene expression studies. Adult total RNA samples, except pancreatic islets and flow-sorted beta cells, were purchased from Clontech (Clontech-Takara Bio Europe, Saint-Germain-en-Laye, France). Adult human islets (n = 2) were available through existing collections at Oxford University and were obtained with full ethical consent. Flow-sorted beta cells were obtained from two brain-dead adult donors (preparations >92% insulin-positive cells), in accordance with French legislation and the local ethical committee, as previously described.

Tissue panel (Oxford). Samples were treated with DNase I (Ambion) to ensure that residual genomic contamination was removed. For each tissue, 1 µg of total RNA was used to generate cDNA by random primed first strand synthesis (Applied Biosystems) according to the manufacturer’s protocol. Reverse transcription was also performed on all samples in the absence of the enzyme reverse transcriptase, and these samples were used as negative controls. Primers were designed to cover all RefSeq transcripts. Resulting cDNA for each tissue was diluted 1:100 and 4 µl used in a 10 µl qRT-PCR reaction with 5.5 µl gene expression master mix (Applied Biosystems) and 0.5 µl gene specific assay (Applied Biosystems). All samples were run in triplicate. A standard curve was generated by pooling 1 µl of each cDNA and serially diluting (1:50, 1:100, 1:200, 1:400, 1:800) and running as above. Expression levels were determined with respect to the mean of four endogenous controls (β-actin, B2M, HPRT, TOP1) and normalizing to the mean of the 1:100 standard for the assay of interest. For ease of presentation, the maximum gene expression has been set to equal 1 and all other tissue expressions reported as a fraction of this.

Tissue panel (Cambridge). Adult human total RNA samples (cerebellum, cortex, spleen, pancreas, lung, kidney, liver, skeletal muscle, heart, testes, adipocyte and total brain) were obtained from Clontech. Random-primed first-strand cDNA synthesis was performed with 100 ng RNA using Super Script II (Invitrogen) according to manufacturer’s instructions. Primers were design to cover the majority of protein coding transcripts. For the standard curve, 200 ng of a pool of all RNA samples was amplified using the same protocol. The resulting cDNA for each tissue was diluted fivefold and 5 µl of each sample were used in a 12 µl SYBR Green PCR Master Mix (Applied Biosystem). The cDNA from the standard curve was diluted twofold and used as above. Primers (SIGMA) were designed to anneal to all annotated isoforms of any given gene. Quantitative PCR reactions were done in triplicate on an ABI 7900HT (Applied Biosystems). Expression levels were calculated from their average crossing points, expressed relative to the control gene Top1 (encoding topoisomerase 1), and normalized to gene-specific expression in pancreas. For the purpose of presentation, for each gene the maximal expression was set to equal one and the rest reported as fraction of this.
number. The results of these duplicate experiments, which largely confirm those reported in the text, are shown in Supplementary Figure 3.  

Flow-sorted beta cells (Lille). Samples were treated with DNase I (Ambion) to ensure that residual genomic contamination was removed. For each tissue, 1 µg of total RNA was used to generate cDNA by random primed first strand synthesis (Applied Biosystems) according to the manufacturer’s protocol. Reverse transcription was also performed on all samples in the absence of the enzyme reverse transcriptase, and these samples were used as negative controls. Total RNA was extracted using Nucleospin RNA II kit (Macherey Nagel) according to the manufacturer’s instructions. Resulting cDNA for each tissue was diluted 1:10, and 4 µl was used in a 20-µl qRT-PCR reaction with 10 µl gene expression master mix (Applied Biosystems) and 1 µl gene-specific assay (Applied Biosystems). Data is presented with the most expressed gene (GLIS3) normalized to 1 and all other genes reported as a fraction of this number.
