Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis

By combining genome-wide association data from 8,130 individuals with type 2 diabetes (T2D) and 38,987 controls of European descent and following up previously unidentified meta-analysis signals in a further 34,412 cases and 59,925 controls, we identified 12 new T2D association signals with combined $P < 5 \times 10^{-8}$. These include a second independent signal at the KCNQ1 locus; the first report, to our knowledge, of an X-chromosomal association (near DUSP9); and a further instance of overlap between loci implicated in monogenic and multifactorial forms of diabetes (at HNF1A). The identified loci affect both beta-cell function and insulin action, and, overall, T2D association signals show evidence of enrichment for genes involved in cell cycle regulation. We also show that a high proportion of T2D susceptibility loci harbor independent association signals influencing apparently unrelated complex traits.

Type 2 diabetes (T2D) is characterized by insulin resistance and deficient beta-cell function\(^1\). The escalating prevalence of T2D and the limitations of currently available preventative and therapeutic options highlight the need for a more complete understanding of T2D pathogenesis. To date, approximately 25 genome-wide significant common variant associations with T2D have been described, mostly through genome-wide association (GWA) analyses\(^2-13\). The identities of the variants and genes mediating the susceptibility effects at most of these signals have yet to be established, and the known variants account for less than 10% of the overall estimated genetic contribution to T2D predisposition. Although some of the unexplained heritability will reflect variants poorly captured by existing GWA platforms, we reasoned that an expanded meta-analysis of existing GWA data would offer augmented power to detect additional common variant signals of modest effect.

RESULTS

GWA meta-analysis and replication

We conducted a meta-analysis of eight T2D GWA studies comprising 8,130 T2D cases and 38,987 controls of European descent. We combined case-control data from the Wellcome Trust Case Control Consortium (WTCCC), Diabetes Genetics Initiative (DGI) and Finland-US Investigation of NIDDM genetics (FUSION) scans (the subjects of a previous joint analysis\(^7\)), with those from scans performed by deCODE genetics\(^6\), the Diabetes Gene Discovery Group\(^3\), the Cooperative Health Research in the Region of Augsburg group (KORAgen), the Rotterdam study and the European Special Population Research Network (EUROSPAN). The effective sample size ($n = 22,044$) of stage 1 of the current (hereafter designated ‘DIAGRAM+’) meta-analysis was more than twice that of the earlier DIAGRAM (DIabetes Genetics Replication and Meta-analysis) study\(^7\). After genomic control correction of each component study, we combined association data for 2,426,886 imputed and genotyped autosomal SNPs into a fixed-effects, additive-model meta-analysis using the inverse-variance method (Online Methods, Fig. 1, Supplementary Tables 1 and 2 and Supplementary Note). We observed only modest genomic control inflation ($\lambda_{gc} = 1.07$), suggesting that the observed results were not due to population stratification. After removing SNPs within established T2D loci (Supplementary Table 3), the resulting quantile-quantile plot was consistent with a modest excess of disease associations of relatively small effect (Supplementary Note). Weak evidence for association at HLA variants strongly associated with autoimmune forms of diabetes (Supplementary Table 3 and Supplementary Note) suggested some case admixture involving subjects with type 1 diabetes or latent autoimmune diabetes of adulthood; however, failure to detect T2D associations at other non-HLA type 1 diabetes susceptibility loci (for example, INS, PTPN22 and IL2RA) indicated that any such misclassification was too modest to drive stage 1 associations outside the HLA. The stage 1 meta-analysis also provided further confirmation of many previously reported signals and, at some of these, refinement of the peak association signal (Fig. 1, Supplementary Table 3 and Supplementary Note).

We selected for stage 2 follow-up the most strongly associated SNP from each of the 23 new autosomal regions showing the most compelling evidence for association (all $P < 10^{-5}$ in stage 1; Supplementary Table 3). We combined exclusively in silico data from three GWA samples (Atherosclerosis Risk in Communities (ARIC) study, Nurses’ Health Study and Framingham Heart Study) not included in the primary meta-analysis (2,832 cases and 15,843 controls) with additional (predominantly de novo) genotyping in up to 31,580 cases and 44,082 controls, for a maximum possible stage 2 sample size of 34,412 cases and 59,925 controls (effective sample size of 79,246), all of European descent (Supplementary Tables 1 and 2).

Stage 2 analyses indicated that the set of 23 signals was enriched for true association signals. In all, 21 showed directional consistency of effect between stage 1 and 2 (binomial test, $P < 3.3 \times 10^{-5}$), and for 15, the stage 2 $P$ value was $< 0.05$ (Supplementary Note). In joint analysis of stage 1 and 2 data (up to 42,542 cases and 98,912 controls),

A full list of author affiliations appears at the end of the paper.

Received 19 January; accepted 26 May; published online 27 June 2010; corrected after print 27 August 2010; doi:10.1038/ng.609
13 autosomal loci exceeded the threshold for genome-wide significance (P ranging from $2.8 \times 10^{-8}$ to $1.4 \times 10^{-12}$) with allele-specific odds ratios (ORs) between 1.06 and 1.14 (Table 1 and Fig. 2). All signals remained close to or beyond genome-wide significance thresholds (the least significant P value was $5.2 \times 10^{-8}$) when we repeated analyses after implementing a second (post meta-analysis) round of genomic control adjustment within stage 1 data (Supplementary Note).

We extended our search for susceptibility signals to the X chromosome, identifying one further signal in the stage 1 discovery samples meeting our criteria for follow-up (represented by rs5945326, near DUSP9, $P = 2.3 \times 10^{-8}$). This SNP showed strong evidence for replication in 8,535 cases and 12,326 controls (OR (allowing for X-inactivation) $1.32$ when we repeated analyses after implementing a second (post meta-analysis) round of genomic control adjustment within stage 1 data (Supplementary Note).

### Fourteen signals reaching genome-wide significance

Two of the 14 signals reaching genome-wide significance on joint analysis (those near MTNR1B and IRS1) represent loci for which T2D associations have been recently reported in samples which partially overlap with those studied here.1,14-16 (Table 1).

A third signal (rs231362) on 11p15 overlaps both intron 11 of KCNQ1 and the KCNQ1OT1 transcript that controls regional imprinting and influences expression of nearby genes including CDKN1C, a known regulator of beta-cell development.18 This signal maps ~150 kb from T2D-associated SNPs in the 3’ end of KCNQ1 first identified in East Asian GWA scans8,9. SNPs within the 3’ signal were also detected in the current DIAGRAM+ meta-analysis (for example, rs163184, $P = 6.8 \times 10^{-5}$), but they failed to meet the threshold for initiating replication. A SNP in the 3’ region (rs2237895) that was reported to reach genome-wide significance in Danish samples8 was neither typed nor imputed in the DIAGRAM+ studies. In our European-descent samples, rs231362 and SNPs in the 3’ signal were not correlated ($r^2 < 0.05$), and conditional analyses (see below) establish these SNPs as independent (Fig. 2 and Supplementary Table 4). Further analysis in Icelandic samples has shown that both associations are restricted to the maternally transmitted allele11. Both T2D loci are independent of the common variant associations with electrocardiographic QT intervals that map at the 5’ end of KCNQ1 ($r^2 < 0.02, D’ < 0.35$ in HapMap European CEU data)19,20 (Supplementary Table 5).

Of the remaining loci, two (near BCL11A and HNF1A) have been highlighted in previous studies but are now shown to reach genome-wide significance. Rare mutations in HNF1A account for a substantial proportion of cases of maturity onset diabetes of the young, and a population-specific variant (G319S) influences T2D risk in Oji-Cree Indians.24 Confirmation of a common variant association at HNF1A brings to five the number of loci known to harbor both rare mutations causal for monogenic forms of diabetes and common variants predisposing to multifactorial diabetes, the others being PPARG, KCNJ11, WFS1 and HNF1B. A T2D association in the BCL11A region was suggested by the earlier DIAGRAM meta-analysis (rs10490072, $P = 3 \times 10^{-5}$), but replication was inconclusive; there is only modest linkage disequilibrium (LD) between rs10490072 and the lead SNP from the present analysis (rs243021, $r^2 = 0.22, D’ = 0.73$ in HapMap CEU).

The remaining nine signals map near the genes HMGA2, CENTD2, KLF14, PRC1, TP53INP1, ZBED3, ZFAND6, CHCHD9 and DUSP9 (Table 1 and Figs. 1 and 2) and represent new T2D risk loci uncovered by the DIAGRAM+ meta-analysis.

### Understanding the genetic architecture of type 2 diabetes

Combining newly identified and previously reported loci and assuming a multiplicative model, the sibling relative risk attributable to the 32 T2D susceptibility variants described in this paper is ~1.14. With addition of the five T2D loci recently identified by the Meta-Analysis of Glucose and Insulin-related traits Consortium (MAGIC) investigators and...
incorporation of estimates of parent-of-origin–specific effect sizes observed at the KCNQ1 and KLF14 signals and at a recently described locus on chromosome 10p15 (which confers substantial risk when paternally inherited but is protective when maternally transmitted), this figure rises to ~1.16. Given estimates of sibling relative risk for T2D in Europeans of ~3 (ref. 25), variant discovery efforts to date have therefore explained only ~10% of observed familial clustering. We used available data to evaluate several mechanisms that might be contributing to that proportion of familiality which reflects residual, unexplained heritability.

Copy number variants (CNVs). We re-examined stage 1 data looking for associations with SNPs known to provide robust, high-LD tags for common CNVs in European populations. After combining four inventories of CNV-tagging SNPs that survey at least 40% of common CNV’s genome-wide >1 kb in size, we found no convincing evidence that this class of variants contributes substantially to T2D risk (Supplementary Note).

Secondary signals revealed by conditional analysis. If there are additional independent susceptibility variants at the loci identified, total genetic variance attributable to these regions will be underestimated. We therefore explained only ~10% of observed familial clustering. We used available data to evaluate several mechanisms that might be contributing to that proportion of familiality which reflects residual, unexplained heritability.

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Position B36 (base pair)</th>
<th>Risk allelea</th>
<th>Nonrisk alleleb</th>
<th>Frequency risk allele (Hapmap CEU)</th>
<th>Nearby genei OR (95%CI)</th>
<th>P value</th>
<th>Stage 2a OR (95% CI)</th>
<th>P value</th>
<th>Stage 1 + 2 OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs243021</td>
<td>2</td>
<td>60,438,323 A G</td>
<td>0.46</td>
<td>BCL11A</td>
<td>1.09 (1.05–1.13)</td>
<td>8.1 × 10−6</td>
<td>1.08 (1.05–1.10)</td>
<td>6.2 × 10−11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs445950</td>
<td>7</td>
<td>76,460,705 G A</td>
<td>0.26</td>
<td>ZBED3</td>
<td>1.16 (1.10–1.23)</td>
<td>4.2 × 10−4</td>
<td>1.07 (1.04–1.10)</td>
<td>2.7 × 10−3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs972283</td>
<td>3</td>
<td>130,117,394 G A</td>
<td>0.55</td>
<td>KLF14</td>
<td>1.10 (1.06–1.15)</td>
<td>1.8 × 10−6</td>
<td>1.06 (1.03–1.09)</td>
<td>6.4 × 10−4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs896854</td>
<td>8</td>
<td>96,029,687 T C</td>
<td>0.48</td>
<td>TP53INP1</td>
<td>1.10 (1.06–1.15)</td>
<td>1.2 × 10−6</td>
<td>1.05 (1.03–1.08)</td>
<td>2.2 × 10−3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs13292136</td>
<td>9</td>
<td>81,141,948 C T</td>
<td>0.93</td>
<td>CHCHD9</td>
<td>1.20 (1.11–1.29)</td>
<td>1.5 × 10−6</td>
<td>1.08 (1.04–1.13)</td>
<td>2.4 × 10−4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Previously known

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Position B36 (base pair)</th>
<th>Risk allelea</th>
<th>Nonrisk alleleb</th>
<th>Frequency risk allele (Hapmap CEU)</th>
<th>Nearby genei OR (95%CI)</th>
<th>P value</th>
<th>Stage 2a OR (95% CI)</th>
<th>P value</th>
<th>Stage 1 + 2 OR (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs7579256</td>
<td>2</td>
<td>226,728,897 A G</td>
<td>0.64</td>
<td>IRS1</td>
<td>1.12 (1.07–1.17)</td>
<td>8.7 × 10−7</td>
<td>1.10 (1.08–1.13)</td>
<td>2.2 × 10−11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs1387153</td>
<td>11</td>
<td>92,313,476 T C</td>
<td>0.28</td>
<td>MNTR1B</td>
<td>1.10 (1.07–1.17)</td>
<td>1.0 × 10−6</td>
<td>1.08 (1.05–1.10)</td>
<td>4.4 × 10−4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OR, odds ratio; CI, confidence interval; Chr., chromosome.
aProxy SNP rs243021 was genotyped in METSIM, FUSION2 and HUNT stage 2 samples. Alleles are indexed to the forward strand of NCBI Build 36.

bGene named for nearest gene (except where there is a very strong positional candidate).

cAll P values reported are two-sided and based on a inverse-variance weighted meta-analysis model (fixed effects).

A number of previously known SNPs are listed. All P values are provided for stage 1 + 2 which exceed a genome-wide threshold (overall P value < 5 × 10−8).

(1) Previously known SNPs were included in the analysis.

(2) Power to detect autosomal SNPs given the stage 1 sample sizes and stage 2 ORs, the risk allele frequency (case and control), and the overall Type I error rate (α = 5 × 10−8).

(3) Positional candidate.

(4) Alleles are indexed to the forward strand of NCBI Build 36.

(5) Maximum stage 2 sample size for the X chromosome SNP was 8,535 cases and 12,326 controls. Allelic ORs calculated assuming 1.07–1.17

(6) P values are reported as two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects).

(7) Allelic ORs calculated assuming 1.05–1.10

(8) Allelic ORs calculated assuming 1.03–1.08

(9) Allelic ORs calculated assuming 1.04–1.09

(10) Allelic ORs calculated assuming 1.07–1.17

(11) Allelic ORs calculated assuming 1.04–1.07

(12) Allelic ORs calculated assuming 1.06–1.10

(13) Allelic ORs calculated assuming 1.04–1.11
Figure 2  Regional plots of the 12 newly discovered T2D loci. Genotyped and imputed SNPs passing quality control measures across all stage 1 studies are plotted with their meta-analysis $P$ values (as $-\log_{10}(P)$ values) as a function of genomic position (NCBI Build 36). In each panel, the index association SNP is represented by a diamond, with stage 1 meta-analysis results denoted by a red diamond and the combined stage 1 and stage 2 meta-analysis results denoted with a clear symbol. Estimated recombination rates (taken from HapMap CEU) are plotted to reflect the local LD structure. Color of remaining SNPs (circles) indicates LD with the index SNP according to a scale from $r^2 = 0$ to $r^2 = 1$ based on pairwise $r^2$ values from HapMap CEU (red, $r^2 = 0.8–1.0$; orange, $r^2 = 0.6–0.8$; green, $r^2 = 0.4–0.6$; blue, $r^2 = 0.2–0.4$; black, $r^2 < 0.2$; gray, no $r^2$ value available). Gene annotations were taken from the University of California Santa Cruz genome browser.
Etiological heterogeneity. To determine whether etiological heterogeneity might have compromised power to detect genuine T2D susceptibility signals, we performed BMI- and age-of-diagnosis (AOD)-stratified analyses within stage 1 data. We compared effect size estimates for all known T2D risk variants in 2,877 obese (defined as BMI > 30 kg/m²) and 4,048 nonobese (BMI ≤ 30 kg/m²) T2D cases when compared to similarly stratified controls (Supplementary Note). Although risk estimates for 23 of the 30 autosomal loci were numerically greater in the nonobese comparison than in the obese comparison (binomial test, \( P = 0.0018 \)), only TCF7L2 (\( P < 0.001 \)) and BCL11A (\( P = 0.02 \)) showed significant (\( P < 0.05 \)) evidence for effect-size heterogeneity. For AOD, we compared risk locus genotypes for 1,317 cases with AOD <45 years of age and 4,283 cases with AOD >45 years of age, as well as continuous analyses of AOD within all cases (\( n = 7,104 \); Supplementary Note), and found no strong evidence of differential effects. Although recognizing that BMI at examination and AOD are imperfect measures of BMI and age at disease onset, we conclude that a focus on more homogeneous subsamples would not have provided more efficient identification of known T2D susceptibility variants. Furthermore, these data argue against the potential for these common variant signals to afford clinically useful subclassification of individuals with T2D.

Overlap with GWA signals for other diseases. We noted that seven of the newly discovered autosomal loci (near BCL11A, ZBED3, KLF14, CHICD9, HMGA2, HNF1A and PRC1) are characterized by strong (\( P < 10^{-6} \)) associations with phenotypes other than T2D (Supplementary Table 5). In each case, these appear to be distinct and independent signals. For example, variants at the 3’ end of HMGA2 (~180 kb distant from the T2D signal) have widely replicated effects on adult height but are weakly correlated with the T2D-associated SNP rs1531343 (\( r^2 < 0.01 \), \( D’ < 0.15 \) in HapMap CEU). The KLF14 region harbors distinct signals for both T2D and basal cell carcinoma\(^{39} \). At HNF1A, previous studies have reported a cluster of associations, with phenotypes including low-density lipoprotein (LDL) cholesterol\(^{30} \) and circulating C-reactive protein levels\(^{31–33} \), mapping ~18–72 kb from the peak T2D signal. Though these two sets of HNF1A signals maintain appreciable LD in European samples (\( r^2 \sim 0.1, D’ \sim 1 \)), they are likely to be independent; the T2D association at the lead SNP for lipids (rs26500000) is far weaker than the association at rs7957197 (\( P = 0.003 \) compared to \( P = 4.6 \times 10^{-7} \) in stage 1 samples), whereas LDL cholesterol shows a reciprocal pattern of association (\( P = 7 \times 10^{-9} \) at rs26500000 compared to \( P = 0.73 \) at rs7957197 in the same lipid meta-analysis data\(^{30} \)).

If we include the KCNQ1 associations described above, previous reports at JAZF1, CDKN2A and CDKAL1 (refs. 34–40) and other signals identified by systematic analysis of the National Human Genome Research Institute (NHGRI) GWA catalog\(^{41} \) (Supplementary Table 5 and Supplementary Note), at least 13 of 30 autosomal T2D loci show this pattern of closely approximated (within 500 kb) but distinct associations with traits other than T2D or related anthropometric and glycemic phenotypes. This is in addition to what appear to be coincident signals involving T2D susceptibility variants at IRS1 (associated with coronary disease), JAZF1 (associated with height) and HNF1B (associated with prostate cancer) (Supplementary Table 5). Simulations conducted using the NHGRI catalog as a reference set indicate that the number of non-T2D signals observed at T2D loci significantly exceeds expectation (\( P \sim 1.6 \times 10^{-3} \) for non-T2D signals within 500 kb of T2D loci, \( P \sim 7.0 \times 10^{-3} (n = 8) \) for non-T2D signals within 100 kb of T2D loci). Many of these instances of colocalization may represent variants within different regulatory domains that result in tissue- and diseasespecific effects mediated through the same genes and pathways.

Understanding the biology of T2D-susceptibility loci

This analysis takes the number of independent loci showing genomewide significant associations with T2D beyond 35. For some, such as those at KCNJ11 and SLC30A8, the molecular mechanisms responsible for the susceptibility effect can be assigned with some confidence\(^{42} \). At others, the identities of the causal variants, the genes through which they act and the pathophysiological processes which they influence remain obscure. We used several approaches designed to link DIAGRAM+ and previously reported T2D association signals to biological insights relevant to T2D pathogenesis.

Physiological analyses. Variants at FTO are known to influence T2D predisposition through an effect on BMI. In ~21,000 population sample individuals from the GWA meta-analysis of adult BMI completed by the Genetic Investigation of ANthrometric Traits (GIANT) consortium\(^3 \), no other autosomal T2D susceptibility locus had the property that the T2D risk allele was significantly associated with higher BMI (Supplementary Note). FTO is therefore the only one of the known T2D signals driven by a strong primary causal association with obesity.

We also examined the effect of T2D susceptibility alleles on continuous glycemic measures in up to 46,186 nondiabetic subjects from the MAGIC meta-analysis\(^{12,13} \). Coefficients for association between the T2D risk allele and higher fasting glucose were positive for 28 of the 31 loci, and 17 of these T2D loci showed significant (\( P < 0.05 \)) directionally consistent associations with fasting glucose (Fig. 3 and Supplementary Note). However, the magnitudes of effect sizes for fasting glucose and T2D were only weakly correlated (Supplementary Fig. 2 and Supplementary Note), indicating that the mechanisms influencing normal glucose homeostasis and those responsible for the development of T2D are not entirely congruent. T2D risk alleles at four loci (at PPARG, FTO, IRS1 and KLF14) were associated (\( P < 0.05 \)) with higher fasting insulin, consistent with a primary effect on insulin action, whereas at three other loci (at TCF7L2, CNTD2 and CDKAL1), the association with reduced fasting insulin indicates beta-cell dysfunction (Fig. 3). Indices of beta-cell function (HOMA-B) and insulin sensitivity (HOMA-IR) derived from paired fasting glucose and insulin measurements from ~37,000 individuals supported these mechanistic inferences (Fig. 3). In all, risk alleles at ten loci (the previously reported loci at MTNR1B, SLC30A8, THADA, TCF7L2, KCNQ1, CAMK1D, CDKAL1, IGF2BP2 and HNF1B and the newly discovered locus at CNTD2) were associated (\( P < 0.05 \)) with reduced beta-cell function, and three loci (previously reported loci at PPARG and FTO and the newly discovered locus at KLF14) were associated with reduced insulin sensitivity. The associations with improved insulin sensitivity evident for risk alleles at TCF7L2, IGF2BP2 and CDKAL1 probably reflect truncated ascertainment, as the MAGIC analyses were restricted to nondiabetic individuals. For the previously reported loci, these findings are broadly consistent with those from more detailed physiological studies\(^{6,8,44} \) and suggest that, of the newly discovered loci, the risk alleles at CNTD2 modify T2D susceptibility through a detrimental effect on beta-cell function. In contrast, the risk alleles at KLF14 and possibly HMGA2 (ref. 45), along with those at PPARG, IRS1 (ref. 10) and ADAMTS9 (ref. 46), appear to have a primary effect on insulin action which is not driven by obesity, unlike the alleles at FTO. The MAGIC meta-analysis did not extend to the X chromosome, but analysis of rs5945326, near DUSP9, in a subset of MAGIC samples (\( n = 14,644–21,118 \)), revealed no significant (\( P < 0.05 \)) associations with any fasting glycemic trait. For this signal, as with the other newly identified loci, more detailed phenotypic analyses will be required to determine how these impact T2D risk. Overall, these data are consistent with the impression that common
We used expression and pathway enrichment analyses to identify over-represented molecular processes (an index of beta cell function; x axis) and HOMA-B (an index of insulin sensitivity; y axis). Each point refers to a single T2D association signal, with colors denoting the strength of the association to either the x-axis variable (left-hand of each pair of plots) or y-axis variable (right-hand of each pair) (red, P < 10^{-3}; orange, 10^{-3} < P < 10^{-2}; yellow, 0.01 < P < 0.05; green, 0.05 < P < 0.20; blue, P > 0.20). The two KCNQ1 associations are distinguished by the notation KCNQ1 for rs163184 and KCNQ1* for rs231362. The gene names associated with each signal have been chosen on the basis of proximity to the index SNP and should not be presumed to indicate causality.

T2D risk alleles more often act through beta-cell dysfunction\(^{13,44}\), but they provide further examples of T2D risk variants that exert their primary effects on insulin action.

**Expression analyses.** We used expression data to seek clues to the genes mediating the T2D susceptibility effects we had detected. First, we examined expression-QTL (eQTL) data (in 23,720 transcripts) for subcutaneous adipose tissue \((n = 603\) with GWA data) and blood \((n = 745\) samples typed with the Illumina 300K chip\(^{47}\) (Table 2 and Supplementary Note). Among the newly identified loci, the most compelling signal was at rs972283, strongly associated with expression of KLF14 in adipose tissue and correlated \((r^2 = 0.3\) in HapMap CEU) with the SNP \((rs738134\) with the strongest KLF14 cis expression signal. Both the T2D and cis eQTL associations at this locus showed strong parent-of-origin effects\(^{11}\). At the TP53INP1 locus, the cis-eQTL data suggest the T2D susceptibility effect is mediated via altered CCNE2 expression. In contrast, the significant cis-eQTL associations at the ZBED3, CENTD2, HNF1A and PRC1 T2D susceptibility signals are likely to be misleading, as the patterns of conditional association indicate that the T2D association and cis eQTL signals are not coincident. At previously reported T2D association signals, we found strong overlap with cis eQTL effects for IRS1 (consistent with data on IRS1 protein expression and function in skeletal muscle\(^{14}\), JAZF1 and CAMK1D\(^{7}\).

We also explored the tissue expression profiles of 27 autosomal genes mapping to the newly discovered regions of association and performed quantitative RT-PCR analyses across a panel of human tissues relevant to T2D pathogenesis (Supplementary Note). The broad expression of many of the transcripts, including 24 transcripts with evidence of beta-cell transcription (Supplementary Note), limited our ability to prioritize among candidate transcripts on the basis of static patterns of transcript expression.

**Pathway and protein-protein interaction analyses.** Reasoning that the additional T2D susceptibility loci would amplify our ability to identify over-represented molecular processes\(^{48}\), we deployed several complementary approaches to detect evidence of pathway or network enrichment (Supplementary Note). Using GRAIL\(^{49}\), we found that genes within T2D-associated regions showed evidence of increased connectivity within PubMed abstracts, though this largely reflects shared roles in monogenic or syndromic diabetes (involving HNF1A, HNF1B and WFS1). We also showed that the extent of protein-protein interaction between the products of genes mapping to the association signals substantially exceeded expectation (Supplementary Note). Pathway enrichment analyses using the PANTHER database\(^{50}\) uncovered some evidence of over-representation of signal transduction and protein metabolism and modification, and Reactome\(^{51}\) highlighted a separate set of pathways including metabolism of lipids and lipoproteins, endothelins and beta-arrestins (for details, see Supplementary Note).

The only consistent signal to emerge across multiple analyses involved cell-cycle regulation. Network analyses based on protein-protein interaction data detected (unadjusted P < 0.004) an 18-member subnetwork characterized by enhanced protein-protein interaction connectivity and highly enriched for genes implicated in cell cycle regulation \((P = 2.8 \times 10^{-7})\). A smaller (five, only partly overlapping genes) cell-cycle network independently emerged from the Reactome analyses, and gene-set enrichment analysis of selected candidate pathways\(^{52}\) also detected over-representation of association signals \((P \sim 0.006\) among cell-cycle genes (Supplementary Note). Because many genes within these networks are expressed in pancreatic islets and T2D-association effects at several of these loci are mediated primarily through beta-cell dysfunction\(^{44}\), these findings highlight the contribution of regulation of beta-cell mass to the long-term maintenance of normal glucose homeostasis.

In addition, these analyses highlighted notable biological connections between sets of genes within confirmed T2D-association regions. For example, HMGA2 emerges as a key transcriptional regulator of IGF2BP2 (refs. 53,54). However, because Hmga1/Hmga2 knockout mice are deficient in adipocyte differentiation\(^{43}\), and the IGF2BP2 risk allele is associated with reduced beta-cell function\(^{55}\), further work is required to establish the relevance of this regulatory
interaction to T2D pathogenesis. Our analyses also revealed that TLE4 (at the CHCHD9 locus) encodes a homolog of Groucho that forms complexes with TCF proteins, including TCF7L2, to modulate transcription at target sites\(^8\). Finally, FURIN, one of the genes mapping to the newly identified PRC1 locus, encodes a paired basic amino acid cleaving enzyme; both NOTCH2 and ADAMTS9 (ref. 7) are known targets of FURIN cleavage\(^7\)–\(^9\).

Notably, these global approaches failed to provide any consistent support for many other mechanisms previously promoted on the basis of biochemical or physiological evidence as likely contributors to T2D pathogenesis\(^9\) (Supplementary Note). Overall, the relative paucity of signals from these analyses—particularly when contrasted with the compelling patterns of enrichment seen for other complex traits\(^8\)—indicates, either that T2D pathogenesis is characterized by...

---

### Table 2 Expression QTL results for T2D-associated variants in blood and adipose tissue

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Position (bp)</th>
<th>Nearby Gene(s)</th>
<th>Risk Allele(s)</th>
<th>Gene (transcript)</th>
<th>Tissue</th>
<th>Effect (standard error)</th>
<th>Adj. P</th>
<th>SNP with strongest correlation with trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs4457053</td>
<td>5</td>
<td>76,460,705</td>
<td>ZBED3</td>
<td>G</td>
<td>PDE8B (NM_003719)</td>
<td>Adipose</td>
<td>0.302 (0.070)</td>
<td>2.8 x 10^{-6}</td>
<td>0.80</td>
</tr>
<tr>
<td>rs972283</td>
<td>7</td>
<td>130,117,394</td>
<td>KLF14</td>
<td>G</td>
<td>KLF14 (NM_138693)</td>
<td>Adipose</td>
<td>−0.387 (0.058)</td>
<td>8.1 x 10^{-11}</td>
<td>0.058</td>
</tr>
<tr>
<td>rs896854</td>
<td>8</td>
<td>96,029,687</td>
<td>TPS3IPN1</td>
<td>C</td>
<td>CCNE2 (NM_057749)</td>
<td>Blood</td>
<td>−0.225 (0.053)</td>
<td>3.8 x 10^{-6}</td>
<td>0.78</td>
</tr>
<tr>
<td>rs1552224</td>
<td>11</td>
<td>72,110,746</td>
<td>CENTD2</td>
<td>A</td>
<td>STAR010 (NM_006645)</td>
<td>Blood</td>
<td>0.337 (0.066)</td>
<td>8.6 x 10^{-7}</td>
<td>0.026</td>
</tr>
<tr>
<td>rs7957197</td>
<td>12</td>
<td>119,945,069</td>
<td>HNF1A</td>
<td>T</td>
<td>ACADS (NM_000017)</td>
<td>Adipose</td>
<td>−0.248 (0.067)</td>
<td>3.7 x 10^{-6}</td>
<td>0.29</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chr.</th>
<th>Position (bp)</th>
<th>Nearby Gene(s)</th>
<th>Risk Allele(s)</th>
<th>Gene (transcript)</th>
<th>Tissue</th>
<th>Effect (standard error)</th>
<th>Adj. P</th>
<th>SNP with strongest correlation with trait</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8042680</td>
<td>15</td>
<td>89,322,341</td>
<td>PRC1</td>
<td>A</td>
<td>VPS33B (NM_018668)</td>
<td>Blood</td>
<td>−0.371 (0.057)</td>
<td>2.9 x 10^{-10}</td>
<td>0.50</td>
</tr>
<tr>
<td>rs758326</td>
<td>2</td>
<td>226,722,897</td>
<td>IRS1</td>
<td>A</td>
<td>IRS1 (Contig50189_RC)</td>
<td>Adipose</td>
<td>−0.251 (0.059)</td>
<td>3.7 x 10^{-6}</td>
<td>0.89</td>
</tr>
<tr>
<td>rs13081389</td>
<td>3</td>
<td>12,264,800</td>
<td>PPARG</td>
<td>A</td>
<td>IGSEC1 (NM_014869)</td>
<td>Adipose</td>
<td>−0.630 (0.131)</td>
<td>2.9 x 10^{-6}</td>
<td>1.4 x 10^{-4}</td>
</tr>
<tr>
<td>rs6795735</td>
<td>3</td>
<td>64,680,405</td>
<td>ADAMTS9C</td>
<td>C</td>
<td>BC040632 (AK022320)</td>
<td>Blood</td>
<td>−0.229 (0.056)</td>
<td>7.6 x 10^{-5}</td>
<td>0.28</td>
</tr>
<tr>
<td>rs849134</td>
<td>7</td>
<td>28,162,747</td>
<td>JAZF1</td>
<td>A</td>
<td>JAZF1 (NM_175061)</td>
<td>Blood</td>
<td>−0.346 (0.055)</td>
<td>1.4 x 10^{-9}</td>
<td>0.70</td>
</tr>
<tr>
<td>rs12779790</td>
<td>10</td>
<td>12,368,016</td>
<td>CAMK1D</td>
<td>G</td>
<td>CAMK1D (AL137340)</td>
<td>Blood</td>
<td>0.668 (0.062)</td>
<td>1.6 x 10^{-5}</td>
<td>0.052</td>
</tr>
<tr>
<td>rs5015480</td>
<td>10</td>
<td>94,455,539</td>
<td>HHEX/IDE</td>
<td>C</td>
<td>KIF11 (NM_004523)</td>
<td>Adipose</td>
<td>0.202 (0.059)</td>
<td>9.4 x 10^{-4}</td>
<td>0.54</td>
</tr>
<tr>
<td>rs972283</td>
<td>7</td>
<td>130,117,394</td>
<td>KLF14</td>
<td>G</td>
<td>KLF14 (NM_138693)</td>
<td>Adipose</td>
<td>−0.387 (0.058)</td>
<td>8.1 x 10^{-11}</td>
<td>0.058</td>
</tr>
</tbody>
</table>

**Notes:**

- Chr.: chromosome.
- Gene(s): genes within 300 kb of the SNP.
- Risk Allele(s): minor allele(s) identified in GWAS.
- Gene (transcript): RefSeq transcript ID (NM_174999).
- Effect (standard error): mean-log expression ratio (log2).
- Adj. P: adjusted p-value.
- SNP with strongest correlation with trait: SNPs with strongest correlation with T2D expression trait.
substantial etiological heterogeneity or that the processes critical to T2D development are poorly represented in existing pathway and interaction databases.

DISCUSSION

By increasing the discovery sample size, our study has substantially expanded the number of loci for which there is strong statistical evidence indicating a role in T2D predisposition. When combined with recent reports of additional T2D susceptibility loci arising from studies of continuous glycemic traits and parent-of-origin effects, the number of confirmed loci for T2D currently stands at 38.

Although these discoveries represent new opportunities to explore the biology of T2D predisposition, the challenges inherent in translating these common variant association signals into biological mechanisms of disease causation are clear. Nevertheless, the analyses we report have generated several mechanistic hypotheses that can direct future efforts at functional evaluation and genetic refinement. At some loci, particularly those near HNF1A, HMGA2 and KLF14, existing biology, coupled with phenotypic and expression data presented here, highlight the named genes as prime candidates for mediating the susceptibility effect. For example, the T2D susceptibility effect near KLF14, which maps within an imprinted region on chromosome 7q32 and which, on the basis of the MAGIC meta-analysis data, appears to be driven by reduced insulin action, is restricted to the maternally transmitted allele. As KLF14 is maternally expressed, and the eQTL association between rs972283 and KLF14 expression (see above) is similarly restricted to the maternal allele, KLF14 (a widely expressed, intronless member of the Krüppel-like family of transcription factors) emerges as the main regional candidate. At the X-chromosome signal, evidence implicating DUSP9 (mitogen-activated protein kinase phosphatase-4) in the regulation of insulin action in mice gives DUSP9 particular salience as an association candidate. However, as described above, failure to detect associations with continuous glycemic phenotypes (including fasting insulin and HOMA-IR) means that the functional connection with DUSP9 remains speculative.

In other regions, such as those near PRC1, TPS31NP1 and CHCHD9, the functional connections and/or eQTL associations of particular genes mapping within or close to the respective association intervals show FURIN, CCNE2 and TLE4, respectively, to be promising biological candidates. At yet other loci, such as those centered around ZBED3, CENTD2 and ZFAND6, existing data provide little, if any, basis for strong inferences concerning the genes likely to mediate the T2D susceptibility effect. Accumulation of new data—through deep sequencing of the regions, fine-mapping and functional studies in humans and in animal models—will be required to characterize the specific variants responsible and the genes and pathways through which they execute their effect on T2D risk.

One theme emerging from this work is the high frequency with which loci implicated in T2D susceptibility harbor variants that influence other common traits. This colocalization of common risk variants exceeds chance expectation, often connects diseases with little obvious mechanistic overlap and typically involves statistically independent susceptibility signals. Recent evidence that tissue-specific eQTL signals are preferentially located in regulatory sequences some distance from transcriptional start sites—in common with many complex trait association signals—suggests that further dissection of these regions should improve understanding of the genomic organization of tissue- and/or developmental-stage-specific regulation.

A further conclusion is that common SNP signals are likely to fall short in explaining the observed familial aggregation of T2D, at least in European descent populations. The limited power of our study (Table 1) to detect several of the genome-wide significant variants we report here (based on the stage 1 sample size and stage 2 odds ratios that minimize ‘winner’s curse’ effects) indicates that there are likely to be many additional common variant signals of similar effect that could be detected by further expansion of the GWA meta-analysis approach. However, it seems unlikely that these will explain a substantial proportion of unexplained heritability. Based on the data presented, the same is likely to be true for common CNVs and for variation on the sex chromosomes. As a result, the attention of researchers in the field is increasingly directed toward evaluation of the contribution of low frequency and rare variants to complex trait susceptibility. Several lines of evidence—the overlap in loci implicated in monogenic and multifactorial diabetes, the congregation of multiple disease signals at a limited number of loci and the conditional analyses—point toward the importance of obtaining complete descriptions of causal genetic variation (of all types and frequencies) at the loci uncovered by this and other GWA studies. Such loci are likely to represent hotspots at which the overall contribution to T2D predisposition and biology may be considerably greater than that estimated using the discovered common variants alone.

METHODS

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

ACKNOWLEDGMENTS

We acknowledge funding from: the Academy of Finland (no. 124243); Agence Nationale de la Recherche (France); American Diabetes Association (1-05-RA-140, 7-08-MN-OK, 7-06-MN-05); Ardis Medical; Association Diabète Risque Vasculaire; Association de Langue Française pour l’Étude du Diabète et des Maladies Métaboliques; Association Française des Diabétiques; Bayer Diagnostics; British Diabetic Association Research; Becton Dickinson; Broad Institute of Harvard and Massachusetts Institute of Technology; The Burroughs Wellcome Fund; Cardionics; Center for Inherited Disease Research (USA); Centre for Medical Systems Biology (The Netherlands); Centre of Excellence Meteorological Disorders Baden-Württemberg (Germany); Caisse Nationale Assurance Maladie des Travailleurs Salariés (France); Clinical Research Institute HUCH Ltd; Deutsche Forschungsgemeinschaft (DFG GrK 1041, DFG RA459, SFB 518); the Danish Diabetes Association; the Danish Health Research Council; Diabetes UK; Doris Duke Charitable Foundation; Erasmus Medical Center (The Netherlands); the Dutch Diabetes Foundation; European Community (HEALTH-F4-2007-201413, HEALTH-2007-B-223211, LSHG-CT-2006-01947, LSHM-CT-2004-512013, LSHM-CT-2004-005272, LSHM-CT-2006-518153); the European Foundation for the Study of Diabetes; the Federal Ministry of Health (Germany); the Federal Ministry of Education and Research (Germany) (FKZ 01G085023 and DZD e.V.); Fédération Française de Cardiologie; The Finnish Diabetes Research Foundation; The Folkhalsan Research Foundation; The Foundation for Strategic Research (Sweden); The Foundation of Bristol-Myers Squibb; the German National Genome Research Network; Helmholz Zentrum München-Research Center for Environment and Health; INSERM (France); La Fondation de France; Lilly; The Linnaeus Centre for Bioinformatics (Sweden); the Lundbeck Foundation Centre of Applied Medical Genomics for Personalized Disease Prediction, Prevention and Care; the Medical Research Council UK (G0601261, G0006491, 881696); Munich Center of Health Sciences–LMU Innovativ (Germany); Merck Santé; the Ministry of Health and Department of Educational Assistance, University and Research of the Autonomous Province of Bolzano (Italy); the Ministry of Innovation, Science, Research and Technology of the State of North Rhine-Westphalia (Germany); the Ministry of Science, Education and Sport (Croatia); the National Heart, Lung, and Blood Institute (N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, N01-HC-55022, N01-HC-55025, N01-HC-55026, N01-HC-55027, N01-HC-55028, N01-HC-55029); National Human Genome Research Institute (U01HG004402, U01HG004399, U01HG004171, 1 Z01 HG000024); the National Institute of Diabetes, Digestive and Kidney Diseases (DK70816, K24-DK080140, U54 DA021519, DK38845, DK60922, DK602370, DK073490, K23-DK65978 and DK072193); the US National Institutes of Health (NN争夺26829, 052226C, HHSN236200625226C, 1K08AR055688, U11R020505, 1K9HL094535-01A1); the Netherlands Foundation for Scientific Research (175.010.2005.011, 047.017.043);
Nord-Pas-de-Calais region (France); Novartis Pharma; Novo Nordisk; the Oxford National Institute for Health Research (NIHR) Biomedical Research Centre (UK); Office National Inter-professionnel des Vins; Peninsula Medical School, Exeter UK; Pfizer Inc; Pierre Fabre laboratory (France); Programme National de Recherche sur le Diabète (France); Richard and Susan Family Foundation/American Diabetes Association Pinnacle Program Project; Roche; the Royal Society (UK); Russian Foundation for Basic Research (047.017.043); Sanofi-Aventis; Sarnoff Cardiovascular Research; Scottish Government Chief Scientist Office; SenterNovem (IOP Genomics grant I0G09012); Sigrid Juselius Foundation; the Skaraborg Institute, Skövde, Sweden; South Tyrolean Sparkasse Foundation; the Swedish Natural Sciences Research Council; The Swedish Research Council (349 2006-2377); the Association Diabète Risque Vasculaire (France); Topcon; the Wallenberg Foundation; and the Wellcome Trust (072960; 076113; 083270; 088885; 079557; 081682; 08596; 077016; 074591). A more complete list of acknowledgments is provided in the Supplementary Note.

AUTHOR CONTRIBUTIONS


ONLINE METHODS

Stage 1 samples, genotyping and analysis. We combined results from eight genome-wide T2D-case control association studies involving European-descent samples: DGDG, deCODE, DGI, Rotterdam, EUROSPAN, FUSION, KORAgen and WTCCC. Sample characteristics are provided in Supplementary Table 2. Details of the SNP genotyping platforms, genotype and sample quality-control measures and autosomal T2D association analysis approach (additive model on ln (OR) scale) are provided in the Supplementary Note. Imputation on the X chromosome was performed using IMPUTE (see URLs) with males coded as homozygotes. For X chromosome SNPs, ORs were calculated under the assumption of X inactivation; in other words, the per-allele OR cited is based on the between-homozygote difference. We performed autosomal and X-chromosomal association analysis on directly genotyped SNPs where available, or on imputed data using the expected allele counts or genotype probabilities.

Stage 1 meta-analysis. The combined stage 1 sample included 8,130 cases and 38,987 controls with an effective sample size of 22,044 (that is, a sample size with power equivalent to 11,022 cases and 11,022 controls). We excluded <1,000 SNPs where the allele frequencies were <0.40 in at least one study and >0.60 in others to minimize possible difficulties due to mislabeled alleles. A total of 2,426,886 autosomal and 60,771 chromosome X chromosome SNPs were available for ≥3 studies; of these, 2,255,857 autosomal and 52,946 chromosome X SNPs were available for ≥17,000 effective total samples. We performed genomic control correction26 of autosomal data from each individual study (separately for directly genotyped and for imputed data) by inflating the standard error of the estimated ln (OR) so that the significance of the SNP matched that of the genomic control P value. For those autosomal SNPs with data on ≥17,000 effective total samples, we estimated the genomic control inflation factor to be 1.074 (1.069 after removing 9,939 SNPs from regions of known T2D association). We used inverse variance-weighted meta-analysis to combine association results for stage 1 and investigated evidence for heterogeneity of ORs using Cochran's Q statistic and I² (ref. 65). We present autosomal meta-analysis results based on individual study genomic control correction in the main paper. The Supplementary Note presents association results after a second round of genomic control adjustment based on the results of the meta-analysis. Individual loci were plotted using LocusZoom (see URLs).

Selection of stage 1 SNPs for stage 2 genotyping. We selected the most strongly associated SNP from each region containing ≥1 SNPs with a stage 1 fixed effects meta-analysis P ≤ 1 × 10⁻⁵ based on data from ≥3 studies. We removed SNPs within 250 kb of the index SNPs for previously identified T2D associations at TCF7L2, PPARG, KCNJ11, CDKAL1, CDKN2A, IGF2BP2, FTO, HHEX, SLC30A8, JAZF1, THADA, CDC123, TSPAN8, NOTCH2, ADAMTS9, HNF1B and WFS1. We did not exclude KCNQ1, as we had evidence of a signal independent of the association previously identified in East Asian samples6,9. Of the 24 SNPs selected for follow-up, all had data from ≥3 studies; 22 had data from ≥17,000 effective total samples (Supplementary Table 3).

Stage 2 samples, genotyping and analysis. We followed up the 23 most strongly associated autosomal SNPs from stage 1 in 19 stage 2 studies by performing in silico replication (for 3 studies), de novo genotyping (for 15 studies) or a combination of both (for 1 study). The stage 2 sample for autosomal signal follow-up comprised up to 34,412 cases and 59,925 controls with an effective sample size of 79,246. For the single X-chromosome signal of interest (identified some time after the autosomal signals), we performed a fixed effects meta-analysis of study-specific AOD effects. In addition, for the five studies with ≥100 cases with AOD <45 years, we meta-analyzed data comparing genotype distributions between early onset (<45 years) and later-onset (≥245 years) cases.

Calculation of sibling relative risk. The contribution to the sibling relative risk of a SNP with SNP allele frequency p and corresponding allelic odds ratio β is assuming a multiplicative genetic model66. Relative risk estimates were combined as products. We estimated sibling relative risks for the 32 established and newly discovered T2D-associated SNPs based on the stage 2 ORs. We also optionally included the five T2D-associated SNPs identified by MAGIC investigators27,13. To capture the parent-of-origin-specific effects at KCNQ1, KLF14 and the additional locus on 11p15, we used previous data11.

Meta-analysis results for T2D SNPs for insulin and glucose-related traits. To establish whether the known and newly discovered T2D susceptibility loci were associated with other continuous glycemic phenotypes, we obtained meta-analysis association data (regression coefficients and P values) from the MAGIC analysis of 21 studies of nondiabetic European-descent individuals12,13. The MAGIC meta-analysis comprised ∼2.5 million genotyped or imputed autosomal SNPs and included up to 46,263 individuals for fasting glucose and up to 38,413 for fasting insulin. Surrogate estimates of beta-cell function (HOMA-B) and insulin resistance (HOMA-IR) derived from fasting variables by homeostasis model assessment47 were also analyzed.

Analysis of CNV-tagging SNPs. We combined autosomal CNV-tagging SNPs from a range of sources (see Supplementary Note). The union of these lists provided 5,219 unique CNV-tagging SNPs for which we had GWA data for ≥17,000 stage 1 samples.

Stage 1 conditional analysis. To identify additional signals after accounting for the effects of known T2D loci, we repeated the GWA analysis after conditioning simultaneously on 30 established and newly discovered T2D-associated SNPs (see Supplementary Note). We did not include SNPs representing the X chromosome DUSP9 locus (as it had not yet been verified as genome-wide significant when the analysis was performed), nor the KCNQ1 locus first identified in East Asian samples6,9 (as rs2237895—genome-wide significant in European ancestry samples—was not present in our tested SNPs or in HaMap); we also did not include the recent addition of six further susceptibility loci11–13 (as these analyses preceded those reports). Analyses were performed as described for the initial stage 1 analyses, with inclusion of study-specific covariates in addition to covariates for the 30 SNPs. These conditional analyses were performed in studies that accounted for 97% of the original stage 1 effective sample size, the exceptions being the non-Erasmus Rucphen Family (ERF) study components of EUROSPAN. To maximize the number of samples analyzed, where genotypes were missing, we replaced them at the 30 conditioned SNPs with the expected allele count based on imputation. We used a P < 10⁻⁴ to identify putative additional signals of interest in identified T2D loci and P < 10⁻³ in other regions of the genome.

Etiological heterogeneity (BMI and age-of-diagnosis–stratified analyses). For established and newly discovered autosomal T2D-associated variants, we performed a T2D case-control analysis within two BMI strata: BMI ≤ 30 kg/m² (4,048 T2D cases and 25,096 controls from 7 studies; see Supplementary Note for details) and BMI > 30 kg/m² (2,877 T2D cases, 6,774 T2D controls from 6 studies). We performed a fixed effects meta-analysis for each BMI stratum and tested for heterogeneity of the meta-analysis BMI strata-specific OR using Cochran's Q statistic65. We performed two analyses to examine the relationship between known T2D-associated variants and AOD of T2D. First, we analyzed AOD as an untransformed continuous trait using linear regression analysis. For the stage 2 meta-analysis (including 2,255,857 autosomal SNPs), ORs were calculated under the assumption of X inactivation; in other words, the per-allele OR cited is based on the between-homozygote difference. For X chromosome SNPs, ORs were calculated under the assumption of X inactivation; in other words, the per-allele OR cited is based on the between-homozygote difference. We performed autosomal and X-chromosomal association analysis on directly genotyped SNPs where available, or on imputed data using the expected allele counts or genotype probabilities.
BMI meta-analysis results for T2D SNPs from the GIANT Consortium. To establish whether the T2D susceptibility loci were associated with BMI, we examined BMI association data using z score, direction of association and P values) for these variants from the GIANT Consortium. Because inclusion of datasets ascertained for T2D case or control status could, at T2D loci, lead to distortion of the relationship with BMI, we restricted these analyses to eight population-based studies (including 21,233 individuals).

eQTL data. To identify T2D-associated SNPs that might influence gene expression, we used previously described mRNA expression data (23,720 mRNA transcripts measured in 603 subcutaneous adipose tissue and 745 peripheral blood samples typed on the Illumina 300K Beadchip). For SNPs and RNA transcripts within a 2-Mb window centered on each T2D index SNP, we tested for association between the log of the average expression ratio of two fluorophores and the allele count (genotype data) or expected allele count (imputed data) using linear regression, with adjustment for age, sex and, in the case of blood samples, differential cell count. For each locus and transcript, we performed additional conditional analyses by including in the regression model either the most strongly associated eQTL SNP in the region or the T2D index SNP. These conditional analyses were designed to test whether a T2D signal and any detected cis eQTL association at the T2D index SNP are likely to reflect the same underlying association. P values were adjusted for relatedness of the individuals by simulating genotypes through the corresponding Icelandic family structure.

Tissue expression data. Adult total RNA samples were purchased from Clontech. Adult human islets (n = 2) were available following extraction from pancreases obtained from cadaveric donors in accordance with national transplant regulations and under ethical approval from the Oxfordshire Research Ethics Committee B. Gene expression assays were purchased from Applied Biosystems. Genes at each locus were chosen on the basis of proximity to the index SNP and biological credibility; the probe chosen for each gene was designed to cover the widest range of known transcripts. Samples were treated with DNase1 (Ambion) to ensure 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 manufacturer's protocol. Resulting cDNA for each tissue was diluted 1:100 and 4 μl was used in a 10 μl quantitative RT-PCR reaction with 5.5 μl gene expression mastermix (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, serially diluting (1:50, 1:100, 1:200, 1:400 and 1:800) and running as above. Expression levels were determined with respect to the mean of three endogenous controls (R2M, 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 one and all other tissue expressions are reported as a fraction of this. In addition, flow-sorted pancreatic beta-cells were available from two adult donors (see above); preparations contained >95% insulin-positive cells. cDNA was generated from 150 ng RNA and treated as above. Resulting cDNA was diluted 1:50 and 4 μl was used in a 10 μl RT-PCR reaction. Gene expression was measured as described above. Assays that failed to demonstrate expression in human islets were excluded from this experiment. Expression levels were calculated through normalization to two endogenous controls (HPRT and B2M) and with respect to the average 1:50 standard curve dilution.


68. Stolerman, E.S. et al. TCF7L2 variants are associated with increased proinsulin/insulin ratios but not obesity traits in the Framingham Heart Study. Diabetologia 52, 614–620 (2009).
Corrigendum: Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis


Nat. Genet. 42, 579–589 (2010); published online 27 June 2010; corrected after print 27 August 2010

In the version of this article initially published, there was an error in Table 1. Specifically, for rs5945326, the risk and non-risk alleles were reversed. The correct risk allele at rs5945326 is A, the non-risk allele is G and the risk allele frequency in HapMap CEU is 0.79. These errors have been corrected in the HTML and PDF versions of the article.