nature
genetics

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[2], 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, $P \sim 3.3 \times 10^{-5}$), 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),
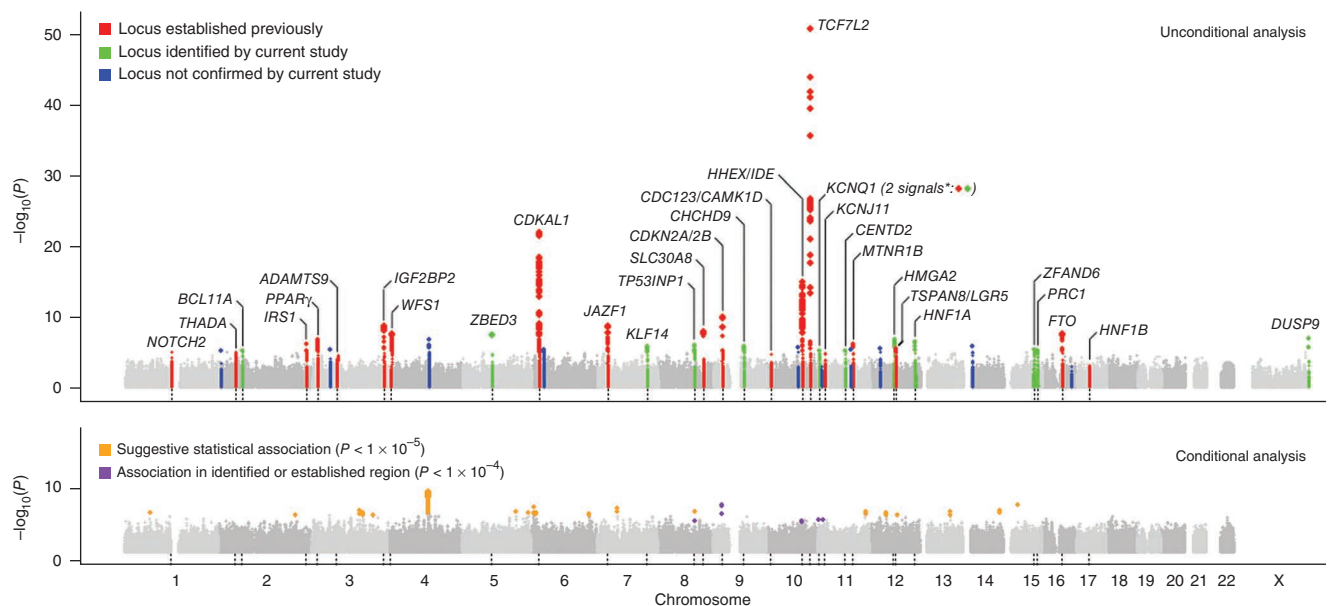
Figure 1 Genome-wide Manhattan plots for the DIAGRAM+ stage 1 meta-analysis. Top panel summarizes the results of the unconditional meta-analysis. Previously established loci are denoted in red and loci identified by the current study are denoted in green. The ten signals in blue are those taken forward but not confirmed in stage 2 analyses. The genes used to name signals have been chosen on the basis of proximity to the index SNP and should not be presumed to indicate causality. The lower panel summarizes the results of equivalent meta-analysis after conditioning on 30 previously established and newly identified autosomal T2D-associated SNPs (denoted by the dotted lines below these loci in the upper panel). Newly discovered conditional signals (outside established loci) are denoted with an orange dot if they show suggestive levels of significance ($P < 10^{-5}$), whereas secondary signals close to already confirmed T2D loci are shown in purple ($P < 10^{-4}$).

13 autosomal loci exceeded the threshold for genome-wide significance (P ranging from 2.8×10^{-8} to 1.4×10^{-22}) 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 5.2×10^{-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*, $P = 2.3 \times 10^{-6}$). This SNP showed strong evidence for replication in 8,535 cases and 12,326 controls (OR (allowing for X-inactivation) 1.32 (95% CI 1.16–1.49), $P = 2.3 \times 10^{-5}$), for a combined association P value of 3.0×10^{-10} (OR 1.27 (95% CI 1.18–1.37)) (**Table 1** and **Fig. 2**).

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[10,14–16] (**Table 1**).

A third signal (rs231362) on 11p15 overlaps both intron 11 of *KCNQ1* and the *KCNQ1OT1* transcript that controls regional imprinting[17] 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 scans[8,9]. 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 samples[9] 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 allele[11]. 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[7,21–23] 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[7]; 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[12,13] and

Table 1 Association results for stage 1 + 2 which exceed a genome-wide threshold (overall P value $< 5 \times 10^{-8}$)

SNP	Chr.	Position B36 (base pair)	Risk allele[b]	Nonrisk allele[b]	Frequency risk allele (Hapmap CEU)	Nearby gene[c]	Stage 1[d] OR (95%CI)	P value	Stage 2[d] OR (95% CI)	P value	Stage 1 + 2[d] OR (95% CI)	P value	Power[e]
							up to 8,130 cases and 38,987 controls		up to 34,412 cases and 59,925 controls[f]		up to 42,542 cases and 98,912 controls		
New T2D susceptibility loci													
rs243021	2	60,438,323	A	G	0.46	BCL11A	1.09 (1.05–1.13)	8.1×10^{-6}	1.08 (1.06–1.10)	6.2×10^{-11}	1.08 (1.06–1.10)	2.9×10^{-15}	0.60
rs4457053	5	76,460,705	G	A	0.26	ZBED3	1.16 (1.10–1.23)	4.2×10^{-8}	1.07 (1.04–1.10)	2.7×10^{-7}	1.08 (1.06–1.11)	2.8×10^{-12}	0.25
rs972283	7	130,117,394	G	A	0.55	KLF14	1.10 (1.06–1.15)	1.8×10^{-6}	1.06 (1.03–1.09)	6.4×10^{-6}	1.07 (1.05–1.10)	2.2×10^{-10}	0.19
rs896854	8	96,029,687	T	C	0.48	TP53INP1	1.10 (1.06–1.15)	1.2×10^{-6}	1.05 (1.03–1.08)	2.2×10^{-5}	1.06 (1.04–1.09)	9.9×10^{-10}	0.08
rs13292136	9	81,141,948	C	T	0.93	CHCHD9	1.20 (1.11–1.29)	1.5×10^{-6}	1.08 (1.04–1.13)	2.4×10^{-4}	1.11 (1.07–1.15)	2.8×10^{-8}	0.02
rs231362	11	2,648,047	G	A	0.52	KCNQ1	1.11 (1.06–1.16)	6.4×10^{-6}	1.07 (1.05–1.09)	3.2×10^{-9}	1.08 (1.06–1.10)	2.8×10^{-13}	0.38
rs1552224	11	72,110,746	A	C	0.88	CENTD2	1.13 (1.07–1.19)	7.0×10^{-6}	1.14 (1.11–1.18)	3.2×10^{-18}	1.14 (1.11–1.17)	1.4×10^{-22}	0.58
rs1531343[a]	12	64,461,161	C	G	0.10	HMGA2	1.20 (1.12–1.29)	1.7×10^{-7}	1.08 (1.04–1.12)	1.1×10^{-4}	1.10 (1.07–1.14)	3.6×10^{-9}	0.07
rs7957197	12	119,945,069	T	A	0.85	HNF1A	1.14 (1.08–1.19)	4.6×10^{-7}	1.05 (1.02–1.09)	4.6×10^{-4}	1.07 (1.05–1.10)	2.4×10^{-8}	0.01
rs11634397	15	78,219,277	G	A	0.60	ZFAND6	1.11 (1.06–1.16)	5.1×10^{-6}	1.05 (1.03–1.08)	1.2×10^{-5}	1.06 (1.04–1.08)	2.4×10^{-9}	0.07
rs8042680	15	89,322,341	A	C	0.22	PRC1	1.10 (1.06–1.15)	8.2×10^{-6}	1.06 (1.03–1.08)	1.6×10^{-6}	1.07 (1.05–1.09)	2.4×10^{-10}	0.09
rs5945326	X	152,553,116	G	A	0.21	DUSP9	1.25 (1.14–1.37)	2.3×10^{-6}	1.32 (1.16–1.49)[f]	2.3×10^{-5}	1.27 (1.18–1.37)	3.0×10^{-10}	0.99[f]
Previously known													
rs7578326	2	226,728,897	A	G	0.64	IRS1	1.12 (1.07–1.17)	8.7×10^{-7}	1.10 (1.08–1.13)	2.2×10^{-15}	1.11 (1.08–1.13)	5.4×10^{-20}	0.83
rs1387153	11	92,313,476	T	C	0.28	MTNR1B	1.12 (1.07–1.17)	1.0×10^{-6}	1.08 (1.05–1.10)	4.4×10^{-10}	1.09 (1.06–1.11)	7.8×10^{-15}	0.46

OR, odds ratio; CI, confidence interval; Chr., chromosome.
[a]Proxy SNP rs2612067 was genotyped in METSIM, FUSION2 and HUNT stage 2 samples. [b]Alleles are indexed to the forward strand of NCBI Build 36. [c]Gene region named for nearest gene (except where there is a very strong positional candidate). [d]All P values reported are two-sided and based on an inverse-variance weighted meta-analysis model (fixed effects). [e]Power to detect autosomal SNPs given the stage 1 sample sizes and stage 2 ORs, the risk AF in Hapmap (CEU), for $\alpha = 1 \times 10^{-5}$. Additive model assumed, 8% population prevalence. [f]Maximum stage 2 sample size for the X chromosome SNP was 8,535 cases and 12,326 controls. Allelic odds ratios calculated assuming X-chromosome inactivation in females. Power to detect SNP given stage 1 sex distribution and X-chromosome inactivation in females, otherwise as for autosomes.

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 11p15 (which confers substantial risk when paternally inherited but is protective when maternally transmitted[11]), 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[26].

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 CNVs 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 when based on the lead common variants alone. To explore the potential for independent secondary alleles, we repeated the stage 1 meta-analysis after simultaneously conditioning on 30 known and newly discovered autosomal loci (**Supplementary Note**). Using a cutoff

of $P < 1 \times 10^{-4}$ (to reflect approximate adjustment for the number of independent SNPs in a ~2 Mb interval), we found preliminary evidence for secondary signals at five loci (TP53INP1, CDKN2A, HHEX-IDE and KCNJ11, in addition to that at KCNQ1; **Fig. 1**, **Supplementary Fig. 1** and **Supplementary Table 4**). At CDKN2A, the secondary signal is consistent with evidence that haplotype-based analyses generate considerably stronger evidence for association than either signal alone[3,27]. Further fine-mapping efforts will be required to confirm the secondary signals at TP53INP1, HHEX-IDE and KCNJ11.

The conditional analysis also provided preliminary ($P < 10^{-5}$, our stage 1 threshold) evidence for 19 signals outside known loci (**Fig. 1** and **Supplementary Table 6**). The most notable signal (rs1481279, conditional $P = 8.4 \times 10^{-9}$) maps near NHEDC1 and corresponds to one of the signals of interest following stage 1 (rs7674212, $P = 1.7 \times 10^{-7}$ unconditioned). Failure to replicate that signal ($P = 0.3$ in 21,889 stage 2 cases and 39,568 controls) suggests this was a false positive (**Supplementary Note**). Several regions showed substantial incremental evidence for association in the conditional analysis as compared to unadjusted analyses and represent potential targets for large-scale replication and gene-gene interaction analyses. Indeed, one of these regions (at rs11708067 near ADCY5, unadjusted $P = 1.7 \times 10^{-4}$, conditional $P = 2.2 \times 10^{-6}$) has recently been shown, following initial identification through GWA analysis of continuous glucose measures, to have genome-wide significant associations with T2D in large-scale case-control analyses that involved several DIAGRAM+ samples[12,13].
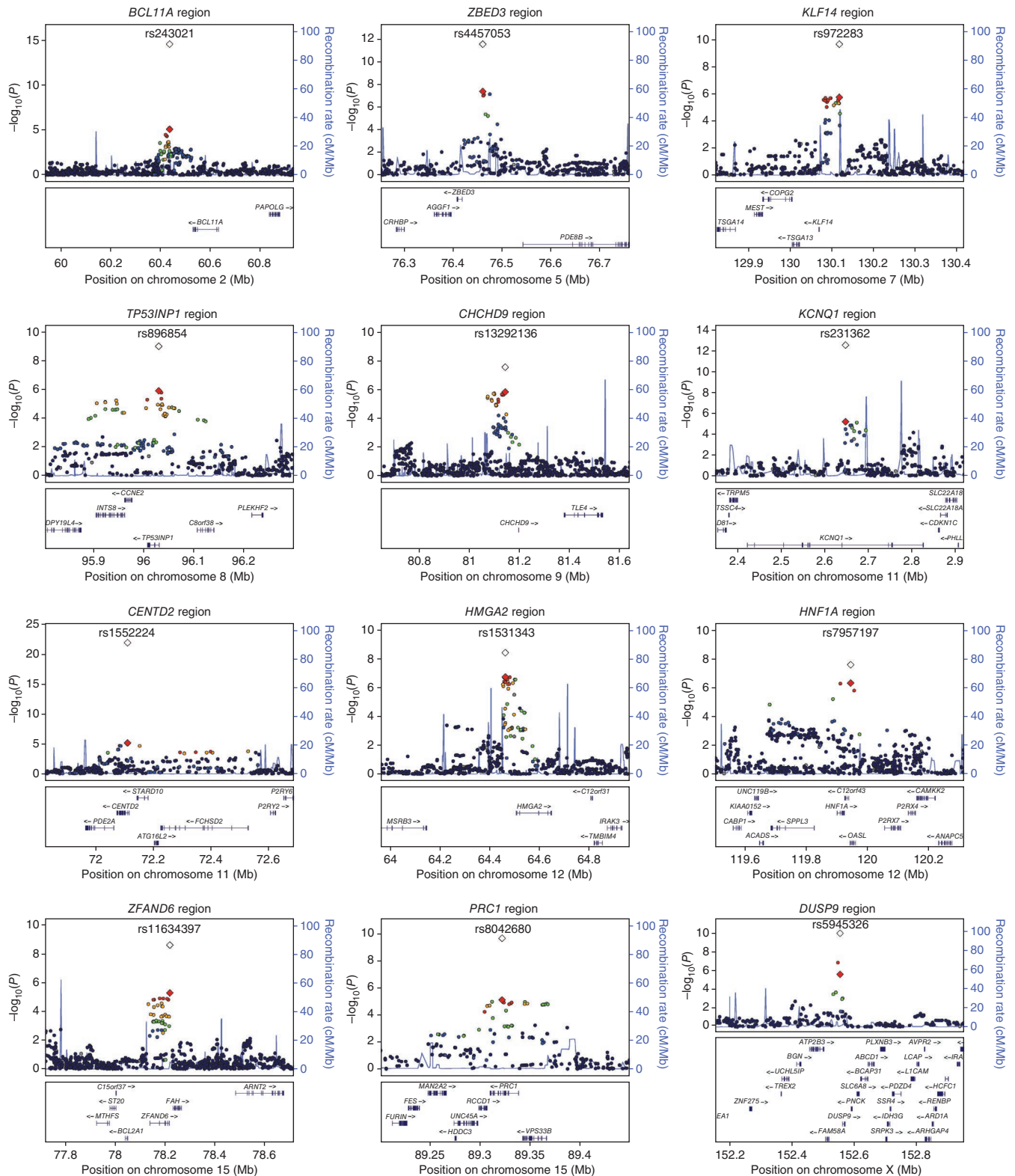
Figure 2 Regional plots of the 12 newly discovered T2D loci. Genotyped and imputed SNPs passing quality control measures across all stage 1 studies are plotted with their meta-analysis *P* values (as −log$_{10}$ values) as a function of genomic position (NCBI Build 36). In each panel, the index association 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 kgm^{-2}) and 4,048 nonobese (BMI ≤ 30 kgm^{-2}) 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 $BCL11A$, $ZBED3$, $KLF14$, $CHCHD9$, $HMGA2$, $HNF1A$ 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[28] 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[29]. At $HNF1A$, previous studies have reported a cluster of associations, with phenotypes including low-density lipoptrotein (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 ~0.1, D' ~1), they are likely to be independent; the T2D association at the lead SNP for lipids (rs2650000) 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 rs2650000 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 ~1.6×10^{-3} for non-T2D signals within 500 kb of T2D loci, P ~7.0×10^{-5} ($n = 8$) 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 disease-specific effects mediated through the same genes and pathways.

Understanding the biology of T2D-susceptibility loci

This analysis takes the number of independent loci showing genome-wide 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 ANthropometric Traits (GIANT) consortium[43], 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 $TCF7L2$, $CENTD2$ 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 measures 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 $CENTD2$) 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 $CENTD2$ 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

Figure 3 Plots of fasting blood glucose, insulin and derived indices for the established and new T2D loci. (**a,b**) Plots of fasting glucose (*x* axis) and fasting insulin (*y* axis). (**c,d**) Plots of HOMA-B (an index of beta cell function; *x* axis) and HOMA-IR (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.
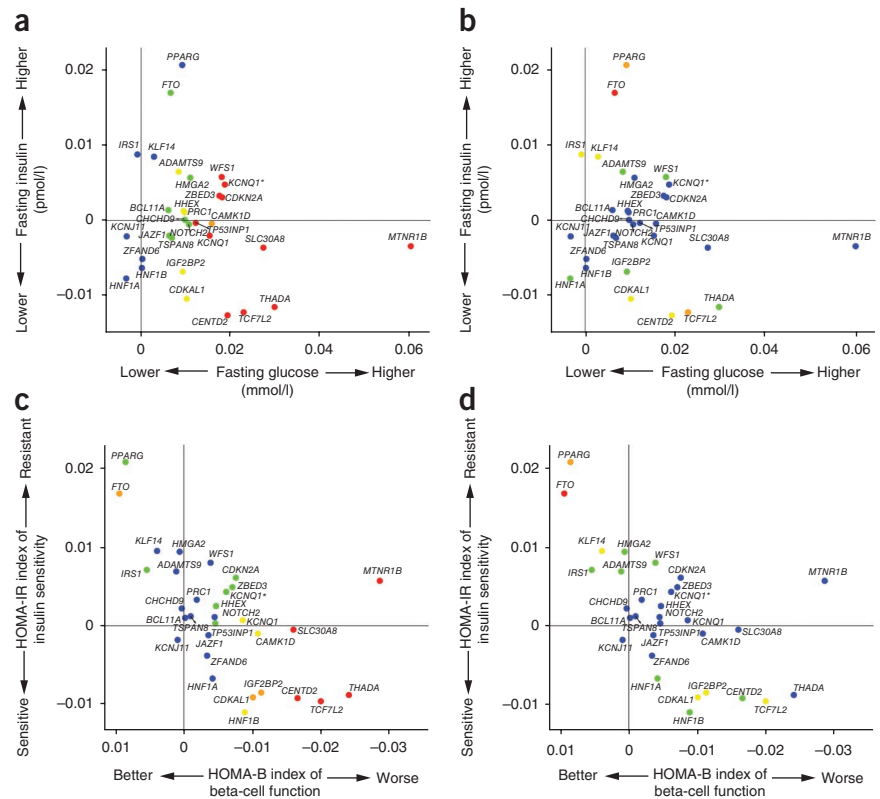
T2D risk alleles more often act through beta-cell dysfunction[12,44], but they provide further examples of T2D risk variants that exert their primary effects on insulin action.

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[10]), *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 $P \sim 0.004$) 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 *Hmga/Hmg1c* knockout mice are deficient in adipocyte differentiation[45], 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[56]. 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[57–59].

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[1] (**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[48]—indicates, either that T2D pathogenesis is characterized by

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

SNP	Chr.	Position B36 (bp)	Nearby gene[a]	Risk allele[b]	Gene (transcript)	Tissue	Effect (s.e.m.)[c]	P value	P_{adj}[d]	SNP (r^2)[f]	P value	P_{adj}[g]
										SNP with strongest correlation with trait[e]		
Novel loci reported in this study												
rs4457053	5	76,460,705	*ZBED3*	G	*PDE8B* (NM_003719)	Adipose	0.302 (0.070)	2.8×10^{-5}	0.80	rs6864250 (0.18)	3.1×10^{-17}	5.8×10^{-13}
					ZBED3 (NM_032367)	Adipose	0.429 (0.068)	1.0×10^{-9}	0.011	rs4704389 (0.20)	3.9×10^{-16}	6.0×10^{-9}
rs972283	7	130,117,394	*KLF14*	G	*KLF14* (NM_138693)	Adipose	−0.387 (0.058)	8.1×10^{-11}	0.058	rs738134 (0.30)	2.2×10^{-12}	0.0014
rs896854	8	96,029,687	*TP53INP1*	T	*CCNE2* (NM_057749)	Blood	−0.225 (0.053)	3.8×10^{-5}	0.78	rs4735339 (0.61)	5.8×10^{-7}	0.0051
rs1552224	11	72,110,746	*CENTD2*	A	*STARD10* (NM_006645)	Blood	0.337 (0.066)	8.6×10^{-7}	0.026	rs519790 (0.04)	2.7×10^{-24}	1.6×10^{-19}
rs7957197	12	119,945,069	*HNF1A*	T	*ACADS* (NM_000017)	Adipose	−0.248 (0.067)	3.7×10^{-4}	0.29	rs9204 (0.02)	1.3×10^{-53}	5.9×10^{-50}
					PSMD9 (NM_002813)	Blood	0.240 (0.065)	3.9×10^{-4}	0.0088	rs3741593 (0.00)	8.3×10^{-8}	1.7×10^{-6}
					OASL (NM_003733)	Adipose	0.318 (0.068)	6.4×10^{-6}	0.13	rs2259883 (0.19)	1.1×10^{-7}	0.0018
					OASL (NM_003733)	Blood	0.319 (0.064)	1.3×10^{-6}	0.37	rs4556628 (0.21)	4.4×10^{-22}	1.4×10^{-16}
					COQ5 (NM_032314)	Blood	0.248 (0.065)	2.1×10^{-4}	0.92	rs10774561 (0.02)	8.7×10^{-39}	4.9×10^{-35}
					UNC119B (NM_032661)	Blood	−0.254 (0.064)	1.4×10^{-4}	0.048	rs11065202 (0.09)	7.8×10^{-12}	2.3×10^{-9}
					CAMKK2 (NM_172215)	Adipose	−0.497 (0.068)	1.2×10^{-12}	0.18	rs11065504 (0.08)	2.7×10^{-117}	3.8×10^{-98}
					CAMKK2 (NM_172215)	Blood	−0.360 (0.063)	3.4×10^{-8}	0.68	rs11065504 (0.08)	7.0×10^{-105}	5.7×10^{-94}
					P2RX4 (NM_175568)	Blood	0.312 (0.065)	3.4×10^{-6}	2.0×10^{-6}	rs25644 (0.03)	3.4×10^{-17}	1.9×10^{-17}
rs8042680	15	89,322,341	*PRC1*	A	*VPS33B* (NM_018668)	Blood	−0.371 (0.057)	2.9×10^{-10}	0.50	rs12595616 (0.57)	2.3×10^{-21}	4.5×10^{-12}
Previously reported loci												
rs7578326	2	226,728,897	*IRS1*	A	*IRS1* (Contig50189_RC)	Adipose	−0.251 (0.059)	3.7×10^{-5}	0.89	rs2943653 (0.93)	3.4×10^{-5}	0.69
					IRS1 (NM_005544)	Adipose	−0.331 (0.059)	5.7×10^{-8}	0.58	rs2176040 (0.74)	7.8×10^{-10}	0.0042
rs13081389	3	12,264,800	*PPARG*	A	*IQSEC1* (NM_014869)	Adipose	−0.630 (0.131)	2.9×10^{-6}	1.4×10^{-4}	rs9211 (0.01)	1.1×10^{-96}	7.4×10^{-94}
rs6795735	3	64,680,405	*ADAMTS9*	C	BC040632 (AK022320)	Adipose	−0.229 (0.056)	7.6×10^{-5}	0.28	rs4521216 (0.02)	3.0×10^{-13}	8.7×10^{-10}
					ADAMTS9 (NM_182920)	Adipose	−0.274 (0.055)	1.5×10^{-6}	0.036	rs7372321 (0.11)	1.1×10^{-9}	2.3×10^{-5}
rs849134	7	28,162,747	*JAZF1*	A	*JAZF1* (NM_175061)	Adipose	−0.346 (0.055)	1.4×10^{-9}	0.70	rs1635852 (1.00)	5.2×10^{-10}	0.17
rs12779790	10	12,368,016	*CAMK1D*	G	*CAMK1D* (AL137430)	Blood	0.668 (0.062)	1.6×10^{-25}	0.052	rs11257655 (0.83)	1.2×10^{-25}	0.04
					CDC123 (NM_006023)	Blood	0.498 (0.064)	3.9×10^{-14}	0.012	rs11257600 (0.57)	1.4×10^{-16}	4.9×10^{-5}
					CAMK1D (NM_020397)	Blood	0.384 (0.064)	7.6×10^{-9}	0.44	rs11257655 (0.83)	3.3×10^{-9}	0.13
					CAMK1D (NM_153498)	Blood	0.643 (0.062)	2.6×10^{-23}	n/a	rs12779790 (1.00)	2.6×10^{-23}	n/a
rs5015480	10	94,455,539	*HHEX/IDE*	C	*KIF11* (NM_004523)	Adipose	0.202 (0.059)	9.4×10^{-4}	0.54	rs10882095 (0.53)	1.6×10^{-4}	0.059
					KIF11 (NM_004523)	Blood	0.185 (0.054)	9.7×10^{-4}	0.62	rs6583826 (0.43)	7.9×10^{-6}	0.0025
					MARCH5 (NM_017824)	Adipose	−0.262 (0.060)	2.5×10^{-5}	0.31	rs10748579 (0.04)	7.4×10^{-36}	3.2×10^{-31}

Chr., chromosome.
[a]Gene region named for nearest gene (except where there is a very strong positional candidate). [b]Alleles are indexed to the forward strand of NCBI Build 36. [c]Effect of T2D risk allele measured in units of standard deviation of the standardized distribution of the mean-log expression ratio. [d]P value for the expression trait association at the lead T2D SNP after conditioning on the local SNP that shows the most significant correlation with the expression trait. [e]The SNP in the 2-Mb window that shows the strongest correlation with the expression trait. [f]Correlation r^2 between the T2D-associated SNP and the SNP that shows the strongest correlation with the expression trait (in CEU HapMap). [g]P value for the expression trait association after conditioning on the strongest T2D associated SNP in the region. A more detailed explanation of the table contents and analyses is provided in the **Supplementary Note**.

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[12,13] and parent-of-origin effects[11], 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[11]. 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[60]) 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[61,62]. 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*, *TP53INP1* 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 resequencing 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[63]—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/.

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

AUTHOR CONTRIBUTIONS

Manuscript writing: B.F.V., L.J.S., V.S., A.P.M., C.D., E.Z., T.F., T.M.F., R. Sladek, U.T., D.A., M.B., M.I.M.

Clinical samples: A.P.M., H.G., C.L., L.Q., M.v.H., P. Navarro, K.A., B. Balkau, R. Benediktsson, R. Blagieva, L.L.B., K.B.B., B. Bravenboer, N.P.B., G. Charpentier, M.C., A.S.F.D., C.S. Fox, C. Gieger, N.G., S.G., S.H., C. Herder, B.I., T.J., P.K., J.K., T.L., A.L., V.L., M.M., T.M., K.M., P. Nilsson, K.R.O., C. Platou, W.R., M.R., M.J.S., B.M.S., G.S., T.S., K. Strassburger, Q.S., B.T., J. Tichet, T.T., R.M.v.D., T.W.v.H., T.v.H., J.V.v.V.-O., C.W., R.N.B., F.S.C., U.G., T.H., G.A.H., D.J.H., K.H., M.Laakso, K.L.M., A.D.M., C.N.A.P., P.P.P., I.R., E.S., J. Tuomilehto, M.W., N.J.W., B.O.B., H.C., A.T.H., F.B.H., J.B.M., J.S.P., O.P., T.M.F., L.G., R. Sladek, U.T., H.-E.W., J.F.W., T.I., P.F., C.M.v.D., K. Stefansson, D.A., M.B., M.I.M.

Stage 1 genotyping and analysis: B.F.V., L.J.S., V.S., A.P.M., C.D., E.Z., C. Huth, Y.S.A., G.T., T.F., H.G., N.A., C.J.W., C.L., A.V.S., M.v.H., P. Navarro, K.A., R. Benediktsson, A.J.B., L.L.B., K.B.B., S.B., N.P.B., G. Charpentier, P.S.C., M.C., G. Crawford, M.R.E., M.G., N.G., C.J.G., C. Guiducci, C. Herder, B.I., A.U.J., N.K., T.L., C.M.L., V.L., M.M., T.M., M.A.M., N.N., P. Nilsson, F.P., G.R., R. Saxena, T.S., K. Strassburger, H.M.S., A.J.S., T.T., R.M.v.D., G.B.W., J.W., R.N.B., S.C., F.S.C., U.G., K.L.M., I.R., E.S., J. Tuomilehto, A.U., N.J.W., H.C., F.B.H., T.M.F., L.G., R. Sladek, U.T., H.-E.W., J.F.W., T.I., P.F., C.M.v.D., K. Stefansson, D.A., M.B., M.I.M.

Stage 2 genotyping and analysis: B.F.V., L.J.S., V.S., A.P.M., C.D., C. Huth, Y.S.A., G.T., H.G., N.A., C.J.W., C.L., J.D., L.Q., M.v.H., P. Navarro, K.A., A.J.B., E.B., L.L.B., K.B.B., S.B., N.P.B., P.S.C., M.C., D.J.C., G. Crawford, A.S.F.D., M.R.E., C.S. Franklin, M.G., C. Gieger, N.G., S.G., C.J.G., C. Guiducci, N.H., C. Herder, B.I., A.U.J., T.J., W.H.L.K., N.K., A.K., P.K., J.K., T.L., M. Li, C.M.L., V.L., T.M., K.M., M.A.M., N.N., P. Nilsson, F.P., A.-K.P., C. Proença, I.P., W.R., N.W.R., N.R.R., G.R., M.R., M.J.S., P.S., T.S., K. Strassburger, H.M.S., Q.S., A.J.S., T.T., R.M.v.D., T.W.v.H., J.V.v.V.-O., G.B.W., M.N.W., C.W., J.W., R.N.B., S.C., F.S.C., U.G., T.H., D.J.H., K.H., M. Laakso, K.L.M., A.D.M., C.N.A.P., P.P.P., I.R., E.S., J. Tuomilehto, A.U., N.J.W., H.C., M.J.D., F.B.H., J.S.P., O.P., I.B., J.C.F., T.M.F., L.G., R. Sladek, H.-E.W., U.T., J.F.W., T.I., P.F., C.M.v.D., D.A., M.B., M.I.M.

Analysis group: B.F.V., L.J.S., V.S., A.P.M., R.P.W., C.D., E.Z., C. Huth, Y.S.A., G.T., T.F., H.G., N.A., C.J.W., J.D., M.v.H., M.G., C. Gieger, A.U.J., N.K., A.K., J.R.B.P., A.-K.P., N.W.R., N.R.R., R. Saxena, M.J.D., P.F., M.B., M.I.M.

Biological analyses: V.S., G.T., L.J.M., S.A.M., J.D., K.S.E., A.L.E., P.R.V.J., V.L., I.P., A.L.G., J.B.M., U.T., K. Stefansson, M.I.M.

Informatics analyses: B.F.V., V.S., G.W., S.R., O.M.H., A.V.S., T.G., W.A.H., L.D.S.

DIAGRAM consortium management: B.F.V., L.J.S., V.S., A.P.M., C.D., E.Z., R.N.B., S.C., F.S.C., A.H., K.L.M., E.S., J. Tuomilehto, R.M.W., G.R.A., H.C., M.J.D., A.T.H., T.M.F., L.G., R. Sladek, U.T., H.-E.W., J.F.W., T.I., P.F., C.M.v.D., K. Stefansson, D.A., M.B., M.I.M.

COMPETING FINANCIAL INTERESTS
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/.

1. Stumvoll, M., Goldstein, B.J. & van Haeften, T.W. Type 2 diabetes: principles of pathogenesis and therapy. *Lancet* **365**, 1333–1346 (2005).

2. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
3. Zeggini, E. *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336–1341 (2007).
4. Saxena, R. *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316**, 1331–1336 (2007).
5. Scott, L.J. *et al.* A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. *Science* **316**, 1341–1345 (2007).
6. Steinthorsdottir, V. *et al.* A variant in *CDKAL1* influences insulin response and risk of type 2 diabetes. *Nat. Genet.* **39**, 770–775 (2007).
7. Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nat. Genet.* **40**, 638–645 (2008).
8. Yasuda, K. *et al.* Variants in *KCNQ1* are associated with susceptibility to type 2 diabetes mellitus. *Nat. Genet.* **40**, 1092–1097 (2008).
9. Unoki, H. *et al.* SNPs in *KCNQ1* are associated with susceptibility to type 2 diabetes in East Asian and European populations. *Nat. Genet.* **40**, 1098–1102 (2008).
10. Rung, J. *et al.* Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* **41**, 1110–1115 (2009).
11. Kong, A. *et al.* Parental origin of sequence variants associated with complex diseases. *Nature* **462**, 868–874 (2009).
12. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
13. Saxena, R. *et al.* Genetic variation in gastric inhibitory polypeptide receptor (GIPR) impacts the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
14. Prokopenko, I. *et al.* Variants in *MTNR1B* influence fasting glucose levels. *Nat. Genet.* **41**, 77–81 (2009).
15. Bouatia-Naji, N. *et al.* A variant near *MTNR1B* is associated with increased fasting plasma glucose levels and type 2 diabetes risk. *Nat. Genet.* **41**, 89–94 (2009).
16. Lyssenko, V. *et al.* Common variant in *MTNR1B* associated with increased risk of type 2 diabetes and impaired early insulin secretion. *Nat. Genet.* **41**, 82–88 (2009).
17. Fitzpatrick, G.V., Soloway, P.D. & Higgins, M.J. Regional loss of imprinting and growth deficiency in mice with a targeted deletion of KvDMR1. *Nat. Genet.* **32**, 426–431 (2002).
18. Kassem, S.A. *et al.* p57(KIP2) expression in normal islet cells and in hyperinsulinism of infancy. *Diabetes* **50**, 2763–2769 (2001).
19. Newton-Cheh, C. *et al.* Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* **41**, 399–406 (2009).
20. Pfeufer, A. *et al.* Common variants at ten loci modulate the QT interval duration in the QTSCD Study. *Nat. Genet.* **41**, 407–414 (2009).
21. Winckler, W. *et al.* Association of common variation in the HNF1α gene region with risk of type 2 diabetes. *Diabetes* **54**, 2336–2342 (2005).
22. Weedon, M.N. *et al.* A large-scale association analysis of common variation of the HNF1α gene with type 2 diabetes in the U.K. Caucasian population. *Diabetes* **54**, 2487–2491 (2005).
23. Bonnycastle, L.L. *et al.* Common variants in maturity-onset diabetes of the young genes contribute to risk of type 2 diabetes in Finns. *Diabetes* **55**, 2534–2540 (2006).
24. Hegele, R.A., Cao, H., Harris, S.B., Hanley, A.J. & Zinman, B. The hepatic nuclear factor-1alpha G319S variant is associated with early-onset type 2 diabetes in Canadian Oji-Cree. *J. Clin. Endocrinol. Metab.* **84**, 1077–1082 (1999).
25. Kobberling, J. & Tillil, H. Empirical risk figures for first degree relatives of non-insulin-dependent diabetics. in *The Genetics of Diabetes Mellitus* (eds. Kobberling, J. & Tattersall, R.) 201–209 (Academic Press, New York, 1982).
26. Manolio, T.A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).
27. Browning, B.L. & Browning, S.R. Haplotypic analysis of Wellcome Trust Case Control Consortium data. *Hum. Genet.* **123**, 273–280 (2008).
28. Weedon, M.N. *et al.* A common variant of *HMGA2* is associated with adult and childhood height in the general population. *Nat. Genet.* **39**, 1245–1250 (2007).
29. Stacey, S.N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat. Genet.* **41**, 909–914 (2009).
30. Kathiresan, S. *et al.* Common variants at 30 loci contribute to polygenic dyslipidemia. *Nat. Genet.* **41**, 56–65 (2009).
31. Reiner, A.P. *et al.* Polymorphisms of the *HNF1A* gene encoding hepatocyte nuclear factor-1 alpha are associated with C-reactive protein. *Am. J. Hum. Genet.* **82**, 1193–1201 (2008).
32. Ridker, P.M. *et al.* Loci related to metabolic-syndrome pathways including *LEPR*, *HNF1A*, *IL6R*, and *GCKR* associate with plasma C-reactive protein: the Women's Genome Health Study. *Am. J. Hum. Genet.* **82**, 1185–1192 (2008).
33. Elliott, P. *et al.* Genetic loci associated with C-reactive protein levels and risk of coronary heart disease. *J. Am. Med. Assoc.* **302**, 37–48 (2009).
34. Helgadottir, A. *et al.* The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat. Genet.* **40**, 217–224 (2008).
35. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
36. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
37. Thomas, G. *et al.* Multiple loci identified in a genome-wide association study of prostate cancer. *Nat. Genet.* **40**, 310–315 (2008).

38. Barrett, J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.* **40**, 955–962 (2008).
39. Johansson, A. *et al.* Common variants in the *JAZF1* gene associated with height identified by linkage and genome-wide association analysis. *Hum. Mol. Genet.* **18**, 373–380 (2009).
40. Quaranta, M. *et al.* Differential contribution of *CDKAL1* variants to psoriasis, Crohn's disease and type II diabetes. *Genes Immun.* **10**, 654–658 (2009).
41. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367 (2009).
42. Nicolson, T.J. *et al.* Insulin storage and glucose homeostasis in mice null for the granule zinc transporter ZnT8 and studies of the type 2 diabetes-associated variants. *Diabetes* **58**, 2070–2083 (2009).
43. Willer, C.J. *et al.* Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).
44. Florez, J.C. Newly identified loci highlight beta cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* **51**, 1100–1110 (2008).
45. Anand, A. & Chada, K. *In vivo* modulation of *Hmgic* reduces obesity. *Nat. Genet.* **24**, 377–380 (2000).
46. Boesgaard, T.W. *et al.* Variant near *ADAMTS9* known to associate with type 2 diabetes is related to insulin resistance in offspring of type 2 diabetes patients–EUGENE2 study. *PLoS One* **4**, e7236 (2009).
47. Emilsson, V. *et al.* Genetics of gene expression and its effect on disease. *Nature* **452**, 423–428 (2008).
48. Wang, K. *et al.* Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. *Am. J. Hum. Genet.* **84**, 399–405 (2009).
49. Raychaudhuri, S. *et al.* Identifying relationships among genomic disease regions: predicting genes at pathogenic SNP associations and rare deletions. *PLoS Genet.* **5**, e1000534 (2009).
50. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–D288 (2005).
51. Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* **37**, D619–D622 (2009).
52. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
53. Brants, J.R. *et al.* Differential regulation of the insulin-like growth factor II mRNA-binding protein genes by architectural transcription factor HMGA2. *FEBS Lett.* **569**, 277–283 (2004).
54. Cleynen, I. *et al.* HMGA2 regulates transcription of the *Imp2* gene via an intronic regulatory element in cooperation with nuclear factor-kappaB. *Mol. Cancer Res.* **5**, 363–372 (2007).
55. Groenewoud, M.J. *et al.* Variants of *CDKAL1* and *IGF2BP2* affect first-phase insulin secretion during hyperglycaemic clamps. *Diabetologia* **51**, 1659–1663 (2008).
56. Brantjes, H., Roose, J., van De Wetering, M. & Clevers, H. All Tcf HMG box transcription factors interact with Groucho-related co-repressors. *Nucleic Acids Res.* **29**, 1410–1419 (2001).
57. Logeat, F. *et al.* The Notch1 receptor is cleaved constitutively by a furin-like convertase. *Proc. Natl. Acad. Sci. USA* **95**, 8108–8112 (1998).
58. Gordon, W.R. *et al.* Effects of S1 cleavage on the structure, surface export, and signaling activity of human Notch1 and Notch2. *PLoS One* **4**, e6613 (2009).
59. Somerville, R.P. *et al.* Characterization of ADAMTS-9 and ADAMTS-20 as a distinct ADAMTS subfamily related to *Caenorhabditis elegans* GON-1. *J. Biol. Chem.* **278**, 9503–9513 (2003).
60. Parker-Katiraee, L. *et al.* Identification of the imprinted KLF14 transcription factor undergoing human-specific accelerated evolution. *PLoS Genet.* **3**, e65 (2007).
61. Emanuelli, B., Eberlé, D., Suzuki, R. & Kahn, C.R. Overexpression of the dual-specificity phosphatase MKP-4/DUSP-9 protects against stress-induced insulin resistance. *Proc. Natl. Acad. Sci. USA* **105**, 3545–3550 (2008).
62. Xu, H. *et al.* Dual specificity mitogen-activated protein (MAP) kinase phosphatase-4 plays a potential role in insulin resistance. *J. Biol. Chem.* **278**, 30187–30192 (2003).
63. Dimas, A.S. *et al.* Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* **325**, 1246–1250 (2009).

Benjamin F Voight[1–3,100], Laura J Scott[4,100], Valgerdur Steinthorsdottir[5,100], Andrew P Morris[6,100], Christian Dina[7,8,100], Ryan P Welch[9], Eleftheria Zeggini[6,10], Cornelia Huth[11,12], Yurii S Aulchenko[13], Gudmar Thorleifsson[5], Laura J McCulloch[14], Teresa Ferreira[6], Harald Grallert[11,12], Najaf Amin[13], Guanming Wu[15], Cristen J Willer[4], Soumya Raychaudhuri[1,2,16], Steve A McCarroll[1,17], Claudia Langenberg[18], Oliver M Hofmann[19], Josée Dupuis[20,21], Lu Qi[22–24], Ayellet V Segrè[1,2,17], Mandy van Hoek[25], Pau Navarro[26], Kristin Ardlie[1], Beverley Balkau[27,28], Rafn Benediktsson[29,30], Amanda J Bennett[14], Roza Blagieva[31], Eric Boerwinkle[32], Lori L Bonnycastle[33], Kristina Bengtsson Boström[34], Bert Bravenboer[35], Suzannah Bumpstead[10], Noisël P Burtt[1], Guillaume Charpentier[36], Peter S Chines[33], Marilyn Cornelis[24], David J Couper[37], Gabe Crawford[1], Alex S F Doney[38,39], Katherine S Elliott[6], Amanda L Elliott[1,17,40], Michael R Erdos[33], Caroline S Fox[21,41], Christopher S Franklin[42], Martha Ganser[4], Christian Gieger[11], Niels Grarup[43], Todd Green[1,2], Simon Griffin[18], Christopher J Groves[14], Candace Guiducci[1], Samy Hadjadj[44], Neelam Hassanali[14], Christian Herder[45], Bo Isomaa[46,47], Anne U Jackson[4], Paul R V Johnson[48], Torben Jørgensen[49,50], Wen H L Kao[51,52], Norman Klopp[11], Augustine Kong[5], Peter Kraft[22,23], Johanna Kuusisto[53], Torsten Lauritzen[54], Man Li[51], Aloysius Lieverse[55], Cecilia M Lindgren[6], Valeriya Lyssenko[56], Michel Marre[57,58], Thomas Meitinger[59,60], Kristian Midthjell[61], Mario A Morken[33], Narisu Narisu[33], Peter Nilsson[56], Katharine R Owen[14], Felicity Payne[10], John R B Perry[62,63], Ann-Kristin Petersen[11], Carl Platou[61], Christine Proença[7], Inga Prokopenko[6,14], Wolfgang Rathmann[64], N William Rayner[6,14], Neil R Robertson[6,14], Ghislain Rocheleau[65–67], Michael Roden[45,68], Michael J Sampson[69], Richa Saxena[1,2,40], Beverley M Shields[62,63], Peter Shrader[3,70], Gunnar Sigurdsson[29,30], Thomas Sparsø[43], Klaus Strassburger[64], Heather M Stringham[4], Qi Sun[22,23], Amy J Swift[33], Barbara Thorand[11], Jean Tichet[71], Tiinamaija Tuomi[46,72], Rob M van Dam[24], Timon W van Haeften[73], Thijs van Herpt[25,55], Jana V van Vliet-Ostaptchouk[74], G Bragi Walters[5], Michael N Weedon[62,63], Cisca Wijmenga[75], Jacqueline Witteman[13], The MAGIC investigators[99], The GIANT Consortium[99], Richard N Bergman[76], Stephane Cauchi[7], Francis S Collins[77], Anna L Gloyn[14], Ulf Gyllensten[78], Torben Hansen[43,79], Winston A Hide[19], Graham A Hitman[80], Albert Hofman[13], David J Hunter[22,23], Kristian Hveem[61,81], Markku Laakso[53], Karen L Mohlke[82], Andrew D Morris[38,39], Colin N A Palmer[38,39], Peter P Pramstaller[83], Igor Rudan[42,84,85], Eric Sijbrands[25], Lincoln D Stein[15], Jaakko Tuomilehto[86–88], Andre Uitterlinden[25], Mark Walker[89], Nicholas J Wareham[18], Richard M Watanabe[76,90], Gonçalo R Abecasis[4], Bernhard O Boehm[31], Harry Campbell[42], Mark J Daly[1,2], Andrew T Hattersley[62,63], Frank B Hu[22–24], James B Meigs[3,70], James S Pankow[91], Oluf Pedersen[43,92,93], H-Erich Wichmann[11,12,94], Inês Barroso[10], Jose C Florez[1–3,95], Timothy M Frayling[62,63], Leif Groop[56,72], Rob Sladek[65–67], Unnur Thorsteinsdottir[5,96], James F Wilson[42], Thomas Illig[11], Philippe Froguel[17,97], Cornelia M van Duijn[13], Kari Stefansson[5,96], David Altshuler[1–3,17,40,95], Michael Boehnke[4] & Mark I McCarthy[6,14,98]

[1]Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA. [2]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [3]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [4]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [5]deCODE Genetics, Reykjavik, Iceland. [6]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [7]CNRS-UMR-8090, Institute of Biology and Lille 2 University, Pasteur Institute, Lille, France. [8]INSERM UMR915 CNRS ERL3147, Nantes, France. [9]Bioinformatics Program, University of Michigan, Ann Arbor, Michigan, USA. [10]Wellcome Trust Sanger Institute, Hinxton, UK. [11]Institute of Epidemiology, Helmholtz Zentrum Muenchen, Neuherberg, Germany. [12]Institute of Medical Informatics, Biometry and Epidemiology, Ludwig-Maximilians-Universität, Munich, Germany. [13]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands. [14]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [15]Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [16]Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [17]Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA. [18]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [19]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [20]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [21]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [22]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [23]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [24]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [25]Department of Internal Medicine, Erasmus University Medical Centre, Rotterdam, The Netherlands. [26]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh, UK. [27]INSERM, CESP Centre for Research in Epidemiology and Population Health, U1018, Epidemiology of Diabetes, Obesity and Chronic Kidney Disease over the Lifecourse, Villejuif, France. [28]University Paris-Sud 11, UMRS 1018, Villejuif, France. [29]Landspitali University Hospital, Reykjavik, Iceland. [30]Icelandic Heart Association, Kopavogur, Iceland. [31]Division of Endocrinology, Diabetes and Metabolism, Ulm University, Ulm, Germany. [32]The Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA. [33]National Human Genome Research Institute, National Institute of Health, Bethesda, Maryland, USA. [34]Research and Development Centre, Skaraborg Primary Care, Skövde, Sweden. [35]Department of Internal Medicine, Catharina Hospital, Eindhoven, The Netherlands. [36]Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes, France. [37]Department of Biostatistics and Collaborative Studies Coordinating Center, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [38]Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK. [39]Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK. [40]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [41]Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [42]Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK. [43]Hagedorn Research Institute, Gentofte, Denmark. [44]Centre Hospitalier Universitaire de Poitiers, Endocrinologie Diabetologie, CIC INSERM 0801, INSERM U927, Université de Poitiers, UFR, Médecine Pharmacie, Poitiers Cedex, France. [45]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [46]Folkhälsan Research Center, Helsinki, Finland. [47]Malmska Municipal Health Center and Hospital, Jakobstad, Finland. [48]Diabetes Research and Wellness Foundation Human Islet Isolation Facility and Oxford Islet Transplant Programme, University of Oxford, Oxford, UK. [49]Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup, Denmark. [50]Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark. [51]Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland, USA. [52]Department of Medicine and Welch Center for Prevention, Epidemiology and Clinical Research, Johns Hopkins University, Baltimore, Maryland, USA. [53]Department of Medicine, University of Kuopio and Kuopio University Hospital, Kuopio, Finland. [54]Department of General Medical Practice, University of Aarhus, Aarhus, Denmark. [55]Department of Internal Medicine, Maxima Medical Center, Eindhoven, The Netherlands. [56]Department of Clinical Sciences, Diabetes and Endocrinology Research Unit, University Hospital Malmö, Lund University, Malmö, Sweden. [57]Department of Endocrinology, Diabetology and Nutrition, Bichat-Claude Bernard University Hospital, Assistance Publique des Hôpitaux de Paris, Paris, France. [58]INSERM U695, Université Paris 7, Paris, France. [59]Institute of Human Genetics, Helmholtz Zentrum Muenchen, Neuherberg, Germany. [60]Institute of Human Genetics, Klinikum rechts der Isar, Technische Universität München, München, Germany. [61]Nord-Trøndelag Health Study (HUNT) Research Center, Department of Community Medicine and General Practice, Norwegian University of Science and Technology, Trondheim, Norway. [62]Genetics of Complex Traits, Institute of Biomedical and Clinical Science, Peninsula Medical School, University of Exeter, Exeter, UK. [63]Diabetes Genetics, Institute of Biomedical and Clinical Science, Peninsula Medical School, University of Exeter, Exeter, UK. [64]Institute of Biometrics and Epidemiology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [65]Department of Human Genetics, McGill University, Montreal, Canada. [66]Department of Medicine, Faculty of Medicine, McGill University, Montreal, Canada. [67]McGill University and Genome Quebec Innovation Centre, Montreal, Canada. [68]Department of Metabolic Diseases, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. [69]Department of Endocrinology and Diabetes, Norfolk and Norwich University Hospital National Health Service Trust, Norwich, UK. [70]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA. [71]Institut interrégional pour la Santé (IRSA), La Riche, France. [72]Department of Medicine, Helsinki University Hospital, University of Helsinki, Helsinki, Finland. [73]Department of Internal Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. [74]Molecular Genetics, Medical Biology Section, Department of Pathology and Medical Biology, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [75]Department of Genetics, University Medical Center Groningen and University of Groningen, Groningen, The Netherlands. [76]Department of Physiology and Biophysics, University of Southern California School of Medicine, Los Angeles, California, USA. [77]National Institute of Health, Bethesda, Maryland, USA. [78]Department of Genetics and Pathology, Rudbeck Laboratory, Uppsala University, Uppsala, Sweden. [79]University of Southern Denmark, Odense, Denmark. [80]Centre for Diabetes, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [81]Department of Medicine, The Hospital of Levanger, Levanger, Norway. [82]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [83]Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy. [84]Croatian Centre for Global Health, Faculty of Medicine, University of Split, Split, Croatia. [85]Institute for Clinical Medical Research, University Hospital 'Sestre Milosrdnice', Zagreb, Croatia. [86]Department of Public Health, University of Helsinki, Helsinki, Finland. [87]South Ostrobothnia Central Hospital, Seinäjoki, Finland. [88]Red RECAVA Grupo RD06/0014/0015, Hospital Universitario La Paz, Madrid, Spain. [89]Diabetes Research Group, Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne, UK. [90]Department of Preventative Medicine, Keck Medical School, University of Southern California, Los Angeles, California, USA. [91]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, USA. [92]Department of Biomedical Science, Panum, Faculty of Health Science, University of Copenhagen, Copenhagen, Denmark. [93]Faculty of Health Science, University of Aarhus, Aarhus, Denmark. [94]Klinikum Grosshadern, Munich, Germany. [95]Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. [96]Faculty of Medicine, University of Iceland, Reykjavík, Iceland. [97]Genomic Medicine, Imperial College London, Hammersmith Hospital, London, UK. [98]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [99]A full list of members is provided in the **Supplementary Note**. [100]These authors contributed equally. Correspondence should be addressed to D.A. (altshuler@molbio.mgh.harvard.edu), M.B. (boehnke@umich.edu), P.F. (froguel@good.ibl.fr), T.I. (illig@helmholtz-muenchen.de), M.I.M. (mark.mccarthy@drl.ox.ac.uk), K. Stefansson (kstefans@decode.is), C.M.v.D. (c.vanduijn@erasmusmc.nl) or J.F.W. (jim.wilson@hgu.mrc.ac.uk).

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 correction[64] 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 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 I^2 (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 samples[8,9]. 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 *de novo* genotyping in 4 studies including 8,535 cases and 12,326 controls. **Supplementary Tables 1** and **2** provide sample counts and summary characteristics. Details of SNP genotyping platforms, genotype quality control, imputation methods (where appropriate) and T2D association analysis approaches are summarized in the **Supplementary Note**.

Stage 1 and 2 meta-analysis. For autosomal SNPs, the maximum sample size for the combination of stages 1 and 2 was 42,542 cases and 98,912 controls, yielding a total effective sample size of 101,290. For the X-chromosome SNP of interest, the equivalent numbers were 16,665 cases and 51,313 controls. We performed a fixed effects meta-analysis to combine association results for the two stages.

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 samples[8,9] (as rs2237895—genome-wide significant in European ancestry samples[9]—was not present in our tested SNPs or in HapMap); we also did not include the recent addition of six further susceptibility loci[11–13] (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 $P < 10^{-5}$ 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/m^2 (4,048 T2D cases and 25,096 controls from 7 studies; see **Supplementary Note** for details) and BMI > 30 kg/m^2 (2,877 T2D cases, 6,764 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 statistic[65]. 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 under an additive genetic model with inclusion of study-specific covariates (see **Supplementary Note** for details of samples), performing 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 (≥45 years) cases.

Calculation of sibling relative risk. The contribution to the sibling relative risk of a SNP with risk allele frequency p and corresponding allelic odds ratio β is

$$\lambda_{sib} = \left[1 + \frac{p(1-p)(\beta-1)^2}{2[(1-p)+\beta p]^2} \right]^2$$

assuming a multiplicative genetic model[66]. 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 investigators[12,13]. To capture the parent-of-origin–specific effects at *KCNQ1*, *KLF14* and the additional locus on 11p15, we used previous data[11].

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 individuals[12,13]. 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 assessment[67] were also analyzed.

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[68], we restricted these analyses to eight population-based studies (including 21,233 individuals[43]).

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[47]). 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 genealogy[69].

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 (*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 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.

URLs. IMPUTE, http://mathgen.stats.ox.ac.uk/impute/impute.html/; LocusZoom, http://csg.sph.umich.edu/locuszoom/.

64. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
65. Higgins, J.P., Thompson, S.G., Deeks, J.J. & Altman, D.G. Measuring inconsistency in meta-analyses. *Br. Med. J.* **327**, 557–560 (2003).
66. Lin, S., Chakravarti, A. & Cutler, D.J. Exhaustive allelic transmission disequilibrium tests as a new approach to genome-wide association studies. *Nat. Genet.* **36**, 1181–1188 (2004).
67. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–409 (1985).
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).
69. Stefansson, H. *et al.* A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129–137 (2005).