nature genetics

Genetic fine mapping and genomic annotation defines causal mechanisms at type 2 diabetes susceptibility loci

We performed fine mapping of 39 established type 2 diabetes (T2D) loci in 27,206 cases and 57,574 controls of European ancestry. We identified 49 distinct association signals at these loci, including five mapping in or near *KCNQ1*. 'Credible sets' of the variants most likely to drive each distinct signal mapped predominantly to noncoding sequence, implying that association with T2D is mediated through gene regulation. Credible set variants were enriched for overlap with FOXA2 chromatin immunoprecipitation binding sites in human islet and liver cells, including at *MTNR1B*, where fine mapping implicated rs10830963 as driving T2D association. We confirmed that the T2D risk allele for this SNP increases FOXA2-bound enhancer activity in islet- and liver-derived cells. We observed allele-specific differences in NEUROD1 binding in islet-derived cells, consistent with evidence that the T2D risk allele increases islet *MTNR1B* expression. Our study demonstrates how integration of genetic and genomic information can define molecular mechanisms through which variants underlying association signals exert their effects on disease.

Genome-wide association studies (GWAS) of common variants, defined by minor allele frequency (MAF) ≥5%, have been successful in identifying loci contributing to T2D susceptibility[1–5]. GWAS-identified loci are typically represented by a 'lead' SNP with the strongest signal of association in the region. However, lead SNPs may not directly influence disease susceptibility but instead be proxies for causal variants because of linkage disequilibrium (LD). Interpretation may be further complicated by the presence of more than one causal variant at a locus, possibly acting through the joint effects of alleles on the same haplotype. This complex genetic architecture would result in multiple 'distinct' association signals at the same locus, which could only be delineated, statistically, through conditional analyses.

With the exception of loci where the lead SNP is a protein-altering variant, including *PPARG*[6], *KCNJ11-ABCC8* (ref. 7), *SLC30A8* (ref. 8) and *GCKR*[9], the mechanisms by which associated alleles influence T2D susceptibility are largely unknown. At other loci, direct biological interpretation of the effect of genetic variation on T2D is more challenging because the association signals mostly map to noncoding sequence. Although recent reports have demonstrated a relationship between T2D-associated variants and transcriptional enhancer activity, particularly in human pancreatic islets, liver cells, adipose tissue and muscle[10–14], the DNA-binding proteins through which these effects are mediated remain obscure. Localization of noncoding causal variants may highlight the specific regulatory elements they perturb and potentially the genes through which they operate, providing valuable insights into the pathophysiological basis of T2D susceptibility at GWAS-identified loci.

To improve the localization of potential causal variants for T2D and to characterize the mechanisms through which these variants alter disease risk, we performed comprehensive fine mapping of 39 established susceptibility loci through high-density imputation into 27,206 cases and 57,574 controls from 23 studies of European ancestry genotyped with the Metabochip[15] (**Supplementary Tables 1** and **2**). Within each locus, we aimed to (i) evaluate the evidence for multiple distinct association signals through conditional analyses; (ii) undertake fine mapping by defining credible sets of variants that account for ≥99% of the probability of driving each distinct association signal; and (iii) interrogate credible sets for functional and regulatory annotation to provide insight into the mechanisms through which variants driving association signals influence disease risk.

RESULTS

Imputation into Metabochip fine-mapping regions

The Metabochip includes high-density coverage of 257 'fine-mapping regions' that have previously been associated with 23 metabolic, cardiovascular and anthropometric traits[15]. SNPs in these regions were selected using reference data from the HapMap Project[16] and the 1000 Genomes Project[17]. At design, 27 T2D susceptibility loci were selected for fine mapping. However, subsequent T2D GWAS efforts have identified additional loci that overlap 12 further fine-mapping regions that were initially selected for other traits (**Supplementary Table 3**). To enhance coverage of variation in the fine-mapping regions, we undertook imputation of the Metabochip scaffold up to the 1000 Genomes Project phase 1 integrated reference panel (March 2012 release)[18], including multi-ancestry haplotypes to reduce error rates[19] (Online Methods).

The quality of imputation was variable across studies, particularly for variants with MAF <5%, and was dependent on the scaffold sample size (**Supplementary Table 4**). We defined variants to be 'well imputed' at widely used thresholds[20] of IMPUTEv2 (ref. 21) info ≥0.4 or minimac[22] r^2 ≥0.3 in at least 80% of the total effective sample size (N_e ≥ 59,122) across studies. With this definition, 99.4% and 89.0%, respectively, of common and low-frequency (0.5% ≤ MAF < 5%) variants in 1000 Genomes Project European-ancestry haplotypes were

A full list of authors and affiliations appears at the end of the paper.

Received 1 December 2014; accepted 7 October 2015; published online 9 November 2015; doi:10.1038/ng.3437

Table 1 Established T2D susceptibility loci with multiple distinct signals of association at locus-wide significance in the GCTA joint regression model ($P_J < 1 \times 10^{-5}$)

						Metabochip GCTA joint model 27,206 cases and 57,574 controls			Validation GCTA joint model 19,662 cases and 115,140 controls			Combined GCTA joint model 46,868 cases and 172,714 controls	
Locus	Index variant	Chr.	Position (Build 37)	Risk allele	Other allele	RAF	OR (95% CI)	P_J	RAF	OR (95% CI)	P_J	OR (95% CI)	P_J
DGKB	rs10276674	7	14,922,007	C	T	0.183	1.08 (1.04–1.11)	4.5×10^{-6}	0.216	1.09 (1.05–1.12)	1.3×10^{-6}	1.08 (1.06–1.11)	2.8×10^{-11}
	rs1974620	7	15,065,467	T	C	0.519	1.06 (1.04–1.09)	1.6×10^{-6}	0.515	1.05 (1.03–1.08)	0.00014	1.06 (1.04–1.08)	1.0×10^{-9}
CDKN2A-	rs10811660	9	22,134,068	G	A	0.830	1.32 (1.27–1.38)	2.4×10^{-44}	0.817	1.21 (1.17–1.26)	2.6×10^{-21}	1.27 (1.23–1.30)	1.1×10^{-61}
CDKN2B	rs10757283	9	22,134,172	T	C	0.437	1.14 (1.10–1.17)	7.4×10^{-18}	0.436	1.11 (1.07–1.14)	1.3×10^{-10}	1.12 (1.10–1.14)	3.6×10^{-26}
KCNQ1	chr11:2692322:D	11	2,692,322	D	R	0.374	1.08 (1.05–1.10)	3.5×10^{-8}	0.413	1.09 (1.06–1.12)	1.2×10^{-8}	1.08 (1.06–1.10)	2.3×10^{-15}
	rs2283220	11	2,755,548	A	G	0.661	1.06 (1.03–1.09)	0.000016	0.710	1.05 (1.02–1.08)	0.0031	1.06 (1.03–1.08)	2.4×10^{-7}
	rs2237895	11	2,857,194	C	A	0.428	1.08 (1.05–1.11)	6.6×10^{-7}	0.433	1.07 (1.03–1.10)	2.8×10^{-4}	1.07 (1.05–1.10)	5.3×10^{-10}
	rs74046911	11	2,858,636	C	T	0.951	1.32 (1.24–1.40)	1.7×10^{-17}	0.943	1.25 (1.17–1.34)	4.8×10^{-10}	1.29 (1.23–1.35)	9.6×10^{-26}
	rs458069	11	2,858,800	G	C	0.707	1.06 (1.03–1.10)	0.00026	0.707	1.07 (1.03–1.11)	0.00085	1.06 (1.04–1.09)	1.0×10^{-6}
HNF1A	rs1169288	12	121,416,650	C	A	0.334	1.10 (1.07–1.13)	5.4×10^{-10}	0.316	1.08 (1.05–1.12)	2.8×10^{-6}	1.09 (1.07–1.12)	8.1×10^{-15}
	rs1800574	12	121,416,864	T	C	0.027	1.21 (1.11–1.31)	5.2×10^{-6}	0.020	1.23 (1.12–1.35)	0.000026	1.22 (1.14–1.29)	5.1×10^{-10}
	chr12:121440833:D	12	121,440,833	R	D	0.416	1.06 (1.03–1.09)	0.000028	0.382	1.08 (1.04–1.11)	2.5×10^{-6}	1.07 (1.05–1.09)	2.9×10^{-10}
MC4R	chr18:57739289:D	18	57,739,289	D	R	0.234	1.05 (1.02–1.09)	0.00079	0.254	1.07 (1.03–1.10)	0.000059	1.06 (1.04–1.08)	1.9×10^{-7}
	rs17066842	18	58,040,624	G	A	0.961	1.13 (1.06–1.21)	0.00033	0.948	1.11 (1.04–1.19)	0.0012	1.12 (1.07–1.17)	1.4×10^{-6}
GIPR	rs4399645	19	46,166,073	T	C	0.395	1.07 (1.04–1.10)	4.4×10^{-7}	0.441	1.05 (1.01–1.08)	0.0046	1.06 (1.04–1.08)	1.4×10^{-8}
	rs2238689	19	46,178,661	C	T	0.425	1.09 (1.07–1.12)	9.7×10^{-12}	0.424	1.07 (1.04–1.10)	9.0×10^{-6}	1.08 (1.06–1.11)	8.3×10^{-16}
HNF4A[a]	rs1800961	20	43,042,364	T	C	0.034	1.16 (1.09–1.24)	0.000011	0.041	1.16 (1.08–1.25)	0.000051	1.16 (1.10–1.22)	2.3×10^{-9}

Each distinct association signal was represented by an index variant in the GCTA joint regression model based on (i) summary statistics from a combined meta-analysis of 46,868 cases and 172,714 controls of European ancestry and (ii) reference genotype data from GoDARTS (3,298 cases and 3,708 controls of European ancestry from the UK) to approximate LD across fine-mapping regions. Insertion-deletion alleles are coded as reference (R) or deletion (D). Chr., chromosome; RAF, risk allele frequency; OR, odds ratio for the risk allele; CI, confidence interval.

[a]The previously reported T2D GWAS SNP at the *HNF4A* locus (rs4812829) is not included in the fine-mapping region. However, the reported index variant, rs1800961, is independent of the GWAS-identified SNP and thus represents a new, distinct association signal at this locus.

well imputed and therefore retained for downstream association analyses. Within studies, imputation quality was consistent across loci, despite the differential priority of fine-mapping regions and their coverage of variation at design (**Supplementary Table 5**). Thus, 1000 Genomes Project imputation into the Metabochip scaffold provides nearly complete coverage of common and low-frequency variation across the 39 T2D susceptibility loci and supports direct interrogation of the majority of variants with MAF ≥0.5% in European-ancestry populations.

Distinct association signals at T2D susceptibility loci

The first step in fine mapping GWAS loci is to delineate distinct association signals arising from multiple causal variants in the same region, which can efficiently be achieved through approximate conditioning with GCTA[23]. Within each T2D fine-mapping region, we identified distinct signals attaining 'locus-wide' significance (represented by an index variant with $P_J < 1 \times 10^{-5}$ in the joint association model) by applying GCTA in two stages (Online Methods). First, we selected index variants on the basis of fixed-effects meta-analysis across Metabochip studies. Second, we performed *in silico* replication of the index variants in a validation meta-analysis of an additional 19,662 T2D cases and 115,140 controls from ten GWAS of European ancestry (**Supplementary Tables 1**, **2** and **6**). Finally, because GCTA is only an approximation, we confirmed the association of each index variant through exact conditional analysis across Metabochip studies (Online Methods and **Supplementary Table 7**).

The most dramatic delineation of distinct association signals was observed for the region flanking *KCNQ1*, where five noncoding index variants attained locus-wide significance **(Table 1** and **Supplementary Fig. 1**). Distinct association signals represented by three of the index variants have been reported in previous GWAS of European[4] and East Asian[24] ancestry: rs74046911 ($P_J = 3.6 \times 10^{-26}$, $r^2 = 0.98$ with the East Asian lead SNP, rs2237897) and rs2237895 ($P_J = 2.1 \times 10^{-9}$, $r^2 = 0.75$ with one European lead SNP, rs163184), both of which map to a <50-kb intronic recombination interval of *KCNQ1*, and chr11:2692322:D ($P_J = 7.2 \times 10^{-16}$, $r^2 = 0.59$ with a second European lead SNP, rs231361), which resides in the *KCNQ1OT1* long noncoding RNA gene that controls regional imprinting[25]. The two remaining distinct association signals at this locus are new. The first, indexed by rs458069 ($P_J = 3.2 \times 10^{-6}$), maps to the same <50-kb recombination interval as rs74046911 and rs2237895 but is in only weak LD with both ($r^2 = 0.02$ and 0.25, respectively). The second, indexed by rs2283220 ($P_J = 2.2 \times 10^{-7}$), resides in a neighboring intron of *KCNQ1*, outside of the <50-kb recombination interval (**Supplementary Fig. 1**).

At the *HNF1A* locus, we observed three distinct association signals **(Table 1** and **Supplementary Fig. 2**), represented by index variants that are in only weak LD with the previously reported lead GWAS SNP, rs12427353. The index variants include two nonsynonymous variants, rs1169288 ($P_J = 4.4 \times 10^{-14}$, $r^2 = 0.09$, encoding HNF1A p.Ile27Leu) and rs1800574 ($P_J = 4.2 \times 10^{-10}$, $r^2 = 0.01$, encoding HNF1A p.Ala98Val), and one intergenic SNP, chr12:121440833:D ($P_J = 2.9 \times 10^{-10}$, $r^2 = 0.19$).

We also observed four loci that each had two distinct association signals (*CDKN2A-CDKN2B*, *DGKB*, *MC4R* and *GIPR*), with each locus represented by noncoding index variants **(Table 1** and **Supplementary Fig. 3**). The index variants at the *CDKN2A-CDKN2B* locus represent the known T2D haplotype association signal mapping to a 12-kb intergenic recombination interval[26–28]. Previous European-ancestry GWAS meta-analyses[4] have highlighted

Table 2 Distinct association signals at established T2D susceptibility loci for which the 99% credible set contains no more than ten variants

									99% credible set		
Locus	Index variant	Chr.	Position (Build 37)	Risk allele	Other allele	RAF	*P* value	OR (95% CI)	SNPs	Interval length (bp)	Interval position (bp)
MTNR1B	rs10830963	11	92,708,710	G	C	0.283	2.9×10^{-12}	1.10 (1.07–1.13)	1	1	92,708,710–92,708,710
TCF7L2	rs7903146	10	114,758,349	T	C	0.260	5.8×10^{-120}	1.39 (1.35–1.43)	3	4,279	114,754,071–114,758,349
KCNQ1	rs74046911	11	2,858,636	C	T	0.951	5.9×10^{-18}	1.33 (1.25–1.42)	3	197	2,858,440–2,858,636
ZBED3	rs7732130	5	76,435,004	G	A	0.278	6.4×10^{-10}	1.09 (1.06–1.12)	5	10,056	76,424,949–76,435,004
CDKN2A-CDKN2B	rs10757283	9	22,134,172	T	C	0.437	2.8×10^{-19}	1.14 (1.11–1.18)	5	1,007	22,133,645–22,134,651
SLC30A8	rs13266634	8	118,184,783	C	T	0.676	1.3×10^{-18}	1.13 (1.10–1.16)	6	33,133	118,184,783–118,217,915
CDKN2A-CDKN2B	rs10811660	9	22,134,068	G	A	0.830	7.0×10^{-43}	1.32 (1.27–1.37)	6	1,397	22,132,698–22,134,094
HNF1B	rs4430796	17	36,098,040	G	A	0.455	6.3×10^{-12}	1.09 (1.07–1.12)	7	5,791	36,097,775–36,103,565
CDKAL1	rs35261542	6	20,675,792	A	C	0.280	9.6×10^{-23}	1.15 (1.12–1.18)	8	30,073	20,673,880–20,703,952
GLIS3	chr9:4294707:I	9	4,294,707	I	R	0.360	6.5×10^{-8}	1.07 (1.05–1.10)	10	15,453	4,283,137–4,298,589

Association summary statistics and credible set construction are based on the meta-analysis of Metabochip studies in 27,206 cases and 57,574 controls of European ancestry. In loci with multiple distinct signals of association, results are presented from exact conditional analysis after adjusting for all other index variants in the fine-mapping region. In loci with a single signal of association, results are presented from unconditional analysis. Chr., chromosome; RAF, risk allele frequency; OR, odds ratio for the risk allele; CI, confidence interval.

a potential distinct association signal, located upstream of the recombination interval in the noncoding *CDKN2B-AS1* (*ANRIL*) transcript. However, our conditional analyses indicate that the association in this region can be fully explained by the two index SNPs in the recombination interval, which, when considered together, fully extinguish the *CDKN2B-AS1* signal (**Supplementary Fig. 4**). The index variants at *DGKB* and *MC4R* also confirm previously reported distinct association signals at these loci in European-ancestry GWAS meta-analyses[4]. At the *GIPR* locus, the two index variants (rs2238689, $P_J = 8.3 \times 10^{-16}$; rs4399645, $P_J = 1.4 \times 10^{-8}$) are not in strong LD with the previously reported[4] lead SNP (rs8108269; $r^2 = 0.43$ with rs2238689 and 0.00 with rs4399645) but together can better explain the T2D association signal in this region.

Finally, we observed a new distinct association signal at the *HNF4A* locus, represented by the coding index variant rs1800961 ($P_J = 1.4 \times 10^{-9}$, encoding HNF4A p.Thr139Ile, referred to as p.Thr130Ile in some previous studies[29]). Unfortunately, this fine-mapping region was included on the Metabochip for high-density lipoprotein cholesterol[15,30] (**Supplementary Table 3**) and does not include the previously reported[4] lead T2D SNP at this locus, rs4812829, precluding conditional analyses in these data. However, rs4812829 is not in LD with our index variant ($r^2 = 0.02$), suggesting that there are at least two distinct T2D association signals at the *HNF4A* locus.

Of the 49 distinct association signals achieving locus-wide significance across the T2D loci represented on the Metabochip (five at *KCNQ1*, three at *HNF1A*, two each at *CDKN2A-CDKN2B*, *DGKB*, *MC4R* and *GIPR*, and one each at the remaining loci), only three index variants are not common (**Supplementary Fig. 5** and **Supplementary Table 6**): rs1800574 (MAF = 2.2%, odds ratio (OR) = 1.21) for one signal at the *HNF1A* locus; rs1800961 (MAF = 3.9%, OR = 1.16) at the *HNF4A* locus; and rs17066842 (MAF = 4.8%, OR = 1.12) for one signal at the *MC4R* locus.

Localizing variants driving T2D association signals

We used statistical evidence of association from the meta-analysis of Metabochip studies to construct 99% credible sets of variants[28] that are most likely to drive the 49 distinct signals (Online Methods, **Supplementary Fig. 6** and **Supplementary Table 8**). For ten distinct association signals mapping to nine loci, the 99% credible set included no more than ten variants (**Table 2** and **Supplementary Table 9**). The greatest refinement was observed at the *MTNR1B* locus, where the credible set included only the index variant, rs10830963, accounting for more than 99.8% of the posterior probability of driving the association signal (π_C). Small credible sets were also observed for the association at *TCF7L2* (three variants, indexed by rs7903146, mapping to 4.3 kb) and one signal at *KCNQ1* (three variants, indexed by rs74046911, mapping to just 200 bp). The 99% credible sets for both distinct association signals at *CDKN2A-CDKN2B* together included just 11 variants in total and map to less than 2 kb.

We performed functional annotation of credible set variants to search for evidence that association signals are driven by coding alleles. Across the 49 signals, only nine coding variants attained π_C >1% (**Supplementary Table 10**), including six previously reported nonsynonymous T2D risk alleles at *PPARG*[6], *KCNJ11-ABCC8* (refs. 7,31,32), *SLC30A8* (refs. 8,33) and *GCKR*[9,34]. The remaining three coding alleles were the index variants for association signals mapping to *HNF4A* (rs1800961, encoding p.Thr139Ile, $\pi_C = 97.4\%$) and *HNF1A* (rs1169288, encoding p.Ile27Leu, $\pi_C = 75.5\%$; rs1800574, encoding p.Ala98Val, $\pi_C = 34.0\%$). Our findings are supported by earlier studies, which reported nominal evidence for association of these three coding variants with T2D and defects in insulin secretion *in vivo* and demonstrated reduced transcriptional activity of HNF1A target genes using *in vitro* assays[29,35]. These data provide strong evidence that *HNF4A* and *HNF1A* are T2D effector transcripts at these loci, a view further supported by the known impact of rare, loss-of-function mutations in these genes on maturity-onset diabetes of the young[36,37]. Given the nearly complete coverage of common and low-frequency variants in fine-mapping regions after 1000 Genomes Project imputation, it is unlikely that additional distinct signals in the established T2D susceptibility loci represented on the Metabochip are driven by coding variation with MAF ≥0.5%, confirming reports that these associations are most likely mediated by effects on gene regulation[10,13,14,38].

Regulatory mechanisms underlying T2D association signals

We sought to understand the regulatory mechanisms through which variants at the 39 established T2D susceptibility loci influence disease by intersecting the 99% credible sets for each distinct association signal with chromatin immunoprecipitation sequence (ChIP-seq) data for 165 transcription factors, chromatin state maps from 12 cell types and long noncoding RNA transcripts from 25 cell types (Online

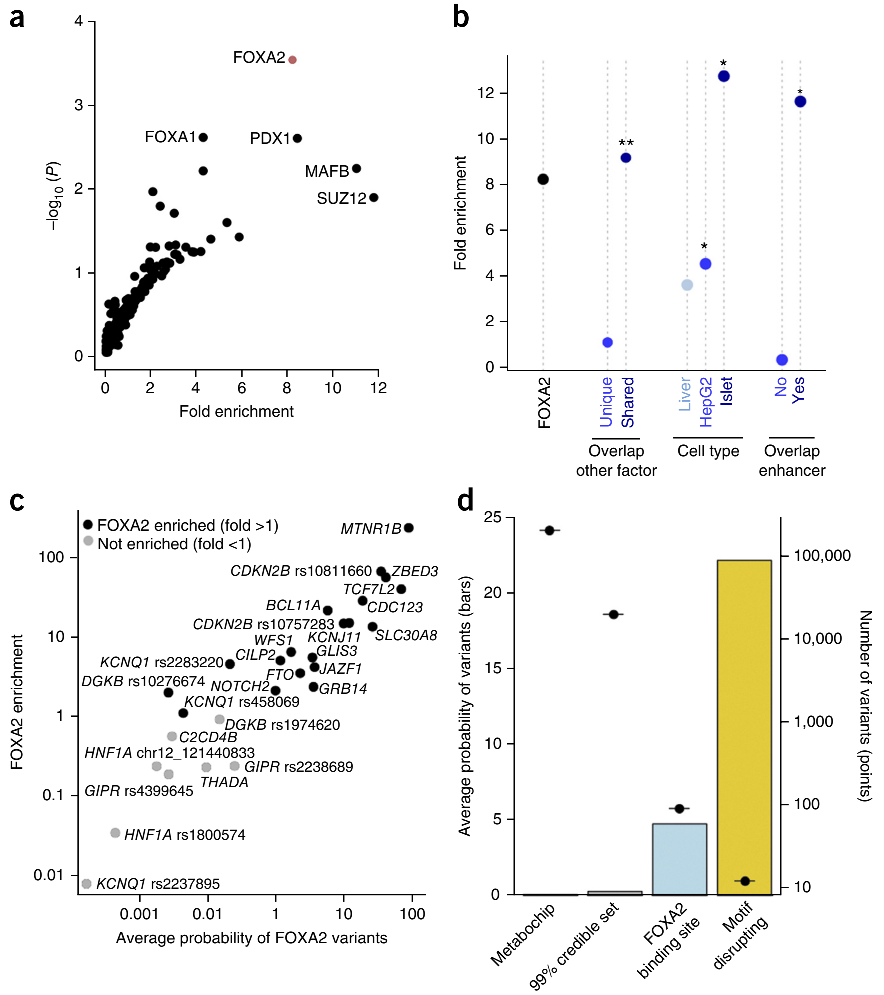

Figure 1 FOX2A-bound sites are a genomic marker of T2D risk variants. (**a**) Variants in ChIP-seq binding sites for 165 proteins were tested for enrichment of posterior probabilities of driving association signals in comparison to variants in shifted sites. Variants in FOXA2 ChIP-seq sites were significantly enriched ($P < 0.00030$). (**b**) FOXA2 ChIP-seq sites were partitioned on the basis of overlap with other genomic features. There was stronger enrichment for (i) FOXA2 sites overlapping a ChIP-seq site for another protein in comparison to unique sites; (ii) sites identified in primary islets in comparison to HepG2 or primary liver cells; and (iii) sites that overlapped islet enhancers in comparison to those that did not ($^{**}P < 0.00030$, $^{*}P < 0.05$). (**c**) Variants at each signal were tested for enrichment for FOXA2 binding sites. Nineteen signals had greater enrichment than expected in comparison to shifted sites; at 15 signals, this enrichment was nominally significant ($P < 0.05$). (**d**) FOXA2-bound variants disrupting recognition motifs have an increased probability of being causal.

Methods and **Supplementary Table 11**). We applied an enrichment procedure that compared the mean posterior probability of driving the association signal for credible set variants directly overlapping sites for each regulatory annotation with a null distribution obtained from randomly shifted site locations within 100 kb in either direction of the original element location.

We first applied this procedure to chromatin state and noncoding RNA elements using the 19,266 credible set variants for all 49 distinct association signals (**Supplementary Fig. 7**). Using Bonferroni correction for the 37 tested cell type annotations ($P < 0.0014$), variants in pancreatic islet enhancer elements[14] had significantly higher posterior probability of driving association signals than expected from the null distribution (1.97-fold enrichment, $P = 0.00022$). We also observed nominal evidence for enrichment of the posterior probability of driving association signals among variants in human islet and hepatocellular carcinoma (HepG2) promoter elements[10,14] ($P = 0.0052$ and 0.0064, respectively). However, there was no corresponding enrichment of variants in regulatory elements for other cell types or in noncoding transcripts. These results are consistent with previous studies supporting a contribution of regulatory enhancer and promoter variants to T2D susceptibility in specific cell types[11–14].

We next sought to gain insight into the transcription factors these regulatory variants perturb and applied the same procedure to ChIP-seq binding data for 165 proteins (**Fig. 1** and **Supplementary Fig. 8**). Using Bonferroni correction for the 165 proteins tested ($P < 0.00030$), the 89 credible set variants overlapping 57 FOXA2 ChIP-seq binding sites, assayed in human HepG2 (ref. 10) and islet[14] cells, had significantly higher posterior probability of driving association signals than expected from the null distribution (8.24-fold enrichment, $P = 0.00028$). The enrichment of FOXA2 ChIP-seq sites was exclusive to sites shared with at least one other factor (9.18-fold enrichment, $P = 0.00028$) in comparison to those that were not shared (1.12-fold enrichment, $P = 0.11$). Enrichment for FOXA2 binding was also more pronounced among sites identified in pancreatic islets (15.43-fold enrichment, $P = 0.00045$) than in those identified in HepG2 cells (4.55-fold enrichment, $P = 0.011$). To exclude the possibility that the enrichment in HepG2 cells was driven by artifacts due to HepG2 being a cultured cell line, we compared the FOXA2 sites in HepG2 cells to those previously assayed in primary liver[39]. We observed significant intersection of the HepG2 and liver FOXA2 sites that overlapped credible set variants ($P = 1.5 \times 10^{-9}$). Consequently, we detected similar FOXA2 binding site enrichment among sites detected in liver (3.63-fold enrichment, $P = 0.061$) to that observed in HepG2 cells. We also compared FOXA2 ChIP-seq sites, across the genome, from liver, HepG2 and islet cells (**Supplementary Fig. 9**). The number of sites varied across the cell types (8,023 for liver, 40,866 for HepG2 cells and 27,291 for islets), likely owing, in part, to technical differences, including in sequencing platform, depth and read length. However, the intersection of FOXA2 sites between each pair of cell types was highly significant ($P < 2.2 \times 10^{-16}$).

Given the preponderance of T2D-associated variants in islet enhancers, we next tested to what extent enrichment of FOXA2 binding is driven by colocalization of variants with these genomic features[14]. Variants in FOXA2-bound sites were not enriched for posterior probability of driving association signals after removing enhancer sites (0.36-fold enrichment, $P = 0.69$). Conversely, variants in islet enhancers remained nominally enriched when FOXA2-bound sites were removed (1.65-fold enrichment, $P = 0.014$). These results suggest that FOXA2 binding assayed by ChIP-seq, at a

Table 3 Motif-altering credible set variants in FOXA2 sites

Locus	Index variant	Motif-altering variant	Chr.	Position (Build 37)	Posterior probability (π_C)	Motif allele	Chromatin state
MTNR1B	rs10830963	rs10830963	11	92,708,710	0.998	G	Islet enhancer, HepG2 enhancer
TCF7L2	rs7903146	rs7903146	10	114,758,349	0.78	T	Islet enhancer
SLC30A8	rs13266634	rs13266634	8	118,184,783	0.29	T	Islet enhancer
CDKN2B	rs10811660	rs10811660	9	22,134,068	0.24	A	Islet enhancer
CDC123	rs11257658	rs11257655	10	12,307,894	0.21	T	Islet enhancer, HepG2 enhancer
JAZF1	rs1513272	rs849133	7	28,192,280	0.042	T	Islet enhancer
KCNQ1	rs2283220	rs231907	11	2,752,130	0.031	T	HepG2 enhancer
FTO	rs9927317	rs9940128	16	53,800,754	0.027	G	Islet enhancer, HepG2 enhancer
FTO	rs9927317	rs9939973	16	53,800,568	0.025	G	Islet enhancer, HepG2 enhancer
KCNQ1	rs458069	rs78688069	11	2,752,183	0.0006	A	HepG2 enhancer
KCNQ1	rs458069	rs190728714	11	2,813,084	0.00042	G	Islet enhancer
DGKB	rs10276674	rs7798360	7	15,055,972	0.00005	G	–

Chr., chromosome.

subset of enhancer element locations that are often shared by other proteins, is a genomic marker of variants with an increased posterior probability of driving T2D association signals.

Having demonstrated global over-representation for FOXA2 ChIP-seq binding by considering all loci simultaneously, we applied the same procedure to the 99% credible sets of each distinct association signal separately to identify those with the strongest evidence for local enrichment (**Fig. 1**). We observed over-representation of credible set variants in islet or HepG2 FOXA2-bound sites for 19 association signals, 15 of which attained nominal significance ($P < 0.05$). A total of 41 credible set variants at these 19 distinct association signals overlapped a FOXA2 ChIP-seq site in at least one of the two cell types (**Supplementary Table 12**). Of these, 12 variants were predicted to disrupt *de novo* recognition motifs (for FOXA2 and other factors) that were enriched in FOXA2-bound sequence (**Table 3** and **Supplementary Table 13**). The mean posterior probability of driving the association (π_C) for these 12 variants was 22.0% on the basis of genetic fine mapping (**Fig. 1**), more than four times greater than for those in FOXA2 ChIP-seq sites that were not motif disrupting at the same signals (mean π_C = 5.2%, P = 0.024). Furthermore, 11 of these 12 variants also overlapped an enhancer element in islets (nine variants) or HepG2 cells (six variants), indicating that they are in transcriptionally active regions (**Table 3**). They include two variants with experimentally validated differences in regulatory activity: rs7903146 (π_C = 77.6%) at *TCF7L2* (ref. 40) and rs11257655 (π_C = 21.1%) at *CDC123* (ref. 41). They also include rs10830963, the index variant at the *MTNR1B* locus, which accounts for 99.8% of the posterior probability of driving the association signal on the basis of genetic fine mapping. These results suggest that FOXA2 binding patterns can be used to highlight specific variants that are potentially causal for T2D susceptibility through altered regulatory binding.

Altered regulatory activity of the *MTNR1B* credible variant

To demonstrate how local enrichment of FOXA2 binding can be used to highlight regulatory mechanisms through which credible set variants might influence T2D susceptibility, we focused on the *MTNR1B*

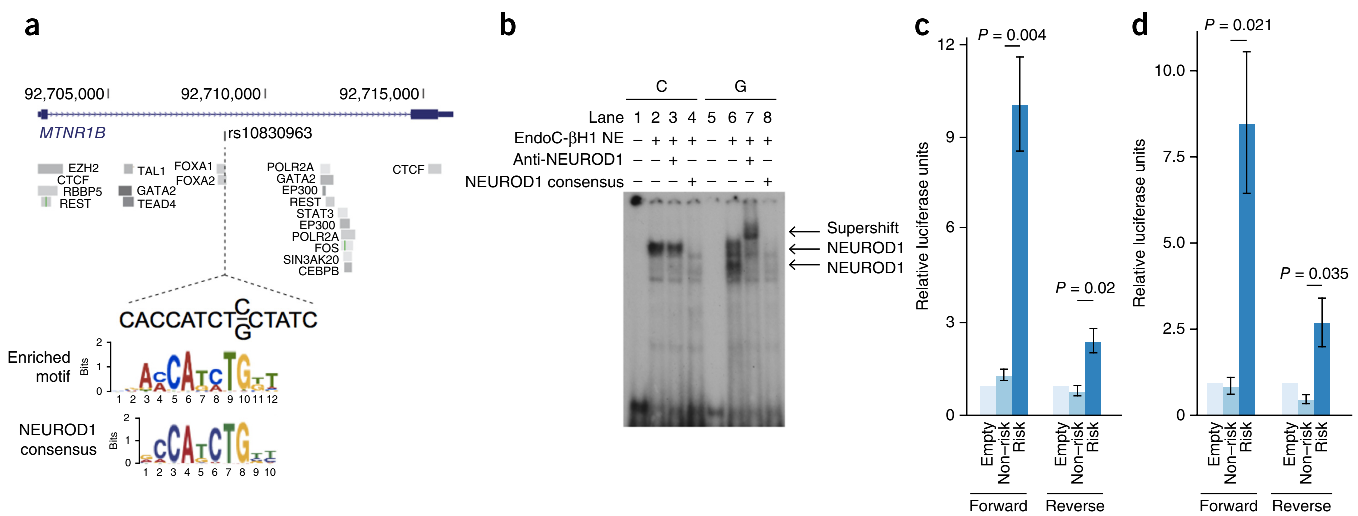

Figure 2 The lone variant in the 99% credible set at the *MTNR1B* locus affects FOXA2-bound enhancer activity. (**a**) The intronic variant rs10830963 has 99.8% posterior probability of driving the association signal at the *MTRN1B* locus. This variant overlaps a FOXA2 binding site, and the G risk allele is predicted to create a *de novo* recognition motif that closely matches the NEUROD1 consensus motif. (**b**) EMSA of 25-bp fragments encompassing either allele of rs10830963 in EndoC-βH1 cell extracts. Proteins were bound to both alleles. In the presence of an antibody to NEUROD1, only the signal corresponding to the risk allele was supershifted; in the presence of an unlabeled probe with the NEUROD1 consensus motif, the signal was competed away. NE, nuclear extract. (**c**,**d**) The 224-bp sequence surrounding each allele of rs10830963 was cloned into a luciferase reporter construct containing a minimal promoter and tested for luciferase activity in EndoC-βH1 (**c**) and HepG2 (**d**) cells (n = 3 independent experiments performed in triplicate for each cell type). Results are presented as means ± s.e.m. The risk allele had significantly increased enhancer activity over the protective allele in both forward and reverse orientations of the element in both cell types.

locus. Variants mapping to this region have among the strongest known effects on both T2D risk[6] and fasting plasma glucose concentration[42], and physiological data indicate an effect of *MTNR1B* on both insulin secretion and insulin action[43]. The lone credible variant in the *MTNR1B* region, rs10830963, overlaps a FOXA2 ChIP-seq binding site, and the risk allele, G, is predicted to create a recognition motif that matches the consensus sequences of NEUROD1 and several other factors (**Fig. 2** and **Supplementary Table 13**). We tested *in silico* predictions of protein binding at rs10830963 via electrophoretic mobility shift assay (EMSA) with 25-bp probe fragments centered on each allele in human pancreatic islet β cell (EndoC-βH1)[44] or human liver HepG2 cell extracts. We observed allele-specific binding patterns with extracts from both cell lines (**Fig. 2** and **Supplementary Fig. 10**).

To determine the specific protein(s) bound at each allele, we then performed supershift experiments using antibodies directed against NEUROD1, FOXA2 and three other factors (TAL1, PTF1A and YY1) whose consensus binding sequences resemble the recognition motif (Online Methods). We observed a shift in the presence of antibody to NEUROD1 for the signal corresponding to the risk allele in EndoC-βH1 extracts, which could be competed away by an excess of unlabeled probe with the NEUROD1 consensus sequence (**Fig. 2**). None of the antibodies tested (including to NEUROD1) shifted the signal corresponding to the risk allele in HepG2 cell extracts (**Supplementary Fig. 10**). These results demonstrate that, *in vitro*, the risk allele of rs10830963 preferentially binds NEUROD1 in islet-derived cells and binds a protein not identified from known recognition motifs in liver-derived cells.

To relate allelic differences in protein binding to genomic activity at this site, we cloned the 224-bp region surrounding rs10830963 into a luciferase reporter vector containing a minimal promoter and tested its enhancer activity in the EndoC-βH1 and HepG2 cell lines. Consistent with *in silico* predictions, we observed a significant ($P < 0.05$) increase in luciferase expression for the element with the risk allele in comparison to the one with the protective allele in both cell lines (**Fig. 2**). Furthermore, RNA sequencing (RNA-seq) data reported from human islets have linked the T2D risk allele of rs10830963 to increased expression of *MTNR1B*[45,46]. Taken together, these results suggest that the G allele of rs10830963 increases T2D risk through increased FOXA2-bound enhancer activity, potentially mediated through NEUROD1 binding in islets, and consequently higher expression of *MTNR1B*.

Candidate effector genes at FOXA2-enriched T2D signals

We hypothesized that the locus-specific effects in mouse transcription factor knockout models would mimic patterns of binding enrichment at human disease-associated loci. We thus attempted to relate FOXA2 binding at the 19 FOXA2-enriched association signals (**Fig. 1**) to target effector genes using previously reported pancreatic islet expression profiles from wild-type and *Foxa1* and *Foxa2* double-null mice[47] (Online Methods). Syntenic genes mapping within 500 kb of the credible sets for the 19 FOXA2-enriched signals were significantly downregulated (45.2% decrease) in *Foxa1* and *Foxa2* double-knockout mice (**Supplementary Fig. 11**) in comparison to genes across the genome (0.021% increase; $P = 0.012$), whereas those mapping within 500 kb of the other 30 T2D association signals were not significantly downregulated (2.25% decrease; $P = 0.20$). We observed a consistent downregulation (39.6% decrease) when considering only the genes mapping closest to each FOXA2-enriched signal, in comparison to those across the genome (0.021% increase; $P = 0.0021$). Thus, data related to altered gene expression in *Foxa1* and *Foxa2* double-knockout mice support patterns of FOXA2 binding site enrichment in humans.

We next identified specific genes at the 19 FOXA2-enriched association signals whose mouse counterparts were downregulated in *Foxa1* and *Foxa2* double-knockout mice, as these might represent effector transcripts for these loci (**Supplementary Table 14**). Several of these genes have previously been implicated as likely effector transcripts in humans, including *TCF7L2* (refs. 48,49) (57% decrease), *KCNJ11* (refs. 7,50) (38% decrease) and *SLC30A8* (ref. 51) (135% decrease). These data also implicate new candidate effector genes at FOXA2-enriched association signals (**Supplementary Table 14**). For example, in *Foxa1* and *Foxa2* double-knockout mice, there was marked downregulation of *Reg4* (1,415% decrease), which maps to a syntenic region of the FOXA2-enriched *NOTCH2* GWAS locus, highlighting *REG4* as a likely effector transcript in humans. Additional examples of candidate effector genes include *IGF2* at the *KCNQ1* locus (135% decrease) and *CAMK1D* at the *CDC123* locus (81% decrease). Together, these results provide additional support for the importance of FOXA2 binding at a subset of T2D susceptibility loci and further highlight specific genes through which regulatory variants in these regions may operate.

DISCUSSION

We have undertaken comprehensive fine mapping of 39 established T2D susceptibility loci in 27,206 cases and 57,574 controls of European ancestry and have demonstrated that the existence of multiple distinct association signals in these regions is a common phenomenon. Index variants for just three of the 49 distinct association signals are not common, despite nearly complete coverage of variation with MAF ≥0.5% in fine-mapping regions after 1000 Genomes Project imputation. Although we cannot evaluate the impact of rare variation (MAF <0.5%) in established T2D susceptibility loci without large-scale resequencing, our data strongly argue against a role for low-frequency variants of large effect via synthetic association[52]. We have demonstrated that seven distinct association signals, mapping to six T2D susceptibility loci represented on the Metabochip, are likely to be driven by coding alleles, including new index variants mapping to *HNF1A* and *HNF4A*. Outside of these regions, our fine mapping confirms previous reports that T2D association signals are primarily driven by noncoding alleles, with effects that are mediated through gene regulation[10,13,14,38].

We have demonstrated, by genomic annotation and functional assays, that FOXA2 binding assayed by ChIP-seq can be used to pinpoint candidate causal regulatory elements, providing routes to understanding the biology of specific T2D susceptibility loci. These elements highlight variants and effector transcripts through which association signals are mediated, via altered binding of either FOXA2, directly, or another transcription factor. For example, at the *MTNR1B* locus, the risk allele of the lone credible set variant, rs10830963, preferentially binds NEUROD1 in islet-derived cells *in vitro* and increases FOXA2-bound enhancer activity in human islet and liver-derived cells. These data are consistent with previous reports correlating the risk allele with higher *MTNR1B* expression[45,46] and not loss of function[53], and they suggest that altered NEUROD1 binding in islets contributes to T2D susceptibility at this locus. Further experiments will be required to confirm our *in vitro* findings regarding NEUROD1 binding *in vivo*. However, our attempts to perform ChIP-seq in primary islet samples of the defined *MTNR1B* genotype were repeatedly unsuccessful, owing to lack of a suitable antibody to NEUROD1. These studies are further complicated by the limited availability of primary human islets, and the slow division rate of human islet-derived cell lines is an impediment to the implementation of genome editing technologies.

FOXA2 is a pioneer factor that binds native chromatin and bookmarks genomic regions for transcriptional activity[54], and it is involved in pancreatic and hepatic development[55,56]. FOXA2 is also expressed in other T2D-relevant cell types, such as adipocytes. Future studies will be required to elucidate the extent to which FOXA2 binding events across cell types influence disease risk. *Foxa2*-null mice have impaired insulin secretion[47], and common variants at the *FOXA2* locus are associated with fasting plasma glucose concentrations[42,57]. Our findings are thus consistent with the involvement of FOXA2 in maintaining normal glucose homeostasis. Common T2D-associated variants at *FOXA2* have also been reported in South Asians[58], although these variants did not attain genome-wide significant association in the largest GWAS for the disease from multiple ancestry groups[1–5] and therefore require further replication. Enrichment of FOXA2 binding has also been reported within genomic intervals containing GWAS signals for endocrine, neuropsychiatric, cardiovascular and cancer traits[59]. Our study has the advantage that we consider only the FOXA2 sites that directly overlap variants driving association signals by first fine mapping GWAS loci, thereby providing more targeted credible sets for functional enrichment. Nevertheless, the results of these studies, taken together, suggest a possible role for FOXA2 across a broad spectrum of complex human phenotypes.

In conclusion, we have highlighted that FOXA2 binding patterns can be used to inform future hypothesis-driven investigation of the variants, genes and molecular mechanisms underlying T2D association signals mapping to noncoding sequence. Continued identification of the effector transcripts at these noncoding association signals will require the use of expression quantitative trait locus (eQTL) data and knockout models, in combination with high-throughput experimental data derived from chromatin conformation capture techniques such as capture C. Our findings support the use of transcription factor binding events as a means to partition susceptibility loci, potentially residing in distinct pathways, within disease-relevant cell types. Finally, our study demonstrates the usefulness of fine mapping through integration of genetic and genomic information from relevant tissues and cellular models to elucidate the pathophysiology of complex human diseases, thus offering a promising avenue for the translation of GWAS findings into clinical use.

URLs. DIAGRAM Consortium, http://diagram-consortium.org/; Endocells, http://www.endocells.com/.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Association summary statistics are available for download from the DIAGRAM Consortium website (see URLs).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

Funding for the research undertaken in this study has been received from the Academy of Finland (including grants 77299, 102318, 10493, 118065, 123885, 124243, 129293, 129680, 136895, 139635, 211119, 213506, 251217 and 263836); Agence National de la Recherche; Association de Langue Française pour l'Etude du Diabète et des Maladies Métaboliques; Association Diabèete Risque Vasculaire; Association Française des Diabétiques; the Association of Danish Pharmacies; the Augustinus Foundation; the Becket Foundation; the British Diabetes Association (BDA) Research; the British Heart Foundation; the Central Norway Health Authority; the Central Finland Hospital District; the Center for Inherited Disease Research (CIDR); the City of Kuopio; the City of Leutkirch; Copenhagen County; the Danish Centre for Evaluation and Health Technology Assessment; the Danish Council for Independent Research; the Danish Heart Foundation; the Danish Research Councils; Deutsche Forschungsgemeinschaft (including project ER 155/6-2); the Diabetes Research Foundation; Diabetes UK; the Doris Duke Charitable Foundation; Erasmus Medical Center; Erasmus University; the Estonian government (SF0180142s08); the European Commission (including ENGAGE HEALTH-F4-2007-201413, FP7-201413, FP7-245536, EXGENESIS LSHM-CT-2004-005272, FP6 LSHM_CT_2006_037197, LSHM-CT-2007-037273, Directorate C-Public Health 2004310, DG XII); the European Regional Development Fund; the Federal Ministry of Education and Research, Germany (including FKZ 01GI1128 and FKZ 01EO1001); the Federal Ministry of Health, Germany; the Finnish Diabetes Association; the Finnish Diabetes Research Foundation; the Finnish Foundation for Cardiovascular Research; the Finnish Medical Society; the Folkhalsan Research Foundation; the Foundation for Life and Health in Finland; the Foundation for Old Servants; the Fredrick och Ingrid Thuring Foundation; the French region of Nord-Pas-de-Calais (Contrat de Projets Etat-Région); the German Center for Diabetes Research; the German Research Council (including grant GRK1041); the German National Genome Research Network; Groupe d'Etude des Maladies Métaboliques et Systémiques; the Health Care Centers in Vasa, Närpes and Korsholm, Finland; the Health Foundation; the Heinz Nixdorf Foundation; Helmholtz Zentrum München; the Helsinki University Central Hospital Research Foundation; the Hospital District of Southwest Finland; the Ib Henriksens Foundation; IngaBritt and Arne Lundberg's Research Foundation (including grant 359); Karolinska Institutet; the Knut and Alice Wallenberg Foundation (including grant KAW 2009.0243); Kuopio University Hospital; the Lundbeck Foundation; the Magnus Bergvall Foundation; the Medical Faculty of University Duisburg-Essen; the Medical Research Council, UK (including grants G0000649 and G0601261); the Ministry for Health, Welfare and Sports, the Netherlands; the Ministry of Education and Culture, Finland (including grants 722 and 627; 2004-2011); the Ministry of Education, Culture and Science, the Netherlands; the Ministry of Health and Prevention, Denmark; the Ministry of Social Affairs and Health, Finland; the Ministry of Innovation, Science, Research and Technology of North Rhine-Westphalia, Germany; the Munich Center of Health Sciences; the Municipal Health Care Center and Hospital in Jakobstad, Finland; the municipality of Rotterdam, the Netherlands; the Närpes Health Care Foundation; the National Health Screening Service of Norway; the National Heart, Lung, and Blood Institute, USA (including grant numbers/contracts HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C, HHSN268201100012C, N01HC25195, N02HL64278, R01HL087641, R01HL59367 and R01HL086694); the National Human Genome Research Institute, USA (including grant numbers/contracts U01HG004402 and N01HG65403); the National Institute for Diabetes and Digestive and Kidney Diseases, USA (including grants R01DK078616, U01DK085526, K24DK080140 and R01DK073490); the National Institute for Health and Welfare, Finland; the National Institutes of Health, USA (including grant numbers/contracts HHSN268200625226C, UL1RR025005, R01DK062370, R01DK072193, 1Z01HG000024, AG028555, AG08724, AG04563, AG10175, AG08861, U01HG004399, DK58845, CA055075, DK085545 and DK098032); the Netherlands Genomics Initiative; the Netherlands Organisation for Health Research and Development; the Netherlands Organisation of Scientific Research NOW Investments (including grants 175.010.2005.011, 911-03-012 and 050-060-810); the Nord-Trondelag County Council; the Nordic Center of Excellence in Disease Genetics; the Norwegian Institute of Public Health; the Norwegian Research Council; the Novo Nordisk Foundation; the Ollquist Foundation; the Oxford National Institute for Health Research (NIHR) Biomedical Research Centre; the Paavo Nurmi Foundation; the Paivikki and Sakari Sohlberg Foundation; the Perklen Foundation; the Pirkanmaa Hospital District, Finland; Programme Hospitalier de Recherche Clinique; Programme National de Recherche sur la Diabète; the Research Institute for Diseases in the Elderly (including grant 014-93-015); the Robert Dawson Evans Endowment, Department of Medicine, Boston University School of Medicine and Boston Medical Center; the Royal Swedish Academy of Sciences; Sarstedt, Germany; the Signe and Ane Gyllenberg Foundation; the Sigrid Juselius Foundation; the Slottery Machine Association, Finland; the Social Insurance Institution of Finland; the South OstroBothnia Hospital District; the state of Baden-Württemberg, Germany; the Stockholm County Council (including grant 560183); the Swedish Cultural Foundation, Finland; the Swedish Diabetes Foundation; the Swedish e-science Research Center; the Swedish Foundation for Strategic Research; the Swedish Heart-Lung Foundation; the Swedish Research Council (including grants SFO EXODIAB 2009-1039, 521-2010-3490, 521-2007-4037, 521-2008-2974, ANDIS 825-2010-5983, LUDC 349-2008-6589 and 8691); the Swedish Society of Medicine; the Tore

Nilsson Foundation; the Torsten and Ragnar Soderbergs Stiftelser (including grant MT33/09); University Hospital Essen; University of Tromsø; the University College London NIHR Biomedical Research Centre; the UK NIHR Cambridge Biomedical Research Centre; Uppsala University; Uppsala University Hospital; the Vaasa Hospital District; the Velux Foundation; and the Wellcome Trust (including the Biomedical Collections Grant GR072960 and grants 076113, 083948, 090367, 090532, 083270, 086596, 098017, 095101, 098051 and 098381). We are grateful to R. Scharfmann (INSERM U1016, Cochin Institute Paris) for the gift of EndoC βH1 cells and for providing technical support with their maintenance. We thank P. Johnson and the Oxford NIHR Biomedical Research Centre–funded Islet Isolation facility for providing human islets for this study. Detailed acknowledgments are provided in the **Supplementary Note.**

AUTHOR CONTRIBUTIONS

Writing group. K.J.G., T.F., Y. Lee, A.R., R.M., M.E.R., A.L.G., D.A., M. Boehnke, T.M.T., M.I.M., A.P.M. **Central meta-analysis group.** K.J.G., T.F., Y. Lee, R.M., A. Mahajan, A. Locke, N.W.R., N.R., T.M.T., M.I.M., A.P.M. **Annotation and functional analysis group.** K.J.G., A.R., M.E.R., S.K.T., J.K.R., N.L.B., M.v.d.B., A.C., I.D., E. Birney, L. Pasquali, J. Ferrer, C.A.O'C., A.L.G., M.I.M. **Validation meta-analysis group.** R.M., R.A.S., I.P., L.J.S., A.P.M. **Metabochip cohort-level primary analysis.** Y. Lee, T.G., T.S., D.T., L.Y., H.G., S. Wahl, M.F., R.J.S., H. Kestler, H. Chheda, L.E., S.G., T.M.T., A.P.M. **Validation cohort-level primary analysis.** V. Steinthorsdottir, G.T., L.Q., L.C.K., E.M.v.L., S.M.W., M. Li, H. Chen, C. Fuchsberger, P. Kwan, C.M., M. Linderman, Y. Lu. **Metabochip design.** H.M.K., B.F.V. **Cohort sample collection, genotyping, phenotyping or additional analysis.** B.F.V., G.R.A., P.A., D.B., B.B., R.B., M. Blüher, H.B., L.L.B., E.P.B., N.P.B., J.C., G.C., P.S.C., M.C.C., D.J.C., A.T.C., R.M.v.D., A.S.F.D., M.D., S.E., J.G.E., T.E., E.E., J. Fadista, J. Flannick, P. Fontanillas, C. Fox, P.W.F., K.G., C.G., B.G., O.G., G.B.G., N.G., C.J.G., M.H., C.T.H., C.H., O.L.H., A.B.H., S.E.H., D.J.H., A.U.J., A.J., M.E.J., T.J., W.-H.L.K., N.D.K., L.K., N.K., A.K., P. Kovacs, P. Kraft, J. Kravic, C. Langford, K.L., L. Liang, P.L., C.M.L., E.L., A. Linneberg, C.-T.L., S.L., J.L., V.L., S. Männistö, O. McLeod, J.M., E.M., G.M., T.W.M., M.M.-N., C.N., M.M.N., N.N.O., K.R.O., D.P., S.P., L. Peltonen, J.R.B.P., C.G.P.P., M.R., D. Ruderfer, D. Rybin, Y.T.v.d.S., B.S., G. Sigurðsson, A.S., G. Steinbach, P.S., K. Strauch, H.M.S., Q.S., B.T., E. Tikkanen, A.T., J. Trakalo, E. Tremoli, T.T., R.W., S. Wiltshire, A.R.W., E.Z. **Validation cohort principal investigators.** R.J.F.L., J.D., J.C.F., E. Boerwinkle, J.S.P., C.v.D., E.S., J.B.M., F.B.H., U.T., K. Stefansson, P.D., P.J.D., T.M.F., A.T.H., I.B., C. Langenberg, N.J.W., M. Boehnke, M.I.M. **Metabochip cohort principal investigators.** T.A.L., R.R., M.S., N.L.P., L. Lind, S.M.K.-K., E.K.-H., T.E.S., J.S., J. Kuusisto, M. Laakso, A. Metspalu, R.E., K.-H.J., S. Moebus, S.R., V. Salomaa, E.I., B.O.B., R.N.B., F.S.C., K.L.M., H. Koistinen, J. Tuomilehto, K.H., I.N., P.D., P.J.D., T.M.F., A.T.H., U.d.F., A.H., T.I., A.P., S.C., R.S., P. Froguel, O.P., T.H., A.D.M., C.N.A.P., S.K., O. Melander, P.M.N., L.C.G., I.B., C. Langenberg, N.J.W., D.A., M. Boehnke, M.I.M. **Project management.** K.J.G., A.L.G., D.A., M. Boehnke, T.M.T., M.I.M., A.P.M. **DIAGRAM Consortium management.** D.A., M. Boehnke, M.I.M.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

1. Kooner, J.S. *et al.* Genome-wide association study in individuals of South Asian ancestry identifies six new type 2 diabetes susceptibility loci. *Nat. Genet.* **43**, 984–989 (2011).
2. Cho, Y.S. *et al.* Meta-analysis of genome-wide association studies identifies eight new loci for type 2 diabetes in East Asians. *Nat. Genet.* **44**, 67–72 (2012).
3. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
4. Morris, A.P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44**, 981–990 (2012).
5. Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46**, 234–244 (2014).
6. Altshuler, D. *et al.* The common PPARγ Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat. Genet.* **26**, 76–80 (2000).
7. Gloyn, A.L. *et al.* Large-scale association studies of variants in genes encoding the pancreatic-cell K_{ATP} channel subunits Kir6.2 (*KCNJ11*) and SUR1 (*ABCC8*) conrm that the *KCNJ11* E23K variant is associated with type 2 diabetes. *Diabetes* **52**, 568–572 (2003).
8. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
9. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes. *Nat. Genet.* **42**, 105–116 (2010).
10. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
11. Elbein, S.C. *et al.* Genetic risk factors for type 2 diabetes: a *trans*-regulatory genetic architecture? *Am. J. Hum. Genet.* **91**, 466–477 (2012).
12. Trynka, G. *et al.* Chromatin marks identify critical cell types for fine-mapping complex trait variants. *Nat. Genet.* **45**, 124–130 (2013).
13. Parker, S.C.J. *et al.* Chromatin stretch enhancer states drive cell-specific gene regulation and harbour human disease risk variants. *Proc. Natl. Acad. Sci. USA* **110**, 17921–17926 (2013).
14. Pasquali, L. *et al.* Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
15. Voight, B.F. *et al.* The Metabochip, a custom genotyping array for genetic studies of metabolic, cardiovascular, and anthropometric traits. *PLoS Genet.* **8**, e1002793 (2012).
16. International HapMap Project Consortium. A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
17. 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
18. 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
19. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* **11**, 499–511 (2010).
20. Winkler, T.W. *et al.* Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
21. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **5**, e1000529 (2009).
22. Howie, B., Fuchsberger, C., Stephens, M., Marchini, J. & Abecasis, G.R. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* **44**, 955–959 (2012).
23. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).
24. 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).
25. 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).
26. 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).
27. Shea, J. *et al.* Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. *Nat. Genet.* **43**, 801–805 (2011).
28. Maller, J.B. *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
29. Jafar-Mohammadi, B. *et al.* A role for coding functional variants in *HNF4A* in type 2 diabetes susceptibility. *Diabetologia* **54**, 111–119 (2011).
30. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
31. Florez, J.C. *et al.* Haplotype structure and genotype-phenotype correlations of the sulfonylurea receptor and the islet ATP-sensitive potassium channel gene. *Diabetes* **53**, 1360–1368 (2004).
32. Hamming, K.S. *et al.* Co-expression of the type 2 diabetes susceptibility gene variants *KCNJ11* E23K and *ABCC8* S1369A alter the ATP and sulfonylurea sensitivities of the ATP-sensitive K^+ channel. *Diabetes* **58**, 2419–2424 (2009).
33. 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).
34. Beer, N.L. *et al.* The P446L variant in *GCKR* associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.* **18**, 4081–4088 (2009).
35. Holmkvist, J. *et al.* Common variants in HNF-1α and risk of type 2 diabetes. *Diabetologia* **49**, 2882–2891 (2006).
36. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-1α gene in maturity-onset diabetes of the young (MODY3). *Nature* **384**, 455–458 (1996).
37. Yamagata, K. *et al.* Mutations in the hepatocyte nuclear factor-4α gene in maturity-onset diabetes of the young (MODY1). *Nature* **384**, 458–460 (1996).
38. Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *Am. J. Hum. Genet.* **95**, 535–552 (2014).
39. Soccio, R.E. *et al.* Species-specific strategies underlying conserved functions of metabolic transcription factors. *Mol. Endocrinol.* **25**, 694–706 (2011).
40. Gaulton, K.J. *et al.* A map of open chromatin in human pancreatic islets. *Nat. Genet.* **42**, 255–259 (2010).
41. Fogarty, M.P., Cannon, M.E., Vadlamudi, S., Gaulton, K.J. & Mohlke, K.L. Identification of a regulatory variant that binds FOXA1 and FOXA2 at the *CDC123/CAMK1D* type 2 diabetes GWAS locus. *PLoS Genet.* **10**, e1004633 (2014).
42. Manning, A.K. *et al.* A genome-wide approach accounting for body-mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44**, 659–669 (2012).
43. Dimas, A.S. *et al.* Impact of type 2 diabetes susceptibility variants on quantitative glycemic traits reveals mechanistic heterogeneity. *Diabetes* **63**, 2158–2171 (2014).

44. Ravassard, P. *et al.* A genetically engineered human pancreatic β cell line exhibiting glucose-inducible insulin secretion. *J. Clin. Invest.* **121**, 3589–3597 (2011).
45. Fadista, J. *et al.* Global genomic and transcriptomic analysis of human pancreatic islets reveals novel genes influencing glucose metabolism. *Proc. Natl. Acad. Sci. USA* **111**, 13924–13929 (2014).
46. 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).
47. Gao, N. *et al.* Foxa1 and Foxa2 maintain the metabolic and secretory features of the mature β-cell. *Mol. Endocrinol.* **24**, 1594–1604 (2010).
48. Zhou, Y. *et al. TCF7L2* is a master regulator of insulin production and processing. *Hum. Mol. Genet.* **23**, 6419–6431 (2014).
49. Lyssenko, V. *et al.* Mechanisms by which common variants in the *TCF7L2* gene increase risk of type 2 diabetes. *J. Clin. Invest.* **117**, 2155–2163 (2007).
50. Gloyn, A.L. *et al.* Activating mutations in the gene encoding the ATP-sensitive potassium-channel subunit Kir6.2 and permanent neonatal diabetes. *N. Engl. J. Med.* **350**, 1838–1849 (2004).
51. Flannick, J. *et al.* Loss-of-function mutations in *SLC30A8* protect against type 2 diabetes. *Nat. Genet.* **46**, 357–363 (2014).
52. Dickson, S.P., Wang, K., Krantz, I., Hakonarson, H. & Goldstein, D.B. Rare variants create synthetic genome-wide associations. *PLoS Biol.* **8**, e1000294 (2010).
53. Bonnefond, A. *et al.* Rare *MTNR1B* variants impairing melatonin receptor 1B function contribute to type 2 diabetes. *Nat. Genet.* **44**, 297–301 (2012).
54. Zaret, K.S. & Carroll, J.S. Pioneer transcription factors: establishing a competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
55. Gao, N. *et al.* Dynamic regulation of *Pdx1* enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev.* **22**, 3435–3448 (2008).
56. Lee, C.S., Friedman, J.R., Fulmer, J.T. & Kaestner, K.H. The initiation of liver development is dependent on Foxa transcription factors. *Nature* **435**, 944–947 (2005).
57. Scott, R.A. *et al.* Large-scale association analyses identify new loci influencing glycaemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44**, 991–1005 (2012).
58. Tabassum, R., Chavali, S., Dwivedi, O.P., Tandon, N. & Bharadwaj, D. Genetic variants of *FOXA2*: risk of type 2 diabetes and effect on metabolic traits in North Indians. *J. Hum. Genet.* **53**, 957–965 (2008).
59. Johnson, M.E., Schug, J., Wells, A.D., Kaestner, K.H. & Grant, S.F. Genome-wide analyses of ChIP-Seq derived FOXA2 DNA occupancy in liver points to genetic networks underpinning multiple complex traits. *J. Clin. Endocrinol. Metab.* **99**, E1580–E1585 (2014).

Kyle J Gaulton[1,2,161], Teresa Ferreira[1,161], Yeji Lee[3,161], Anne Raimondo[4,161], Reedik Mägi[5,161], Michael E Reschen[6,161], Anubha Mahajan[1], Adam Locke[3], N William Rayner[1,4,7], Neil Robertson[1,4], Robert A Scott[8], Inga Prokopenko[9], Laura J Scott[3], Todd Green[10], Thomas Sparso[11], Dorothee Thuillier[12], Loic Yengo[12], Harald Grallert[13–15], Simone Wahl[13–15], Mattias Frånberg[16–18], Rona J Strawbridge[16], Hans Kestler[19,20], Himanshu Chheda[21], Lewin Eisele[22], Stefan Gustafsson[23], Valgerdur Steinthorsdottir[24], Gudmar Thorleifsson[24], Lu Qi[25–28], Lennart C Karssen[29], Elisabeth M van Leeuwen[29], Sara M Willems[8,29], Man Li[30], Han Chen[31,32], Christian Fuchsberger[3], Phoenix Kwan[3], Clement Ma[3], Michael Linderman[33], Yingchang Lu[34], Soren K Thomsen[4], Jana K Rundle[4], Nicola L Beer[1,4], Martijn van de Bunt[1,4], Anil Chalisey[6], Hyun Min Kang[3], Benjamin F Voight[35], Gonçalo R Abecasis[3], Peter Almgren[36], Damiano Baldassarre[37,38], Beverley Balkau[39,40], Rafn Benediktsson[41,42], Matthias Blüher[43,44], Heiner Boeing[45], Lori L Bonnycastle[46], Erwin P Bottinger[47], Noël P Burtt[10], Jason Carey[10], Guillaume Charpentier[48], Peter S Chines[46], Marilyn C Cornelis[49], David J Couper[50], Andrew T Crenshaw[10], Rob M van Dam[26,51], Alex S F Doney[52,53], Mozhgan Dorkhan[54], Sarah Edkins[7], Johan G Eriksson[55–58], Tonu Esko[5,59,60], Elodie Eury[61], João Fadista[36], Jason Flannick[10], Pierre Fontanillas[10], Caroline Fox[62,63], Paul W Franks[26,36,64,65], Karl Gertow[16], Christian Gieger[13,14], Bruna Gigante[66], Omri Gottesman[47], George B Grant[10], Niels Grarup[11], Christopher J Groves[4], Maija Hassinen[67], Christian T Have[11], Christian Herder[68,69], Oddgeir L Holmen[70], Astradur B Hreidarsson[42], Steve E Humphries[71], David J Hunter[25–27,72], Anne U Jackson[3], Anna Jonsson[36], Marit E Jørgensen[73], Torben Jørgensen[74–76], Wen-Hong L Kao[30,160], Nicola D Kerrison[8], Leena Kinnunen[55], Norman Klopp[13,77], Augustine Kong[24], Peter Kovacs[43,44], Peter Kraft[25,32,72], Jasmina Kravic[36], Cordelia Langford[7], Karin Leander[66], Liming Liang[25,32], Peter Lichtner[78], Cecilia M Lindgren[1,10], Eero Lindholm[36], Allan Linneberg[74,79,80], Ching-Ti Liu[31], Stéphane Lobbens[61], Jian'an Luan[8], Valeriya Lyssenko[36,73], Satu Männistö[55], Olga McLeod[16], Julia Meyer[81], Evelin Mihailov[5], Ghazala Mirza[82], Thomas W Mühleisen[83–85], Martina Müller-Nurasyid[81,86–88], Carmen Navarro[89–91], Markus M Nöthen[83,84], Nikolay N Oskolkov[36], Katharine R Owen[4,92], Domenico Palli[93], Sonali Pechlivanis[22], Leena Peltonen[7,10,21,55,160], John R B Perry[8], Carl G P Platou[70,94], Michael Roden[68,69,95], Douglas Ruderfer[96], Denis Rybin[97], Yvonne T van der Schouw[98], Bengt Sennblad[16,17], Gunnar Sigurðsson[42,99], Alena Stančáková[100], Gerald Steinbach[101], Petter Storm[36], Konstantin Strauch[81,87], Heather M Stringham[3], Qi Sun[26,27], Barbara Thorand[14,15], Emmi Tikkanen[21,102], Anke Tonjes[43,44], Joseph Trakalo[1], Elena Tremoli[37,38], Tiinamaija Tuomi[21,58,103,104], Roman Wennauer[105], Steven Wiltshire[1,160], Andrew R Wood[106], Eleftheria Zeggini[7], Ian Dunham[107], Ewan Birney[107], Lorenzo Pasquali[108–110], Jorge Ferrer[111,112], Ruth J F Loos[8,34,47,113], Josée Dupuis[31,62], Jose C Florez[60,114–116], Eric Boerwinkle[117,118], James S Pankow[119], Cornelia van Duijn[29,120], Eric Sijbrands[105], James B Meigs[114,121], Frank B Hu[25–27], Unnur Thorsteinsdottir[24,41], Kari Stefansson[24,41], Timo A Lakka[67,122,123], Rainer Rauramaa[67,123], Michael Stumvoll[43,44], Nancy L Pedersen[124], Lars Lind[125], Sirkka M Keinanen-Kiukaanniemi[126,127], Eeva Korpi-Hyövälti[128], Timo E Saaristo[129,130], Juha Saltevo[131], Johanna Kuusisto[100], Markku Laakso[100], Andres Metspalu[5,132], Raimund Erbel[133], Karl-Heinz Jöckel[22], Susanne Moebus[22], Samuli Ripatti[7,21,102,134], Veikko Salomaa[55], Erik Ingelsson[1,23], Bernhard O Boehm[135,136], Richard N Bergman[137], Francis S Collins[46], Karen L Mohlke[138], Heikki Koistinen[55,139,140], Jaakko Tuomilehto[55,141–143], Kristian Hveem[70], Inger Njølstad[144], Panagiotis Deloukas[7,145], Peter J Donnelly[1,146],

Timothy M Frayling[106], Andrew T Hattersley[147], Ulf de Faire[66], Anders Hamsten[16], Thomas Illig[13,77], Annette Peters[14,15,88], Stephane Cauchi[12], Rob Sladek[148,149], Philippe Froguel[9,12,61], Torben Hansen[11,150], Oluf Pedersen[11], Andrew D Morris[151], Collin N A Palmer[52,53], Sekar Kathiresan[10,115,152], Olle Melander[36], Peter M Nilsson[36], Leif C Groop[21,36], Inês Barroso[7,153,154], Claudia Langenberg[8], Nicholas J Wareham[8], Christopher A O'Callaghan[6], Anna L Gloyn[1,4,92,162], David Altshuler[10,114–116,155,156,162], Michael Boehnke[3,162], Tanya M Teslovich[3,162], Mark I McCarthy[1,4,92,162] & Andrew P Morris[1,5,157,158,162] for the DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium[159]

[1]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [2]Department of Genetics, Stanford University, Stanford, California, USA. [3]Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA. [4]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [5]Estonian Genome Center, University of Tartu, Tartu, Estonia. [6]Centre for Cellular and Molecular Physiology, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [7]Wellcome Trust Sanger Institute, Hinxton, UK. [8]Medical Research Council (MRC) Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge, UK. [9]Genomics of Common Disease, Imperial College London, London, UK. [10]Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [11]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [12]Lille Institute of Biology, European Genomics Institute of Diabetes, Lille, France. [13]Research Unit of Molecular Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [14]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [15]German Center for Diabetes Research, Neuherberg, Germany. [16]Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. [17]Science for Life Laboratory, Stockholm, Sweden. [18]Department for Numerical Analysis and Computer Science, Stockholm University, Stockholm, Sweden. [19]Leibniz Institute for Age Research, Fritz Lipmann Institute, Jena, Germany. [20]Medical Systems Biology, Ulm University, Ulm, Germany. [21]Finnish Institute for Molecular Medicine (FIMM), Helsinki, Finland. [22]Institute for Medical Informatics, Biometry and Epidemiology, University Hospital of Essen, Essen, Germany. [23]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [24]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [25]Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA. [26]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [27]Channing Division of Network Medicine, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [28]Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, Louisiana, USA. [29]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, the Netherlands. [30]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, USA. [31]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [32]Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts, USA. [33]Icahn Institute for Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [34]Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [35]Department of Pharmacology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [36]Lund University Diabetes Centre, Department of Clinical Science Malmo, Scania University Hospital, Lund University, Malmo, Sweden. [37]Centro Cardiologico Monzino, Istituto di Ricovero e Cura a Carattere Scientifico (IRCCS), Milan, Italy. [38]Dipartimento di Scienze Farmacologiche e Biomolecolari, Università di Milano, Milan, Italy. [39]INSERM Centre de Recherche Epidémiologie et Santé des Populations (CESP) U1018, Villejuif, France. [40]University Paris Sud 11, UMRS 1018, Villejuif, France. [41]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [42]Landspitali University Hospital, Reykjavik, Iceland. [43]Integrated Treatment and Research (IFB) Center for Adiposity Diseases, University of Leipzig, Leipzig, Germany. [44]Department of Medicine, University of Leipzig, Leipzig, Germany. [45]German Institute of Human Nutrition, Potsdam-Rehbruecke, Germany. [46]National Human Genome Research Institute, US National Institutes of Health, Bethesda, Maryland, USA. [47]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [48]Endocrinology-Diabetology Unit, Corbeil-Essonnes Hospital, Corbeil-Essonnes, France. [49]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA. [50]Collaborative Studies Coordinating Center, Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [51]Saw Swee Hock School of Public Health, National University of Singapore, Singapore. [52]Diabetes Research Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK. [53]Pharmacogenomics Centre, Biomedical Research Institute, University of Dundee, Ninewells Hospital, Dundee, UK. [54]Lund University Diabetes Centre, Department of Clinical Science Malmo, Novo Nordisk Scandinavia, Malmo, Sweden. [55]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [56]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [57]Unit of General Practice, Helsinki University General Hospital, Helsinki, Finland. [58]Folkhalsan Research Center, Helsinki, Finland. [59]Division of Endocrinology, Children's Hospital, Boston, Massachusetts, USA. [60]Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. [61]CNRS UMR 8199, Institute of Biology and Lille 2 University, Pasteur Institute, Lille, France. [62]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [63]Division of Endocrinology and Metabolism, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [64]Department of Clinical Sciences, Lund University, Malmo, Sweden. [65]Department of Public Health and Clinical Medicine, Umea University, Umea, Sweden. [66]Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [67]Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [68]Institute for Clinical Diabetology, German Diabetes Center, Leibniz Center for Diabetes Research at Heinrich Heine University Dusseldorf, Dusseldorf, Germany. [69]German Center for Diabetes Research, partner site Dusseldorf, Dusseldorf, Germany. [70]Nord-Trøndelag Health Study (HUNT) Research Center, Department of Public Health and General Practice, Norwegian University of Science and Technology, Levanger, Norway. [71]Cardiovascular Genetics, British Heart Foundation (BHF) Laboratories, Institute of Cardiovascular Sciences, University College London, London, UK. [72]Program in Genetic Epidemiology and Statistical Genetics, Harvard School of Public Health, Boston, Massachusetts, USA. [73]Steno Diabetes Center, Gentofte, Denmark. [74]Research Centre for Prevention and Health, Capital Region of Denmark, Copenhagen, Denmark. [75]Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [76]Faculty of Medicine, University of Aalborg, Aalborg, Denmark. [77]Hannover Unified Biobank, Hannover Medical School, Hannover, Germany. [78]Institute of Human Genetics, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [79]Copenhagen University Hospital, Rigshospitalet, Copenhagen, Denmark. [80]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [81]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [82]Biomedical Research Centre Genomics Core Facility, Guy's and St Thomas' National Health Service (NHS) Foundation Trust, Guy's & St Thomas' Hospital, London, UK. [83]Institute of Human Genetics, University of Bonn, Bonn, Germany. [84]Department of Genomics, Life and Brain Center, University of Bonn, Bonn, Germany. [85]Institute of Neuroscience and Medicine (INM-1), Research Centre Jülich, Jülich, Germany. [86]Department of Medicine I, University Hospital Grosshadern, Ludwig Maximilians Universität, Munich, Germany. [87]Institute of Medical Informatics, Biometry and Epidemiology, Chair of Genetic Epidemiology, Ludwig Maximilians Universität, Neuherberg, Germany. [88]DZHK (German Center for Cardiovascular Research), partner site Munich Heart Alliance, Munich, Germany. [89]Department of Epidemiology, Murcia Regional Health Council, Instituto Murciano de Investigación Biosanitaria Virgen de la Arrixaca (IMIB-Arrixaca), Murcia, Spain. [90]Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [91]Department of Health and Social Sciences, Universidad de Murcia, Murcia, Spain. [92]Oxford National Institute for Health Research Biomedical Research Centre, Churchill Hospital, Oxford, UK. [93]Cancer Research and Prevention Institute (ISPO), Florence, Italy. [94]Department of Internal Medicine, Levanger Hospital, Nord-Trondelag Health Trust, Levanger, Norway. [95]Department of Endocrinology and Diabetology, University Hospital Dusseldorf, Dusseldorf, Germany. [96]Division of Psychiatric Genomics, Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [97]Boston University Data Coordinating Center, Boston, Massachusetts, USA. [98]University Medical Center Utrecht, Utrecht, the Netherlands. [99]Icelandic Heart Association, Kopavogur, Iceland. [100]Department of Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [101]Department of Clinical Chemistry and Central Laboratory, University of Ulm, Ulm, Germany. [102]Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland. [103]Department of Endocrinology, Abdominal Center, Helsinki University Hospital, Helsinki, Finland. [104]Research Program for Diabetes and Obesity, University of Helsinki, Helsinki, Finland. [105]Department of Internal

Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. [106]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [107]European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Hinxton, UK. [108]Division of Endocrinology, Germans Trias i Pujol University Hospital and Research Institute, Badalona, Spain. [109]Josep Carreras Leukaemia Research Institute, Badalona, Spain. [110]CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Barcelona, Spain. [111]Department of Medicine, Imperial College London, London, UK. [112]Institut d'Investigacions Biomèdiques August Pi i Sunyer, Centre Esther Koplowitz, Barcelona, Spain. [113]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [114]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [115]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [116]Diabetes Research Center, Diabetes Unit, Massachusetts General Hospital, Boston, Massachusetts, USA. [117]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [118]Human Genome Sequencing Center at Baylor College of Medicine, Houston, Texas, USA. [119]Division of Epidemiology and Community Health, University of Minnesota, Minneapolis, Minnesota, USA. [120]Netherlands Genomics Initiative, Netherlands Consortium for Healthy Ageing and Center for Medical Systems Biology, Rotterdam, the Netherlands. [121]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA. [122]Institute of Biomedicine/Physiology, University of Eastern Finland, Kuopio, Finland. [123]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [124]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [125]Department of Medical Sciences, Uppsala University Hospital, Uppsala, Sweden. [126]Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland. [127]Unit of General Practice, Oulu University Hospital, Oulu, Finland. [128]South Ostrobothnia Central Hospital, Seinajoki, Finland. [129]Finnish Diabetes Association, Tampere, Finland. [130]Pirkanmaa District Hospital, Tampere, Finland. [131]Department of Medicine, Central Finland Central Hospital, Jyvasklya, Finland. [132]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [133]Clinic of Cardiology, West German Heart Centre, University Hospital of Essen, University Duisdurg-Essen, Essen, Germany. [134]Public Health Genomics Unit, National Institute for Health and Welfare, Helsinki, Finland. [135]Division of Endocrinology and Diabetes, Department of Internal Medicine, University Medical Centre Ulm, Ulm, Germany. [136]Lee Kong Chian School of Medicine, Imperial College London and Nanyang Technological University, Singapore. [137]Diabetes and Obesity Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [138]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA. [139]Division of Endocrinology, Department of Medicine, Helsinki University Central Hospital, Helsinki, Finland. [140]Minerva Foundation Institute for Medical Research, Helsinki, Finland. [141]Instituto de Investigación Sanitaria del Hospital Universitario La Paz, Madrid, Spain. [142]Centre for Vascular Prevention, Danube University Krems, Krems, Austria. [143]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [144]Department of Community Medicine, Faculty of Health Sciences, University of Tromso, Tromso, Norway. [145]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [146]Department of Statistics, University of Oxford, Oxford, UK. [147]Institute of Biomedical and Clinical Science, University of Exeter Medical School, Exeter, UK. [148]Montreal Diabetes Research Center, Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, Quebec, Canada. [149]McGill University and Génome Québec Innovation Centre, Montreal, Quebec, Canada. [150]Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark. [151]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [152]Cardiovascular Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [153]University of Cambridge Metabolic Research Laboratories, Wellcome Trust–MRC Institute of Metabolic Science, Cambridge, UK. [154]National Institute for Health Research Cambridge Biomedical Research Centre, Cambridge, UK. [155]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [156]Department of Molecular Biology, Harvard Medical School, Boston, Massachusetts, USA. [157]Department of Biostatistics, University of Liverpool, Liverpool, UK. [158]Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK. [159]A list of members and affiliations appears in the **Supplementary Note**. [160]Deceased. [161]These authors contributed equally to this work. [162]These authors jointly supervised this work. Correspondence should be addressed to K.J.G. (kgaulton@gmail.com), M.I.M. (mark.mccarthy@drl.ox.ac.uk) or A.P.M. (apmorris@liverpool.ac.uk).

ONLINE METHODS

Ethics statement. All human research was approved by the relevant institutional review boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent.

Metabochip imputation and association analysis. We considered a total of 27,206 T2D cases and 57,574 controls from 23 studies from populations of European ancestry (**Supplementary Table 1**), all genotyped with the Metabochip. Sample and variant quality control was performed within each study (**Supplementary Table 2**). To improve the quality of the genotype scaffold in each study, variants were subsequently removed if (i) allele frequencies differed from those for European-ancestry haplotypes from the 1000 Genomes Project Consortium phase 1 integrated reference panel (March 2012 release)[18] by more than 20%; AT/GC variants had MAF >40% because of potential undetected errors in strand alignment; or (iii) MAF was <1%, because of difficulties in calling rare variants. Each scaffold was then imputed up to the phase 1 integrated reference panel (all ancestries; March 2012 release) from the 1000 Genomes Project Consortium[18], using IMPUTEv2 (ref. 21) or minimac[22]. Within each study, well-imputed variants (IMPUTEv2 info >0.4 or minimac r^2 >0.3) were tested for T2D association under an additive model after adjustment for study-specific covariates (**Supplementary Table 2**), including principal components to adjust for population structure. Association summary statistics for each variant for each study were corrected for residual population structure using the genomic control inflation factor[60] obtained from 3,598 independent (r^2 <0.05) QT-interval variants, which were not expected to be associated with T2D[4] (**Supplementary Table 2**). We then combined association summary statistics for each variant across studies via fixed-effects inverse variance–weighted meta-analysis. The results of the meta-analysis were subsequently corrected by a second round of QT-interval genomic control (λ_{QT} = 1.18) to account for structure between studies. Variants were excluded from downstream analyses if they were reported in less than 80% of the total effective sample size, defined as $N_e = 4 \times N_{cases} \times N_{controls}/(N_{cases} + N_{controls})$, thus removing those that were not well imputed in the majority of studies.

Identification of distinct association signals in established GWAS loci. We used GCTA[23] to select index variants in each of the 39 established loci represented on Metabochip with nominal evidence of association ($P_J < 0.001$) with T2D in an approximate joint regression model. The GCTA model made use of (i) summary statistics from the fixed-effects meta-analysis Metabochip studies and (ii) genotype data for 3,298 T2D cases and 3,708 controls of UK ancestry from GoDARTS as a reference for LD across each fine-mapping region. For comparison, we also obtained association summary statistics for the selected index variants from the GCTA joint regression model on the basis of genotype data from an alternative reference consisting of 4,435 T2D cases and 5,757 controls of Finnish ancestry from FUSION (**Supplementary Fig. 12** and **Supplementary Table 15**). Selected index variants were then carried forward for *in silico* follow-up in validation meta-analysis.

The validation meta-analysis consisted of 19,662 T2D cases and 115,140 controls from ten GWAS from populations of European ancestry, genotyped with a range of genome-wide arrays (**Supplementary Table 1**). Sample and variant quality control was performed within each study (**Supplementary Table 2**). Each scaffold was then imputed up to the phase 1 integrated reference panel (all ancestries; March 2012 release) from the 1000 Genomes Project Consortium[18], using IMPUTEv2 or minimac. Within each study, well-imputed variants (IMPUTEv2 info ≥0.4 or minimac r^2 ≥0.3) were tested for T2D association under an additive model after adjustment for study-specific covariates (**Supplementary Table 2**), including principal components to adjust for population structure. Association summary statistics for each variant for each study were corrected for residual population structure using the genomic control inflation factor[60] (**Supplementary Table 2**). We then combined association summary statistics for each variant across studies via fixed-effects inverse variance–weighted meta-analysis.

Association summary statistics for the selected index variants from the Metabochip and validation meta-analyses were next combined via fixed-effects inverse variance–weighted meta-analysis. In each of the 39 established loci represented on the Metabochip, GCTA[23] was used to select index variants with locus-wide evidence of association ($P_J < 1 \times 10^{-5}$) in the approximate joint regression model on the basis of (i) summary statistics from the combined meta-analysis and (ii) genotype data for 3,298 T2D cases and 3,708 controls from GoDARTS as a reference for LD across in each fine-mapping region.

For established loci with multiple index variants selected at locus-wide significance from the GCTA approximate joint regression model in combined meta-analysis, we performed exact conditioning within each Metabochip study (**Supplementary Table 7**). To obtain the association signal attributed to a specific index variant, high-quality variants (IMPUTEv2 info >0.4 or minimac r^2 >0.3) were tested for T2D association under an additive model after adjustment for study-specific covariates (**Supplementary Table 2**) and genotypes at other selected index variants in the fine-mapping region. Association summary statistics for each study were corrected for residual population structure using the QT-interval genomic control inflation factor obtained in the Metabochip meta-analysis. For each association signal, summary statistics for each variant were then combined across discovery studies via fixed-effects inverse variance meta-analysis and subsequently corrected by a second round of QT-interval genomic control (λ_{QT} = 1.18).

Credible set construction. In an ideal fine-mapping experiment, we would calculate the posterior probability of driving each distinct association signal for all variants mapping to a locus. However, the posterior probability is determined by the association signal effect size of the variant and the corresponding standard error, which is also affected by the quality of imputation across studies, among other factors. To minimize the impact of imputation quality on fine mapping, we therefore retained only variants that were directly typed and/or well imputed in at least 80% of the total effective sample size. Assuming that the variant driving an association signal meets these quality criteria, the probability that it would be contained within the 99% credible set would be ~0.99.

For each distinct signal, we first calculated the posterior probability, π_{Cj}, that the jth variant is driving the association, given by

$$\pi_{Cj} = \frac{\Lambda_j}{\sum_k \Lambda_k}$$

where the summation is over all retained variants in the fine-mapping region. In this expression, Λ_j is the approximate Bayes factor[61] for the jth variant, given by

$$\Lambda_j = \sqrt{\frac{V_j}{V_j + \omega}} \exp\left[\frac{\omega \beta_j^2}{2V_j\left(V_j + \omega\right)}\right]$$

where β_j and V_j denote the estimated allelic effect (log(OR)) and corresponding variance from the meta-analysis across Metabochip studies. In loci with multiple distinct signals of association, results are presented from exact conditional meta-analysis after adjusting for all other index variants in the fine-mapping region. In loci with a single association signal, results are presented from unconditional meta-analysis. The parameter ω denotes the prior variance in allelic effects, taken here to be 0.04 (ref. 61). The 99% credible set[29] for each signal was then constructed by (i) ranking all variants according to their Bayes factor Λ_j and (ii) including ranked variants until their cumulative posterior probability of driving the association attained or exceeded 0.99.

Genomic annotation data and enrichment analyses. We obtained genomic annotation data for transcription factor binding sites assayed through ChIP experiments from multiple sources. We used sites from the ENCODE Project Consortium[10] for 161 proteins available from the UCSC human genome browser. We also obtained raw ChIP and input sequence data for additional factors assayed in primary pancreatic islets[14]. We then processed these additional factors using protocols employed by the ENCODE Project Consortium[10]. First, sequence reads were aligned to the human genome (hg19) using the Burrows-Wheeler Aligner (BWA)[62] with sex-specific references and were then converted to BAM files using SAMtools[63] after removing duplicate reads and those not uniquely mapped. Binding sites were called from reads of each replicate, as well as reads pooled across all replicates, using SPP[64]. The raw sites from each replicate of a protein were compared using an irreproducible

doi:10.1038/ng.3437

discovery rate[65] (IDR) threshold of 0.02. The resulting number of sites passing this IDR threshold was then used to filter the pooled sites of a protein. The set of sites was further filtered for artifacts using a blacklist of genomic regions from the ENCODE Project Consortium. Sites from all sources for each protein, including ENCODE, were then combined. The complete set of 165 proteins employed in these analyses is presented in **Supplementary Table 11**. In addition, we obtained FOXA2 ChIP-seq sites that were previously identified in human liver[39] and converted their positions to hg19 coordinates.

We obtained annotation data for five histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K36me3 and H3K27me3) and CTCF binding assayed from ChIP experiments. We used data from nine cell types from ENCODE[10] (GM12878, K562, HepG2, HSMM, HUVEC, NHEK, NHLF, H1-hESC and HMEC); we also obtained raw ChIP data assayed in primary pancreatic islets[14] and premature and mature human adipose stromal cells[66]. We mapped reads to the hg19 reference genome using BWA[62] and used the resulting mapped reads from these 12 cell types as input to ChromHMM[67]. We assigned states on the basis of the following chromatin signatures: active promoter (H3K4me3 and H3K27ac); strong enhancer 1 (H3K4me3, H3K27ac and H3K4me1); strong enhancer 2 (H3K27ac and H3K4me1); weak enhancer (H3K4me1); poised promoter (H3K27me3, H3K4me3 and H3K4me1); repressed (H3K27me3); insulator (CTCF); and transcription (H3K36me3). For each cell type, we pooled the three enhancer states into one enhancer category and the two promoter states into one promoter category. We also identified long noncoding RNA data from the Human Body Map (UCSC Genome Browser) and from pancreatic islets[68].

For each genomic annotation, we tested for overall enrichment of the posterior probability that overlapping variants in the 99% credible sets are driving distinct association signals (π_C). We first calculated the mean posterior probability (mean π_C) over the set of variants overlapping a given annotation. We then generated a null distribution of the mean posterior probability (mean π_C) by (i) shifting the genomic locations of binding sites a random distance within 100 kb in either direction; (ii) recalculating the mean posterior probability for 99% credible set variants overlapping shifted sites; and (iii) repeating this procedure 100,000 times. We estimated the fold enrichment of each overlap by calculating the expected null posterior probability and dividing the observed probability by the expected probability. We calculated a P value for the enrichment by the proportion of permutations for which the expected posterior probability of driving the association signal was greater than or equal to that observed. We considered cell type annotations to be significantly enriched if the P value was less than 0.05/37 = 0.0014 (Bonferroni correction for 37 annotations). We considered transcription factor binding site annotations to be significantly enriched if the P value was less than 0.05/165 = 0.00030 (Bonferroni correction for 165 factors). We next partitioned binding sites into those that are shared with another factor (where the genomic interval intersects a site for at least one other factor) and those that are unique. We also partitioned binding sites on the basis of overlap with islet enhancer elements. For each factor with significant enrichment across all credible sets (FOXA2), we applied the same enrichment analysis but restricted to credible set variants for each distinct association signal, separately.

We assessed the evidence for intersection in FOXA2 ChIP-seq sites from islets[14], HepG2 cells[10] and liver[39], across the genome and overlapping credible set variants, using BEDtools[69].

Motif analysis. We conducted recognition motif enhancement analyses for the set of FOXA2 ChIP-seq binding sites. First, we obtained the repeat-masked genomic sequence underlying each site using the UCSC human genome browser. We scanned sequences for enrichment in these motifs using MEME-ChIP[70], which uses up to 100 bp surrounding the midpoint of each site. This resulted in 198 enriched motifs with an E value (expected number of hits) less than 0.05 (**Supplementary Table 16**). We compared each motif to those known from JASPAR[71], ENCODE[10] and Homer[72] using Tomtom[73].

Second, we identified variants in FOXA2 ChIP-seq sites predicted to disrupt an enriched recognition motif by (i) scanning the 25 bp of sequence flanking each variant allele using FIMO[74] ($P < 0.0001$) and (ii) retaining variants in highly conserved positions (entropy less than 0.5). For the 12 variants at FOXA2-enriched signals disrupting at least one recognition motif (**Table 3** and **Supplementary Table 14**), we compared their posterior probabilities of driving the association (π_C) with those for non-disrupting variants in FOXA2 ChIP-seq sites at the same signals using a two-sided Wilcoxon rank-sum test.

Electrophoretic mobility shift assays. EMSA was performed using nuclear extracts from human HepG2 and EndoC-βH1 cells. HepG2 cells were the generous gift of the Ratcliffe laboratory[75] and were authenticated by genotyping in the major histocompatibility complex (MHC) region. Endo-βH1 cells were obtained from Endocells and have been previously authenticated[44]. Both cell lines were tested and found negative for mycoplasma contamination. Nuclear extracts were incubated with [γ-^{32}P]ATP end-labeled double-stranded DNA probes (PerkinElmer). The forward-strand probe sequences used are presented in **Supplementary Table 17**.

For each lane of the EMSA, 5 μg of nuclear extract was incubated with 100 fmol labeled probes in a 10-μl binding reaction containing 10 mM Tris-HCl, pH 7.5, 4% glycerol, 1 mM $MgCl_2$, 0.5 mM EDTA, 0.5 mM DTT, 50 mM NaCl and 1 μg poly(dI-dC). For competition assays, unlabeled probe at 100-fold excess was added to the binding reaction before addition of labeled probes. For supershift assays, the nuclear extract was preincubated with 1 μg of antibody for 30 min on ice before the probe was added. The following antibodies were used: anti-NEUROD1 (sc-1084X, Santa Cruz Biotechnology), anti-PTF1A (sc-98612X, Santa Cruz Biotechnology), anti-HNF3B (FOXA2) (sc-6554X, Santa Cruz Biotechnology), anti-YY1 (sc-281X, Santa Cruz Biotechnology), anti-TAL1 (sc-12984X, Santa Cruz Biotechnology), normal rabbit immunoglobulin (sc-2027, Santa Cruz Biotechnology) and normal goat immunoglobulin (sc-2028, Santa Cruz Biotechnology).

Luciferase activity. We synthesized 224-bp sequences containing either the risk or protective allele of the *MTNR1B* enhancer sequence at rs10830963 in either the forward or reverse orientation by GeneArt (Life Technologies). Complementary single-stranded oligonucleotides were then annealed and subcloned into the minimal promoter–driven luciferase vector pGL4.23 (Promega) using the NheI and XhoI restriction sites. Isolated clones were verified by sequencing.

For luciferase assays, HepG2 human liver cells and EndoC-βH1 human β cells[45] were counted and seeded into 24-well trays (Corning) at 1.5×10^5 (HepG2) or 1.4×10^5 (EndoC-βH1) cells/well. Transfections were performed in triplicate with either Lipofectamine 2000 (HepG2) or FuGENE 6 (EndoC-βH1) according to the manufacturer's instructions. Cells were transfected with 700 ng of pGL4.23 DNA containing the protective or risk *MTNR1B* enhancer sequence in either the forward or reverse orientation or with an equivalent amount of empty vector DNA, plus 10 ng of pRL-SV40 DNA (Promega) as a transfection control, per well. Cells were lysed 48 h after transfection and analyzed for firefly and *Renilla* luciferase activities using the Dual-Luciferase Assay System (Promega) according to the manufacturer's instructions, in half-volume 96-well tray format on an Enspire Multimode Plate Reader (PerkinElmer). Firefly luciferase activity was normalized to *Renilla* luciferase activity for each well, and the results were expressed as a mean normalized activity relative to cells transfected with empty vector. All experiments were performed three times in triplicate. A two-sided unpaired t test was used to compare luciferase activity between alleles.

Mouse gene expression analysis. We obtained fold changes in pancreatic islet gene expression in wild-type compared to *Foxa1* and *Foxa2* double-null mice[47]. We used Ensembl to map mouse genes to human orthologs. We filtered for human genes annotated as protein coding in GENCODE. This filtering resulted in 4,629 human protein-coding genes for analysis.

First, we calculated the genomic interval spanned by the variants in each credible set. We expanded this interval for 500 kb on either side and identified the set of genes overlapping this region using BEDtools[69]. To account for syntenic differences in gene order between species, we retained only genes that were (i) on the same chromosome and (ii) in exactly the same relative order in both the mouse and human genomes. At the *GIPR* locus, one of the genes was ordered differently and thus removed from the analysis. At two loci, *KCNJ11* and *HNF1A*, at least one of the genes was located on a different part of the same chromosome, and, at another locus, *GCK*, the genes were located on different chromosomes. For these three loci, we retained only the genes that were at the same chromosomal location to the interval covered by the credible set for the association signal for that locus (by lifting over from hg19 to mouse build mm10). Second, for each distinct association signal, we identified the closest gene to the index variant using BEDtools[69]. We then partitioned distinct association signals into those with evidence for enriched FOXA2 binding (fold enrichment >1) and those without, counting each gene only once in a given group. For each analysis, we converted the fold changes to percentages and compared the percent change in expression using a one-sided Wilcoxon rank-sum test between genes in each partition and all 4,629 protein-coding genes.

60. Devlin, B. & Roeder, K. Genomic control for association studies. *Biometrics* **55**, 997–1004 (1999).
61. Wakefield, J. Bayesian measure of the probability of false discovery in genetic epidemiology studies. *Am. J. Hum. Genet.* **81**, 208–227 (2007).
62. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
63. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
64. Kharchenko, P.V., Tolstorukov, M.Y. & Park, P.J. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* **26**, 1351–1359 (2008).
65. Li, Q., Brown, J.B., Huang, H. & Bickel, P.J. Measuring reproducibility of high-throughput experiments. *Ann. App. Stat.* **5**, 1752–1779 (2011).
66. Mikkelsen, T.S. *et al.* Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156–169 (2010).
67. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat. Biotechnol.* **28**, 817–825 (2010).
68. Morán, I. *et al.* Human β cell transcriptome analysis uncovers lncRNAs that are tissue-specific, dynamically regulated, and abnormally expressed in type 2 diabetes. *Cell Metab.* **16**, 435–448 (2012).
69. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
70. Machanick, P. & Bailey, T.L. MEME-ChiP. Motif analysis of large DNA datasets. *Bioinformatics* **27**, 1696–1697 (2011).
71. Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **42**, D142–D147 (2014).
72. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B-cell identifies. *Mol. Cell* **38**, 576–589 (2010).
73. Bailey, T.L. *et al.* MEME Suite: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
74. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
75. Pugh, C.W., Tan, C.C., Jones, R.W. & Ratcliffe, P.J. Functional analysis of an oxygen-regulated transcriptional enhancer lying 3′ to the mouse erythropoietin gene. *Proc. Natl. Acad. Sci. USA* **88**, 10553–10557 (1991).

 doi:10.1038/ng.3437