nature genetics

A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance

Recent genome-wide association studies have described many loci implicated in type 2 diabetes (T2D) pathophysiology and β-cell dysfunction but have contributed little to the understanding of the genetic basis of insulin resistance. We hypothesized that genes implicated in insulin resistance pathways might be uncovered by accounting for differences in body mass index (BMI) and potential interactions between BMI and genetic variants. We applied a joint meta-analysis approach to test associations with fasting insulin and glucose on a genome-wide scale. We present six previously unknown loci associated with fasting insulin at $P < 5 \times 10^{-8}$ in combined discovery and follow-up analyses of 52 studies comprising up to 96,496 non-diabetic individuals. Risk variants were associated with higher triglyceride and lower high-density lipoprotein (HDL) cholesterol levels, suggesting a role for these loci in insulin resistance pathways. The discovery of these loci will aid further characterization of the role of insulin resistance in T2D pathophysiology.

In contrast to recent progress in the discovery of genetic variants underlying T2D pathophysiology and β-cell function, the understanding of the genetic basis of insulin resistance remains limited[1]. Partly because early case-control studies of T2D were designed to maximize the likelihood of detecting variants that directly increase T2D risk rather than those that affect risk through the mediation of adiposity, most of the associated loci discovered in these studies mapped to genes related to β-cell dysfunction[2]. More recently, we have shown that the genetic architectures of quantitative indices of β-cell function and of insulin resistance differ markedly: given the same individuals, sample sizes and biochemical measurements, we described a larger number of signals for β-cell function than for insulin resistance[3,4]. Although this observation is consistent with the higher reported heritability of insulin secretion compared to resistance, overall heritability estimates of insulin resistance in individuals of European ancestry of 25–44% suggest that many loci remain to be discovered and that new strategies are required for their identification[5].

Obesity is an important determinant of insulin resistance[6]. It was postulated that adiposity might modulate the genetic determinants of insulin resistance and contribute to the heterogeneity of T2D etiology. It has been shown that the heritability of insulin resistance increases with higher BMI[7], and some candidate gene studies have observed that genetic effect size varies with adiposity level[8–10], findings that are compatible with the presence of an underlying interaction between BMI and genetic variants for insulin resistance. Furthermore, the adipokine hormones and proinflammatory cytokines that are produced by adipose tissue can influence insulin signaling via diverse mechanisms[11,12], and these processes may interact with genetic variants influencing insulin resistance pathways. Therefore, to identify variants associated with insulin resistance, it may also be important to account for gene variant by BMI interaction, which would allow for the potential for adiposity levels to perturb the physiological milieu in which genetic variants in insulin signaling pathways operate. Adiposity may also hinder the identification of genetic variants influencing insulin resistance by introducing variance in the outcome that is not attributable to genetic variation[5], suggesting that adjustment for adiposity *per se* may be necessary.

A joint test that investigates the association between an outcome and a genetic variant, while allowing for possible effect modification by an environmental variable, has been proposed[13]. Moreover, a statistical method was developed that extends this joint test to a meta-analysis context[14]. This enabled us to simultaneously test both the main genetic effect, adjusted for BMI, and potential interaction between each genetic variant and BMI. This joint meta-analysis (JMA) approach can provide increased power for detecting genetic loci when underlying interaction effects are suspected but unknown[13], and, notably, as shown in simulation studies, this approach does not reduce power to detect the main genetic effects in the absence of interaction[14]. Within the Meta-Analyses of Glucose- and Insulin-related traits Consortium (MAGIC), we implemented this approach and performed a genome-wide JMA to search for SNPs significantly associated with glycemic traits, while simultaneously adjusting for BMI and allowing for interaction with BMI. Using this method, we successfully identified loci that are associated with fasting insulin levels at genome-wide significance levels.

RESULTS

Study overview

As a first phase, we conducted a discovery genome-wide JMA of the main effects of SNPs and of SNP by BMI (SNP × BMI) interaction for four diabetes-related quantitative traits: fasting insulin levels, fasting glucose levels and surrogate measures of β-cell function

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

Received 7 July 2011; accepted 13 April 2012; published online 13 May 2012; doi:10.1038/ng.2274

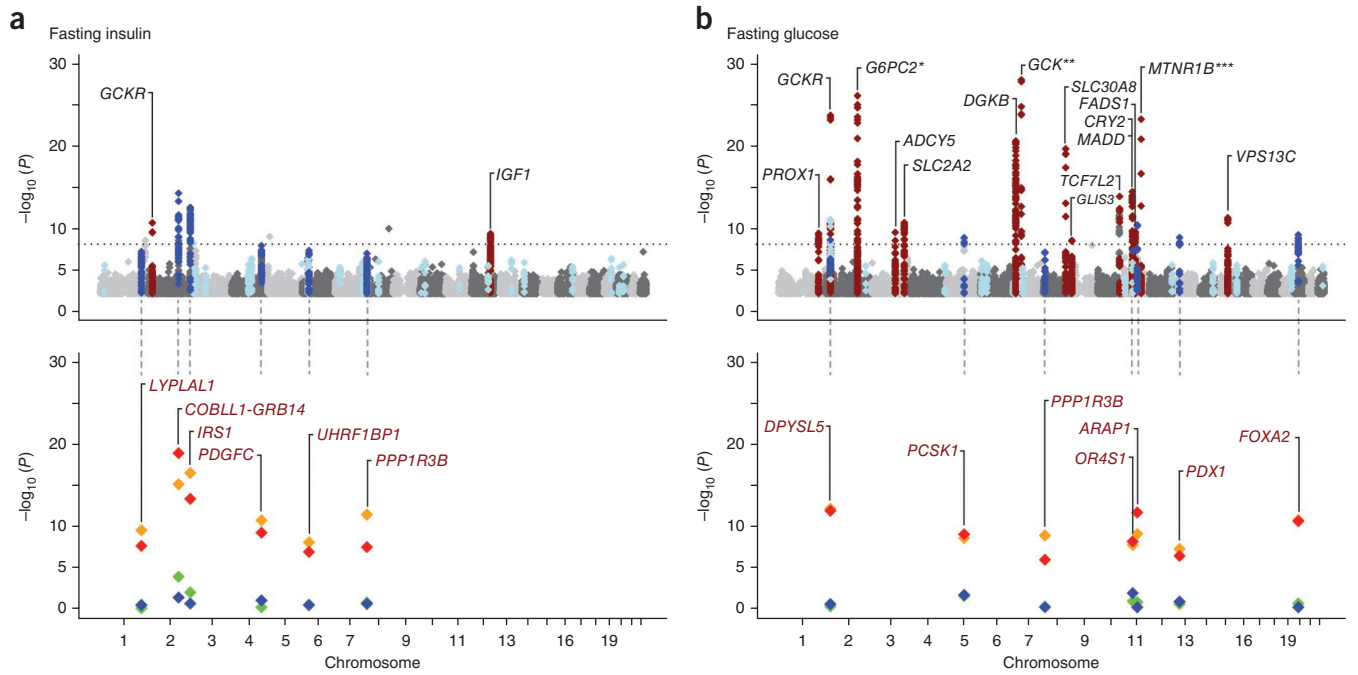

Figure 1 Genome-wide association plots of the discovery JMA. Association results are shown for fasting insulin (**a**) and fasting glucose (**b**) levels. We observed 17 loci with known associations (red) and took 50 loci forward to follow-up analysis (light and dark blue). Of these loci, 12 reached genome-wide significance in the combined discovery and follow-up JMA (dark blue). The *P* values of these 12 loci from the models fit in the combined discovery and follow-up analyses are shown below the plots: red, JMA; orange, main genetic effects, adjusting for BMI; green, interaction, with continuous BMI; blue, interaction, with dichotomous BMI. **G6PC2* JMA $P = 1.7 \times 10^{-113}$; ***GCK* JMA $P = 8.3 \times 10^{-56}$; ****MTNR1B* JMA $P = 4.38 \times 10^{-105}$.

(HOMA-B) and insulin resistance (HOMA-IR)[15]. Measures of fasting insulin, HOMA-B and HOMA-IR were log transformed.

The discovery-stage JMA for fasting insulin levels of approximately 2.4 million SNPs in 51,750 non-diabetic individuals from 29 studies (**Supplementary Table 1**) showed previously reported associations of variants in *IGF1* and *GCKR* with fasting insulin at genome-wide significance (**Fig. 1a** and **Supplementary Table 2**) and identified 31 previously unreported loci tentatively associated ($P < 1 \times 10^{-5}$) with fasting insulin (**Fig. 1a**, **Supplementary Fig. 1** and **Supplementary Table 3a**)[3]. The discovery JMA for fasting glucose in 58,074 individuals showed association of 16 loci previously described in MAGIC meta-analyses of fasting glucose levels (**Fig. 1b**, **Supplementary Fig. 1** and **Supplementary Table 2**) and identified 20 previously unreported loci tentatively associated with fasting glucose ($P < 1 \times 10^{-5}$) (**Fig. 1b** and **Supplementary Table 3b**). The JMA approach identified an excess of previously undetected signals relative to the number expected by chance (**Supplementary Fig. 1**). The SNPs that showed association at $P < 1 \times 10^{-5}$ in the discovery JMA for HOMA-IR and HOMA-B largely overlapped with those identified for fasting insulin and fasting glucose levels, without identifying additional loci. We therefore focused all subsequent analyses on fasting insulin and fasting glucose levels. Results for associations with HOMA-IR and HOMA-B are provided (**Supplementary Table 4**).

At each locus that reached $P < 1 \times 10^{-5}$ in the discovery JMA, we chose an index SNP to represent the association signal in the corresponding genomic region. We selected a total of 50 SNPs (30 SNPs associated with fasting insulin levels, 19 SNPs associated with fasting glucose levels and 1 SNP associated with both) for the second, follow-up phase of analysis. The follow-up analysis included 23 studies with fasting glucose data, 22 of which also had measurements of fasting insulin levels, and comprised up to 38,422 and 33,823 individuals for fasting glucose and fasting insulin analyses, respectively (**Supplementary Table 1**). We combined study-specific results from both the discovery (phase 1) and the follow-up (phase 2) studies for the 50 index SNPs (**Supplementary Table 3**) for total sample sizes of 96,496 for fasting glucose levels and 85,573 for fasting insulin levels. In the combined JMA, we identified six loci associated with fasting insulin levels and seven loci associated with fasting glucose levels at genome-wide levels of significance ($P < 5 \times 10^{-8}$), including *PPP1R3B*, which was associated with both fasting insulin and fasting glucose levels (**Fig. 1**, **Table 1** and **Supplementary Table 3**). Descriptions of noteworthy genes around each association signal are given (**Boxes 1** and **2**).

As the JMA simultaneously tested both main genetic effects and interaction effects, in order to further characterize the association signals observed using the JMA, we performed additional analyses examining main effects and interaction effects separately. These additional analyses included meta-analysis of regression models of the main effects, with and without adjustment for BMI, univariate meta-analysis of interaction effects using continuous BMI and meta-analyses of the main effects of SNPs in strata defined by a BMI cutoff of 28 kg/m^2, which was chosen on the basis of the median BMI of the largest cohorts included in our discovery stage JMA (BMI strata were categorized as leaner, where BMI <28 kg/m^2, and heavier, where BMI ≥28 kg/m^2).

Associations with fasting insulin levels

The loci examined in follow-up analysis that associated with fasting insulin levels (natural log transformed) at genome-wide levels of significance in the combined JMA are presented in **Table 1** and described in **Box 1**. The rs7607980 SNP, located near *COBLL1-GRB14*, provided the strongest signal for fasting insulin levels, with a JMA

Table 1 Genetic loci associated at genome-wide significance ($P < 5 \times 10^{-8}$) with fasting insulin or glucose level from the JMA in the combined discovery and follow-up analysis

Fasting insulin levels

		Alleles		Stage 1: discovery (26 cohorts) $n = 51,750$			Stage 2: follow-up (22 cohorts) $n = 31,450$			Combined (48 cohorts) $n = 83,116$		
Nearest gene	Index SNP	R (freq.)	O (freq.)	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})
COBLL1-GRB14	rs7607980	T (0.88)	C (0.12)	0.021 (0.004)	0.04 (0.005)	6.1×10^{-14} (0.25)	0.026 (0.006)	0.039 (0.008)	4.9×10^{-7} (0.04)	0.023 (0.003)	0.039 (0.004)	4.3×10^{-20} (0.06)
IRS1	rs2943634	C (0.66)	A (0.34)	0.018 (0.003)	0.026 (0.004)	1.7×10^{-12} (0.62)	0.018 (0.008)	0.021 (0.010)	3.6×10^{-3} (0.53)	0.018 (0.003)	0.025 (0.004)	2.5×10^{-14} (0.70)
PPP1R3B	rs4841132	A (0.10)	G (0.90)	0.02 (0.004)	0.031 (0.006)	1.4×10^{-7} (0.50)	0.034 (0.013)	0.032 (0.016)	7.3×10^{-4} (0.45)	0.021 (0.004)	0.031 (0.006)	1.7×10^{-10} (0.52)
PDGFC	rs4691380	C (0.67)	T (0.33)	0.016 (0.003)	0.021 (0.004)	1.5×10^{-8} (0.36)	0.017 (0.008)	0.020 (0.010)	7.2×10^{-2} (0.31)	0.016 (0.003)	0.021 (0.004)	5.3×10^{-9} (0.37)
UHRF1BP1	rs4646949	T (0.75)	G (0.25)	0.016 (0.003)	0.021 (0.004)	5.9×10^{-8} (0.08)	0.0009 (0.008)	0.009 (0.010)	1.6×10^{-1} (0.24)	0.014 (0.003)	0.020 (0.004)	3.7×10^{-8} (0.05)
LYPLAL1	rs2785980	T (0.67)	C (0.33)	0.016 (0.003)	0.017 (0.004)	7.7×10^{-8} (0.03)	0.011 (0.009)	0.018 (0.010)	9.7×10^{-2} (0.20)	0.016 (0.003)	0.017 (0.004)	2.0×10^{-8} (0.03)

Fasting glucose levels

		Alleles		Stage 1: discovery (29 cohorts) $n = 58,074$			Stage 2: follow-up (23 cohorts) $n = 38,422$			Combined (52 cohorts) $n = 96,496$		
Nearest gene	Index SNP	R (freq.)	O (freq.)	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})	BMI = 25 effect (SE)	BMI = 30 effect (SE)	JMA P (P_{het})
ARAP1	rs11603334	G (0.83)	A (0.17)	0.025 (0.004)	0.033 (0.005)	4.4×10^{-11} (0.25)	0.016 (0.005)	0.026 (0.007)	2.2×10^{-4} (0.004)	0.022 (0.003)	0.030 (0.004)	2.4×10^{-14} (0.01)
FOXA2	rs6048205	A (0.95)	G (0.05)	0.044 (0.007)	0.033 (0.009)	7.1×10^{-10} (0.14)	0.033 (0.009)	0.023 (0.012)	1.4×10^{-3} (0.03)	0.040 (0.005)	0.029 (0.007)	1.6×10^{-12} (0.03)
DPYSL5	rs1371614	T (0.25)	C (0.75)	0.020 (0.004)	0.026 (0.005)	2.9×10^{-9} (0.19)	0.019 (0.005)	0.015 (0.006)	2.1×10^{-4} (0.47)	0.020 (0.003)	0.022 (0.004)	2.3×10^{-12} (0.23)
PCSK1	rs13179048	C (0.69)	A (0.31)	0.021 (0.003)	0.016 (0.004)	1.6×10^{-9} (0.24)	0.028 (0.010)	0.027 (0.013)	2.2×10^{-2} (0.16)	0.022 (0.003)	0.018 (0.004)	1.6×10^{-10} (0.18)
PDX1	rs2293941	A (0.22)	G (0.78)	0.022 (0.004)	0.015 (0.005)	1.3×10^{-8} (0.47)	0.013 (0.005)	0.016 (0.006)	7.8×10^{-3} (0.04)	0.019 (0.003)	0.016 (0.004)	5.3×10^{-10} (0.10)
PPP1R3B	rs4841132	A (0.10)	G (0.90)	0.026 (0.005)	0.028 (0.007)	7.9×10^{-7} (0.94)	0.033 (0.015)	0.054 (0.021)	3.1×10^{-3} (0.02)	0.027 (0.005)	0.030 (0.007)	7.6×10^{-9} (0.55)
OR4S1	rs1483121	G (0.86)	A (0.14)	0.027 (0.005)	0.022 (0.006)	6.5×10^{-8} (0.70)	0.014 (0.006)	0.006 (0.008)	3.4×10^{-2} (0.42)	0.021 (0.004)	0.015 (0.005)	1.6×10^{-8} (0.56)

Directly genotyped and imputed SNPs were tested for association with fasting glucose or fasting insulin level by JMA of the SNP and SNP × BMI effects. The effect estimates presented are derived from the JMA for the specified BMI (in kg/m^2). A test of heterogeneity (P_{het}) was performed for each locus. From the discovery analysis, 50 loci were taken forward to the follow-up analysis (31 for fasting insulin and 20 for fasting glucose levels; *PPP1R3B* was followed up in both traits). R, trait-raising allele; O, other allele; freq., allele frequency; SE, standard error, *n*, sample size presented is the maximum observed among the SNPs presented here.

P value of 4.3×10^{-20}. Analyses performed to characterize this association further showed that the significant association of rs7607980 with fasting insulin levels was driven primarily by a BMI-adjusted main genetic effect ($P = 1.7 \times 10^{-16}$) and showed some suggestion of interaction with BMI (*P* for interaction = 1.6×10^{-4}; **Fig. 2** and **Supplementary Table 3a**). Modeled from the combined JMA, the estimated effect size of the genetic variant was greater at a BMI of 30 kg/m^2 (JMA $\beta = 0.039$) compared to at a BMI of 25 kg/m^2 (JMA $\beta = 0.023$) (**Table 1**). These estimates were supported by BMI-stratified analyses: the genetic effect was larger in heavier (main genetic effect $\beta = 0.041$, $P = 2.98 \times 10^{-10}$) than in leaner (main genetic effect $\beta = 0.018$, $P = 1.77 \times 10^{-5}$) individuals (*P* value for the difference between strata = 0.02; **Fig. 2**).

In addition to *COBLL1-GRB14*, the combined JMA showed five loci to be associated with fasting insulin (*IRS1*, *PPP1R3B*, *PDGFC*, *LYPLAL1* and *UHRF1BP1*) at genome-wide significance levels (**Table 1** and **Supplementary Fig. 2**). For these loci, the additional models to characterize association showed that the JMA association signals were largely driven by main genetic effects adjusted for BMI ($P < 5.5 \times 10^{-9}$) and showed little evidence of interaction ($P > 0.02$; **Supplementary Fig. 3** and **Supplementary Table 3a**). Additionally, the combined JMA at the *PEPD* locus showed evidence suggestive of association, albeit just below the conventional significance threshold ($P = 8.69 \times 10^{-8}$). The results of the JMA and the additional meta-analyses for all loci included in the follow-up (phase 2) analysis are presented (**Supplementary Table 3a**).

As we were interested in identifying loci associated with insulin resistance, we investigated associations between variants that were associated with genome-wide significance and other traits related to insulin resistance in genome-wide meta-analysis results provided by other consortia, including the Diabetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium[4], the Genomewide Investigation of ANThropometric Measures (GIANT) Consortium[16,17] and the Global Lipid Genetic Consortium (GLGC)[18] (**Table 2**). At five out of six loci associated with fasting insulin, the insulin-raising allele was associated with both lower HDL cholesterol and higher triglycerides, a dyslipidemic profile typical of insulin resistance. The alleles associated with increased insulin levels

Box 1 Genes of biological interest within 500 kb of SNPs associated with fasting insulin index

COBLL1-GRB14: The index SNP (rs7607980) identifies a coding variant that induces an amino-acid change (p.Asn939Asp) in cordon-bleu protein-like 1, encoded by *COBBL1*. Cordon-bleu is a protein involved in neural tube development[31]. The allele associated with increased fasting insulin levels at the index SNP was associated with higher expression of *COBLL1* ($P = 0.02$) in skeletal muscle in the Malmö Exercise Study (personal communication, O. Hansson). A SNP at this locus (rs10195252) is associated with waist-hip-ratio[17] but is in low linkage disequilibrium (LD) with the index SNP ($r^2 = 0.148$). *GRB14* encodes growth factor receptor–bound protein 14, which inhibits signaling of the insulin receptor (MIM 601524)[32]. In addition to being associated with triglyceride levels[33], rs10195252 is reported to have a *cis*-acting association with *GRB14* transcript level in omental fat ($P = 1.0 \times 10^{-13}$)[18].

***IRS1*:** *IRS1* encodes insulin receptor substrate 1, a critical docking protein in the insulin signaling cascade, which, when phosphorylated by the insulin receptor, activates downstream signaling pathways[34]. SNPs in or near *IRS1* are associated with T2D, HOMA-IR, fasting insulin and coronary artery disease (CAD)[3,35,36] but have not previously been shown to be associated fasting insulin at the genome-wide significance. The index SNP reported here (rs2943634) is the same intergenic SNP associated with CAD[36] and is in LD with the nearby rs2943641 SNP that is associated with insulin resistance[3] and T2D[4] ($r^2 = 0.782$) and the rs7578326 SNP associated with T2D ($r^2 = 0.815$). The variant identified by the lipid GWAS (rs2972146; associated with both triglycerides and HDL) is also reported to have a *cis*-acting association with *IRS1* transcript level in omental adipose tissue ($P = 2.0 \times 10^{-8}$)[17], and this SNP is in strong LD with the index SNP ($r^2 = 0.751$). Furthermore, the index SNP (rs2943634) is in moderate LD ($r^2 = 0.438$) with three SNPs (rs1849878, rs2673148 and rs2713547) that tag a known copy-number variant (CNV) (CNVR1152.1) ($r^2 = 0.51$, 0.52 and 0.51, respectively).

***PPP1R3B-TNKS*:** This locus is associated with both fasting insulin and fasting glucose in the present study. *PPP1R3B* encodes protein phosphatase 1 regulatory (inhibitor) subunit 3B, which prevents glycogen breakdown by regulating the interaction of phosphorylase protein 1 (PP1) with glycogen metabolism enzymes (MIM 610541)[32]. Two SNPs in or near *PPP1R3B* are also associated with lipids[18,37] (rs9987289 and rs2126259) and C-reactive protein (rs9987289)[21], and these are both in strong LD with the index SNP ($r^2 = 1.0$ and 0.803, respectively). rs9987289 is also reported to have a *cis*-acting association with *PPP1R3B* transcript level in human liver ($P = 1 \times 10^{-14}$)[18]. In the present study, we observed that rs19334, associated with expression of *PPP1R3B* ($P = 2.34 \times 10^{-12}$) in a liver eQTL data set[38], is in low LD with the index SNP ($r^2 = 0.001$) but showed moderate association with fasting insulin in the discovery sample ($P = 2.87 \times 10^{-4}$). *TNKS* encodes TRF1-interacting, ankyrin–related ADP-ribose polymerase, which interacts with telomeric repeat–binding factor 1 (TRF1) to regulate telomere length[39], and also interacts with insulin-responsive amino peptidase (IRAP) in GLUT4 vesicles (MIM 603303)[40]. Tankyrases are also thought be involved in the Wnt signaling pathway[41].

***PEPD-CEBPA-KCTD15*:** *PEPD* encodes peptidase D, an enzyme responsible for the recycling of proline and likely essential for collagen production (MIM 613230)[42]. A SNP in *PEPD* (rs731839) is associated with adiponectin levels (B. Richards, personal communication) and is in strong LD with the index SNP ($r^2 = 0.77$). The index SNP is associated with *PEPD* expression in the adipocyte eQTL data set ($P = 9.96 \times 10^{-10}$)[43], but, most likely, this association is driven by the association between rs17226118 and *PEPD* expression ($P = 2.2 \times 10^{-55}$), despite the weak LD between the index SNP and rs17226118 ($r^2 = 0.145$). *CEBPA* encodes CCAAT/enhancer-binding protein α (C/EBPα), which may control the expression of leptin, an adipokine implicated in weight regulation (MIM 116897)[44]. *KCTD15* encodes potassium channel tetramerisation domain–containing 15. A SNP in or near this gene (rs29941) is associated with BMI[16] but is in very low LD with the index SNP ($r^2 = 0.015$).

***UHRF1BP1-PPARD*:** *UHRF1BP1* encodes ubiquitin-like containing PHD and RING finger domains 1 (UHRF1)-binding protein 1. UHRF1 is a protein that influences DNA methylation (MIM 607990)[45]. The index SNP in *UHRF1BP1* (rs4646949) is in strong LD ($r^2 = 0.724$) with a SNP (rs2293242) that causes a nonsense change in *ANKS1A*. The same index SNP is in moderate LD ($r^2 = 0.363$) with a coding SNP that results in a missense change in *UHRF1BP1* (encoding p.Met1098Thr) that is considered damaging on the basis of analysis of homologous sequences[46] (**Supplementary Table 7**). Index SNP rs4646949 in *UHRF1BP1* is in strong LD ($r^2 = 0.95$) with two perfect proxies (rs2477508 and rs2814922) for a known CNV (CNVR2857.1). In a liver tissue eQTL data set, rs12173920 was associated with *UHRF1BP1* and *STEAP4* (encoding six transmembrane epithelial antigen of prostate 4) expression levels ($P = 9.56 \times 10^{-6}$ and 2.91×10^{-8}, respectively) and is in moderate LD ($r^2 = 0.334$) with the index SNP[38]. Another SNP in the region, rs2814944, is associated with HDL levels and has been reported to have a *cis*-acting association with the transcript level of *UHRF1BP1* in both omental ($P = 3.0 \times 10^{-25}$) and subcutaneous ($P = 2.0 \times 10^{-18}$) adipose tissues[18]. *PPARD* encodes peroxisome proliferator-activated receptor-δ, a protein involved in the breakdown of fat (MIM 600409)[47].

***PDGFC-GLRB*:** *PDGFC* encodes platelet-derived growth factor C, a ligand that binds to specific PDGF receptors (α-α and α-β) and has a growth factor domain homologous to VEGF (MIM 608452)[48]. Tyrosine phosphorylation is activated by the binding of these growth factors to receptors. *GLRB* encodes glycine receptor β, a subunit of glycine receptors (neurotransmitter-gated ion channels) that are likely involved in glycine receptor structure (MIM 138492)[49].

LYPLAL1-SLC30A10: *LYPLAL1* encodes lysophospholipase-like 1, a protein that may be involved in adiposity and fat distribution[17,50–52]. A SNP near *LYPLAL1* (rs4846567) is associated with WHR and is in strong LD with the index SNP ($r^2 = 0.796$). Furthermore, a variant in this vicinity (rs3001032) is associated with adiponectin levels (B. Richards, personal communication) and is in strong LD with the index SNP ($r^2 = 0.828$). *SLC30A10* encodes solute carrier family 30, member 10 of the zinc transporter subfamily of cation-diffusion facilitators that allow for the outward flow of zinc from cells (MIM 611146)[53].

near *COBLL1-GRB14*, *LYPLAL1* and *PPP1R3B* were also associated with a greater waist-to-hip ratio (WHR; adjusted for BMI). Individuals carrying alleles associated with higher fasting insulin levels at the index SNP in or near *COBLL1-GRB14*, *LYPLAL1*, *IRS1* or *PDGFC* were at increased risk for T2D, albeit with relatively small odds ratios (**Table 2**).

Associations with fasting glucose levels

Genetic variants associated with fasting glucose levels at genome-wide significance in the combined JMA are presented in **Table 1** and described in **Box 2**; index SNPs were in or near *ARAP1*, *PCSK1*, *FOXA2*, *DPYSL5*, *OR4S1*, *PDX1* and *PPP1R3B*. The combined JMA results for the index SNP near *GRB10* showed suggestive association with fasting glucose levels, but the *P* value did not reach genome-wide significance ($P = 8.3 \times 10^{-8}$). Of note, a SNP in *GRB10* (rs2237457) was previously found to be associated with T2D ($P = 1.1 \times 10^{-5}$) in the Old Order Amish[19].

Additional analyses revealed that main genetic effects were the primary contributors to these associations ($P < 3.6 \times 10^{-8}$) (**Supplementary Table 3b**), with little evidence observed for

Box 2 Genes of biological interest within 500 kb of index SNPs associated with fasting glucose levels

***PCSK1*:** *PCSK1* encodes proprotein convertase, subtilisin/kexin-type 1. This protein initiates proinsulin processing to insulin. The expression of this protein is regulated by glucose (MIM 162150)[54]. The nonsynonymous variant rs6232 (encoding p.Asn221Asp) and rs6234 and rs6235 (a pair of variants encoding p.Gln665Glu and p.Ser690Thr) have been shown to be associated with obesity in children and adults[55]. The index SNP is in strong LD with rs6234 (r^2 = 0.814). A report from MAGIC showed that rs6235 in *PCSK1* is associated with proinsulin levels at a genome-wide level of significance[29].

***OR4S1-PTPRJ*:** *OR4S1* encodes olfactory receptor, family 4, subfamily S, member 1. This family of receptors sets off neuronal responses for smell perception. *PTPRJ* encodes protein tyrosine phosphatase (PTP), receptor type J. By hindering phosphorylation, this protein is thought to inhibit T-cell receptor signaling.

***ARAP1-INPPL1-STARD10*:** A SNP in this region, rs1552224, is associated with T2D[4] and is in perfect LD with the index SNP (r^2 = 1.0). The genetic variant associated with increased risk of T2D is also associated with lower fasting proinsulin levels (adjusted for insulin levels), suggesting that the variant might cause a defect in the early steps of insulin production[29]. *ARAP1* encodes ankyrin repeat and pleckstrin homology domains–containing protein 1 that controls ARF-, RHO- and CDC42-dependent cell functions[56]. *INPPL1* encodes inositol polyphosphate phosphatase like–1. *In vivo* mouse studies indicated that this protein is a negative regulator of insulin signaling and sensitivity (MIM 600829)[57]. *STARD10* encodes StAR-related lipid transfer (START) domain containing 10, a protein thought to be involved in sperm cell maturation but highly expressed in liver and pancreas tissues[58] (NF Expression Atlas 2 Data from U133A and GNF1H Chips). Levels of *STARD10* expression are higher in pancreatic and islet tissues than in any other human tissue type[29].

***FOXA2*:** *FOXA2* encodes forkhead box A2, a DNA-binding protein that regulates the expression of key genes active in glucose sensing in β cells (MIM 600288). When *FOXA2* was continually activated in mice, this led to increases in neuronal MCH and orexin expression, insulin sensitivity, food consumption and metabolism[59]. It has previously been suggested that mutations in *FOXA2* may affect glucose homeostasis[60].

***GRB10*:** *GRB10* encodes growth factor receptor–bound protein 10 (MIM 601523). This protein forms a complex with insulin receptors and inhibits their signaling[61]. This protein also interacts with Grb10-interacting GYF proteins (GIGYFs) to regulate insulin-like growth factor receptor 1 signaling[62]. The risk allele at rs2237457 showed an association with T2D ($P = 1.1 \times 10^{-5}$) and glucose excursion during an oral glucose tolerance test (OGTT) ($P = 0.001$) in the Old Order Amish Study[19].

***DPYSL5-KHK-PPM1G*:** *DPYSL5* encodes dihydropyrimidinase-like 5 and is thought to be involved in directing neuronal growth cones in development (MIM 608383). *KHK* encodes ketohexokinase and is the first enzyme that acts in the breakdown of fructose (MIM 229800). *PPM1G* encodes protein phosphatase, magnesium-dependent 1, which carries out a dephosphorylation event essential for forming the spliceosome (MIM 605119)[63]. There is moderate LD (r^2 = 0.404) between the index SNP in *DPYSL5* (rs1371614) and a SNP (rs2384572) that causes a missense change (encoding p.Ile116Met or p.Ile20Met, depending on the transcript) in *CGREF1* (which encodes cell growth regulator with EF-hand domain 1) that is considered damaging based on analysis of homologous and orthologous sequences[46] (**Supplementary Table 7**).

PDX1: *PDX1* encodes pancreatic and duodenal homeobox 1 and is responsible for the development of the pancreas, determining maturation and differentiation of common pancreatic precursor cells (MIM 600733). As pancreatic morphogenesis proceeds, *PDX1* action is eventually restricted to β and δ cells of the islets, where it seems to regulate expression of the *INS* (encoding insulin) and *SST* (encoding somatostatin) genes, respectively. A deletion and missense changes in *PDX1* (causing p.Glu164Asp and p.Glu178Lys alterations) are associated with pancreas agenesis[64]. Whereas the deletion causes pancreas agenesis in homozygotes, heterozygosity is associated with maturity-onset diabetes of the young, type 4 (MODY4)[65]. The p.Glu224Lys variant was also shown to cause MODY4 (ref. 66). Other variants were suspected to result in increased risk for T2D in some family studies.

PPP1R3B-TNKS: see summary in **Box 1**.

interaction at these loci (*P* values for interaction >0.06; **Supplementary Fig. 3**). The combined results of the JMA and the additional meta-analyses for all loci included in the follow-up (phase 2) analysis are presented (**Supplementary Table 3b**).

In contrast to associations observed at fasting insulin loci, SNPs associated with the fasting glucose index did not have compelling associations with metabolic traits related to insulin resistance (**Table 2**), but alleles associated with increased fasting glucose levels near *PCSK1* and *PPP1R3B* showed nominal associations with lower levels of glucose 2 h after a glucose tolerance test (2-hour glucose) (**Supplementary Table 4).**

Functional exploration, the results of expression and expression quantitative trait locus (eQTL) searches and conditional analyses using available databases are presented (**Supplementary Table 5** and **Supplementary Note**).

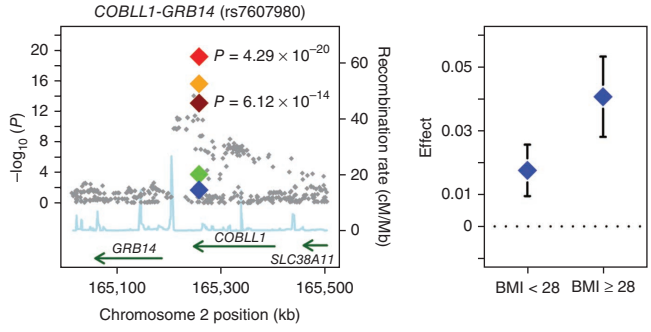

Figure 2 Regional plot of the *COBLL1-GRB14* genomic locus. Left, discovery JMA *P* values are shown in the background. For the SNP taken forward to follow-up analyses (rs7607980), the color-coded *P* values for the different analyses are shown: dark red, discovery JMA; light red, combined discovery and follow-up JMA; orange, main effects, adjusting for BMI; green, interaction, with continuous BMI; blue, interaction, with dichotomous BMI. Right, effect estimates with 95% confidence intervals of the allele associated with increased trait levels in the two BMI strata. Effect = β estimates from regression models of ln(fasting insulin), adjusting for age, sex and other study-specific covariates.

DISCUSSION

Our study shows that loci associated with insulin resistance exist and can be identified when appropriate analysis methods are used to control for the influence of adiposity on insulin resistance. In a discovery data set that was approximately 35% larger than that for

Table 2 Associations with T2D, lipid profile and anthropometric measures for loci significantly associated with fasting insulin or fasting glucose level

						Lipid profile								Anthropometric measures					
				T2D		TC		LDL		HDL		Triglycerides		BMI (kg/m^2)			WHR (adjusted BMI)		
Trait	Index SNP	Nearest gene(s)	Alleles R/O	OR	*P*	Dir.	*P*	Dir.	*P*	Dir.	*P*	Dir.	*P*	β	SE	*P*	β	SE	*P*
FI	rs7607980	*COBLL1-GRB14*	T/C	1.08	0.02	+	0.09	+	0.05	–	3.9×10^{-10}	+	4.1×10^{-8}	–0.016	0.007	0.03	0.031	0.007	2.7×10^{-5}
FI	rs2943634	*IRS1*	C/A	1.09	2.7×10^{-5}	+	0.91	+	0.14	–	2.3×10^{-9}	+	5.2×10^{-8}	–0.014	0.005	0.01	0.005	0.005	0.33
FI/FG	rs4841132	*PPP1R3B*	A/G	1.04	0.24	–	2.2×10^{-19}	–	1.5×10^{-12}	–	8.4×10^{-23}	+	0.017	0.007	0.008	0.40	0.020	0.008	0.02
FI	rs4691380	*PDGFC*	C/T	1.05	0.03	–	0.14	+	0.70	–	1.4×10^{-4}	+	0.008	–0.005	0.005	0.34	0.007	0.005	0.17
FI	rs4646949	*UHRF1BP1*	T/G	1.04	0.13	–	0.09	–	0.04	+	0.37	+	0.81	–0.009	0.005	0.09	0.008	0.005	0.14
FI	rs2785980	*LYPLAL1*	T/C	1.04	0.05	+	0.22	+	0.05	–	9.0×10^{-4}	+	0.0023	–0.007	0.005	0.19	0.030	0.005	2.3×10^{-9}
FG	rs11603334	*ARAP1*	G/A	1.13	7.8×10^{-6}	–	0.87	–	0.97	+	0.32	–	0.28	–0.012	0.006	0.07	0.001	0.006	0.83
FG	rs6048205	*FOXA2*	A/G	1.05	0.39	+	0.07	+	0.15	+	0.16	+	0.21	–0.020	0.012	0.09	0.007	0.011	0.55
FG	rs1371614	*DPSYL5*	T/C	0.98	0.32	–	0.29	+	0.51	–	0.11	–	5.5×10^{-4}	–0.008	0.006	0.13	0.001	0.005	0.84
FG	rs13179048	*PCSK1*	C/A	1.02	0.34	–	0.09	–	0.09	+	0.64	–	0.44	–0.011	0.005	0.03	0.006	0.005	0.25
FG	rs2293941	*PDX1*	A/G	1.01	0.76	+	0.48	–	0.97	+	0.02	–	0.34	–0.007	0.006	0.23	0.006	0.006	0.28
FG	rs1483121	*OR4S1*	G/A	1.00	0.91	–	0.14	–	0.05	–	0.77	+	0.23	0.017	0.007	0.02	0.004	0.007	0.63

Associations with T2D were assessed in the DIAGRAM meta-analysis of up to 8,130 cases and 38,987 controls. Associations with lipid profile were performed using publicly available data and included up to 99,900 individuals. Associations with anthropometric measures were performed in the GIANT data set including up to 123,685 individuals. R/O alleles, trait-raising allele/other allele; TC, total cholesterol; LDL, LDL cholesterol; HDL, HDL cholesterol; OR, odds ratio; dir., direction of effect; SE, standard error; FI, fasting insulin; FG, fasting glucose.

our previous meta-analysis, we describe six loci not previously known to be associated with fasting insulin levels, adding substantially to the two loci that were observed using standard association analyses without adjustment for BMI. We also describe seven additional loci associated with fasting glucose levels. Further, we detected all previously reported associations for fasting glucose (16 loci) and insulin (2 loci) levels (**Supplementary Table 2**)[3]. With the initial explosion in the number of loci found to be associated with T2D[1], we expected to discover similar numbers of loci implicated in insulin resistance and secretion, which has not been the case until now. Our approach was successful at identifying loci implicated in insulin resistance in numbers proportional to those implicated in insulin secretion. Our results underscore the importance of taking adiposity into account in understanding the heterogeneity of T2D etiology.

The association of fasting insulin levels with the genetic variant located at the *COBLL1-GRB14* locus is of particular interest because of its concomitant and directionally consistent association with insulin resistance–related traits and its plausible mechanism of action. *GRB14* is a tissue-specific negative regulator of insulin receptor signaling (**Box 2**). The risk allele at the rs7607980 index SNP is associated with higher triglyceride and lower HDL cholesterol levels, which constitute a dyslipidemic profile that is characteristic of the insulin-resistant state, suggesting a putative role in insulin resistance pathways. Furthermore, this pattern of association was observed for nearly all index SNPs associated with fasting insulin levels, increasing our confidence that our approach identified genetic loci involved in insulin resistance pathogenesis. As shown in a previous GIANT report[17], risk alleles at *COBLL1-GRB14* and *LYPLAL1* were also associated with increased WHR adjusted for BMI (**Table 2**), suggesting that they could influence the regulation of adipose tissue distribution to induce insulin resistance. In line with plausible biological actions in adipose tissue, eQTL results in omental adipose tissue showed that expression of *GRB14* is associated with the rs10195252 SNP ($r^2 = 0.17$ with the rs7607980 index SNP) in the report from GLGC[18].

Our results suggest a SNP × BMI interaction with fasting insulin levels at *COBLL1-GRB14*, with a larger genetic effect estimate in heavier individuals, which is compatible with the notion that some genetic variants implicated in insulin resistance pathways are more likely to be found when studying genetic effects in an 'obesogenic' environment characterized by high levels of energy intake and low levels of expenditure that promote a positive energy balance and the accumulation of excess adiposity. Our observation supports findings from the TwinsUK study reporting greater heritability estimates for insulin resistance traits at higher BMI[7]. Moreover, some candidate gene studies showed that genetic variants implicated in insulin signaling pathways are more readily observable in heavier populations[20]. It was proposed that higher BMI potentiates the effect of genetic variants on insulin resistance pathways and suggested that this effect could be due to tissue-specific response to the obesogenic environment[5]. It is currently unknown how a genetic variant could contribute to a different response to an obesogenic exposure within specific tissues (liver, muscle, adipose or even central nervous system), but our data add to the current literature by identifying one potential example and argue for functional studies to pursue the previously proposed hypothesis[5].

Among the other loci, *PPP1R3B* is likely to act via hepatic metabolism to influence fasting insulin and glucose levels, as well as the lipid profile (HDL, low-density lipoprotein (LDL) and total cholesterol) and C-reactive protein (CRP) levels[21]. *PEPD* encodes peptidase D and is likely to have a role in adipokine biology, an idea that is supported by eQTL data in adipose tissue (**Supplementary Table 6**) and an observed association with adiponectin levels (ADIPOgen Consortium; B. Richards, personal communication). Adiponectin is suspected to act as an insulin sensitizer, with low levels conferring T2D risk only among individuals who are also insulin resistant[22]; the independent *PEPD* associations with this adipokine are consistent with a direct role in insulin resistance. Additionally, other genes located near the reported loci (**Table 1**) also represent plausible biologic candidates for factors involved in various processes related to insulin resistance (**Box 1**).

The JMA approach has the potential to identify genetic variants whose effects differ depending on levels of adiposity, with interactions being tested in either direction. As noted previously, a search for T2D-associated genes in leaner populations is more likely to identify

genes implicated in β-cell function by focusing on pathophysiological mechanisms independent of adiposity and insulin resistance[23]. As a corollary of this observation, genetic loci previously shown to be associated with β-cell function in non-diabetic individuals may show stronger effect sizes in lean compared to obese participants, although our results provided no evidence to support this for fasting glucose levels (**Supplementary Table 2**).

There have been few previous efforts investigating SNP × BMI interaction and its impact on the risk for T2D or related glycemic traits at a genome-wide level. Genetic variants associated with T2D in the Wellcome Trust Case Control Consortium were investigated by dividing the cases into strata of obese and non-obese individuals: in addition to *TCF7L2* (of larger effect in non-obese cases), the only other signal they detected and replicated was the well-known obesity-mediated *FTO* association with T2D in the obese strata[24]. As expected, we did not detect an association between *FTO* and fasting insulin or glucose levels because the JMA approach included an intrinsic adjustment for BMI.

The major strengths of our study are that we used a large sample size, that all cohorts were composed of individuals of European descent and, notably, that we used a validated statistical approach and successfully applied it at a genome-wide level for the first time. We identified all previously described associations with fasting insulin and glucose levels, which highlights the usefulness of the JMA approach, even in the absence of underlying gene-environment interaction.

We used a two-phase analytical approach with the conventional $P < 5.0 \times 10^{-8}$ significance threshold in combined analyses to determine genome-wide significance for the identified SNPs. This approach has been shown to have greater power when the size of the follow-up sample is smaller than that of the discovery phase[25,26] and has previously been successful applied by our group[27–29] and others[30]. All 12 loci described in **Table 1** showed consistent directions of effect in the follow-up phase. However, in employing a replication strategy that declares statistical significance on the basis of the follow-up phase alone, only 4 of the 12 loci would have reached conventional statistical significance after accounting for multiple testing ($P < 0.05/51 = 0.00098$; with significance threshold of α/N_1, where N_1 is the number of associations tested based on 50 SNPs taken forward for replication including 1 SNP associated with both fasting glucose and insulin levels).

Only 12 significantly associated loci were identified after the second stage, despite the fact that 50 loci were taken forward from the discovery stage and that only 2 or 3 associations at $P < 0.00001$ would be expected per trait under the hypothesis of no association. Among the reasons for the fewer than expected significant associations, the size of the follow-up sample is most noteworthy. To maximize sample size, we took advantage of studies that offered either *de novo* genotyping or access to genome-wide SNP arrays or had performed genotyping on the Metabochip. Although our study represents the largest effort of its kind, it should be acknowledged that the follow-up sample size was still approximately one-half of the size of the discovery sample, and false negatives might remain among those loci not reaching genome-wide significance.

We used the JMA as the primary model and further characterized only those loci that reached genome-wide significance in additional models. However, consideration of results from the follow-up stage alone showed that the interaction term at the *COBBL1-GRB14* locus did not reach statistical significance in additional models, although we observed consistency in direction and effect size in phases 1 and 2.

Our main JMA model assumes a linear interaction, where the per-allele effect of a SNP changes across the continuous spectrum of BMI. If the interaction effect is nonlinear or a threshold effect exists, in which case the association would only be present in one extreme of the BMI distribution, the results of the interaction test with continuous BMI might not agree with those from a model with dichotomous BMI distribution. This could explain the inconsistencies of the interaction results for *IRS1*, *PCSK1* and *OR4S1*, for which we did observe some suggestions of interaction in the stratified models, which were not supported by models including BMI as a continuous variable (**Supplementary Fig. 3**).

We used the JMA approach to test for the main genetic effects of SNPs with adjustment for BMI and to allow for interaction between BMI and SNPs, but the results we observed might be due to one of many factors that are correlated with BMI, including lifestyle. Additionally, although we identified genes that highlight potential new pathways to insulin resistance and T2D development, we recognize that we have not localized the associations to specific genes.

Previous attempts to identify loci associated with insulin resistance have been hindered in part by methodological limitations. In the present study, by using the newly developed JMA approach, we observed six additional loci associated with fasting insulin levels and other insulin resistance–associated traits. We used the JMA approach, as we hypothesized that the degree of adiposity might mask, positively confound or modify the associations between genetic variants and insulin resistance traits. Although the associations we observed resulted mainly from adjustment for BMI, the JMA method allowed the flexibility and power to observe these main effect–driven associations, while allowing us to simultaneously test our hypothesis of interaction. The identification of these loci offers the potential to further characterize the etiology and causality of T2D and the role of insulin resistance in that process.

URLs. QUICKTEST, http://toby.freeshell.org/software/quicktest.shtml; METAL, http://www.sph.umich.edu/csg/abecasis/Metal/; International HapMap Project, http://hapmap.ncbi.nlm.nih.gov/; SNPper, http://snpper.chip.org/; SNAP, http://www.broadinstitute.org/mpg/snap/ldsearch.php; SIFT dbSNP, http://sift.jcvi.org/www/SIFT_dbSNP.html; OMIM, http://www.ncbi.nlm.nih.gov/omim; GeneCards, http://www.genecards.org/; Microarray data, http://genome.ucsc.edu/cgi-bin/hgNear; R, http://www.rproject.org/; SAS, http://sas.com; ProbABEL, http://www.genabel.org/packages/ProbABEL.

METHODS

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

Note: Supplementary information is available in the online version of the paper.

ACKNOWLEDGMENTS

A full list of acknowledgments is provided in the **Supplementary Note**.

AUTHOR CONTRIBUTIONS

A.K.M., J.D. and J.B.M. conceived of the study, A.K.M. and R.A.S. performed the analysis, A.K.M., M.F.H. and R.A.S. wrote the manuscript, J.B.M. and C. Langenberg directed the work, and J.L.G., N.B.-N., H.C., D.R., C.-T.L., L.F.B., I.P., R.M.W., J.C.F., J.D., J.B.M. and C. Langenberg provided analytical advice and revised the manuscript.

Individuals involved in the management and design of studies contributing to this project were A.A.H., A.F.W., A. Hofman, A. Hamsten, A.Hingorani, A.J., A.P., A.R., A.R.S., A.T., B.A.O., B.F.V., B.P., C. Bouchard, C. Langenberg, C.M.V., D. Boomsma, D.C., D.M.W., D.B.S., D.S.S., E. Boerwinkle, E.d.G., E.I., E. Widen, F.S.C., G.E., G.H., G.V.D., G. Waeber, G. Willamsen, H. Campbell, H.W., I.B.B., I.M., I.R., J.B.M., J.C.F., J.D., J.F.W., J.G.E., J.I.R., J. Kaprio, J. Kuusisto, J. Lindstrom, J.P., J.S.B., J.S.P., J.T., J.W.J., K.L.M., K.O.O., K.O.K., K. Raikkonen, K.S.S., L.F., L.G., L.J.L., L.J.P., L. Peltonen, L. Perusse, M.A.P., M. Perola, M.A., M. Boban,

M. Boehnke, M.G.F., M.I.M., M.-R.J., M.K., M.M., M.S., M.S.R., M.T.M.L., M.U., N.B.N., N.G.F., N.J.W., N.L.G., N.L.P., N.L.S., N.S., O.H., O.P., P.A., P.A.P., P.D., P.G., P. Magnusson, P. Meneton, P.P.P., P.V., P.W.F., R.A.S., R.C., R.M.W., R.N.B., R.J.F.L., R.R., S.B., S. Hercberg, S.H.W., S.L.R.K., S.M.W., S.T., T.B.H., T.L., T.M.F., T.S., U.d.F., U.S., V.G., V.M., V.S., W.K., Y.A. and Y.L.

Individuals involved in sample collection and phenotyping for this project were A.C.A., A. Hamsten, A. Hofman, A.J., A.R., A.S., A.T., B.B., C. Bouchard, C.H., C.M.V., C.S.F., D. Boomsma, D. Budmir, D.C., D.S.S., E. Brunner, E.I., E.J.G.S., G.E., G.H., G. Willamsen, H.W., I.K., I.M., I.R., J.B.M., J.C.M.W., J.F.W., J.G.E., J. Hung, J. Kaprio, J. Lahti, J.M.E., J.P.B., J.H.S., K.L., K.O.K., L.F., L. Lannfelt, L. Lind, L.Z., M.D., M.G.F., M.-R.J., M.K., M. Laakso, M. Pehlic, M. Pericic, M.T.M.L., N.J.W., N.L.P., O.C., O.P., P.F., P.G., P.K., P. Magnusson, P.P.P., P.V., P.W.F., R.C., R.J.L., R.S., S.B., S. Hercberg, S.H.W., S.K., S.K.-K., T.B.H., T.M.F., T.Z., U.d.F., U.S., V.G., V. Lyssenko, V.S. and Y.L.

Individuals involved with genotyping for this project were A.A.H., A.G.U., A.J.S., A.P., B.G., B.S., C. Bellis, C.H., C. Langenberg, C.M.V., C.S.F., D. Boomsma, D.C., D.Z., E. Boerwinkle, E.E., E.I., E. Wheeler, E. Widen, G.M.L., H. Campbell, I.B., J.B., J.D., J.F.W., J. Hui, J.I.R., J.J.H., J. Kaprio, J.P., J.P.B., K.L., K.O.K., K. Stirrups, L.F., L.J.P., L.L.B., L. Perusse, M.A.M., M.C.V., M.D.A., M.-R.J., M.K., M. Pehlic, M. Perola, M.R.E., M.T.M.L., N.B.N., N.L.P., N.N., P.F., P.K., P. Magnusson, P.P.P., P.S.C., R.J.S., S.B., S. Heath, S.L., T.M.F., T.Z., V.V., Y.B. and Y.L.

Individuals involved with statistical analysis for this project were A.G., A.K.M., A.U.J., A.V.S., B.G., B.S., C. Ladenvall, C. Lecoeur, C.T.L., D. Barnes, D.R., D. Shungin, E.I., E.R., E. Wheeler, G.C., G.L., H. Chen, H.Y., I.F., I.P., J.D., J.J.H., J. Lahti, J.O., J.P., J.R.B.P., J.S.A., K.L., K. Rice, L.F.B., L.J.R.T., L. Perusse, M.A.J., M.E.M., M.P., M.S.L., M.T.M.L., M.V., N.A., N.B.N., N.L.G., P.C.D.S., P.N., P.S., P.W.F., R.A.S., R.J.S., R.M., S.K., T.A., T.J., T.T., V. Lagou, V.V., Y.A., Y.L. and Y.V.S.

COMPETING FINANCIAL INTERESTS

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

Published online at http://www.nature.com/doifinder/10.1038/ng.2274.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Billings, L.K. & Florez, J.C. The genetics of type 2 diabetes: what have we learned from GWAS? *Ann. NY Acad. Sci.* **1212**, 59–77 (2010).
2. Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881–885 (2007).
3. Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nat. Genet.* **42**, 105–116 (2010).
4. Voight, B.F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42**, 579–589 (2010).
5. Prudente, S., Morini, E. & Trischitta, V. Insulin signaling regulating genes: effect on T2DM and cardiovascular risk. *Nat. Rev. Endocrinol.* **5**, 682–693 (2009).
6. Kahn, S.E., Hull, R.L. & Utzschneider, K.M. Mechanisms linking obesity to insulin resistance and type 2 diabetes. *Nature* **444**, 840–846 (2006).
7. Wang, X. *et al.* Heritability of insulin sensitivity and lipid profile depend on BMI: evidence for gene-obesity interaction. *Diabetologia* **52**, 2578–2584 (2009).
8. Florez, J.C. *et al.* Effects of the type 2 diabetes–associated *PPARG* P12A polymorphism on progression to diabetes and response to troglitazone. *J. Clin. Endocrinol. Metab.* **92**, 1502–1509 (2007).
9. Ludovico, O. *et al.* Heterogeneous effect of peroxisome proliferator–activated receptor γ2 Ala12 variant on type 2 diabetes risk. *Obesity (Silver Spring)* **15**, 1076–1081 (2007).
10. Cauchi, S. *et al.* The genetic susceptibility to type 2 diabetes may be modulated by obesity status: implications for association studies. *BMC Med. Genet.* **9**, 45 (2008).
11. Trujillo, M.E. & Scherer, P.E. Adipose tissue–derived factors: impact on health and disease. *Endocr. Rev.* **27**, 762–778 (2006).
12. Shoelson, S.E., Lee, J. & Goldfine, A.B. Inflammation and insulin resistance. *J. Clin. Invest.* **116**, 1793–1801 (2006).
13. Kraft, P., Yen, Y.C., Stram, D.O., Morrison, J. & Gauderman, W.J. Exploiting gene-environment interaction to detect genetic associations. *Hum. Hered.* **63**, 111–119 (2007).
14. Manning, A.K. *et al.* Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP × environment regression coefficients. *Genet. Epidemiol.* **35**, 11–18 (2011).
15. Matthews, D.R. *et al.* Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. *Diabetologia* **28**, 412–419 (1985).
16. Speliotes, E.K. *et al.* Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
17. Heid, I.M. *et al.* Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nat. Genet.* **42**, 949–960 (2010).
18. Teslovich, T.M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
19. Rampersaud, E. *et al.* Identification of novel candidate genes for type 2 diabetes from a genome-wide association scan in the Old Order Amish: evidence for replication from diabetes-related quantitative traits and from independent populations. *Diabetes* **56**, 3053–3062 (2007).
20. Stolerman, E.S. *et al.* Haplotype structure of the *ENPP1* gene and nominal association of the K121Q missense single nucleotide polymorphism with glycemic traits in the Framingham Heart Study. *Diabetes* **57**, 1971–1977 (2008).
21. Dehghan, A. *et al.* Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. *Circulation* **123**, 731–738 (2011).
22. Hivert, M.F. *et al.* Insulin resistance influences the association of adiponectin levels with diabetes incidence in two population-based cohorts: the Cooperative Health Research in the Region of Augsburg (KORA) S4/F4 study and the Framingham Offspring Study. *Diabetologia* **54**, 1019–1024 (2011).
23. Florez, J.C. Newly identified loci highlight β cell dysfunction as a key cause of type 2 diabetes: where are the insulin resistance genes? *Diabetologia* **51**, 1100–1110 (2008).
24. Timpson, N.J. *et al.* Adiposity-related heterogeneity in patterns of type 2 diabetes susceptibility observed in genome-wide association data. *Diabetes* **58**, 505–510 (2009).
25. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Optimal designs for two-stage genome-wide association studies. *Genet. Epidemiol.* **31**, 776–788 (2007).
26. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.* **38**, 209–213 (2006).
27. Saxena, R. *et al.* Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
28. Soranzo, N. *et al.* Common variants at 10 genomic loci influence hemoglobin A_{1C} levels via glycemic and nonglycemic pathways. *Diabetes* **59**, 3229–3239 (2010).
29. Strawbridge, R.J. *et al.* Genome-wide association identifies nine common variants associated with fasting proinsulin levels and provides new insights into the pathophysiology of type 2 diabetes. *Diabetes* **60**, 2624–2634 (2011).
30. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467**, 832–838 (2010).
31. Carroll, E.A. *et al.* Cordon-bleu is a conserved gene involved in neural tube formation. *Dev. Biol.* **262**, 16–31 (2003).
32. Depetris, R.S. *et al.* Structural basis for inhibition of the insulin receptor by the adaptor protein Grb14. *Mol. Cell* **20**, 325–333 (2005).
33. Ridker, P.M. *et al.* Polymorphism in the *CETP* gene region, HDL cholesterol, and risk of future myocardial infarction: genomewide analysis among 18 245 initially healthy women from the Women's Genome Health Study. *Circ. Cardiovasc. Genet.* **2**, 26–33 (2009).
34. White, M.F. The IRS-signalling system: a network of docking proteins that mediate insulin action. *Mol. Cell. Biochem.* **182**, 3–11 (1998).
35. Rung, J. *et al.* Genetic variant near *IRS1* is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nat. Genet.* **41**, 1110–1115 (2009).
36. Samani, N.J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
37. Waterworth, D.M. *et al.* Genetic variants influencing circulating lipid levels and risk of coronary artery disease. *Arterioscler. Thromb. Vasc. Biol.* **30**, 2264–2276 (2010).
38. Schadt, E.E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
39. Smith, S., Giriat, I., Schmitt, A. & de Lange, T. Tankyrase, a poly(ADP-ribose) polymerase at human telomeres. *Science* **282**, 1484–1487 (1998).
40. Chi, N.W. & Lodish, H.F. Tankyrase is a golgi-associated mitogen-activated protein kinase substrate that interacts with IRAP in GLUT4 vesicles. *J. Biol. Chem.* **275**, 38437–38444 (2000).
41. Huang, S.M. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
42. Royce, P.M. & Steinmann, B. Prolidase deficiency. *Connective Tissue and Its Heritable Disorders* 727–743 (Wiley-Liss, New York, 2002).
43. Nica, A.C. *et al.* The architecture of gene regulatory variation across multiple human tissues: the MuTHER study. *PLoS Genet.* **7**, e1002003 (2011).
44. The ARIC investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am. J. Epidemiol.* **129**, 687–702 (1989).
45. Bostick, M. *et al.* UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760–1764 (2007).
46. Ng, P.C. & Henikoff, S. Predicting deleterious amino acid substitutions. *Genome Res.* **11**, 863–874 (2001).
47. Evans, R.M., Barish, G.D. & Wang, Y.X. PPARs and the complex journey to obesity. *Nat. Med.* **10**, 355–361 (2004).
48. Reigstad, L.J. *et al.* Platelet-derived growth factor (PDGF)-C, a PDGF family member with a vascular endothelial growth factor–like structure. *J. Biol. Chem.* **278**, 17114–17120 (2003).
49. Handford, C.A. *et al.* The human glycine receptor β subunit: primary structure, functional characterisation and chromosomal localisation of the human and murine genes. *Brain Res. Mol. Brain Res.* **35**, 211–219 (1996).
50. den Hoed, M. *et al.* Genetic susceptibility to obesity and related traits in childhood and adolescence: influence of loci identified by genome-wide association studies. *Diabetes* **59**, 2980–2988 (2010).

51. Hotta, K. *et al.* Polymorphisms in *NRXN3*, *TFAP2B*, *MSRA*, *LYPLAL1*, *FTO* and *MC4R* and their effect on visceral fat area in the Japanese population. *J. Hum. Genet.* **55**, 738–742 (2010).
52. Lindgren, C.M. *et al.* Genome-wide association scan meta-analysis identifies three loci influencing adiposity and fat distribution. *PLoS Genet.* **5**, e1000508 (2009).
53. Seve, M., Chimienti, F., Devergnas, S. & Favier, A. *In silico* identification and expression of *SLC30* family genes: an expressed sequence tag data mining strategy for the characterization of zinc transporters' tissue expression. *BMC Genomics* **5**, 32 (2004).
54. Ohagi, S. *et al.* Human prohormone convertase 3 gene: exon-intron organization and molecular scanning for mutations in Japanese subjects with NIDDM. *Diabetes* **45**, 897–901 (1996).
55. Benzinou, M. *et al.* Common nonsynonymous variants in *PCSK1* confer risk of obesity. *Nat. Genet.* **40**, 943–945 (2008).
56. Miura, K. *et al.* ARAP1: a point of convergence for Arf and Rho signaling. *Mol. Cell* **9**, 109–119 (2002).
57. Clément, S. *et al.* The lipid phosphatase SHIP2 controls insulin sensitivity. *Nature* **409**, 92–97 (2001).
58. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
59. Silva, J.P. *et al.* Regulation of adaptive behaviour during fasting by hypothalamic Foxa2. *Nature* **462**, 646–650 (2009).
60. Xing, C., Cohen, J.C. & Boerwinkle, E. A weighted false discovery rate control procedure reveals alleles at *FOXA2* that influence fasting glucose levels. *Am. J. Hum. Genet.* **86**, 440–446 (2010).
61. Liu, F. & Roth, R.A. Grb-IR: a SH2-domain-containing protein that binds to the insulin receptor and inhibits its function. *Proc. Natl. Acad. Sci. USA* **92**, 10287–10291 (1995).
62. Giovannone, B. *et al.* Two novel proteins that are linked to insulin-like growth factor (IGF-I) receptors by the Grb10 adapter and modulate IGF-I signaling. *J. Biol. Chem.* **278**, 31564–31573 (2003).
63. Murray, M.V., Kobayashi, R. & Krainer, A.R. The type 2C Ser/Thr phosphatase PP2Cγ is a pre-mRNA splicing factor. *Genes Dev.* **13**, 87–97 (1999).
64. Schwitzgebel, V.M. *et al.* Agenesis of human pancreas due to decreased half-life of insulin promoter factor 1. *J. Clin. Endocrinol. Metab.* **88**, 4398–4406 (2003).
65. Stoffers, D.A., Ferrer, J., Clarke, W.L. & Habener, J.F. Early-onset type-II diabetes mellitus (MODY4) linked to *IPF1*. *Nat. Genet.* **17**, 138–139 (1997).
66. Cockburn, B.N. *et al.* Insulin promoter factor–1 mutations and diabetes in Trinidad: identification of a novel diabetes-associated mutation (E224K) in an Indo-Trinidadian family. *J. Clin. Endocrinol. Metab.* **89**, 971–978 (2004).

Alisa K Manning[1–4,163], Marie-France Hivert[5,6,163], Robert A Scott[7,163], Jonna L Grimsby[5,8], Nabila Bouatia-Naji[9,10], Han Chen[1], Denis Rybin[11], Ching-Ti Liu[1], Lawrence F Bielak[12], Inga Prokopenko[13,14], Najaf Amin[15], Daniel Barnes[7], Gemma Cadby[16,17], Jouke-Jan Hottenga[18,19], Erik Ingelsson[20], Anne U Jackson[21], Toby Johnson[22], Stavroula Kanoni[23], Claes Ladenvall[24,25], Vasiliki Lagou[13,14], Jari Lahti[26], Cecile Lecoeur[9,10], Yongmei Liu[27], Maria Teresa Martinez-Larrad[28], May E Montasser[29], Pau Navarro[30], John R B Perry[14,31,32], Laura J Rasmussen-Torvik[33], Perttu Salo[34], Naveed Sattar[35], Dmitry Shungin[24,25,36,37], Rona J Strawbridge[38], Toshiko Tanaka[39], Cornelia M van Duijn[15,40,41], Ping An[42], Mariza de Andrade[43], Jeanette S Andrews[44], Thor Aspelund[45,46], Mustafa Atalay[47], Yurii Aulchenko[15], Beverley Balkau[48,49], Stefania Bandinelli[50], Jacques S Beckmann[51,52], John P Beilby[53–55], Claire Bellis[56], Richard N Bergman[57], John Blangero[56], Mladen Boban[58], Michael Boehnke[21], Eric Boerwinkle[59], Lori L Bonnycastle[60], Dorret I Boomsma[18,19], Ingrid B Borecki[42], Yvonne Böttcher[61], Claude Bouchard[62], Eric Brunner[63], Danijela Budimir[58], Harry Campbell[64], Olga Carlson[65], Peter S Chines[60], Robert Clarke[66], Francis S Collins[60], Arturo Corbatón-Anchuelo[28], David Couper[67], Ulf de Faire[68], George V Dedoussis[69], Panos Deloukas[23], Maria Dimitriou[69], Josephine M Egan[65], Gudny Eiriksdottir[45], Michael R Erdos[60], Johan G Eriksson[70–74], Elodie Eury[9,10], Luigi Ferrucci[75], Ian Ford[76], Nita G Forouhi[7], Caroline S Fox[77,78], Maria Grazia Franzosi[79], Paul W Franks[24,25,36,80,81], Timothy M Frayling[31], Philippe Froguel[9,10,82], Pilar Galan[83], Eco de Geus[18,19], Bruna Gigante[68], Nicole L Glazer[84,85], Anuj Goel[14,86], Leif Groop[24,25], Vilmundur Gudnason[45,46], Göran Hallmans[87], Anders Hamsten[38], Ola Hansson[24,25], Tamara B Harris[88], Caroline Hayward[30], Simon Heath[89], Serge Hercberg[83], Andrew A Hicks[90], Aroon Hingorani[91], Albert Hofman[15,41], Jennie Hui[53–55,92], Joseph Hung[55,93], Marjo-Riitta Jarvelin[94–97], Min A Jhun[12], Paul C D Johnson[76], J Wouter Jukema[98,99], Antti Jula[34], W H Kao[100], Jaakko Kaprio[74,101,102], Sharon L R Kardia[12], Sirkka Keinanen-Kiukaanniemi[103,104], Mika Kivimaki[63], Ivana Kolcic[105], Peter Kovacs[106], Meena Kumari[91], Johanna Kuusisto[107], Kirsten Ohm Kyvik[108,109], Markku Laakso[107], Timo Lakka[47,110], Lars Lannfelt[111], G Mark Lathrop[89], Lenore J Launer[88], Karin Leander[68], Guo Li[112], Lars Lind[113], Jaana Lindstrom[114], Stéphane Lobbens[9,10], Ruth J F Loos[7], Jian'an Luan[7], Valeriya Lyssenko[24,25], Reedik Mägi[13,14,115], Patrik K E Magnusson[20], Michael Marmot[63], Pierre Meneton[116], Karen L Mohlke[117], Vincent Mooser[118], Mario A Morken[60], Iva Miljkovic[119], Narisu Narisu[60], Jeff O'Connell[29], Ken K Ong[7], Ben A Oostra[120], Lyle J Palmer[16,17], Aarno Palotie[2,23,101,121], James S Pankow[122], John F Peden[14,86], Nancy L Pedersen[20], Marina Pehlic[123], Leena Peltonen[23,101,162], Brenda Penninx[124–127], Marijana Pericic[128], Markus Perola[34], Louis Perusse[129], Patricia A Peyser[12], Ozren Polasek[105], Peter P Pramstaller[90], Michael A Province[42], Katri Räikkönen[26], Rainer Rauramaa[110,130], Emil Rehnberg[20], Ken Rice[131], Jerome I Rotter[132], Igor Rudan[64,133], Aimo Ruokonen[134], Timo Saaristo[135,136], Maria Sabater-Lleal[38], Veikko Salomaa[34], David B Savage[137], Richa Saxena[2,138], Peter Schwarz[139], Udo Seedorf[140], Bengt Sennblad[38], Manuel Serrano-Rios[28], Alan R Shuldiner[29,141], Eric J G Sijbrands[142], David S Siscovick[112,143,144], Johannes H Smit[126], Kerrin S Small[32], Nicholas L Smith[143–146], Albert Vernon Smith[45,46], Alena Stančáková[107], Kathleen Stirrups[23], Michael Stumvoll[61,147], Yan V Sun[148], Amy J Swift[60], Anke Tönjes[61,147], Jaakko Tuomilehto[114,149–151], Stella Trompet[98], Andre G Uitterlinden[15,41,142], Matti Uusitupa[152,153], Max Vikström[68], Veronique Vitart[30], Marie-Claude Vohl[154], Benjamin F Voight[2], Peter Vollenweider[155], Gerard Waeber[155], Dawn M Waterworth[118], Hugh Watkins[14,87], Eleanor Wheeler[156],

Elisabeth Widen[101], Sarah H Wild[64], Sara M Willems[15], Gonneke Willemsen[18,19], James F Wilson[64], Jacqueline C M Witteman[15,41], Alan F Wright[30], Hanieh Yaghootkar[31], Diana Zelenika[89], Tatijana Zemunik[123], Lina Zgaga[64,157], DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium[158], The Multiple Tissue Human Expression Resource (MUTHER) Consortium[158], Nicholas J Wareham[7], Mark I McCarthy[13,14,159], Ines Barroso[137,156], Richard M Watanabe[57,160], Jose C Florez[2,8,138,161], Josée Dupuis[1,77], James B Meigs[5,8,164] & Claudia Langenberg[7,164]

[1]Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts, USA. [2]Program in Medical and Population Genetics, Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [3]Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [4]Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. [5]General Medicine Division, Massachusetts General Hospital, Boston, Massachusetts, USA. [6]Department of Medicine, Universite de Sherbrooke, Sherbrooke, Québec, Canada. [7]Medical Research Council (MRC) Epidemiology Unit, Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [8]Department of Medicine, Harvard Medical School, Boston, Massachusetts, USA. [9]Institut Pasteur de Lille, Lille, France. [10]Université Lille Nord de France, Lille, France. [11]Boston University Data Coordinating Center, Boston, Massachusetts, USA. [12]Department of Epidemiology, University of Michigan, Ann Arbor, Michigan, USA. [13]Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford, UK. [14]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [15]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [16]Genetic Epidemiology and Biostatistics Platform, Ontario Institute for Cancer Research, Toronto, Ontario, Canada. [17]Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Toronto, Ontario, Canada. [18]Netherlands Twin Register, VU University, Amsterdam, The Netherlands. [19]Department of Biological Psychology, VU University, Amsterdam, The Netherlands. [20]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [21]Department of Biostatistics and Center for Statistical Genetics, University of Michigan School of Public Health, Ann Arbor, Michigan, USA. [22]Clinical Pharmacology and The Genome Centre, William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [23]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [24]Department of Clinical Sciences, Diabetes and Endocrinology, Lund University, Malmö, Sweden. [25]Lund University Diabetes Centre, Malmö, Sweden. [26]Institute of Behavioural Sciences, University of Helsinki, Helsinki, Finland. [27]Department of Epidemiology and Prevention, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [28]Spanish Biomedical Research Centre in Diabetes and Associated Metabolic Disorders, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos, Madrid, Spain. [29]Division of Endocrinology, Diabetes, and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore, Maryland, USA. [30]MRC Human Genetics Unit, MRC Institute of Genetics & Molecular Medicine (IGMM), University of Edinburgh, Edinburgh, UK. [31]Genetics of Complex Traits, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter, UK. [32]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [33]Department of Preventive Medicine, Northwestern University, Feinberg School of Medicine, Chicago, Illinois, USA. [34]Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [35]Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow, UK. [36]Department of Public Health & Clinical Medicine, Genetic Epidemiology & Clinical Research Group, Umeå University Hospital, Umeå, Sweden. [37]Department of Odontology, Umeå University, Umeå, Sweden. [38]Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden. [39]Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA. [40]Centre for Medical Systems Biology, Netherlands Genomics Initiative, The Hague, The Netherlands. [41]Netherlands Consortium for Healthy Aging, Netherlands Genomics Initiative, Rotterdam, The Netherlands. [42]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, Saint Louis, Missouri, USA. [43]Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, USA. [44]Department of Biostatistical Sciences, Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina, USA. [45]Icelandic Heart Association, Kopavogur, Iceland. [46]University of Iceland, Reykjavik, Iceland. [47]Institute of Biomedicine/Physiology, University of Eastern Finland, Kuopio Campus, Kuopio, Finland. [48]Institut de la Santé et de la Recherche Médicale (INSERM), Centre de recherche en Épidémiologie et Santé des Populations (CESP) Centre for Research in Epidemiology and Population Health, Villejuif, France. [49]University Paris–Sud 11, Unité Mixte de Recherche en Santé (UMRS) 1018, Villejuif, France. [50]Geriatric Unit, Azienda Sanitaria di Firenze, Florence, Italy. [51]Department of Medical Genetics, University of Lausanne, Lausanne, Switzerland. [52]Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [53]PathWest Laboratory of Western Australia, Department of Molecular Genetics, J Block, Queen Elizabeth II (QEII) Medical Centre, Nedlands, Western Australia, Australia. [54]School of Pathology and Laboratory Medicine, The University of Western Australia, Nedlands, Western Australia, Australia. [55]Busselton Population Medical Research Foundation, B Block, QEII Medical Centre, Nedlands, Western Australia, Australia. [56]Texas Biomedical Research Institute, San Antonio, Texas, USA. [57]Department of Physiology & Biophysics, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA. [58]Department of Pharmacology, Faculty of Medicine, University of Split, Split, Croatia. [59]Human Genetics Center, University of Texas Health Science Center at Houston, Houston, Texas, USA. [60]Genome Technology Branch, National Human Genome Research Institute, US National Institutes of Health, Bethesda, Maryland, USA. [61]Integriertes Forschungs- und Behandlungszentrum (IFB) Adiposity Diseases, University of Leipzig, Leipzig, Germany. [62]Pennington Biomedical Research Center, Baton Rouge, Louisiana, USA. [63]Department of Epidemiology & Public Health, University College London, London, UK. [64]Centre for Population Health Sciences, University of Edinburgh, Edinburgh, UK. [65]Laboratory of Clinical Investigation, National Institute on Aging, Baltimore, Maryland, USA. [66]Clinical Trial Service Unit, University of Oxford, Oxford, UK. [67]Department of Biostatistics, University of North Carolina Gillings School of Global Public Health, Chapel Hill, North Carolina, USA. [68]Division of Cardiovascular Epidemiology, Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden. [69]Department of Nutrition–Dietetics, Harokopio University, Athens, Greece. [70]Department of General Practice and Primary Health Care, University of Helsinki, Helsinki, Finland. [71]Unit of General Practice, Helsinki University Central Hospital, Helsinki, Finland. [72]Folkhalsan Research Centre, Helsinki, Finland. [73]Vaasa Central Hospital, Vaasa, Finland. [74]National Institute for Health and Welfare, Helsinki, Finland. [75]Longitudinal Studies Section, Clinical Research Branch, National Institute on Aging, Baltimore, Maryland, USA. [76]Robertson Centre for Biostatistics, University of Glasgow, Glasgow, UK. [77]National Heart, Lung, and Blood Institute's Framingham Heart Study, Framingham, Massachusetts, USA. [78]Division of Endocrinology, Diabetes, and Hypertension, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA. [79]Department of Cardiovascular Research, Mario Negri Institute for Pharmacological Research, Milan, Italy. [80]Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts, USA. [81]Institut National de la Recherche Agronomique, Université Paris, Bobigny Cedex, France. [82]Genomic Medicine, Hammersmith Hospital, Imperial College London, London, UK. [83]INSERM, Université Paris, Bobigny Cedex, France. [84]Department of Medicine, Section of Preventive Medicine and Epidemiology, Boston University School of Medicine, Boston, Massachusetts, USA. [85]Department of Epidemiology, Boston University School of Public Health, Boston, Massachusetts, USA. [86]Department of Cardiovascular Medicine, University of Oxford, Oxford, UK. [87]Department of Public Health & Clinical Medicine, Nutrition Research, Umeå University, Umeå, Sweden. [88]Intramural Research Program, Laboratory of Epidemiology, Demography, and Biometry, National Institute on Aging, Bethesda, Maryland, USA. [89]Centre National de Génotypage, Commissariat à l'Energie Atomique, Institut de Génomique, Evry, France. [90]Institute of Genetic Medicine, European Academy Bozen/Bolzano (EURAC), Bolzano, Italy. (Affiliated Institute of the University Lübeck). [91]Genetic Epidemiology Group, Department of Epidemiology & Public Health, University College London, London, UK. [92]School of Population Health, The University of Western Australia, Nedlands, Western Australia, Australia. [93]Sir Charles Gairdner Hospital Unit, School of Medicine & Pharmacology, University of Western Australia, Nedlands, Western Australia, Australia. [94]Department of Epidemiology and Biostatistics, School of Public Health, MRC–Health Protection Agency (HPA) Centre for Environment and Health, Faculty of Medicine, Imperial College London, London, UK. [95]Institute of Health Sciences, University of Oulu, Oulu, Finland. [96]Biocenter Oulu, University of Oulu, Oulu, Finland. [97]National Institute of Health and Welfare, Oulu, Finland. [98]Department of Cardiology C5-P, Leiden University Medical Center, Leiden, The Netherlands. [99]Durrer Center for Cardiogenetic Research, Amsterdam, The Netherlands. [100]Division of Epidemiology, Johns Hopkins School of Public Health, Baltimore, Maryland, USA. [101]Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland. [102]Department of Public Health, Hjelt Institute, University of Helsinki, Helsinki, Finland. [103]Faculty of Medicine, Institute of Health Sciences, University of Oulu, Oulu, Finland. [104]Unit of General Practice, Oulu University Hospital, Oulu, Finland. [105]Department of Public Health, Faculty of Medicine, University of Split, Split, Croatia. [106]Interdisciplinary Centre for Clinical Research, University of Leipzig, Leipzig, Germany. [107]Department of Medicine, University of Eastern Finland and

Kuopio University Hospital, Kuopio, Finland. [108]Institute of Regional Health Services Research, University of Southern Denmark, Odense, Denmark. [109]Odense Patient data Explorative Network (OPEN), University of Southern Denmark, Odense, Denmark. [110]Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [111]Department of Public Health and Caring Sciences, Uppsala University, Rudbecklaboratoriet, Uppsala, Sweden. [112]Cardiovascular Health Research Unit, University of Washington, Seattle, Washington, USA. [113]Department of Medical Sciences, University Hospital, Uppsala University, Uppsala, Sweden. [114]Diabetes Prevention Unit, Department of Chronic Disease Prevention, National Institute for Health and Welfare, Helsinki, Finland. [115]Estonian Genome Center, University of Tartu, Tartu, Estonia. [116]INSERM, Centre de Recherche des Cordeliers, Paris, France. [117]Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. [118]Division of Genetics, GlaxoSmithKline, Philadelphia, Pennsylvania, USA. [119]Department of Epidemiology, Center for Aging and Population Health, University of Pittsburgh, Pittsburgh, Pennsylvania, USA. [120]Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands. [121]Department of Medical Genetics, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland. [122]Division of Epidemiology and Community Health, University of Minnesota School of Public Health, Minneapolis, Minnesota, USA. [123]Department of Biology, Faculty of Medicine, University of Split, Split, Croatia. [124]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands. [125]Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands. [126]Department of Psychiatry, Neuroscience Campus Amsterdam, VU University Medical Center, Amsterdam, The Netherlands. [127]Institute for Research in Extramural Medicine (EMGO Institute) for Health and Care Research, VU University Medical Center, Amsterdam, The Netherlands. [128]Institute for Anthropological Research, Zagreb, Croatia. [129]Department of Preventive Medicine, Laval University, Quebec City, Quebec, Canada. [130]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [131]Department of Biostatistics, University of Washington, Seattle, Washington, USA. [132]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, Calfornia, USA. [133]Centre for Global Health, University of Split, Split, Croatia. [134]Institute of Clinical Medicine, University of Oulu, Oulu, Finland. [135]Finnish Diabetes Association, Tampere, Finland. [136]Pirkanmaa Hospital District and Diabetes Center, Tampere, Finland. [137]Metabolic Research Laboratories, Institute of Metabolic Science, University of Cambridge, Addenbrooke's Hospital, Cambridge, UK. [138]Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. [139]Department of Medicine, Division Prevention and Care of Diabetes, University of Dresden, Dresden, Germany. [140]Leibniz Institute for Arteriosclerosis Research, University of Munster, Munster, Germany. [141]Geriatric Research and Education Clinical Center, Veterans Administration Medical Center, Baltimore, Maryland, USA. [142]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [143]Department of Medicine, University of Washington, Seattle, Washington, USA. [144]Department of Epidemiology, University of Washington, Seattle, Washington, USA. [145]Group Health Research Institute, Group Health Cooperative, Seattle, Washington, USA. [146]Seattle Epidemiologic Research and Information Center, Veterans Affairs Office of Research and Development, Seattle, Washington, USA. [147]Division of Endocrinology and Diabetes, Department of Medicine, University of Leipzig, Leipzig, Germany. [148]Department of Epidemiology, Emory University, Atlanta, Georgia, USA. [149]South Ostrobothnia Central Hospital, Seinäjoki, Finland. [150]Hospital Universitario La Paz, Madrid, Spain. [151]Centre for Vascular Prevention, Danube–University Krems, Krems, Austria. [152]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [153]Research Unit, Kuopio University Hospital, Kuopio, Finland. [154]Department of Food Science and Nutrition, Laval University, Quebec City, Quebec, Canada. [155]Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois, Lausanne, Switzerland. [156]Metabolic Disease Group, Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. [157]Department of Medical Statistics, Epidemiology and Medical Informatics, University of Zagreb, Zagreb, Croatia. [158]A full list of members is provided in the **Supplementary Note**. [159]Oxford National Institute for Health Research (NIHR) Biomedical Research Centre, Churchill Hospital, Oxford, UK. [160]Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, California, USA. [161]Diabetes Unit, Diabetes Research Center, Massachusetts General Hospital, Boston, Massachusetts, USA. [162]Deceased. [163]These authors contributed equally to this work. [164]These authors jointly directed this work. Correspondence should be addressed to J.B.M. (jmeigs@partners.org) or C. Langenberg (claudia.langenberg@mrc-epid.cam.ac.uk).

ONLINE METHODS

Cohorts. We recruited 29 studies involved in MAGIC for the first phase of discovery analysis, totaling up to 58,070 individuals of European ancestry. Many of these studies were involved in the previous MAGIC effort on fasting glucose and fasting insulin levels[3]. In the second phase, follow-up included 23 cohorts with fasting glucose data and 22 cohorts with fasting insulin data, comprising up to 38,422 and 33,823 non-diabetic individuals of European ancestry, respectively (detailed discovery and follow-up cohort characteristics are presented in **Supplementary Table 1**). We excluded individuals with T2D on the basis of known T2D or anti-diabetic treatment and/or a fasting glucose level of ≥7 mM, so as to remove their influence on misclassification of trait levels[67]. Local research ethic committees approved all studies, and all participants gave informed consent to each original study.

Phenotypes and quantitative traits. Fasting glucose and fasting insulin levels were measured from whole blood, plasma or serum using assays specific for each cohort (**Supplementary Table 1**). Homeostatic model assessment indices HOMA-B and HOMA-IR were derived from fasting glucose and fasting insulin measures[15]. Fasting insulin, HOMA-B and HOMA-IR measures were log transformed for analyses. Anthropometric measurements, such as BMI (kg/m^2), were obtained in each cohort, following standardized procedures. Trait values were not imputed, and outliers were not excluded.

Phase 1 discovery genotyping. All cohorts included in the discovery phase had genome-wide arrays performed, the details of which are specified (**Supplementary Table 1**). We implemented the following quality control inclusion criteria for the SNPs in all cohorts: (i) Hardy-Weinberg equilibrium P value of $>1 \times 10^{-6}$; (ii) minor allele frequency of >1% and (iii) SNP call rate of >95%. Imputation was performed using MACH[68,69] (r^2 >0.3) or IMPUTE[70,71] (proper_info > 0.3) on the basis of the HapMap Utah residents of Northern and Western European ancestry (CEU) population (Build 36). Both genotyped and imputed SNPs were used in the analysis, with priority given to genotyped SNPs when both were available. Consequently, up to 2.4 million SNPs (genotyped or imputed) were included in the discovery phase meta-analyses.

Phase 2 follow-up genotyping. Studies were invited to participate in the follow-up phase of the analysis through either *in silico* searches with existing genotype data or *de novo* genotyping. A large percentage of the follow-up cohorts used the Illumina Cardio-Metabochip. Study-specific details for follow-up genotyping are described (**Supplementary Table 1**).

Statistical models. Each study submitted regression summary statistics for the meta-analyses, all of which assumed an additive genetic effect and were adjusted at a minimum for age, sex and study-level covariates (**Supplementary Table 1**). The first regression model included SNP, BMI and SNP × BMI terms to allow for SNP × BMI interaction effects (model 1). This regression model was used in the JMA and is considered the primary analysis. Meta-analyses were performed on additional regression models solely to characterize the associations observed in the JMA as being driven by either SNP main effects or by SNP × BMI interactions. Main effects models estimating the effect of each SNP on the dependent variables were performed, both with (model 2) and without (model 3) adjustment for BMI. Additionally, the continuous BMI measure was dichotomized at 28 kg/m^2; this cutoff was chosen on the basis of the median BMI in of the largest cohorts included in our discovery analysis. An interaction model using the dichotomized BMI was fit (model 4), and stratified models were used to obtain stratum-specific estimates in leaner (BMI < 28 kg/m^2; model 5) and heavier (BMI ≥ 28 kg/m^2; model 6) individuals. For models 1 and 4, both the SNP and SNP × BMI regression coefficients and their covariance matrix were reported by each study, and robust variance estimates were used to correct the observed inflation of the false positive rate for the interaction P value[72].

In phase 1, models 1–3 were performed across the genome, whereas models 4–6 were performed on SNPs with discovery JMA P values of $<1 \times 10^{-5}$. For phase 2, models 1–6 were performed using SNPs taken forward to the follow-up stage. Regression statistics were obtained with R, SAS, QUICKTEST or ProbABEL[73] (**Supplementary Table 1**).

The models were fit as

Model 1. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \beta_{BMI} \times BMI + \beta_{SNP \times BMI} \times SNP \times BMI + \varepsilon$ (BMI as continuous variable)

Model 2. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \varepsilon$

Model 3. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \beta_{BMI} \times BMI + \varepsilon$

Model 4. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \beta_{BMI\ dichotomous} \times BMI + \beta_{SNP \times BMI\ dichotomous} \times SNP \times BMI + \varepsilon$ (BMI as a dichotomous variable)

Model 5. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \varepsilon$ (in individuals of BMI < 28)

Model 6. $Y = \beta_0 + \beta_C \times Cov + \beta_{SNP} \times SNP + \varepsilon$ (in individuals of BMI ≥ 28),

where Y is the continuous outcome, Cov is a matrix of covariates, such as age, sex and additional study–specific adjustments, and SNP is a single-nucleotide polymorphism with additive genetic coding.

Meta-analyses in phases 1 and 2. Our objective was to identify genetic loci associated with fasting insulin, as a surrogate for insulin resistance, which may be masked by variation in or interaction with BMI. In the primary analysis, we performed a JMA of both the SNP effect and the SNP × BMI interaction effect from model 1, using a recently developed method[14]. The JMA is an effective screening tool when the underlying interaction model is unknown and, notably, retains power when there is no interaction effect. The JMA provides estimates of β_{SNP} and $\beta_{SNP \times BMI}$ and allows for a 2-degree-of-freedom joint test of the null hypothesis H_0: $\beta_{SNP} = \beta_{SNP \times BMI} = 0$, as well as a test of heterogeneity of regression coefficients. Additionally, the JMA detects association in the presence of either a significant SNP effect, adjusted for BMI, or a SNP × BMI interaction effect. We implemented the JMA method in METAL (v. 2010-02-08, with a provided patch[14]). SNPs that reached a discovery significance threshold in the JMA of $P < 1 \times 10^{-5}$ with $P_{het} > 0.001$ and were available in at least one-third of the total sample size were taken forward to follow-up analyses. In order to characterize the association signals observed in the JMA, additional meta-analyses were performed that examined main effects and interaction effects separately. Inverse-variance weighted meta-analysis[74] and the heterogeneity test[75] were applied in the univariate meta-analyses of the SNP main effects in models 2, 3, 5 and 6 and the SNP × BMI interaction effect from models 1 and 4.

In each genomic region identified from the JMA, the index SNP was chosen as the SNP with the strongest association (lowest P value) unless the region contained a SNP known to be associated with a metabolic condition or trait (with $P < 1 \times 10^{-5}$), in which case the latter SNP was selected. In case a follow-up cohort was not able to provide results for an index SNP, several proxy SNPs were chosen using LD information from HapMap and the 1000 Genomes Project[76]. SNPs identified by the JMA but which had previously reported associations with diabetes and/or diabetes-related quantitative traits in MAGIC were not selected for follow-up analyses and are reported in **Supplementary Table 2**.

Models 1–6 were performed within each follow-up study, and meta-analyses of the SNPs were performed within the follow-up cohorts and in a combined meta-analysis including the discovery cohorts. Meta-analyses were performed with proxy SNPs if the index SNP was not available. A SNP was considered to be significantly associated if the JMA in the combined sample yielded $P < 5 \times 10^{-8}$. Described effect estimates (**Supplementary Table 3**) were obtained from a combined meta-analysis that excluded proxy SNPs.

Two of the index SNPs (near *CHL1* and *TUBA3C*) showed significant associations in the discovery JMA. We excluded these loci from the combined analysis because we observed low P_{het} values and because neither the main effects nor the interaction models supported the associations observed in the JMA at these loci.

Conditional analyses. We conducted conditional analyses for index SNPs that were within approximately 1 Mb of established SNPs (known associations with T2D or glycemic traits). We conditioned on the established SNPs for a chromosome-specific analysis in the Framingham Heart Study, one of the largest cohorts in the meta-analyses in which the associations of both the established SNPs and the index SNP were observed. If the SNP retained nominal significance in the conditional analysis, we considered the association signal to be not solely caused by LD with the established SNP. Because only one study was involved in this conditional analysis, we used a significance threshold of 0.05 to account for the limited power.

doi:10.1038/ng.2274

Associations with related metabolic traits. As we were interested in the effect of the index SNPs on insulin resistance–related traits, associations of each index SNP with related glycemic traits and T2D were sought using results from previous DIAGRAM and MAGIC meta-analyses for T2D[4], 2-hour glucose[27] and glycated hemoglobin[28]. We also performed searches of these SNPs in publicly available data from previous meta-analyses of lipid quantitative traits[18]. Associations with BMI[16] and WHR adjusted for BMI[17] were also sought in published GIANT meta-analyses.

Functional exploration. To determine whether any index SNPs were in LD with nearby functional coding variants, we used SNPper[77] to extract all SNPs that fell within 500 kb of the index SNP (1 Mb total). From this list, we chose all coding SNPs and used SNAP to determine whether any of these coding SNPs were in LD with the index SNP (CEU, 1000 Genomes Project; 500-kb distance). We used the online tool SIFT dbSNP to predict potential damage to the protein.

To establish whether index SNPs were in LD with known CNVs, we used a database of 7,411 SNPs that map to 3,188 CNVs from the Wellcome Trust Case Control Consortium (WTCCC)[78] in addition to a list of 422 SNPs mapping to 261 deletions[79]. All known SNPs in LD with these marker SNPs were retrieved using the SNAP tool (CEU, 1000 Genomes; within 500 kb). We determined whether any index SNPs were in LD with those SNPs mapping to CNVs.

Expression studies and eQTLs. We searched for genes potentially implicated in glycemic regulation and insulin resistance in the flanking regions 500 kb on each side of the index SNPs (1 Mb total). Functions of these genes were investigated using online resources, such as Online Mendelian Inheritance in Man (OMIM) or GeneCards V3.

We examined expression levels of candidate genes nearest the index SNPs and within 500 kb of the index SNPs using available online microarray expression data (Genomics Institute of the Novartis Research Foundation (GNF) Expression Atlas 2 Data from U133A and GNF1H Chips). Potential candidates for association with fasting insulin levels were further pursued with eQTL studies in available data sets.

We queried all SNPs in the 1-Mb region surrounding the index SNPs for fasting insulin levels to see whether they were associated with expression levels of any genes in a liver tissue gene expression database[38], which used a sample size of 427 subjects of European ancestry and where SNPs with association *P* values less than 0.003 were listed. We used the SNAP tool to determine whether eQTLs were in LD with index SNPs for fasting insulin. Index SNPs for fasting insulin were also queried in the Multiple Tissue Human Expression Resource (MuTHER) expression database[43] to determine whether any were eQTLs in subcutaneous adipose tissue ($n = 776$). Further eQTL searches were performed for top eQTL findings for any SNPs located within 1 Mb of each index SNP ($P < 1 \times 10^{-3}$). Data were analyzed in GenABEL and ProbABEL[73]. A genome-wide false discovery rate of 1% corresponded to a *P*-value threshold of 5.1×10^{-5}, and data were corrected for multiple testing.

67. Rasmussen-Torvik, L.J. *et al.* Impact of repeated measures and sample selection on genome-wide association studies of fasting glucose. *Genet. Epidemiol.* **34**, 665–673 (2010).
68. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu. Rev. Genomics Hum. Genet.* **10**, 387–406 (2009).
69. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
70. 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).
71. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
72. Voorman, A., Lumley, T., McKnight, B. & Rice, K. Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS ONE* **6**, e19416 (2011).
73. Aulchenko, Y.S., Struchalin, M.V. & van Duijn, C.M. ProbABEL package for genome-wide association analysis of imputed data. *BMC Bioinformatics* **11**, 134 (2010).
74. Petitti, D.B. Statistical methods in meta-analysis. in *Meta-analysis, Decision Analysis, and Cost-effectiveness Analysis* 90–114 (Oxford University Press, New York, 1994).
75. Higgins, J.P. & Thompson, S.G. Quantifying heterogeneity in a meta-analysis. *Stat. Med.* **21**, 1539–1558 (2002).
76. The 1000 Genomes Project Consortium. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
77. Riva, A. & Kohane, I.S. A SNP-centric database for the investigation of the human genome. *BMC Bioinformatics* **5**, 33 (2004).
78. The Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* **464**, 713–720 (2010).
79. McCarroll, S.A. *et al.* Common deletion polymorphisms in the human genome. *Nat. Genet.* **38**, 86–92 (2006).