LUND UNIVERSITY

Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility.

Wessel, Jennifer; Chu, Audrey Y; Willems, Sara M; Wang, Shuai; Yaghootkar, Hanieh; Brody, Jennifer A; Dauriz, Marco; Hivert, Marie-France; Raghavan, Sridharan; Lipovich, Leonard; Hidalgo, Bertha; Fox, Keolu; Huffman, Jennifer E; An, Ping; Lu, Yingchang; Rasmussen-Torvik, Laura J; Grarup, Niels; Ehm, Margaret G; Li, Li; Baldridge, Abigail S; Stančáková, Alena; Abrol, Ravinder; Besse, Céline; Boland, Anne; Bork-Jensen, Jette; Fornage, Myriam; Freitag, Daniel F; Garcia, Melissa E; Guo, Xiuqing; Hara, Kazuo; Isaacs, Aaron; Jakobsdottir, Johanna; Lange, Leslie A; Layton, Jill C; Li, Man; Hua Zhao, Jing; Meidtner, Karina; Morrison, Alanna C; Nalls, Mike A; Peters, Marjolein J; Sabater-Lleal, Maria; Schurmann, Claudia; Silveira, Angela; Smith, Albert V; Southam, Lorraine; Stoiber, Marcus H; Strawbridge, Rona J; Taylor, Kent D; V Varga, Tibor; Allin, Kristine H

Published in:
Nature Communications

DOI:
10.1038/ncomms6897

2015

Link to publication

Citation for published version (APA):
Wessel, J., Chu, A. Y., Willems, S. M., Wang, S., Yaghootkar, H., Brody, J. A., Dauriz, M., Hivert, M.-F., Raghavan, S., Lipovich, L., Hidalgo, B., Fox, K., Huffman, J. E., An, P., Lu, Y., Rasmussen-Torvik, L. J., Grarup, N., Ehm, M. G., Li, L., ... Goodarzi, M. O. (2015). Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nature Communications*, *6*, Article 5897. https://doi.org/10.1038/ncomms6897

Total number of authors:
204

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 20. Feb. 2026

nature COMMUNICATIONS

ARTICLE

Received 5 Jun 2014 | Accepted 12 Nov 2014 | Published 29 Jan 2015

DOI: 10.1038/ncomms6897

OPEN

Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility

Jennifer Wessel, Audrey Y. Chu, Sara M. Willems, Shuai Wang *et al.*[#]

Fasting glucose and insulin are intermediate traits for type 2 diabetes. Here we explore the role of coding variation on these traits by analysis of variants on the HumanExome BeadChip in 60,564 non-diabetic individuals and in 16,491 T2D cases and 81,877 controls. We identify a novel association of a low-frequency nonsynonymous SNV in *GLP1R* (A316T; rs10305492; MAF = 1.4%) with lower FG ($\beta = -0.09 \pm 0.01\,\mathrm{mmol\,l^{-1}}$, $P = 3.4 \times 10^{-12}$), T2D risk (OR[95%CI] = 0.86[0.76–0.96], $P = 0.010$), early insulin secretion ($\beta = -0.07 \pm 0.035$ $\mathrm{pmol_{insulin}\ mmol_{glucose}^{-1}}$, $P = 0.048$), but higher 2-h glucose ($\beta = 0.16 \pm 0.05\,\mathrm{mmol\,l^{-1}}$, $P = 4.3 \times 10^{-4}$). We identify a gene-based association with FG at *G6PC2* ($p_{\mathrm{SKAT}} = 6.8 \times 10^{-6}$) driven by four rare protein-coding SNVs (H177Y, Y207S, R283X and S324P). We identify rs651007 (MAF = 20%) in the first intron of *ABO* at the putative promoter of an antisense lncRNA, associating with higher FG ($\beta = 0.02 \pm 0.004\,\mathrm{mmol\,l^{-1}}$, $P = 1.3 \times 10^{-8}$). Our approach identifies novel coding variant associations and extends the allelic spectrum of variation underlying diabetes-related quantitative traits and T2D susceptibility.

Correspondence and requests for materials should be addressed to R.A.S. (email: robert.scott@mrc-epid.cam.ac.uk) or to M.O.G. (email: mark.goodarzi@cshs.org). [#]A full list of authors and their affiliations appears at the end of the paper.

Genome-wide association studies (GWAS) highlight the role of common genetic variation in quantitative glycaemic traits and susceptibility to type 2 diabetes (T2D)[1,2]. However, recent large-scale sequencing studies report that rapid expansions in the human population have introduced a substantial number of rare genetic variants[3,4], with purifying selection having had little time to act, which may harbour larger effects on complex traits than those observed for common variants[3,5,6]. Recent efforts have identified the role of low frequency and rare coding variation in complex disease and related traits[7–10], and highlight the need for large sample sizes to robustly identify such associations[11]. Thus, the Illumina HumanExome BeadChip (or exome chip) has been designed to allow the capture of rare ($\text{MAF}<1\%$), low frequency ($\text{MAF}=1\text{–}5\%$) and common ($\text{MAF}\geq 5\%$) exonic single nucleotide variants (SNVs) in large sample sizes.

To identify novel coding SNVs and genes influencing quantitative glycaemic traits and T2D, we perform meta-analyses of studies participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE[12]) T2D-Glycemia Exome Consortium[13]. Our results show a novel association of a low frequency coding variant in *GLP1R*, a gene encoding a drug target in T2D therapy (the incretin mimetics), with FG and T2D. The minor allele is associated with lower FG, lower T2D risk, lower insulin response to a glucose challenge and higher 2-h glucose, pointing to physiological effects on the incretin system. Analyses of non-synonymous variants also enable us to identify particular genes likely to underlie previously identified associations at six loci associated with FG and/or FI (*G6PC2, GPSM1, SLC2A2, SLC30A8, RREB1* and *COBLL1)* and five with T2D (*ARAP1*, *GIPR*, *KCNJ11, SLC30A8* and *WFS1*). Further, we found non-coding variants whose putative functions in epigenetic and post-transcriptional regulation of *ABO* and *G6PC2* are supported by experimental ENCODE Consortium, GTEx and transcriptome data from islets. In conclusion, our approach identifies novel coding and non-coding variants and extends the allelic and functional spectrum of genetic variation underlying diabetes-related quantitative traits and T2D susceptibility.

Results

An overview of the study design is shown in Supplementary Fig. 1, and participating studies and their characteristics are detailed in Supplementary Data 1. We conducted single variant and gene-based analyses for fasting glucose (FG) and fasting insulin (FI), by combining data from 23 studies comprising up to 60,564 (FG) and 48,118 (FI) non-diabetic individuals of European and African ancestry. We followed up associated variants at novel and known glycaemic loci by tests of association with T2D, additional physiological quantitative traits (including post-absorptive glucose and insulin dynamic measures), pathway analyses, protein conformation modelling, comparison with whole-exome sequence data and interrogation of functional annotation resources including ENCODE[14,15] and GTEx[16]. We performed single-variant analyses using additive genetic models of 150,558 SNVs (P value for significance $\leq 3\times 10^{-7}$) restricted to $\text{MAF}>0.02\%$ (equivalent to a minor allele count (MAC) ≥ 20), and gene-based tests using Sequence Kernel Association (SKAT) and Weighted Sum Tests (WST) restricted to variants with $\text{MAF}<1\%$ in a total of 15,260 genes (P value for significance $\leq 2\times 10^{-6}$, based on number of gene tests performed). T2D case/control analyses included 16,491 individuals with T2D and 81,877 controls from 22 studies (Supplementary Data 2).

Novel association of a GLP1R variant with glycaemic traits. We identified a novel association of a nonsynonymous SNV (nsSNV) (A316T, rs10305492, $\text{MAF}=1.4\%$) in the gene encoding the receptor for glucagon-like peptide 1 (*GLP1R),* with the minor (A) allele associated with lower FG ($\beta=-0.09\pm 0.01\,\text{mmol}\,\text{l}^{-1}$ (equivalent to 0.14 SDs in FG), $P=3.4\times 10^{-12}$, variance explained = 0.03%, Table 1 and Fig. 1), but not with FI ($P=0.67$, Supplementary Table 1). GLP-1 is secreted by intestinal L-cells in response to oral feeding and accounts for a major proportion of the so-called 'incretin effect', that is, the augmentation of insulin secretion following an oral glucose challenge relative to an intravenous glucose challenge. GLP-1 has a range of downstream actions including glucose-dependent stimulation of insulin release, inhibition of glucagon secretion from the islet alpha-cells, appetite suppression and slowing of gastrointestinal motility[17,18]. In follow-up analyses, the FG-lowering minor A allele was associated with lower T2D risk (OR [95%CI] = 0.86 [0.76–0.96], $P=0.010$, Supplementary Data 3). Given the role of incretin hormones in post-prandial glucose regulation, we further investigated the association of A316T with measures of post-challenge glycaemia, including 2-h glucose, and 30 min-insulin and glucose responses expressed as the insulinogenic index[19] in up to 37,080 individuals from 10 studies (Supplementary Table 2). The FG-lowering allele was associated with higher 2-h glucose levels (β in SDs per-minor allele [95%CI]: 0.10 [0.04, 0.16], $P=4.3\times 10^{-4}$, $N=37{,}068$) and lower insulinogenic index (-0.09 [-0.19, -0.00], $P=0.048$, $N=16{,}203$), indicating lower early insulin secretion (Fig. 1). Given the smaller sample size, these associations are less statistically compelling; however, the directions of effect indicated by their beta values are comparable to those observed for fasting glucose. We did not find a significant association between A316T and the measure of 'incretin effect', but this was only available in a small sample size of 738 non-diabetic individuals with both oral and intravenous glucose tolerance test data (β in SDs per-minor allele [95%CI]: 0.24 [-0.20–0.68], $P=0.28$, Fig. 1 and Supplementary Table 2). We did not see any association with insulin sensitivity estimated by euglycaemic-hyperinsulinemic clamp or frequently sampled IV glucose tolerance test (Supplementary Table 3). While stimulation of the GLP-1 receptor has been suggested to reduce appetite[20] and treatment with GLP1R agonists can result in

Table 1 | Novel SNPs associated with fasting glucose in African and European ancestries combined.

Gene	Variation type	Chr	Build 37 position	dbSNPID	Alleles		African and European				Proportion of trait variance explained
					Effect	Other	EAF	Beta	s.e.	P	
GLP1R	A316T	6	39046794	rs10305492	A	G	0.01	-0.09	0.013	3.4×10^{-12}	0.0003
ABO	intergenic	9	136153875	rs651007	A	G	0.20	0.02	0.004	1.3×10^{-8}	0.0002

EAF, effect allele frequency.
Fasting glucose concentrations were adjusted for sex, age, cohort effects and up to 10 principal components in up to 60,564 (AF $N=9{,}664$ and EU $N=50{,}900$) non-diabetic individuals. Effects are reported per copy of the minor allele. Beta coefficient units are in $\text{mmol}\,\text{l}^{-1}$.

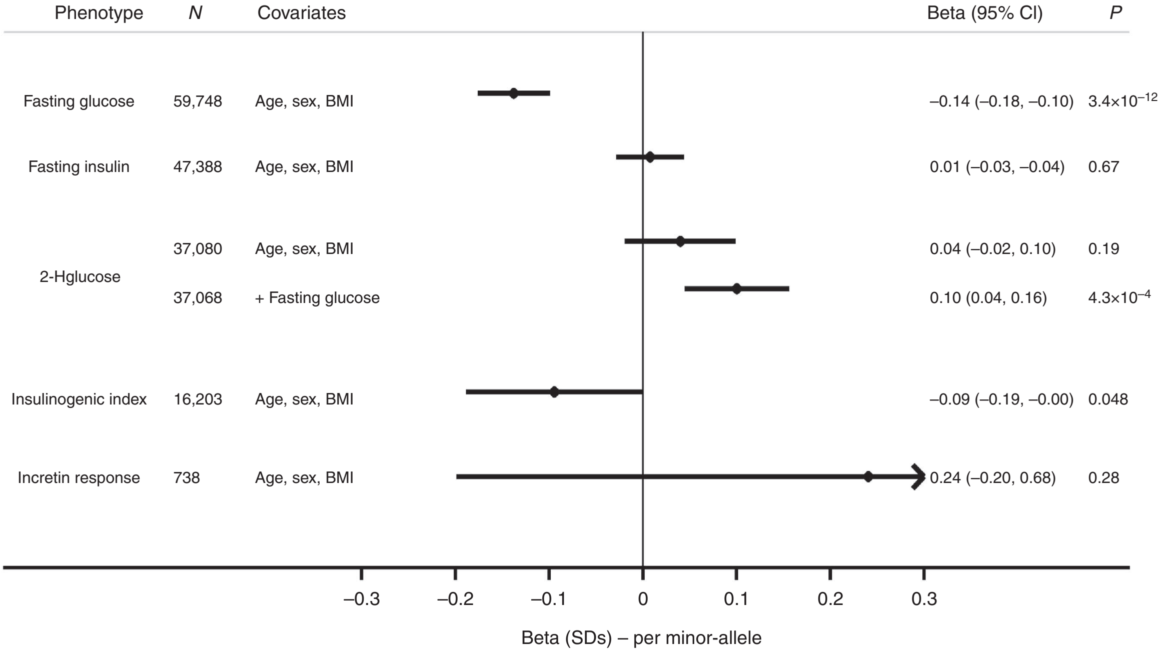


Figure 1 | Glycaemic associations with rs10305492 (*GLP1R* A316T). Glycaemic phenotypes were tested for association with rs10305492 in *GLP1R* (A316T). Each phenotype, sample size (*N*), covariates in each model, beta per s.d., 95% confidence interval (95%CI) and *P* values (*P*) are reported. Analyses were performed on native distributions and scaled to s.d. values from the Fenland or Ely studies to allow comparisons of effect sizes across phenotypes.

reductions in BMI[21], these potential effects are unlikely to influence our results, which were adjusted for BMI.

In an effort to examine the potential functional consequence of the *GLP1R* A316T variant, we modelled the A316T receptor mutant structure based on the recently published[22] structural model of the full-length human GLP-1 receptor bound to exendin-4 (an exogenous GLP-1 agonist). The mutant structural model was then relaxed in the membrane environment using molecular dynamics simulations. We found that the T316 variant (in transmembrane (TM) domain 5) disrupts hydrogen bonding between N320 (in TM5) and E364 (TM6) (Supplementary Fig. 2). In the mutant receptor, T316 displaces N320 and engages in a stable interaction with E364, resulting in slight shifts of TM5 towards the cytoplasm and TM6 away from the cytoplasm (Supplementary Figs 3 and 4). This alters the conformation of the third intracellular loop, which connects TM5 and TM6 within the cell, potentially affecting downstream signalling through altered interaction with effectors such as G proteins.

A targeted Gene Set Enrichment Analysis (Supplementary Table 4) identified enrichment of genes biologically related to *GLP1R* in the incretin signalling pathway ($P = 2 \times 10^{-4}$); after excluding *GLP1R* and previously known loci *PDX1*, *GIPR* and *ADCY5*, the association was attenuated ($P = 0.072$). Gene-based tests at *GLP1R* did not identify significant associations with glycaemic traits or T2D susceptibility, further supported by Fig. 2, which indicates only one variant in the *GLP1R* region on the exome chip showing association with FG.

To more fully characterize the extent of local sequence variation and its association with FG at *GLP1R*, we investigated 150 *GLP1R* SNVs identified from whole-exome sequencing in up to 14,118 individuals available in CHARGE and the GlaxoSmithKline discovery sequence project (Supplementary Table 5). Single-variant analysis identified association of 12 other SNVs with FG ($P < 0.05$; Supplementary Data 4), suggesting that additional variants at this locus may influence FG, including two variants (rs10305457 and rs761386) in close proximity to splice sites that raise the possibility that their functional impact is exerted via effects on *GLP1R* pre-mRNA splicing. However, the smaller sample size of the sequence data limits power for firm conclusions.

Association of noncoding variants in ABO with glycaemic traits. We also newly identified that the minor allele A at rs651007 near the *ABO* gene was associated with higher FG ($\beta = 0.02 \pm 0.004$ mmol l^{-1}, MAF = 20%, $P = 1.3 \times 10^{-8}$, variance explained = 0.02%, Table 1). Three other associated common variants in strong linkage disequilibrium (LD) ($r^2 = 0.95$–1) were also located in this region; conditional analyses suggested that these four variants reflect one association signal (Supplementary Table 6). The FG-raising allele of rs651007 was nominally associated with increased FI ($\beta = 0.008 \pm 0.003$, $P = 0.02$, Supplementary Table 1) and T2D risk (OR [95%CI] = 1.05 [1.01–1.08], $P = 0.01$, Supplementary Data 3). Further, we independently replicated the association at this locus with FG in non-overlapping data from MAGIC[1] using rs579459, a variant in LD with rs651007 and genotyped on the Illumina CardioMetabochip ($\beta = 0.008 \pm 0.003$ mmol l^{-1}, $P = 5.0 \times 10^{-3}$; $N_{\text{MAGIC}} = 88{,}287$). The FG-associated SNV at *ABO* was in low LD with the three variants[23] that distinguish between the four major blood groups O, A_1, A_2 and B (rs8176719 $r^2 = 0.18$, rs8176749 $r^2 = 0.01$ and rs8176750 $r^2 = 0.01$). The blood group variants (or their proxies) were not associated with FG levels (Supplementary Table 7).

Variants in the *ABO* region have been associated with a number of cardiovascular and metabolic traits in other studies (Supplementary Table 8), suggesting a broad role for this locus in cardiometabolic risk. A search of the four FG-associated variants and their associations with metabolic traits using data available through other CHARGE working groups (Supplementary Table 9) revealed a significant association of rs651007 with BMI in women ($\beta = 0.025 \pm 0.01$ kg m^{-2}, $P = 3.4 \times 10^{-4}$) but not in men. As previously reported[24,25], the FG increasing allele of rs651007 was associated with increased LDL and TC (LDL: $\beta = 2.3 \pm 0.28\,\mathrm{mg\,dl^{-1}}$, $P = 6.1 \times 10^{-16}$; TC: $\beta = 2.4 \pm 0.33\,\mathrm{mg\,dl^{-1}}$, $P = 3.4 \times 10^{-13}$). As the FG-associated *ABO* variants were located in non-coding regions (intron 1 or intergenic) we interrogated public regulatory annotation data sets, GTEx[16] (http://www.gtexportal.org/home/) and the ENCODE Consortium resources[14] in the UCSC Genome Browser[15] (http://genome.ucsc.edu/) and identified a number of genomic features coincident with each of the four FG-associated variants. Three of these SNPs, upstream of the ABO promoter, reside in a DNase I hypersensitive site with canonical enhancer marks in ENCODE Consortium data: H3K4Me1 and H3K27Ac (Supplementary Fig. 5). We analysed all SNPs with similar annotations, and found that these three are coincident with DNase, H3K4Me1 and H3K27Ac values each near the genome-wide mode of these assays (Supplementary Fig. 6). Indeed, in haematopoietic model K562 cells, the ENCODE Consortium has identified the region overlapping these SNPs as a putative enhancer[14]. Interrogating the GTEx database ($N = 156$), we found that rs651007 ($P = 5.9 \times 10^{-5}$) and rs579459 ($P = 6.7 \times 10^{-5}$) are eQTLs for *ABO*, and rs635634 ($P = 1.1 \times 10^{-4}$) is an eQTL for *SLC2A6* in whole blood (Supplementary Table 10). The fourth SNP, rs507666, resides near the transcription start site of a long non-coding RNA that is antisense to exon 1 of *ABO* and expressed in pancreatic islets (Supplementary Fig. 5). rs507666 was also an eQTL for the glucose transporter *SLC2A6* ($P = 1.1 \times 10^{-4}$) (Supplementary Fig. 5 and Supplementary Table 10). *SLC2A6* codes for a glucose transporter whose relevance to glycaemia and T2D is largely unknown, but expression is increased in rodent models of diabetes[26]. Gene-based analyses did not reveal significant quantitative trait associations with rare coding variation in *ABO*.

Rare variants in G6PC2 are associated with fasting glucose. At the known glycaemic locus *G6PC2*, gene-based analyses of 15 rare predicted protein-altering variants (MAF < 1%) present on the exome chip revealed a significant association of this gene with FG (cumulative MAF of 1.6%, $p_{\mathrm{SKAT}} = 8.2 \times 10^{-18}$, $p_{\mathrm{WST}} = 4.1 \times 10^{-9}$; Table 2). The combination of 15 rare SNVs remained associated with FG after conditioning on two known common SNVs in LD[27] with each other (rs560887 in intron 1 of *G6PC2* and rs563694 located in the intergenic region between *G6PC2* and *ABCB11*) (conditional $p_{\mathrm{SKAT}} = 5.2 \times 10^{-9}$, $p_{\mathrm{WST}} = 3.1 \times 10^{-5}$; Table 2 and Fig. 3), suggesting that the observed rare variant associations were distinct from known common variant signals. Although *ABCB11* has been proposed to be the causal gene at this locus[28], identification of rare and putatively functional variants implicates *G6PC2* as the much more likely causal candidate. As rare alleles that increase risk for common disease may be obscured by rare, neutral mutations[4], we tested the contribution

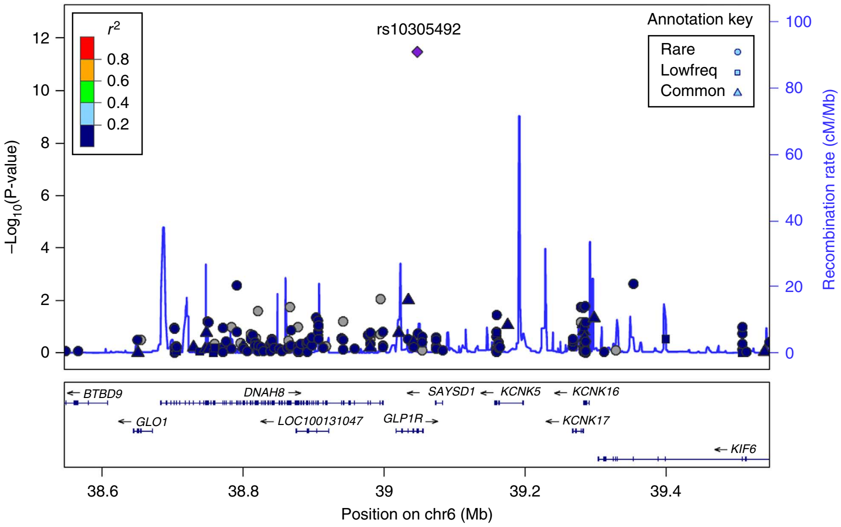
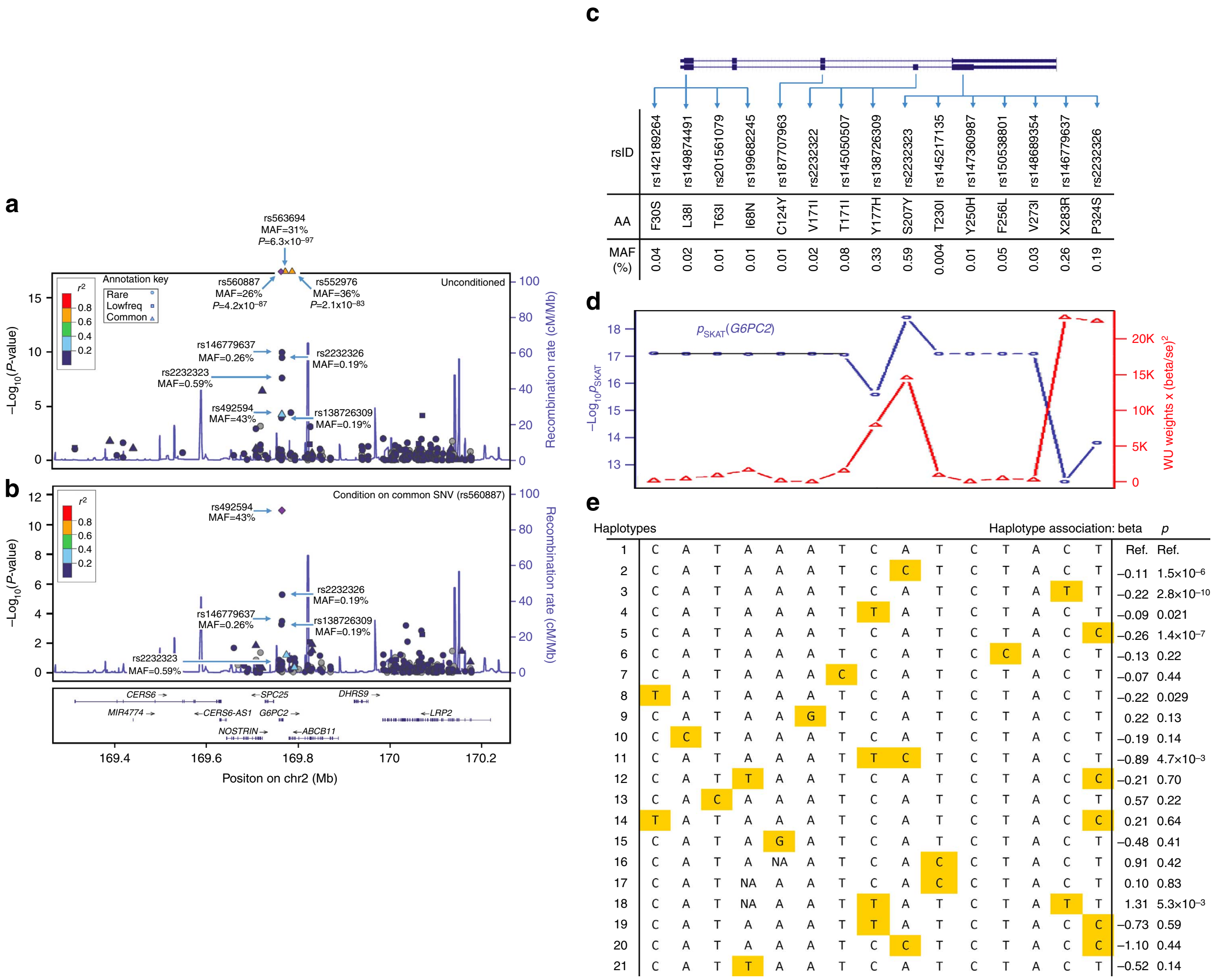


Figure 2 | *GLP1R* regional association plot. Regional association results ($-\log_{10}p$) for fasting glucose of *GLP1R* locus on chromosome 6. Linkage disequilibrium (r^2) indicated by colour scale legend. Triangle symbols indicate variants with MAF > 5%, square symbols indicate variants with MAF1–5% and circle symbols indicate variants with MAF < 1%.

Table 2 | Gene-based associations of *G6PC2* with fasting glucose in African and European ancestries combined.

Gene	Chr: Build 37 position	cMAF*	SNVs (n)†	Weighted sum test (WST)				Sequence Kernel Association Test (SKAT)			
				P	P‡	P§	P‖	P	P‡	P§	P‖
G6PC2	2:169757930-169764491	0.016	15	4.1×10^{-9}	2.6×10^{-5}	2.3×10^{-4}	3.1×10^{-5}	8.2×10^{-18}	4.8×10^{-9}	6.8×10^{-6}	5.2×10^{-9}

Fasting glucose concentrations were adjusted for sex, age, cohort effects and up to 10 principal components in up to 60,564 non-diabetic individuals.
*cMAF = combined minor allele frequency of all variants included in the analysis.
†SNVs(n) = number of variants included in the analysis; variants were restricted to those with MAF < 0.01 and annotated as nonsynonymous, splice-site, or stop loss/gain variants.
‡*P* value for gene-based test after conditioning on rs563694.
§*P* value for gene-based test after conditioning on rs560887.
‖*P* value for gene-based test after conditioning on rs563694 and rs560887.



rsID	rs142189264	rs149874491	rs201561079	rs199682245	rs187707963	rs2232322	rs145050507	rs138726309	rs2232323	rs145217135	rs147360987	rs150538801	rs148689354	rs146779637	rs2232326
AA	F30S	L38I	T63I	I68N	C124Y	V171I	T171I	Y177H	S207Y	T230I	Y250H	F256L	V273I	X283R	P324S
MAF (%)	0.04	0.02	0.01	0.01	0.01	0.02	0.08	0.33	0.59	0.004	0.01	0.05	0.03	0.26	0.19

Haplotypes																Haplotype association: beta	p
1	C	A	T	A	A	A	T	C	A	T	C	T	A	C	T	Ref.	Ref.
2	C	A	T	A	A	A	T	C	**C**	T	C	T	A	C	T	−0.11	1.5×10^{-6}
3	C	A	T	A	A	A	T	C	A	T	C	T	A	**T**	T	−0.22	2.8×10^{-10}
4	C	A	T	A	A	A	T	**T**	A	T	C	T	A	C	T	−0.09	0.021
5	C	A	T	A	A	A	T	C	A	T	C	T	A	C	**C**	−0.26	1.4×10^{-7}
6	C	A	T	A	A	A	T	C	A	T	C	**C**	A	C	T	−0.13	0.22
7	C	A	T	A	A	A	**C**	C	A	T	C	T	A	C	T	−0.07	0.44
8	**T**	A	T	A	A	A	T	C	A	T	C	T	A	C	T	−0.22	0.029
9	C	A	T	A	A	**G**	T	C	A	T	C	T	A	C	T	0.22	0.13
10	C	**C**	T	A	A	A	T	C	A	T	C	T	A	C	T	−0.19	0.14
11	C	A	T	A	A	A	T	**T**	**C**	T	C	T	A	C	T	−0.89	4.7×10^{-3}
12	C	A	T	**T**	A	A	T	C	A	T	C	T	A	C	**C**	−0.21	0.70
13	C	A	**C**	A	A	A	T	C	A	T	C	T	A	C	T	0.57	0.22
14	**T**	A	T	A	A	A	T	C	A	T	C	T	A	C	**C**	0.21	0.64
15	C	A	T	A	**G**	A	T	C	A	T	C	T	A	C	T	−0.48	0.41
16	C	A	T	A	NA	A	T	C	A	**C**	C	T	A	C	T	0.91	0.42
17	C	A	T	NA	A	A	T	C	A	**C**	C	T	A	C	T	0.10	0.83
18	C	A	T	NA	A	A	T	**T**	A	T	C	T	A	**T**	T	1.31	5.3×10^{-3}
19	C	A	T	A	A	A	T	**T**	A	T	C	T	A	C	**C**	−0.73	0.59
20	C	A	T	A	A	A	T	C	**C**	T	C	T	A	C	**C**	−1.10	0.44
21	C	A	T	**T**	A	A	T	C	A	T	C	T	A	C	T	−0.52	0.14

Figure 3 | G6PC2. (**a**) Regional association results ($-\log_{10}p$) for fasting glucose of the *G6PC2* locus on chromosome 2. Minor allele frequencies (MAF) of common and rare *G6PC2* SNVs from single-variant analyses are shown. *P* values for rs560887, rs563694 and rs552976 were artificially trimmed for the figure. Linkage disequilibrium (r^2) indicated by colour scale legend. *y*-Axis scaled to show associations for variant rs560887 (purple dot, MAF = 43%, $P = 4.2\times10^{-87}$). Triangle symbols indicate variants with MAF > 5%, square symbols indicate variants with MAF1–5% and circle symbols indicate variants with MAF <1%. (**b**) Regional association results ($-\log_{10}p$) for fasting glucose conditioned on rs560887 of *G6PC2*. After adjustment for rs560887, both rare SNVs rs2232326 (S324P) and rs146779637 (R283X), and common SNV rs492594 remain significantly associated with FG indicating the presence of multiple independent associations with FG at the *G6PC2* locus. (**c**) Inset of *G6PC2* gene with depiction of exon locations, amino-acid substitutions and MAFs of the 15 SNVs included in gene-based analysis (MAF < 1% and nonsynonymous, splice-site and gain/loss-of-function variation types as annotated by dbNSFPv2.0). (**d**) The contribution of each variant on significance and effect of the SKAT test when one variant is removed from the test. Gene-based SKAT *P* values (blue line) and test statistic (red line) of *G6PC2* after removing one SNV at a time and re-calculating the association. (**e**) Haplotypes and haplotype association statistics and *P* values generated from the 15 rare SNVs from gene-based analysis of *G6PC2* from 18 cohorts and listed in panel (**c**). Global haplotype association, $P = 1.1\times10^{-17}$. Haplotypes ordered by decreasing frequency with haplotype 1 as the reference. Orange highlighting indicates the minor allele of the SNV on the haplotype.

of each *G6PC2* variant by removing one SNV at a time and re-calculating the evidence for association across the gene. Four SNVs, rs138726309 (H177Y), rs2232323 (Y207S), rs146779637 (R283X) and rs2232326 (S324P), each contributed to the association with FG (Fig. 3c and Supplementary Table 11). Each of these SNVs also showed association with FG of larger effect size in unconditional single-variant analyses (Supplementary Data 5), consistent with a recent report in which H177Y was associated with lower FG levels in Finnish cohorts[29]. We developed a novel haplotype meta-analysis method to examine the opposing direction of effects of each SNV. Meta-analysis of haplotypes with the 15 rare SNVs showed a significant global test of association with FG ($p_{\mathrm{global\ test}} = 1.1\times10^{-17}$) (Supplementary Table 12) and supported the findings from the gene-based tests. Individual haplotype tests showed that the most significantly associated haplotypes were those carrying a single rare allele at R283X ($P = 2.8\times10^{-10}$), S324P ($P = 1.4\times10^{-7}$) or Y207S ($P = 1.5\times10^{-6}$) compared with the most common haplotype. Addition of the known common intronic variant (rs560887) resulted in a stronger global haplotype association test ($p_{\mathrm{global\ test}} = 1.5\times10^{-81}$), with the most strongly associated haplotype carrying the minor allele at rs560887 (Supplementary Table 13). Evaluation of regulatory annotation found that this intronic SNV is near the splice acceptor of intron 3 (RefSeq: NM_021176.2) and has been implicated in *G6PC2* pre-mRNA splicing[30]; it is also near the transcription start site of the expressed sequence tag (EST) DB031634, a potential cryptic minor isoform of *G6PC2* mRNA (Supplementary Fig. 7). No associations were observed in gene-based analysis of *G6PC2* with FI or T2D (Supplementary Tables 14 and 15).

Further characterization of exonic variation in *G6PC2* by exome sequencing in up to 7,452 individuals identified 68 SNVs (Supplementary Table 5), of which 4 were individually associated with FG levels and are on the exome chip (H177Y, MAF = 0.3%, $P = 9.6 \times 10^{-5}$; R283X, MAF = 0.2%, $P = 8.4 \times 10^{-3}$; S324P, MAF = 0.1%, $P = 1.7 \times 10^{-2}$; rs560887, intronic, MAF = 40%; $P = 7 \times 10^{-9}$) (Supplementary Data 6). Thirty-six SNVs met criteria for entering into gene-based analyses (each MAF < 1%). This combination of 36 coding variants was associated with FG (cumulative MAF = 2.7%, $p_{\mathrm{SKAT}} = 1.4 \times 10^{-3}$, $p_{\mathrm{WST}} = 5.4 \times 10^{-4}$, Supplementary Table 16). Ten of these SNVs had been included in the exome chip gene-based analyses. Analyses indicated that the 10 variants included on the exome chip data had a stronger association with FG ($p_{\mathrm{SKAT}} = 1.3 \times 10^{-3}$, $p_{\mathrm{WST}} = 3.2 \times 10^{-3}$ vs $p_{\mathrm{SKAT}} = 0.6$, $p_{\mathrm{WST}} = 0.04$ using the 10 exome chip or the 26 variants not captured on the chip, respectively, Supplementary Table 16).

Pathway analyses of FG and FI signals. In agnostic pathway analysis applying MAGENTA (http://www.broadinstitute.org/mpg/magenta/) to all curated biological pathways in KEGG (http://www.genome.jp/kegg/), GO (http://www.geneontology.org), Reactome (http://www.reactome.org), Panther (http://www.pantherdb.org), Biocarta (http://www.biocarta.com) and Ingenuity (http://www.ingenuity.com/) databases, no pathways achieved our Bonferroni-corrected threshold for significance of $P < 1.6 \times 10^{-6}$ for gene set enrichment in either FI or FG data sets (Supplementary Tables 17 and 18). The pathway *P* values were further attenuated when loci known to be associated with either trait were excluded from the analysis. Similarly, even after narrowing the MAGENTA analysis to gene sets in curated databases with names suggestive of roles in glucose, insulin or broader metabolic pathways, we did not identify any pathways that met our Bonferroni-corrected threshold for significance of $P < 2 \times 10^{-4}$ (Supplementary Table 19).

Testing nonsynonomous variants for association in known loci. Owing to the expected functional effects of protein-altering variants, we tested SNVs (4,513 for FG and 1,281 for FI) annotated as nonsynonymous, splice-site or stop gain/loss by dbNSFP[31] in genes within 500 kb of known glycaemic variants[1,27,32] for association with FG and FI to identify associated coding variants, which may implicate causal genes at these loci (Supplementary Table 20). At the *DNLZ-GPSM1* locus, a common nsSNV (rs60980157; S391L) in the *GPSM1* gene was significantly associated with FG (Bonferroni corrected *P* value $< 1.1 \times 10^{-5} = 0.05/4513$ SNVs for FG), and had previously been associated with insulinogenic index[9]. The *GPSM1* variant is common and in LD with the intronic index variant in the *DNLZ* gene (rs3829109) from previous FG GWAS[1] ($r^2_{\mathrm{EU}} = 0.68$; 1000 Genomes EU). The association of rs3829109 with FG was previously identified using data from the Illumina CardioMetabochip, which poorly captured exonic variation in the region[1]. Our results implicate *GPSM1* as the most likely causal gene at this locus (Supplementary Fig. 8a). We also observed significant associations with FG for eight other potentially protein-altering variants in five known FG loci, implicating three genes (*SLC30A8*, *SLC2A2* and *RREB1*) as potentially causal, but still undetermined for two loci (*MADD* and *IKBKAP*) (Supplementary Figs 6f–8b). At the *GRB14/COBLL1* locus, the known GWAS[1,32] nsSNV rs7607980 in the *COBLL1* gene was significantly associated with FI (Bonferroni corrected *P* value $< 3.9 \times 10^{-5} = 0.05/1281$ SNVs for FI), further suggesting *COBLL1* as the causal gene, despite prior functional evidence that GRB14 may represent the causal gene at the locus[33] (Supplementary Fig. 8g).

Similarly, we performed analyses for loci previously identified by GWAS of T2D, but only focusing on the 412 protein-altering variants within the exonic coding region of the annotated gene(s) at 72 known T2D loci[2,34] on the exome chip. In combined ancestry analysis, three nsSNVs were associated with T2D (Bonferroni-corrected *P* value threshold ($P < 0.05/412 = 1.3 \times 10^{-4}$) (Supplementary Data 7). At *WFS1*, *SLC30A8* and *KCNJ11*, the associated exome chip variants were all common and in LD with the index variant from previous T2D GWAS in our population (r^2_{EU}: 0.6–1.0; 1000 Genomes), indicating these coding variants might be the functional variants that were tagged by GWAS SNVs. In ancestry stratified analysis, three additional nsSNVs in *SLC30A8*, *ARAP1* and *GIPR* were significantly associated with T2D exclusively in African ancestry cohorts among the same 412 protein-altering variants (Supplementary Data 8), all with MAF > 0.5% in the African ancestry cohorts, but MAF < 0.02% in the European ancestry cohorts. The three nsSNVs were in incomplete LD with the index variants at each locus ($r^2_{\mathrm{AF}} = 0$, $D'_{\mathrm{AF}} = 1$; 1000 Genomes). SNV rs1552224 at *ARAP1* was recently shown to increase *ARAP1* mRNA expression in pancreatic islets[35], which further supports *ARAP1* as the causal gene underlying the common GWAS signal[36]. The association for nsSNV rs73317647 in *SLC30A8* ($\mathrm{OR}_{\mathrm{AF}}$[95%CI]: 0.45[0.31–0.65], $p_{\mathrm{AF}} = 2.4 \times 10^{-5}$, $\mathrm{MAF}_{\mathrm{AF}} = 0.6\%$) is consistent with the recent report that rare or low frequency protein-altering variants at this locus are associated with protection against T2D[10]. The protein-coding effects of the identified variants indicate all five genes are excellent causal candidates for T2D risk. We did not observe any other single variant nor gene-based associations with T2D that met chip-wide Bonferroni significance thresholds ($P < 4.5 \times 10^{-7}$ and $P < 1.7 \times 10^{-6}$, respectively).

Associations at known FG, FI and T2D index variants. For the previous reported GWAS loci, we tested the known FG and FI SNVs on the exome chip. Overall, 34 of the 38 known FG GWAS index SNVs and 17 of the 20 known FI GWAS SNVs (or proxies, $r^2 \geq 0.8$ 1000 Genomes) were present on the exome chip. Twenty-six of the FG and 15 of the FI SNVs met the threshold for significance ($p_{\mathrm{FG}} < 1.5 \times 10^{-3}$ (0.05/34 FG SNVs), $p_{\mathrm{FI}} < 2.9 \times 10^{-3}$ (0.05/17 FI SNVs)) and were in the direction consistent with previous GWAS publications. In total, the direction of effect was consistent with previous GWAS publications for 33 of the 34 FG SNVs and for 16 of the 17 FI SNVs (binomial probability: $p_{\mathrm{FG}} = 2.0 \times 10^{-9}$, $p_{\mathrm{FI}} = 1.4 \times 10^{-4}$, Supplementary Data 9). Of the known 72 T2D susceptibility loci, we identified 59 index variants (or proxies $r^2 \geq 0.8$ 1000 Genomes) on the exome chip; 57 were in the direction consistent with previous publications (binomial probability: $P = 3.1 \times 10^{-15}$, see Supplementary Data 10). In addition, two of the known MODY variants were on the exome chip. Only *HNF4A* showed nominal significance with FG levels (rs139591750, $P = 3 \times 10^{-3}$, Supplementary Table 21).

Discussion

Our large-scale exome chip-wide analyses identified a novel association of a low frequency coding variant in *GLP1R* with FG and T2D. The minor allele, which lowered FG and T2D risk, was associated with a lower early insulin response to a glucose challenge and higher 2-h glucose. Although the effect size on fasting glucose is slightly larger than for most loci reported to date, our findings suggest that few low frequency variants have a very large effect on glycaemic traits and further demonstrate the need for large sample sizes to identify associations of low frequency variation with complex traits. However, by directly genotyping low frequency coding variants that are poorly captured through imputation, we were able to identify particular genes likely to underlie previously identified associations. Using this approach, we implicate causal genes at six loci associated with fasting glucose and/or FI (*G6PC2, GPSM1, SLC2A2, SLC30A8, RREB1* and *COBLL1)* and five with T2D (*ARAP1*, *GIPR*, *KCNJ11, SLC30A8* and *WFS1*). For example, via gene-based analyses, we identified 15 rare variants in *G6PC2* ($p_{\mathrm{SKAT}} = 8.2 \times 10^{-18}$), which are independent of the common non-coding signals at this locus and implicate this gene as underlying previously identified associations. We also revealed non-coding variants whose putative functions in epigenetic and post-transcriptional regulation of *ABO* and *G6PC2* are supported by experimental ENCODE Consortium, GTEx and transcriptome data from islets and for which future focused investigations using human cell culture and animal models will be needed to clarify their functional influence on glycaemic regulation.

The seemingly paradoxical observation that the minor allele at *GLP1R* is associated with opposite effects on FG and 2-h glucose is not unique to this locus, and is also observed at the *GIPR* locus, which encodes the receptor for gastric inhibitory peptide (GIP), the other major incretin hormone. However, for *GLP1R*, we observe that the FG-lowering allele is associated with lower risk of T2D, while at *GIPR*, the FG-lowering allele is associated with higher risk of T2D (and higher 2-h glucose)[1]. The observation that variation in both major incretin receptors is associated with opposite effects on FG and 2-h glucose is a finding whose functional elucidation will yield new insights into incretin biology. An example where apparently paradoxical findings prompted cellular physiologic experimentation that yielded new knowledge is the *GCKR* variant P446L associated with opposing effects on FG and triglycerides[37,38]. The *GCKR* variant was found to increase active cytosolic GCK, promoting glycolysis and hepatic glucose uptake while increasing substrate for lipid synthesis[39,40].

Two studies have characterized the *GLP1R* A316T variant *in vitro*. The first study found no effect of this variant on cAMP response to full-length GLP-1 or exendin-4 (endogenous and exogenous agonists)[41]. The second study corroborated these findings, but documented as much as 75% reduced cell surface expression of T316 compared with wild-type, with no alteration in agonist binding affinity. Although this reduced expression had little impact on agonist-induced cAMP response or ERK1/2 activation, receptors with T316 had greatly reduced intracellular calcium mobilization in response to GLP-1(7-36NH$_2$) and exendin-4 (ref. 42). Given that GLP-1 induced calcium mobilization is a key factor in the incretin response, the *in vitro* functional data on T316 are consistent with the reduced early insulin response we observed for this variant, further supported by the Glp1r-knockout mouse, which shows lower early insulin secretion relative to wild-type mice[43].

The associations of *GLP1R* variation with lower FG and T2D risk are more challenging to explain, and highlight the diverse and complex roles of *GLP1R* in glycaemic regulation. While future experiments will be needed, here we offer the following hypothesis. Given fasting hyperglycaemia observed in Glp1r-knockout mice[43], A316T may be a gain-of-function allele that activates the receptor in a constitutive manner, causing beta cells to secrete insulin at a lower ambient glucose level, thereby maintaining a lower FG; this could in turn cause downregulation of GLP1 receptors over time, causing incretin resistance and a higher 2-h glucose after an oral carbohydrate load. Other variants in G protein-coupled receptors central to endocrine function such as the TSH receptor (*TSHR*), often in the transmembrane domains[44] (like A316T, which is in a transmembrane helix (TM5) of the receptor peptide), have been associated with increased constitutive activity alongside reduced cell surface expression[45,46], but blunted or lost ligand-dependent signalling[46,47].

The association of variation in *GLP1R* with FG and T2D represents another instance wherein genetic epidemiology has identified a gene that codes for a direct drug target in T2D therapy (incretin mimetics), other examples including *ABCC8/ KCNJ11* (encoding the targets of sulfonylureas) and *PPARG* (encoding the target of thiazolidinediones). In these examples, the drug preceded the genetic discovery. Today, there are over 100 loci showing association with T2D and glycaemic traits. Given that at least three of these loci code for potent antihyperglycaemic targets, these genetic discoveries represent a promising long-term source of potential targets for future diabetes therapies.

In conclusion, our study has shown the use of analysing the variants present on the exome chip, followed-up with exome sequencing, regulatory annotation and additional phenotypic characterization, in revealing novel genetic effects on glycaemic homeostasis and has extended the allelic and functional spectrum of genetic variation underlying diabetes-related quantitative traits and T2D susceptibility.

Methods

Study cohorts. The CHARGE consortium was created to facilitate large-scale genomic meta-analyses and replication opportunities among multiple large population-based cohort studies[12]. The CHARGE T2D-Glycemia Exome Consortium was formed by cohorts within the CHARGE consortium as well as collaborating non-CHARGE studies to identify rare and common functional variation contributing to glycaemic traits and T2D susceptibility (Supplementary Note 1). Up to 23 cohorts participated in this effort representing a maximum total sample size of 60,564 (FG) and 48,118 (FI) participants without T2D for quantitative trait analyses. Individuals were of European (84%) and African (16%) ancestry. Full study characteristics are shown in Supplementary Data 1. Of the 23 studies contributing to quantitative trait analysis, 16 also contributed data on T2D status. These studies were combined with six additional cohorts with T2D case–control status for follow-up analyses of the variants observed to influence FG and FI and analysis of known T2D loci in up to 16,491 T2D cases and 81,877 controls across 4 ancestries combined (African, Asian, European and Hispanic; see Supplementary Data 2 for T2D case–control sample sizes by cohort and ancestry). All studies were approved by their local institutional review boards and written informed consent was obtained from all study participants.

Quantitative traits and phenotypes. FG ($\mathrm{mmol\ l^{-1}}$) and FI ($\mathrm{pmol\ l^{-1}}$) were analysed in individuals free of T2D. FI was log transformed for genetic association tests. Study-specific sample exclusions and detailed descriptions of glycaemic measurements are given in Supplementary Data 1. For consistency with previous glycaemic genetic analyses, T2D was defined by cohort and included one or more of the following criteria: a physician diagnosis of diabetes, on anti-diabetic treatment, fasting plasma glucose $\geq 7\ \mathrm{mmol\ l^{-1}}$, random plasma glucose $\geq 11.1\ \mathrm{mmol\ l^{-1}}$ or haemoglobin A1C$\geq$6.5% (Supplementary Data 2).

Exome chip. The Illumina HumanExome BeadChip is a genotyping array containing 247,870 variants discovered through exome sequencing in ~12,000 individuals, with ~75% of the variants with a MAF<0.5%. The main content of the chip comprises protein-altering variants (nonsynonymous coding, splice-site and stop gain or loss codons) seen at least three times in a study and in at least two studies providing information to the chip design. Additional variants on the chip included common variants found through GWAS, ancestry informative markers (for African and Native Americans), mitochondrial variants, randomly selected synonymous variants, HLA tag variants and Y chromosome variants. In the present study we analysed association of the autosomal variants with glycaemic traits and T2D. See Supplementary Fig. 1 for study design and analysis flow.

Exome array genotyping and quality control. Genotyping was performed with the Illumina HumanExome BeadChipv1.0 (N = 247,870 SNVs) or v1.1 (N = 242,901 SNVs). Illumina's GenTrain version 2.0 clustering algorithm in GenomeStudio or zCall[48] was used for genotype calling. Details regarding genotyping and QC for each study are summarized in Supplementary Data 1. To improve accurate calling of rare variants 10 studies comprising $N = 62,666$ samples participated in joint calling centrally, which has been described in detail elsewhere[13]. In brief, all samples were combined and genotypes were initially auto-called with the Illumina GenomeStudio v2011.1 software and the GenTrain2.0 clustering algorithm. SNVs meeting best practices criteria[13] based on call rates, genotyping quality score, reproducibility, heritability and sample statistics were then visually inspected and manually re-clustered when possible. The performance of the joint calling and best practices approach (CHARGE clustering method) was evaluated by comparing exome chip data to available whole-exome sequencing data ($N=530$ in ARIC). The CHARGE clustering method performed better compared with other calling methods and showed 99.8% concordance between the exome chip and exome sequence data. A total of 8,994 SNVs failed QC across joint calling of studies and were omitted from all analyses. Additional studies used the CHARGE cluster files to call genotypes or used a combination of gencall and zCall[48]. The quality control criteria performed by each study for filtering of poorly genotyped individuals and of low-quality SNVs included a call rate of <0.95, gender mismatch, excess autosomal heterozygosity, and SNV effect estimate s.e. $>10^{-6}$. Concordance rates of genotyping across the exome chip and GWAS platforms were checked in ARIC and FHS and was $>99\%$. After SNV-level and sample-level quality control, 197,481 variants were available for analyses. The minor allele frequency spectrums of the exome chip SNVs by annotation category are depicted in Supplementary Table 22. Cluster plots of *GLP1R* and *ABO* variants are shown in Supplementary Fig. 9.

Whole-exome sequencing. For exome sequencing analyses we had data from up to 14,118 individuals of European ancestry from seven studies, including four studies contributing exome sequence samples that also participated in the exome chip analyses (Atherosclerosis Risk in Communities Study (ARIC, $N=2{,}905$), Cardiovascular Health Study (CHS, $N=645$), Framingham Heart Study (FHS, $N=666$) and Rotterdam Study (RS, $N=702$)) and three additional studies, Erasmus Rucphen Family Study (ERF, $N=1{,}196$), the Exome Sequencing Project (ESP, $N=1{,}338$) and the GlaxoSmithKline discovery sequence project[3] (GSK, $N=6{,}666$). The GlaxoSmithKline (GSK) discovery sequence project provided summary level statistics combining data from GEMS, CoLaus and LOLIPOP collections that added additional exome sequence data at *GLP1R*, including $N=3{,}602$ samples with imputed genotypes. In all studies sequencing was performed using the Illumina HiSeq 2000 platform. The reads were mapped to the GRCh37 Human reference genome (http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/) using the Burrows-Wheeler aligner (BWA[49], http://bio-bwa.sourceforge.net/), producing a BAM[50] (binary alignment/map) file. In ERF, the NARWHAL pipeline[51] was used for this purpose as well. In GSK paired-end short reads were aligned with SOAP[52]. GATK[53] (http://www.broadinstitute.org/gatk/) and Picard (http://picard.sourceforge.net) were used to remove systematic biases and to do quality recalibration. In ARIC, CHS and FHS the Atlas2[54] suite (Atlas-SNP and Atlas-indel) was used to call variants and produce a variant call file (VCF[55]). In ERF and RS genetic variants were called using the Unified Genotyper Tool from GATK, for ESP the University of Michigan's multisample SNP calling pipeline UMAKE was used (H.M. Kang and G. Jun, unpublished data) and in GSK variants were called using SOAPsnp[56]. In ARIC, CHS and FHS variants were excluded if SNV posterior probability was <0.95 (QUAL<22), number of variant reads were <3, variant read ratio was <0.1, $>99\%$ variant reads were in a single strand direction, or total coverage was <6. Samples that met a minimum of 70% of the targeted bases at $\times 20$ or greater coverage were submitted for subsequent analysis and QC in the three cohorts. SNVs with $>20\%$ missingness, >2 observed alleles, monomorphic, mean depth at the site of >500-fold or HWE $P<5\times10^{-6}$ were removed. After variant-level QC, a quality assessment of the final sequence data was performed in ARIC, CHS and FHS based on a number of measures, and all samples with a missingness rate of $>20\%$ were removed. In RS, samples with low concordance to genotyping array ($<$ 95%), low transition/transversion ratio (<2.3) and high heterozygote to homozygote ratio (>2.0) were removed from the data. In ERF, low-quality variants were removed using a QUAL<150 filter. Details of variant and sample exclusion criteria in ESP and GSK have been described before[3,57]. In brief, in ESP these were based on allelic balance (the proportional representation of each allele in likely heterozygotes), base quality distribution for sites supporting the reference and alternate alleles, relatedness between individuals and mismatch between called and phenotypic gender. In GSK these were based on sequence depth, consensus quality and concordance with genome-wide panel genotypes, among others.

Phenotyping glycaemic physiologic traits in additional cohorts. We tested association of the lead signal rs10305492 at *GLP1R* with glycaemic traits in the post absorptive state because it has a putative role in the incretin effect. Cohorts with measurements of glucose and/or insulin levels post 75 g oral glucose tolerance test (OGTT) were included in the analysis (see Supplementary Table 2 for list of participating cohorts and sample sizes included for each trait). We used linear regression models under the assumption of an additive genetic effect for each physiologic trait tested.

Ten cohorts (ARIC, CoLaus, Ely, Fenland, FHS, GLACIER, Health2008, Inter99, METSIM, RISC, Supplementary Table 2) provided data for the 2-h glucose levels for a total sample size of 37,080 individuals. We collected results for 2-h insulin levels in a total of 19,362 individuals and for 30 min-insulin levels in 16,601 individuals. Analyses of 2-h glucose, 2-h insulin and 30 min-insulin were adjusted using three models: (1) age, sex and centre; (2) age, sex, centre and BMI; and (3) age, sex, centre, BMI and FG. The main results in the manuscript are presented using model 3. We opted for the model that included FG because these traits are dependent on baseline FG[1,58]. Adjusting for baseline FG assures the effect of a variant on these glycaemic physiologic traits are independent of FG.

We calculated the insulinogenic index using the standard formula: [insulin 30 min − insulin baseline]/[glucose 30 min − glucose baseline] and collected data from five cohorts with appropriate samples (total $N=16{,}203$ individuals). Models were adjusted for age, sex, centre, then additionally for BMI. In individuals with ≥ 3 points measured during OGTT, we calculated the area under the curve (AUC) for insulin and glucose excursion over the course of OGTT using the trapezoid method[59]. For the analysis of $\mathrm{AUC_{ins}}$ ($N=16{,}126$ individuals) we used three models as discussed above. For the analysis of $\mathrm{AUC_{ins}/AUC_{gluc}}$ ($N=16{,}015$ individuals) we only used models 1 and 2 for adjustment.

To calculate the incretin effect, we used data derived from paired OGTT and intra-venous glucose tolerance test (IVGTT) performed in the same individuals using the formula: ($\mathrm{AUC_{ins}}$ OGTT-$\mathrm{AUC_{ins}}$ IVGTT)/$\mathrm{AUC_{ins}}$ OGTT in RISC ($N=738$). We used models 1 and 2 (as discussed above) for adjustment.

We were also able to obtain lookups for estimates of insulin sensitivity from euglycaemic-hyperinsulinemic clamps and from frequently sampled intravenous glucose tolerance test from up to 2,170 and 1,208 individuals, respectively (Supplementary Table 3).

All outcome variables except 2-h glucose were log transformed. Effect sizes were reported as s.d. values using s.d. values of each trait in the Fenland study[60], the Ely study[61] for insulinogenic index and the RISC study[62] for incretin effects to allow for comparison of effect sizes across phenotypes.

Statistical analyses. The R package seqMeta was used for single variant, conditional and gene-based association analyses[63] (http://cran.r-project.org/web/packages/seqMeta/). We performed linear regression for the analysis of quantitative traits and logistic regression for the analysis of binary traits. For family-based cohorts linear mixed effects models were used for quantitative traits and related individuals were removed before logistic regression was performed. All studies used an additive coding of variants to the minor allele observed in the jointly called data set[13]. All analyses were adjusted for age, sex, principal components calculated from genome-wide or exome chip genotypes and study-specific covariates (when applicable) (Supplementary Data 1). Models testing FI were further adjusted for BMI[32]. Each study analysed ancestral groups separately. At the meta-analysis level ancestral groups were analysed both separately and combined. Meta-analyses were performed by two independent analysts and compared for consistency. Overall quantile-quantile plots are shown in Supplementary Fig. 10.

Bonferroni correction was used to determine the threshold of significance. In single-variant analyses, for FG and FI, all variants with a MAF$>0.02\%$ (equivalent to a MAC≥ 20; $N_{\mathrm{SNVs}}=150{,}558$) were included in single-variant association tests; the significance threshold was set to $P\leq 3\times10^{-7}$ ($P=0.05/150{,}558$), corrected for the number of variants tested. For T2D, all variants with a MAF$>0.01\%$ in T2D cases (equivalent to a MAC≥ 20 in cases; $N_{\mathrm{SNVs}}=111{,}347$) were included in single-variant tests; the significance threshold was set to $P\leq 4.5\times10^{-7}$ ($P=0.05/111{,}347$).

We used two gene-based tests: the Sequence Kernel Association Test (SKAT) and the Weighted Sum Test (WST) using Madsen Browning weights to analyze variants with MAF$<1\%$ in genes with a cumulative MAC≥ 20 for quantitative traits and cumulative MAC≥ 40 for binary traits. These analyses were limited to stop gain/loss, nsSNV, or splice-site variants as defined by dbNSFP v2.0 (ref. 31). We considered a Bonferroni-corrected significance threshold of $P\leq 1.6\times10^{-6}$ (0.05/30,520 tests (15,260 genes × 2 gene-based tests)) in the analysis of FG and FI and $P\leq 1.7\times10^{-6}$ (0.05/29,732 tests (14,866 genes × 2 gene-based tests)) in the analysis of T2D. Owing to the association of multiple rare variants with FG at *G6PC2* from both single and gene-based analyses, we removed one variant at a time and repeated the SKAT test to determine the impact of each variant on the gene-based association effects (Wu weight) and statistical significance.

We performed conditional analyses to control for the effects of known or newly discovered loci. The adjustment command in seqMeta was used to perform conditional analysis on SNVs within 500 kb of the most significant SNV. For *ABO* we used the most significant SNV, rs651007. For *G6PC2* we used the previously reported GWAS variants, rs563694 and rs560887, which were also the most significant SNV(s) in the data analysed here.

The threshold of significance for known FG and FI loci was set at $p_{\mathrm{FG}}\leq 1.5\times10^{-3}$ and $p_{\mathrm{FI}}<2.9\times10^{-3}$ ($=0.05/34$ known FG loci and $=0.05/17$ known FI loci). For FG, FI and T2D functional variant analyses the threshold of significance was computed as $P=1.1\times10^{-5}$ ($=0.05/4513$ protein affecting SNVs at 38 known FG susceptibility loci), $P=3.9\times10^{-5}$ ($=0.05/1281$ protein affecting SNVs at 20 known FI susceptibility loci), $P=1.3\times10^{-4}$ ($=0.05/412$ protein affecting SNVs at 72 known T2D susceptibility loci) and $P=3.5\times10^{-4}$ ($0.05/(72\times2)$) for the gene-based analysis of 72 known T2D susceptibility loci[2,34]. We assessed the associations of glycaemic[1,32,64] and T2D[2,34] variants identified by previous GWAS in our population.

We developed a novel meta-analysis approach for haplotype results based on an extension of Zaykin's method[65]. We incorporated family structure into the basic model, making it applicable to both unrelated and related samples. All analyses were performed in *R*. We developed an *R* function to implement the association test at the cohort level. The general model formula for *K*-observed haplotypes (with the most frequent haplotype used as the reference) is

$$Y = \mu + X\gamma + \beta_2 h_2 + \cdots + \beta_K + b + \varepsilon \qquad (1)$$

Where Y is the trait; X is the covariates matrix; h_m($m = 2,\ldots,$ K) is the expected haplotype dosage: if the haplotype is observed, the value is 0 or 1; otherwise, the posterior probability is inferred from the genotypes; b is the random intercept accounting for the family structure (if it exists), and is 0 for unrelated samples; ε is the random error.

For meta-analysis, we adapted a multiple parameter meta-analysis method to summarize the findings from each cohort[66]. One primary advantage is that this approach allows variation in the haplotype set provided by each cohort. In other words, each cohort could contribute uniquely observed haplotypes in addition to those observed by multiple cohorts.

Associations of *ABO* variants with cardiometabolic traits. Variants in the *ABO* region have been associated with a number of cardiovascular and metabolic traits in other studies (Supplementary Table 8), suggesting a broad role for the locus in cardiometabolic risk. For significantly associated SNVs in this novel glycaemic trait locus, we further investigated their association with other metabolic traits, including systolic blood pressure (SBP, in mm Hg), diastolic blood pressure (DBP, in mm Hg), body mass index (BMI, in $\mathrm{kg\,m^{-2}}$), waist hip ratio (WHR) adjusted for BMI, high-density lipoprotein cholesterol (HDL-C, in $\mathrm{mg\,dl^{-1}}$), low-density lipoprotein cholesterol (LDL-C, in $\mathrm{mg\,dl^{-1}}$), triglycerides (TG, natural log transformed, in % change units) and total cholesterol (TC, in $\mathrm{mg\,dl^{-1}}$). These traits were examined in single-variant exome chip analysis results in collaboration with other CHARGE working groups. All analyses were conducted using the R packages skatMeta or seqMeta[63]. Analyses were either sex stratified (BMI and WHR analyses) or adjusted for sex. Other covariates in the models were age, principal components and study-specific covariates. BMI, WHR, SBP and DBP analyses were additionally adjusted for age squared; WHR, SBP and DBP were BMI adjusted. For all individuals taking any blood pressure lowering medication, 15 mm Hg was added to their measured SBP value and 10 mm Hg to the measured DBP value. As described in detail previously[8] in selected individuals using lipid lowering medication, the untreated lipid levels were estimated and used in the analyses. All genetic variants were coded additively. Maximum sample sizes were 64,965 in adiposity analyses, 56,538 in lipid analyses and 92,615 in blood pressure analyses. Threshold of significance was $P = 6.2 \times 10^{-3}$ ($P = 0.05/8$, where eight is the number of traits tested).

Pathway analyses of *GLP1R*. To examine whether biological pathways curated into gene sets in several publicly available databases harboured exome chip signals below the threshold of exome-wide significance for FG or FI, we applied the MAGENTA gene-set enrichment analysis (GSEA) software as previously described using all pathways in the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Reactome, Panther, BioCarta and Ingenuity pathway databases[67]. Genes in each pathway were scored based on unconditional meta-analysis *P* values for SNVs falling within 40 kb upstream and 110 kb downstream of gene boundaries; we used a 95th percentile enrichment cutoff in MAGENTA, meaning pathways (gene sets) were evaluated for enrichment with genes harbouring signals exceeding the 95th percentile of all genes. As we tested a total of 3,216 pathways in the analysis, we used a Bonferroni-corrected significance threshold of $P < 1.6 \times 10^{-5}$ in this unbiased examination of pathways. To limit the GSEA analysis to pathways that might be implicated in glucose or insulin metabolism, we selected gene sets from the above databases whose names contained the terms 'gluco,' 'glycol,' 'insulin' or 'metabo.' We ran MAGENTA with FG and FI data sets on these 'glucometabolic' gene sets using the same gene boundary definitions and 95th percentile enrichment cutoff as described above; as this analysis involved 250 gene sets, we specified a Bonferroni-corrected significance threshold of $P < 2.0 \times 10^{-4}$. Similarly, to examine whether genes associated with incretin signalling harboured exome chip signals, we applied MAGENTA software to a gene set that we defined comprised genes with putative biologic functions in pathways common to *GLP1R* activation and insulin secretion, using the same gene boundaries and 95th percentile enrichment cutoff described above (Supplementary Table 4). To select genes for inclusion in the incretin pathway gene set, we examined the 'Insulin secretion' and 'Glucagon-like peptide-1 regulates insulin secretion' pathways in KEGG and Reactome, respectively. From these two online resources, genes encoding proteins implicated in GLP1 production and degradation (namely glucagon and *DPP4*), acting in direct pathways common to *GLP1R* and insulin transcription, or involved in signalling pathways shared by *GLP1R* and other incretin family members were included in our incretin signalling pathway gene set; however, we did not include genes encoding proteins in the insulin secretory pathway or encoding cell membrane ion channels as these processes likely have broad implications for insulin secretion independent from *GLP1R* signalling. As this pathway included genes known to be associated with FG, we repeated the MAGENTA analysis excluding genes with known association from our gene set—*PDX1*, *ADCY5*, *GIPR* and *GLP1R* itself.

Protein conformation simulations. The A316T receptor mutant structure was modelled based on the WT receptor structure published previously[22]. First, the Threonine residue is introduced in place of Alanine at position 316. Then, this receptor structure is inserted back into the relaxed membrane-water system from the WT structure[22]. T316 residue and other residues within 5 Å of itself are minimized using the CHARMM force field[68] in the NAMD[69] molecular dynamics (MD) programme. This is followed by heating the full receptor-membrane-water to 310 K and running MD simulation for 50 ns using the NAMD program. Electrostatics are treated by E-wald summation and a time step of 1 fs is used during the simulation. The structure snapshots are saved every 1 ps and the fluctuation analysis (Supplementary Fig. 3) used snapshots every 100 ps. The final snapshot is shown in all the structural figures.

Annotation and functional prediction of variants. Variants were annotated using dbNSFP v2.0 (ref. 31). GTEx (Genotype-Tissue Expression Project) results were used to identify variants associated with gene expression levels using all available tissue types[16]. The Encyclopedia of DNA Elements (ENCODE) Consortium results[14] were used to identify non-coding regulatory regions, including but not limited to transcription factor binding sites (ChIP-seq), chromatin state signatures, DNAse I hypersensitive sites and specific histone modifications (ChIP-seq) across the human cell lines and tissues profiled by ENCODE. We used the UCSC Genome Browser[15,70] to visualize these data sets, along with the public transcriptome data contained in the browser's 'Genbank mRNA' (cDNA) and 'Human ESTs' (Expressed Sequence Tags) tracks, on the hg19 human genome assembly. LncRNA and antisense transcription were inferred by manual annotation of these public transcriptome tracks at UCSC. All relevant track groups were displayed in Pack or Full mode and the Experimental Matrix for each subtrack was configured to display all extant intersections of these regulatory and transcriptional states with a selection of cell or tissue types comprised of ENCODE Tier 1 and Tier 2 human cell line panels, as well as all cells and tissues (including but not limited to pancreatic beta cells) of interest to glycaemic regulation. We visually scanned large genomic regions containing genes and SNVs of interest and selected trends by manual annotation (this is a standard operating procedure in locus-specific in-depth analyses utilizing ENCODE and the UCSC Browser). Only a subset of tracks displaying gene structure, transcriptional and epigenetic data sets from or relevant to T2D, and SNVs in each region of interest was chosen for inclusion in each UCSC Genome Browser-based figure. Uninformative tracks (those not showing positional differences in signals relevant to SNVs or genes of interest) were not displayed in the figures. ENCODE and transcriptome data sets were accessed via UCSC in February and March 2014. To investigate the possible significant overlap between the ABO locus SNPs of interest and ENCODE feature annotations we performed the following analysis. The following data sets were retrieved from the UCSC genome browser: wgEncodeRegTfbsClusteredV3 (TFBS); wgEncodeRegDnaseClusteredV2 (DNase); all H3K27ac peaks (all: wgEncodeBroadHistone*H3k27acStdAln.bed files); and all H3K4me1 peaks (all: wgEncodeBroadHistone*H3k4me1StdAln.bed files). The histone mark files were merged and the maximal score was taken at each base over all cell lines. These features were then overlapped with all SNPs on the exome chip from this study using bedtools (v2.20.1). GWAS SNPs were determined using the NHGRI GWAS catalogue with *P* value $< 5 \times 10^{-8}$. LD values were obtained by the PLINK program based on the Rotterdam Study for SNPs within 100 kB with an r^2 threshold of 0.7. Analysis of these files was completed with a custom R script to produce the fractions of non-GWAS SNPs with stronger feature overlap than the ABO SNPs as well as the Supplementary Figure.

References

1. Scott, R. A. *et al.* Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. *Nat. Genet.* **44,** 991–1005 (2012).
2. DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat. Genet.* **46,** 234–244 (2014).
3. Nelson, M. R. *et al.* An abundance of rare functional variants in 202 drug target genes sequenced in 14,002 people. *Science* **337,** 100–104 (2012).
4. Keinan, A. & Clark, A. G. Recent explosive human population growth has resulted in an excess of rare genetic variants. *Science* **336,** 740–743 (2012).
5. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337,** 64–69 (2012).
6. Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493,** 216–220 (2013).
7. Morrison, A. C. *et al.* Whole-genome sequence-based analysis of high-density lipoprotein cholesterol. *Nat. Genet.* **45,** 899–901 (2013).
8. Peloso, G. M. *et al.* Association of low-frequency and rare coding-sequence variants with blood lipids and coronary heart disease in 56,000 whites and blacks. *Am. J. Hum. Genet.* **94,** 223–232 (2014).
9. Huyghe, J. R. *et al.* Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. *Nat. Genet.* **45,** 197–201 (2013).

10. Flannick, J. *et al.* Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat. Genet.* **46,** 357–363 (2014).
11. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111,** E455–E464 (2014).
12. Psaty, B. M. *et al.* Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium: Design of prospective meta-analyses of genome-wide association studies from 5 cohorts. *Circ. Cardiovasc. Genet.* **2,** 73–80 (2009).
13. Grove, M. L. *et al.* Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS ONE* **8,** e68095 (2013).
14. Bernstein, B. E. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
15. Rosenbloom, K. R. *et al.* ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Res.* **41,** D56–D63 (2013).
16. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45,** 580–585 (2013).
17. Drucker, D. J. & Nauck, M. A. The incretin system: glucagon-like peptide-1 receptor agonists and dipeptidyl peptidase-4 inhibitors in type 2 diabetes. *Lancet.* **368,** 1696–1705 (2006).
18. Garber, A. J. Incretin therapy-present and future. *Rev. Diabet. Stud.* **8,** 307–322 (2011).
19. Seltzer, H. S., Allen, E. W., Herron, Jr. A. L. & Brennan, M. T. Insulin secretion in response to glycemic stimulus: relation of delayed initial release to carbohydrate intolerance in mild diabetes mellitus. *J. Clin. Invest.* **46,** 323–335 (1967).
20. Dailey, M. J. & Moran, T. H. Glucagon-like peptide 1 and appetite. *Trends Endocrinol. Metab.* **24,** 85–91 (2013).
21. Astrup, A. *et al.* Safety, tolerability and sustained weight loss over 2 years with the once-daily human GLP-1 analog, liraglutide. *Int. J. Obes.* **36,** 843–854 (2012).
22. Kirkpatrick, A., Heo, J., Abrol, R. & Goddard, 3rd W. A. Predicted structure of agonist-bound glucagon-like peptide 1 receptor, a class B G protein-coupled receptor. *Proc. Natl Acad. Sci. USA* **109,** 19988–19993 (2012).
23. Olsson, M. L. & Chester, M. A. Polymorphism and recombination events at the ABO locus: a major challenge for genomic ABO blood grouping strategies. *Transfus. Med.* **11,** 295–313 (2001).
24. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat. Genet.* **43,** 333–338 (2011).
25. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466,** 707–713 (2010).
26. Keembiyehetty, C. *et al.* Mouse glucose transporter 9 splice variants are expressed in adult liver and kidney and are up-regulated in diabetes. *Mol. Endocrinol.* **20,** 686–697 (2006).
27. 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).
28. Chen, W. M. *et al.* Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. *J. Clin. Invest.* **118,** 2620–2628 (2008).
29. Service, S. K. *et al.* Re-sequencing expands our understanding of the phenotypic impact of variants at GWAS loci. *PLoS Genet.* **10,** e1004147 (2014).
30. Baerenwald, D. A. *et al.* Multiple functional polymorphisms in the G6PC2 gene contribute to the association with higher fasting plasma glucose levels. *Diabetologia* **56,** 1306–1316 (2013).
31. Liu, X., Jian, X. & Boerwinkle, E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. *Hum. Mutat.* **34,** E2393–E2402 (2013).
32. Manning, A. K. *et al.* A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance. *Nat. Genet.* **44,** 659–669 (2012).
33. Hemming, R. *et al.* Human growth factor receptor bound 14 binds the activated insulin receptor and alters the insulin-stimulated tyrosine phosphorylation levels of multiple proteins. *Biochem. Cell Biol.* **79,** 21–32 (2001).
34. Morris, A. P. *et al.* Large-scale association analysis provides insights into the genetic architecture and pathophysiology of type 2 diabetes. *Nat. Genet.* **44,** 981–990 (2012).
35. Kulzer, J. R. *et al.* A common functional regulatory variant at a type 2 diabetes locus upregulates ARAP1 expression in the pancreatic beta cell. *Am. J. Hum. Genet.* **94,** 186–197 (2014).
36. Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.* **42,** 579–589 (2010).
37. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, Novartis Institutes of BioMedical Research *et al.* Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science* **316,** 1331–1336 (2007).
38. Orho-Melander, M. *et al.* Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. *Diabetes* **57,** 3112–3121 (2008).
39. Rees, M. G. *et al.* Cellular characterisation of the GCKR P446L variant associated with type 2 diabetes risk. *Diabetologia* **55,** 114–122 (2012).
40. Beer, N. L. *et al.* The P446L variant in GCKR associated with fasting plasma glucose and triglyceride levels exerts its effect through increased glucokinase activity in liver. *Hum. Mol. Genet.* **18,** 4081–4088 (2009).
41. Fortin, J. P., Schroeder, J. C., Zhu, Y., Beinborn, M. & Kopin, A. S. Pharmacological characterization of human incretin receptor missense variants. *J. Pharmacol. Exp. Ther.* **332,** 274–280 (2010).
42. Koole, C. *et al.* Polymorphism and ligand dependent changes in human glucagon-like peptide-1 receptor (GLP-1R) function: allosteric rescue of loss of function mutation. *Mol. Pharmacol.* **80,** 486–497 (2011).
43. Scrocchi, L. A. *et al.* Glucose intolerance but normal satiety in mice with a null mutation in the glucagon-like peptide 1 receptor gene. *Nat. Med.* **2,** 1254–1258 (1996).
44. Gozu, H. I., Lublinghoff, J., Bircan, R. & Paschke, R. Genetics and phenomics of inherited and sporadic non-autoimmune hyperthyroidism. *Mol. cCell. Endocrinol.* **322,** 125–134 (2010).
45. Vassart, G. & Costagliola, S. G protein-coupled receptors: mutations and endocrine diseases. *Nat. Rev. Endocrinol.* **7,** 362–372 (2011).
46. Van Sande, J. *et al.* Somatic and germline mutations of the TSH receptor gene in thyroid diseases. *J. Clin. Endocrinol. Metab.* **80,** 2577–2585 (1995).
47. Tonacchera, M. *et al.* Functional characteristics of three new germline mutations of the thyrotropin receptor gene causing autosomal dominant toxic thyroid hyperplasia. *J. Clin. Endocrinol. Metab.* **81,** 547–554 (1996).
48. Goldstein, J. I. *et al.* zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28,** 2543–2545 (2012).
49. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
50. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).
51. Brouwer, R. W., van den Hout, M. C., Grosveld, F. G. & van Ijcken, W. F. NARWHAL, a primary analysis pipeline for NGS data. *Bioinformatics* **28,** 284–285 (2012).
52. Li, R., Li, Y., Kristiansen, K. & Wang, J. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24,** 713–714 (2008).
53. DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43,** 491–498 (2011).
54. Challis, D. *et al.* An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics* **13,** 8 (2012).
55. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27,** 2156–2158 (2011).
56. Li, R. *et al.* SNP detection for massively parallel whole-genome resequencing. *Genome Res.* **19,** 1124–1132 (2009).
57. Lange, L. A. *et al.* Whole-exome sequencing identifies rare and low-frequency coding variants associated with LDL cholesterol. *Am. J. Hum. Genet.* **94,** 233–245 (2014).
58. 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).
59. Matthews, J. N., Altman, D. G., Campbell, M. J. & Royston, P. Analysis of serial measurements in medical research. *BMJ* **300,** 230–235 (1990).
60. Rolfe Ede, L. *et al.* Association between birth weight and visceral fat in adults. *Am. J. Clin. Nutr.* **92,** 347–352 (2010).
61. Forouhi, N. G., Luan, J., Hennings, S. & Wareham, N. J. Incidence of Type 2 diabetes in England and its association with baseline impaired fasting glucose: the Ely study 1990-2000. *Diabet. Med.* **24,** 200–207 (2007).
62. Hills, S. A. *et al.* The EGIR-RISC STUDY (The European group for the study of insulin resistance: relationship between insulin sensitivity and cardiovascular disease risk): I. Methodology and objectives. *Diabetologia* **47,** 566–570 (2004).
63. Voorman, A., Brody, J., Chen, H. & Lumley, T. seqMeta: An R package for meta-analyzing region-based tests of rare DNA variants. *R package version* **1,** 3 (2013).
64. Holmen, O. L. *et al.* Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. *Nat. Genet.* **46,** 345–351 (2014).
65. Zaykin, D. V. *et al.* Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum. Hered.* **53,** 79–91 (2002).
66. Becker, B. J. & Wu, M. J. The synthesis of regression slopes in meta-analysis. *Stat. Sci.* **22,** 414–429 (2007).
67. Segre, A. V., Groop, L., Mootha, V. K., Daly, M. J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLoS Genet.* **6,** e1001058 (2010).
68. Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J. Comput. Chem.* **30,** 1545–1614 (2009).
69. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26,** 1781–1802 (2005).
70. Karolchik, D., Hinrichs, A. S. & Kent, W. J. The UCSC Genome Browser. *Curr. Protoc. Bioinformatics* Chapter 1, Unit 1.4 (2012).

Acknowledgements

CHARGE: Funding support for 'Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium' was provided by the NIH through the American Recovery and Reinvestment Act of 2009 (ARRA) (5RC2HL102419). Sequence data for 'Building on GWAS for NHLBI-diseases: the U.S. CHARGE consortium' was provided by Eric Boerwinkle on behalf of the Atherosclerosis Risk in Communities (ARIC) Study, L. Adrienne Cupples, principal investigator for the Framingham Heart Study, and Bruce Psaty, principal investigator for the Cardiovascular Health Study. Sequencing was carried out at the Baylor Genome Center (U54 HG003273). Further support came from HL120393, 'Rare variants and NHLBI traits in deeply phenotyped cohorts' (Bruce Psaty, principal investigator). Supporting funding was also provided by NHLBI with the CHARGE infrastructure grant HL105756. In addition, M.J.P. was supported through the 2014 CHARGE Visiting Fellow grant—HL105756, Dr Bruce Psaty, PI.

ENCODE: ENCODE collaborators Ben Brown and Marcus Stoiber were supported by the LDRD# 14-200 (B.B. and M.S.) and 4R00HG006698-03 (B.B.) grants.

AGES: This study has been funded by NIA contract N01-AG-12100 with contributions from NEI, NIDCD and NHLBI, the NIA Intramural Research Program, Hjartavernd (the Icelandic Heart Association) and the Althingi (the Icelandic Parliament).

ARIC: The Atherosclerosis Risk in Communities (ARIC) Study is carried out as a collaborative study supported by National Heart, Lung, and Blood Institute (NHLBI) contracts (HHSN268201100005C, HHSN268201100006C, HHSN268201100007C, HHSN268201100008C, HHSN268201100009C, HHSN268201100010C, HHSN268201100011C and HHSN268201100012C), R01HL087641, R01HL59367 and R01HL086694; National Human Genome Research Institute contract U01HG004402; and National Institutes of Health contract HHSN268200625226C. We thank the staff and participants of the ARIC study for their important contributions. Infrastructure was partly supported by Grant Number UL1RR025005, a component of the National Institutes of Health and NIH Roadmap for Medical Research.

CARDIA: The CARDIA Study is conducted and supported by the National Heart, Lung, and Blood Institute in collaboration with the University of Alabama at Birmingham (HHSN268201300025C & HHSN268201300026C), Northwestern University (HHSN268201300027C), University of Minnesota (HHSN268201300028C), Kaiser Foundation Research Institute (HHSN268201300029C), and Johns Hopkins University School of Medicine (HHSN268200900041C). CARDIA is also partially supported by the Intramural Research Program of the National Institute on Aging. Exome chip genotyping and data analyses were funded in part by grants U01-HG004729, R01-HL093029 and R01-HL084099 from the National Institutes of Health to Dr Myriam Fornage. This manuscript has been reviewed by CARDIA for scientific content.

CHES: This work was supported in part by The Chinese-American Eye Study (CHES) grant EY017337, an unrestricted departmental grant from Research to Prevent Blindness, and the Genetics of Latinos Diabetic Retinopathy (GOLDR) Study grant EY14684.

CHS: This CHS research was supported by NHLBI contracts HHSN268201200036C, HHSN268200800007C, N01HC55222, N01HC85079, N01HC85080, N01HC85081, N01HC85082, N01HC85083, N01HC85086; and NHLBI grants HL080295, HL087652, HL103612, HL068986 with additional contribution from the National Institute of Neurological Disorders and Stroke (NINDS). Additional support was provided through AG023629 from the National Institute on Aging (NIA). A full list of CHS investigators and institutions can be found at http://www.chs-nhlbi.org/pi.htm. The provision of genotyping data was supported in part by the National Center for Advancing Translational Sciences, CTSI grant UL1TR000124, and the National Institute of Diabetes and Digestive and Kidney Disease Diabetes Research Center (DRC) grant DK063491 to the Southern California Diabetes Endocrinology Research Center. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The CoLaus Study: We thank the co-primary investigators of the CoLaus study, Gerard Waeber and Peter Vollenweider, and the PI of the PsyColaus Study Martin Preisig. We gratefully acknowledge Yolande Barreau, Anne-Lise Bastian, Binasa Ramic, Martine Moranville, Martine Baumer, Marcy Sagette, Jeanne Ecoffey and Sylvie Mermoud for their role in the CoLaus data collection. The CoLaus study was supported by research grants from GlaxoSmithKline and from the Faculty of Biology and Medicine of Lausanne, Switzerland. The PsyCoLaus study was supported by grants from the Swiss National Science Foundation (#3200B0–105993) and from GlaxoSmithKline (Drug Discovery—Verona, R&D).

CROATIA-Korcula: The CROATIA-Korcula study would like to acknowledge the invaluable contributions of the recruitment team in Korcula, the administrative teams in Croatia and Edinburgh and the people of Korcula. Exome array genotyping was performed at the Wellcome Trust Clinical Research Facility Genetics Core at Western General Hospital, Edinburgh, UK. The CROATIA-Korcula study on the Croatian island of Korucla was supported through grants from the Medical Research Council UK and the Ministry of Science, Education and Sport in the Republic of Croatia (number 108-1080315-0302).

EFSOCH: We are extremely grateful to the EFSOCH study participants and the EFSOCH study team. The opinions given in this paper do not necessarily represent those of NIHR, the NHS or the Department of Health. The EFSOCH study was supported by South West NHS Research and Development, Exeter NHS Research and Development, the Darlington Trust, and the Peninsula NIHR Clinical Research Facility at the University of Exeter. Timothy Frayling, PI, is supported by the European Research Council grant: SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC.

EPIC-Potsdam: We thank all EPIC-Potsdam participants for their invaluable contribution to the study. The study was supported in part by a grant from the German Federal Ministry of Education and Research (BMBF) to the German Center for Diabetes Research (DZD e.V.). The recruitment phase of the EPIC-Potsdam study was supported by the Federal Ministry of Science, Germany (01 EA 9401) and the European Union (SOC 95201408 05 F02). The follow-up of the EPIC-Potsdam study was supported by German Cancer Aid (70-2488-Ha I) and the European Community (SOC 98200769 05 F02). Furthermore, we thank Ellen Kohlsdorf for data management as well as the follow-up team headed by Dr Manuala Bergmann for case ascertainment.

ERF: The ERF study was supported by grants from the Netherlands Organization for Scientific Research (NWO) and a joint grant from NWO and the Russian Foundation for Basic research (Pionier, 047.016.009, 047.017.043), Erasmus MC, and the Centre for Medical Systems Biology (CMSB; National Genomics Initiative). Exome sequencing analysis in ERF was supported by the ZonMw grant (91111025).

For the ERF Study, we are grateful to all participants and their relatives, to general practitioners and neurologists for their contributions, to P. Veraart for her help in genealogy and to P. Snijders for his help in data collection.

FamHS: The Family Heart Study (FamHS) was supported by NIH grants R01-HL-087700 and R01-HL-088215 (Michael A. Province, PI) from NHLBI; and R01-DK-8925601 and R01-DK-075681 (Ingrid B. Borecki, PI) from NIDDK.

FENLAND: The Fenland Study is funded by the Medical Research Council (MC_U106179471) and Wellcome Trust. We are grateful to all the volunteers for their time and help, and to the General Practitioners and practice staff for assistance with recruitment. We thank the Fenland Study Investigators, Fenland Study Co-ordination team and the Epidemiology Field, Data and Laboratory teams. The Fenland Study is funded by the Medical Research Council (MC_U106179471) and Wellcome Trust.

FHS: Genotyping, quality control and calling of the Illumina HumanExome BeadChip in the Framingham Heart Study was supported by funding from the National Heart, Lung and Blood Institute Division of Intramural Research (Daniel Levy and Christopher J. O'Donnell, Principle Investigators). A portion of this research was conducted using the Linux Clusters for Genetic Analysis (LinGA) computing resources at Boston University Medical Campus. Also supported by National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) R01 DK078616, NIDDK K24 DK080140 and American Diabetes Association Mentor-Based Postdoctoral Fellowship Award #7-09-MN-32, all to Dr Meigs, a Canadian Diabetes Association Research Fellowship Award to Dr Leong, a research grant from the University of Verona, Italy to Dr Dauriz, and NIDDK Research Career Award K23 DK65978, a Massachusetts General Hospital Physician Scientist Development Award and a Doris Duke Charitable Foundation Clinical Scientist Development Award to Dr Florez.

FIA3: We are indebted to the study participants who dedicated their time and samples to these studies. We thank Åsa Ågren (Umeå Medical Biobank) for data organization and Kerstin Enquist and Thore Johansson (Västerbottens County Council) for technical assistance with DNA extraction. This particular project was supported by project grants from the Swedish Heart-Lung Foundation, Umeå Medical Research Foundation and Västerbotten County Council.

The Genetics Epidemiology of Metabolic Syndrome (GEMS) Study: We thank Metabolic Syndrome GEMs investigators: Scott Grundy, Jonathan Cohen, Ruth McPherson, Antero Kesaniemi, Robert Mahley, Tom Bersot, Philip Barter and Gerard Waeber. We gratefully acknowledge the contributions of the study personnel at each of the collaborating sites: John Farrell, Nicholas Nikolopoulos and Maureen Sutton (Boston); Judy Walshe, Monica Prentice, Anne Whitehouse, Julie Butters and Tori Nicholls (Australia); Heather Doelle, Lynn Lewis and Anna Toma (Canada); Kari Kervinen, Seppo Poykko, Liisa Mannermaa and Sari Paavola (Finland); Claire Hurrel, Diane Morin, Alice Mermod, Myriam Genoud and Roger Darioli (Switzerland); Guy Pepin, Sibel Tanir, Erhan Palaoglu, Kerem Ozer, Linda Mahley and Aysen Agacdiken (Turkey); and Deborah A. Widmer, Rhonda Harris and Selena Dixon (United States). Funding for the GEMS study was provided by GlaxoSmithKline.

GeneSTAR: The Johns Hopkins Genetic Study of Atherosclerosis Risk (GeneSTAR) Study was supported by NIH grants through the National Heart, Lung, and Blood Institute (HL58625-01A1, HL59684, HL071025-01A1, U01HL72518, HL112064, and HL087698) and the National Institute of Nursing Research (NR0224103) and by M01-RR000052 to the Johns Hopkins General Clinical Research Center. Genotyping services were provided through the RS&G Service by the Northwest Genomics Center at the University of Washington, Department of Genome Sciences, under U.S. Federal Government contract number HHSN268201100037C from the National Heart, Lung, and Blood Institute.

GLACIER: We are indebted to the study participants who dedicated their time, data and samples to the GLACIER Study as part of the Västerbottens hälsoundersökningar (Västerbottens Health Survey). We thank John Hutiainen and Åsa Ågren (Northern Sweden Biobank) for data organization and Kerstin Enquist and Thore Johansson (Västerbottens County Council) for extracting DNA. We also thank M. Sterner, M. Juhas and P. Storm (Lund University Diabetes Center) for their expert technical assistance with genotyping and genotype data preparation. The GLACIER Study was supported by grants from Novo Nordisk, the Swedish Research Council, Påhlssons Foundation, The Heart Foundation of Northern Sweden, the Swedish Heart Lung Foundation, the Skåne Regional Health Authority, Umeå Medical Research Foundation and the Wellcome Trust. This particular project was supported by project grants from the Swedish Heart-Lung Foundation, the Swedish Research Council, the Swedish Diabetes Association, Påhlssons Foundation and Novo nordisk (all grants to P. W. Franks).

GOMAP (Genetic Overlap between Metabolic and Psychiatric Disease): This work was funded by the Wellcome Trust (098051). We thank all participants for their important contribution. We are grateful to Georgia Markou, Laiko General Hospital Diabetes Centre, Maria Emetsidou and Panagiota Fotinopoulou, Hippokratio General Hospital Diabetes Centre, Athina Karabela, Dafni Psychiatric Hospital, Eirini Glezou and Marios Matzioros, Dromokaiteio Psychiatric Hospital, Angela Rentari, Harokopio University of Athens, and Danielle Walker, Wellcome Trust Sanger Institute.

Generation Scotland: Scottish Family Health Study (GS:SFHS): GS:SFHS is funded by the Chief Scientist Office of the Scottish Government Health Directorates, grant number CZD/16/6 and the Scottish Funding Council. Exome array genotyping for GS:SFHS was funded by the Medical Research Council UK and performed at the Wellcome Trust Clinical Research Facility Genetics Core at Western General Hospital, Edinburgh, UK. We also acknowledge the invaluable contributions of the families who took part in the Generation Scotland: Scottish Family Health Study, the general practitioners and Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes academic researchers, IT staff, laboratory technicians, statisticians and research managers. The chief investigators of Generation Scotland are David J. Porteous (University of Edinburgh), Lynne Hocking (University of Aberdeen), Blair Smith (University of Dundee), and Sandosh Padmanabhan (University of Glasgow).

GSK (CoLaus, GEMS, Lolipop): We thank the GEMS Study Investigators: Philip Barter, PhD; Y. Antero Kesäniemi, PhD; Robert W. Mahley, PhD; Ruth McPherson, FRCP; and Scott M. Grundy, PhD. Dr Waeber MD, the CoLaus PI's Peter Vollenweider MD and Gerard Waeber MD, the LOLIPOP PI's, Jaspal Kooner MD and John Chambers MD, as well as the participants in all the studies. The GEMS study was sponsored in part by GlaxoSmithKline. The CoLaus study was supported by grants from GlaxoSmithKline, the Swiss National Science Foundation (Grant 33CSCO-122661) and the Faculty of Biology and Medicine of Lausanne.

Health ABC: The Health, Aging and Body Composition (HABC) Study is supported by NIA contracts N01AG62101, N01AG62103 and N01AG62106. The exome-wide association study was funded by NIA grant 1R01AG032098-01A1 to Wake Forest University Health Sciences and was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (Z01 AG000949-02 and Z01 AG007390-07, Human subjects protocol UCSF IRB is H5254-12688-11). Portions of this study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD. (http:/biowulf.nih.gov).

Health2008: The Health2008 cohort was supported by the Timber Merchant Vilhelm Bang's Foundation, the Danish Heart Foundation (Grant number 07-10-R61-A1754-B838-22392F), and the Health Insurance Foundation (Helsefonden) (Grant number 2012B233).

HELIC: This work was funded by the Wellcome Trust (098051) and the European Research Council (ERC-2011-StG 280559-SEPI). The MANOLIS cohort is named in honour of Manolis Giannakakis, 1978–2010. We thank the residents of Anogia and surrounding Mylopotamos villages, and of the Pomak villages, for taking part. The HELIC study has been supported by many individuals who have contributed to sample collection (including Antonis Athanasiadis, Olina Balafouti, Christina Batzaki, Georgios Daskalakis, Eleni Emmanouil, Chrisoula Giannakaki, Margarita Giannakopoulou, Anastasia Kaparou, Vasiliki Kariakli, Stella Koinaki, Dimitra Kokori, Maria Konidari, Hara Koundouraki, Dimitris Koutoukidis, Vasiliki Mamakou, Eirini Mamalaki, Eirini Mpamiaki, Maria Tsoukana, Dimitra Tzakou, Katerina Vosdogianni, Niovi Xenaki, Eleni Zengini), data entry (Thanos Antonos, Dimitra Papagrigoriou, Betty Spiliopoulou), sample logistics (Sarah Edkins, Emma Gray), genotyping (Robert Andrews, Hannah Blackburn, Doug Simpkin, Siobhan Whitehead), research administration (Anja Kolb-Kokocinski, Carol Smee, Danielle Walker) and informatics (Martin Pollard, Josh Randall).

INCIPE: NIcole Soranzo's research is supported by the Wellcome Trust (Grant Codes WT098051 and WT091310), the EU FP7 (EPIGENESYS Grant Code 257082 and BLUEPRINT Grant Code HEALTH-F5-2011-282510).

Inter99: The Inter99 was initiated by Torben Jørgensen (PI), Knut Borch-Johnsen (co-PI), Hans Ibsen and Troels F. Thomsen. The steering committee comprises the former two and Charlotta Pisinger. The study was financially supported by research grants from the Danish Research Council, the Danish Centre for Health Technology Assessment, Novo Nordisk Inc., Research Foundation of Copenhagen County, Ministry of Internal Affairs and Health, the Danish Heart Foundation, the Danish Pharmaceutical Association, the Augustinus Foundation, the Ib Henriksen Foundation, the Becket Foundation and the Danish Diabetes Association. Genetic studies of both Inter99 and Health 2008 cohorts were funded by the Lundbeck Foundation and produced by The Lundbeck Foundation Centre for Applied Medical Genomics in Personalised Disease Prediction, Prevention and Care (LuCamp, www.lucamp.org). The Novo Nordisk Foundation Center for Basic Metabolic Research is an independent Research Center at the University of Copenhagen partially funded by an unrestricted donation from the Novo Nordisk Foundation (www.metabol.ku.dk).

InterAct Consortium: Funding for the InterAct project was provided by the EU FP6 programme (grant number LSHM_CT_2006_037197). We thank all EPIC participants and staff for their contribution to the study. We thank the lab team at the MRC Epidemiology Unit for sample management and Nicola Kerrison for data management.

IPM Bio*Me* Biobank: The Mount Sinai IPM Bio*Me* Program is supported by The Andrea and Charles Bronfman Philanthropies. Analyses of Bio*Me* data was supported in part through the computational resources and staff expertise provided by the Department of Scientific Computing at the Icahn School of Medicine at Mount Sinai.

The Insulin Resistance Atherosclerosis Family Study (IRASFS): The IRASFS was conducted and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (HL060944, HL061019, and HL060919). Exome chip genotyping and data analyses were funded in part by grants DK081350 and HG007112. A subset of the IRASFS exome chips were contributed with funds from the Department of Internal Medicine at the University of Michigan. Computing resources were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics.

The Insulin Resistance Atherosclerosis Study (IRAS): The IRAS was conducted and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (HL047887, HL047889, HL047890 and HL47902). Exome chip genotyping and data analyses were funded in part by grants DK081350 and HG007112). Computing resources were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics.

JHS: The JHS is supported by contracts HHSN268201300046C, HHSN268201300047C, HHSN268201300048C, HHSN268201300049C, HHSN268201300050C from the National Heart, Lung and Blood Institute and the National Institute on Minority Health and Health Disparities. ExomeChip genotyping was supported by the NHLBI of the National Institutes of Health under award number R01HL107816 to S. Kathiresan. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

The London Life Sciences Prospective Population (LOLIPOP) Study: We thank the co-primary investigators of the LOLIPOP study: Jaspal Kooner, John Chambers and Paul Elliott. The LOLIPOP study is supported by the National Institute for Health Research Comprehensive Biomedical Research Centre Imperial College Healthcare NHS Trust, the British Heart Foundation (SP/04/002), the Medical Research Council (G0700931), the Wellcome Trust (084723/Z/08/Z) and the National Institute for Health Research (RP-PG-0407-10371).

MAGIC: Data on glycaemic traits were contributed by MAGIC investigators and were downloaded from www.magicinvestigators.org.

MESA: The Multi-Ethnic Study of Atherosclerosis (MESA) and MESA SHARe project are conducted and supported by contracts N01-HC-95159 through N01-HC-95169 and RR-024156 from the National Heart, Lung, and Blood Institute (NHLBI). Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. MESA Family is conducted and supported in collaboration with MESA investigators; support is provided by grants and contracts R01HL071051, R01HL071205, R01HL071250, R01HL071251, R01HL071252, R01HL071258, R01HL071259. MESA Air is conducted and supported by the United States Environmental Protection Agency (EPA) in collaboration with MESA Air investigators; support is provided by grant RD83169701. We thank the participants of the MESA study, the Coordinating Center, MESA investigators, and study staff for their valuable contributions. A full list of participating MESA investigators and institutions can be found at http://www.mesa-nhlbi.org. Additional support was provided by the National Institute for Diabetes and Digestive and Kidney Diseases (NIDDK) grants R01DK079888 and P30DK063491 and the National Center for Advancing Translational Sciences grant UL1-TR000124. Further support came from the Cedars-Sinai Winnick Clinical Scholars Award (to M.O. Goodarzi).

METSIM: The METSIM study was funded by the Academy of Finland (grants no. 77299 and 124243). M.L. acknowledges funding from the Academy of Finland. M.B. and K.M. acknowledge grant funding from NIH grants DK062370, DK093757, DK072193.

MRC Ely: The Ely Study was funded by the Medical Research Council (MC_U106179471) and Diabetes UK. We are grateful to all the volunteers, and to the staff of St Mary's Street Surgery, Ely and the study team.

PROCARDIS: We thank all participants in this study. The European Community Sixth Framework Program (LSHM-CT-2007-037273), AstraZeneca, the British Heart Foundation, the Oxford British Heart Foundation Centre of Research Excellence, the Wellcome Trust (075491/Z/04), the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular and Diabetes Programs of Karolinska Institutet and Stockholm County Council, the Foundation for Strategic Research and the Stockholm County Council (560283). Bengt Sennblad acknowledges funding from the Magnus Bergvall Foundation and the Foundation for Old Servants. Rona J. Strawbridge is supported by the Swedish Heart-Lung Foundation, the Tore Nilsson foundation, the Fredrik and Ingrid Thuring foundation and the Foundation for Old Servants. Maria Sabater-Lleal acknowledges funding from Åke-wiberg, Tore Nilsson and Karolinska Institutet Foundations. Mattias Frånberg acknowledges funding from the Swedish e-science Research Center (SeRC).

RISC: We are extremely grateful to the RISC study participants and the RISC study team. The RISC Study is partly supported by EU grant QLG1-CT-2001-01252. Additional support for the RISC Study has been provided by AstraZeneca (Sweden). The RISC Study was supported by European Union grant QLG1-CT-2001-01252 and AstraZeneca. Ele Ferrannini acknowledges grant funding from Boehringer-Ingelheim and Lilly&Co and works as a consultant for Boehringer-Ingelheim, Lilly&Co., MSD, Sanofi, GSK, Janssen, Menarini, Novo Nordisk, AstraZeneca.

Rotterdam Study: The Rotterdam Study is funded by the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/ Netherlands Organization for Scientific Research (NWO) project nr. 050-060-810, CHANCES (nr 242244), Erasmus Medical Center and Erasmus University Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. Abbas Dehghan is supported by NWO grant veni (veni, 916.12.154) and the EUR Fellowship. We are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists.

SCARF: We thank all participants in this study. The study was funded by the Foundation for Strategic Research, the Swedish Heart-Lung Foundation, the Swedish Research Council (8691, 12660, 20653), the European Commission (LSHM-CT-2007-037273), the Knut and Alice Wallenberg Foundation, the Torsten and Ragnar Söderberg Foundation, the Strategic Cardiovascular and Diabetes Programmes of Karolinska Institutet and the Stockholm County Council, and the Stockholm County Council (560183). Bengt Sennblad acknowledges funding from the Magnus Bergvall Foundation and the Foundation for Old Servants. Mattias Frånberg acknowledges funding from the Swedish e-Science Research Center (SeRC).

SCES: The Singapore Chinese Eye Study (SCES) was supported by the National Medical Research Council (NMRC), Singapore (grants 0796/2003, IRG07nov013, IRG09nov014, NMRC 1176/2008, STaR/0003/2008, CG/SERI/2010) and Biomedical Research Council (BMRC), Singapore (08/1/35/19/550 and 09/1/35/19/616).

TEENAGE (TEENs of Attica: Genes and Environment): This research has been co-financed by the European Union (European Social Fund—ESF) and Greek national funds through the Operational Program 'Education and Lifelong Learning' of the National Strategic Reference Framework (NSRF)—Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund. This work was funded by the Wellcome Trust (098051).

We thank all study participants and their families as well as all volunteers for their contribution in this study. We thank the Sample Management and Genotyping Facilities staff at the Wellcome Trust Sanger Institute for sample preparation, quality control and genotyping.

Uppsala Longitudinal Study of Adult Men (ULSAM): The exome chip genotyping and data analyses were supported by Uppsala University, Knut och Alice Wallenberg Foundation, European Research Council, Swedish Diabetes Foundation (grant no. 2013-024), Swedish Research Council (grant no. 2012-1397), and Swedish Heart-Lung Foundation (20120197). C.M.L. is supported by a Wellcome Trust Research Career Development Fellowship (086596/Z/08/Z).

INGI-VB: The Val Borbera study (INGI-VB) thanks the inhabitants of the Val Borbera for participating in the study, the local administrations and the ASL-Novi Ligure for support and Fiammetta Viganò for technical help. We also thank Professor Clara Camaschella, Professor Federico Caligaris-Cappio and the MDs of the Medicine Dept. of the San Raffaele Hospital for help with clinical data collection. The study was supported by funds from Fondazione Compagnia di San Paolo-Torino, Fondazione Cariplo-Milano, Italian Ministry of Health Progetto Finalizzato 2007 and 2012, Italian Ministry of Health Progetto CCM 2010, and PRIN 2009.

WGHS: The WGHS is supported by HL043851 and HL080467 from the National Heart, Lung, and Blood Institute and CA047988 from the National Cancer Institute, the Donald W. Reynolds Foundation and the Fondation Leducq, with collaborative scientific support and funding for genotyping provided by Amgen.

Author contributions

Writing group: J.W., A.Y.C., S.M.W., S.W., H.Y., J.A.B., M.D., M.-F.H., S.R., K.F., L.L., B.H., R.A., J.B.B., M.S., J.C.F., J.D., J.B.M., J.I.R., R.A.S., M.O.G.

Project, design, management and coordination: J.D., B.M.P., D.S.S., J.B.M., J.I.R., R.A.S., M.O.G.

Cohort PI: R.A., A.C., Y.L., D.M.B., L.A.C., G.G., T.J., E.I., A.J.K., C.L., R.A.M., J.M.N., W.H.-H.S., D.T., D.V., R.V., L.E.W., H.B., E.P.B., G.D., E.F., M.F., O.H.F., P.W.F., R.A.G., V.G., A.H., A.T.H., C.H., A. Hofman, J.-H.J., D.L., A.L., B.A.O., C.J.O., S.P., J.S.P., M.A.P., S.S.R., P.M.R., I.R., M.B.S., B.S., A.G.U., M.W., N.J.W., H.W., T.Y.W., E.Z., J.K., M.L., I.B.B., D.I.C., B.M.P., C.M.v.D., D.M.W., E.B., W.H.L.K., R.J.F.L., T.M.F., J.I.R.

Sample collection and phenotyping: M.D., M.-F.H., S.R., L.L., F.K., N.G., A.S., M.G., A.S., T.A., N.A.B., Y.-D.I.C., C.Y.C., A.C., A.D., G.B.E., G.E., S.A.E., A.-E.F., O.G., M.L.G., G.H., M.K.I., M.E.J., T.J., M.K., A.T.K., J.K., I.T.L., W.-J.L., A.S.L., C.L., A.L., A.M., R. McKean-Cowdin, O. McLeod, I.N., A.P., N.W.R., I.S., J.A.S., N.T., M.T., E.T., D.M.B., G.G., E.I., C.L., J.M.N., W.H.-H.S., D.V., R.V., H.B., E.P.B., V.G., T.B.H., C.H., A.H., C.L., L.L., D.L., S.P., O.P., M.A.P., P.M.R., M.B.S., B.S., N.J.W., M.L., B.M.P., E.S.T., C.M.v.D., D.M.W., J.C.F., J.G.W., D.S.S., R.A.S.

Genotyping: A.Y.C., J.B., N.G., J.B.-J., M.F., J.H.Z., A.C.M., L.S., K.D.T., J.B.-J., K.H.A., J.L.A., C.B., D.W.B., Y.-D.I.C., C.Y.C., M.F., F.G., A.G., T.H., P.H., C.C.K., G.M., D.M., I.N., N.D.P., O.P., B.S., N.S., E.K.S., E.A.S., C.B., A.B., K.S., J.C.B., M.B., K.M., E.I., R.A.M., E.P.B., P.D., A.Hofman, C.L., D.L., M.A.P., A.G.U., N.J.W., D.I.C., E.S.T., C.M.v.D., D.M.W., J.I.R., R.A.S., M.O.G.

Statistical Analysis: J.W., A.Y.C., S.M.W., S.W., H.Y., J.B., M.D., M.-F.H., S.R., B.H., F.K., J.E.H., P.A., Y.C.L., L.J.R.-T., N.G., M.G.E., L.L., A.S.B., A.S., R.A., J.B.—J., D.F.F., XG., K.H., A.I., J.J., L.A.L., J.C.L., M.L., J.H.Z., K.M., M.A.N., M.J.P., M.S.-L., C.S., A.V.S., L.S., M.H.S., R.J.S., T.V.V., N.A., C.B., S.M.B., Y.C., J.C., F.G., W.A.G.III, S.G., Y.H., J.H., M.K.I., R.A.J., A.K., A.T.K., E.M.L., J.L., C.L., C.M.L., G.M., N.M.M,, N.D.P., D.P., F.R., K.R., C.F.S., J.A.S., N.S., K.S., M.T., S.J., L.R.Y., J.B., J.B.B., G.M.P., D.I.C., D.M.W., J.D., J.I.R., R.A.S.

Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: J.C.F. has received consulting honoraria from PanGenX and Pfizer; T.F. consulted for Boeringer Ingelheim; J.B.M. serves as a consultant to LipoScience, and Quest Diagnostics; B.P. serves on the DSMB of a clinical trial for a device funded by the manufacturer (Zoll LifeCor) and on the Steering Committee for the Yale Open Data Access Project funded by Johnson & Johnson; D.M.W., M.G.E., L.L. and J.A. are all full time employees of GlaxoSmithKline. P.M.R. and D.I.C. have research grant support from Amgen, AstraZeneca and the NHLBI. The remaining authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Wessel, J. *et al.* Low-frequency and rare exome chip variants associate with fasting glucose and type 2 diabetes susceptibility. *Nat. Commun.* 6:5897 doi: 10.1038/ncomms6897 (2015).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/

Jennifer Wessel[1,2,*], Audrey Y. Chu[3,4,*], Sara M. Willems[5,6,*], Shuai Wang[7,*], Hanieh Yaghootkar[8], Jennifer A. Brody[9,10], Marco Dauriz[11,12,13], Marie-France Hivert[14,15,16], Sridharan Raghavan[11,12], Leonard Lipovich[17,18], Bertha Hidalgo[19], Keolu Fox[10,20], Jennifer E. Huffman[4,21], Ping An[22], Yingchang Lu[23,24], Laura J. Rasmussen-Torvik[25], Niels Grarup[26], Margaret G. Ehm[27], Li Li[27], Abigail S. Baldridge[25], Alena Stančáková[28], Ravinder Abrol[29,30], Céline Besse[31], Anne Boland[31], Jette Bork-Jensen[26], Myriam Fornage[32], Daniel F. Freitag[33,34], Melissa E. Garcia[35], Xiuqing Guo[36], Kazuo Hara[23,24], Aaron Isaacs[5], Johanna Jakobsdottir[37], Leslie A. Lange[38], Jill C. Layton[39], Man Li[40], Jing Hua Zhao[6], Karina Meidtner[41], Alanna C. Morrison[42], Mike A. Nalls[43], Marjolein J. Peters[44,45], Maria Sabater-Lleal[46], Claudia Schurmann[23,24], Angela Silveira[46], Albert V. Smith[37,47], Lorraine Southam[33,48], Marcus H. Stoiber[49], Rona J. Strawbridge[46], Kent D. Taylor[36], Tibor V. Varga[50], Kristine H. Allin[26], Najaf Amin[5], Jennifer L. Aponte[27], Tin Aung[51,52], Caterina Barbieri[53], Nathan A. Bihlmeyer[54,55], Michael Boehnke[56], Cristina Bombieri[57], Donald W. Bowden[58], Sean M. Burns[16], Yuning Chen[7], Yii-Derl Chen[36], Ching-Yu Cheng[51,52,59,60], Adolfo Correa[61], Jacek Czajkowski[22], Abbas Dehghan[62], Georg B. Ehret[63,64], Gudny Eiriksdottir[37], Stefan A. Escher[50], Aliki-Eleni Farmaki[65], Mattias Frånberg[46,66], Giovanni Gambaro[67], Franco Giulianini[3], William A. Goddard III [30], Anuj Goel[68], Omri Gottesman[23], Megan L. Grove[42], Stefan Gustafsson[69], Yang Hai[36], Göran Hallmans[70], Jiyoung Heo[71], Per Hoffmann[72,73,74], Mohammad K. Ikram[51,60,75], Richard A. Jensen[9,10], Marit E. Jørgensen[76], Torben Jørgensen[77,78], Maria Karaleftheri[79], Chiea C. Khor[52,59,80], Andrea Kirkpatrick[30], Aldi T. Kraja[22], Johanna Kuusisto[81], Ethan M. Lange[38,82], I.T. Lee[83,84], Wen-Jane Lee[85], Aaron Leong[11,12], Jiemin Liao[51,52], Chunyu Liu[4], Yongmei Liu[86], Cecilia M. Lindgren[87], Allan Linneberg[77,88,89], Giovanni Malerba[57], Vasiliki Mamakou[90,91], Eirini Marouli[65], Nisa M. Maruthur[92,93], Angela Matchan[33], Roberta McKean-Cowdin[94], Olga McLeod[46], Ginger A. Metcalf[95], Karen L. Mohlke[38], Donna M. Muzny[95], Ioanna Ntalla[65,96], Nicholette D. Palmer[58,97], Dorota Pasko[8], Andreas Peter[98,99], Nigel W. Rayner[33,48,100], Frida Renström[50], Ken Rice[9,101], Cinzia F. Sala[53], Bengt Sennblad[46,102], Ioannis Serafetinidis[103], Jennifer A. Smith[104], Nicole Soranzo[33,105], Elizabeth K. Speliotes[106], Eli A. Stahl[107], Kathleen Stirrups[33,108], Nikos Tentolouris[109], Anastasia Thanopoulou[110], Mina Torres[94], Michela Traglia[53], Emmanouil Tsafantakis[111], Sundas Javad[6], Lisa R. Yanek[112], Eleni Zengini[91,113], Diane M. Becker[112], Joshua C. Bis[9,10], James B. Brown[49,114], L. Adrienne Cupples[4,7], Torben Hansen[26,115], Erik Ingelsson[69,87], Andrew J. Karter[116], Carlos Lorenzo[117], Rasika A. Mathias[112], Jill M. Norris[118], Gina M. Peloso[119,120], Wayne H.-H. Sheu[83,84,121], Daniela Toniolo[53], Dhananjay Vaidya[112], Rohit Varma[94], Lynne E. Wagenknecht[122], Heiner Boeing[123], Erwin P. Bottinger[23], George Dedoussis[65], Panos Deloukas[108,124,125], Ele Ferrannini[126], Oscar H. Franco[62], Paul W. Franks[50,127,128], Richard A. Gibbs[95], Vilmundur Gudnason[37,47], Anders Hamsten[46], Tamara B. Harris[35], Andrew T. Hattersley[129], Caroline Hayward[21], Albert Hofman[62], Jan-Håkan Jansson[128,130], Claudia Langenberg[6], Lenore J. Launer[35], Daniel Levy[131,132], Ben A. Oostra[5], Christopher J. O'Donnell[4,12,133], Stephen O'Rahilly[134], Sandosh Padmanabhan[135], James S. Pankow[136], Ozren Polasek[137], Michael A. Province[22], Stephen S. Rich[138], Paul M. Ridker[3,139], Igor Rudan[140], Matthias B. Schulze[41,99], Blair H. Smith[141], André G. Uitterlinden[44,62], Mark Walker[142], Hugh Watkins[68], Tien Y. Wong[51,52,60], Eleftheria Zeggini[33], The EPIC-InterAct Consortium[§], Markku Laakso[81], Ingrid B. Borecki[22], Daniel I. Chasman[3,143], Oluf Pedersen[26], Bruce M. Psaty[9,10,144,145,146], E. Shyong Tai[59,147], Cornelia M. van Duijn[5,148], Nicholas J. Wareham[6], Dawn M. Waterworth[149], Eric Boerwinkle[42,95], W.H. Linda Kao[40,93,150], Jose C. Florez[12,16,119,120], Ruth J.F. Loos[23,24,151], James G. Wilson[152], Timothy M. Frayling[8], David S. Siscovick[153,154], Josée Dupuis[4,7], Jerome I. Rotter[36], James B. Meigs[11,12], Robert A. Scott[6] & Mark O. Goodarzi[29,155]

[1] Department of Epidemiology, Fairbanks School of Public Health, Indianapolis, Indiana 46202, USA. [2] Department of Medicine, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. [3] Division of Preventive Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA. [4] National Heart, Lung, and Blood Institute (NHLBI) Framingham Heart Study, Framingham, Massachusetts 01702, USA. [5] Genetic Epidemiology Unit, Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000 CE, The Netherlands. [6] MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Cambridge CB2 0SL, UK. [7] Department of Biostatistics, Boston University School of Public Health, Boston, Massachusetts 02118, USA. [8] Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. [9] Cardiovascular Health Research Unit, University of Washington, Seattle, Washington 98101, USA. [10] Department of Medicine, University of Washington, Seattle, Washington 98195, USA. [11] Massachusetts General Hospital, General Medicine Division, Boston, Massachusetts 02114, USA. [12] Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. [13] Division of Endocrinology, Diabetes and Metabolism, Department of Medicine, University of Verona Medical School and Hospital Trust of Verona, Verona 37126, Italy. [14] Harvard Pilgrim Health Care Institute, Department of Population Medicine, Harvard Medical School, Boston, Massachusetts 02215, USA. [15] Division of Endocrinology and Metabolism, Department of Medicine, Université de Sherbrooke, Sherbrooke, Québec, Canada J1K 2R1. [16] Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [17] Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan 48201, USA. [18] Department of Neurology, Wayne State University School of Medicine, Detroit, Michigan 48202, USA. [19] Department of Epidemiology, University of Alabama at Birmingham, Birmingham, Alabama 35233, USA. [20] Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA. [21] MRC Human Genetics Unit, MRC IGMM, University of Edinburgh, Edinburgh, Scotland EH4 2XU, UK. [22] Division of Statistical Genomics and Department of Genetics, Washington University School of Medicine, St. Louis, Missouri 63108, USA. [23] The Charles Bronfman Institute for Personalized Medicine, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [24] The Genetics of Obesity and Related Metabolic Traits Program, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [25] Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, Illinois 60611, USA. [26] The Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark. [27] Quantitative Sciences, PCPS, GlaxoSmithKline, North Carolina 27709, USA. [28] Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland, Kuopio FI-70211, Finland. [29] Department of Medicine and Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. [30] Materials and Process Simulation Center, California Institute of Technology, Pasadena, California 91125, USA. [31] CEA, Institut de Génomique, Centre National de Génotypage, 2 Rue Gaston Crémieux, EVRY Cedex 91057, France. [32] Brown

Foundation Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas 77030, USA. [33] The Wellcome Trust Sanger Institute, Hinxton CB10 1SA, UK. [34] Department of Public Health and Primary Care, Strangeways Research Laboratory, University of Cambridge, Cambridge CB1 8RN, UK. [35] Intramural Research Program, National Institute on Aging, Bethesda, Maryland 21224, USA. [36] Institute for Translational Genomics and Population Sciences, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California 90502, USA. [37] Icelandic Heart Association, Holtasmari 1, Kopavogur IS-201, Iceland. [38] Department of Genetics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [39] Indiana University, Fairbanks School of Public Health, Indianapolis, Indiana 46202, USA. [40] Department of Epidemiology, Johns Hopkins University, Baltimore, Maryland 21205, USA. [41] Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke, Nuthetal DE-14558, Germany. [42] Human Genetics Center, School of Public Health, The University of Texas Health Science Center at Houston, Houston, Texas 77225, USA. [43] Laboratory of Neurogenetics, National Institute on Aging, Bethesda, Maryland 20892, USA. [44] Department of Internal Medicine, Erasmus University Medical Center, Rotterdam 3000 CE, The Netherlands. [45] The Netherlands Genomics Initiative-sponsored Netherlands Consortium for Healthy Aging (NGI-NCHA), Leiden/Rotterdam 2300 RC, The Netherlands. [46] Atherosclerosis Research Unit, Department of Medicine Solna, Karolinska Institutet, Stockholm SE-171 77, Sweden. [47] University of Iceland, Reykjavik IS-101, Iceland. [48] Wellcome Trust Centre for Human Genetics, Oxford OX3 7BN, UK. [49] Department of Genome Dynamics, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [50] Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Skåne University Hospital, Malmö SE-205 02, Sweden. [51] Singapore Eye Research Institute, Singapore National Eye Centre, Singapore 168751, Singapore. [52] Department of Ophthalmology, National University of Singapore and National University Health System, Singapore 119228, Singapore. [53] Division of Genetics and Cell Biology, San Raffaele Research Institute, Milano 20132, Italy. [54] Predoctoral Training Program in Human Genetics, McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Maryland 21205, USA. [55] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [56] Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, Michigan 48109, USA. [57] Section of Biology and Genetics, Department of Life and Reproduction Sciences, University of Verona, Verona 37100, Italy. [58] Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, North Carolina 27157, USA. [59] Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 119228, Singapore. [60] Office of Clinical Sciences, Duke-NUS Graduate Medical School, National University of Singapore, Singapore 169857, Singapore. [61] Department of Medicine, University of Mississippi Medical Center, Jackson, Mississippi 39216, USA. [62] Department of Epidemiology, Erasmus University Medical Center, Rotterdam 3000 CE, The Netherlands. [63] McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA. [64] Division of Cardiology, Geneva University Hospital Geneva 1211, Switzerland. [65] Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens 17671, Greece. [66] Department of Numerical Analysis and Computer Science, SciLifeLab, Stockholm University, Stockholm SE-106 91, Sweden. [67] Division of Nephrology, Department of Internal Medicine and Medical Specialties, Columbus-Gemelli University Hospital, Catholic University, Rome 00168, Italy. [68] Department of Cardiovascular Medicine, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [69] Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala SE-751 85, Sweden. [70] Department of Biobank Research, Umeå University, Umeå SE-901 87, Sweden. [71] Department of Biomedical Technology, Sangmyung University, Chungnam 330-720, Korea. [72] Institute of Human Genetics, Department of Genomics, Life & Brain Center, University of Bonn, Bonn DE-53127, Germany. [73] Human Genomics Research Group, Division of Medical Genetics, University Hospital Basel Department of Biomedicine 4031, Basel, Switzerland. [74] Institute of Neuroscience and Medicine (INM-1) Genomic Imaging Research Center Juelich, Juelich DE-52425, Germany. [75] Memory Aging & Cognition Centre (MACC), National University Health System, Singapore 117599, Singapore. [76] Steno Diabetes Center, Gentofte DK-2820, Denmark. [77] Research Centre for Prevention and Health, Glostrup University Hospital, Glostrup DK-2600, Denmark. [78] Faculty of Medicine, University of Aalborg, Aalborg DK-9220, Denmark. [79] Echinos Medical Centre, Echinos 67300, Greece. [80] Division of Human Genetics, Genome Institute of Singapore, Singapore 138672, Singapore. [81] Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio FI-70211, Finland. [82] Department of Biostatistics, University of North Carolina, Chapel Hill, North Carolina 27599, USA. [83] Division of Endocrine and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 407, Taiwan. [84] School of Medicine, National Yang-Ming University, Taipei 112, Taiwan. [85] Department of Medical Research, Taichung Veterans General Hospital, Taichung 407, Taiwan. [86] Department of Epidemiology & Prevention, Division of Public Health Sciences, Wake Forest University, Winston-Salem, North Carolina 27106, USA. [87] Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. [88] Department of Clinical Experimental Research, Copenhagen University Hospital Glostrup, Glostrup DK-2600, Denmark. [89] Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen DK-2200, Denmark. [90] National and Kapodistrian University of Athens, Faculty of Medicine, Athens 115 27, Greece. [91] Dromokaiteio Psychiatric Hospital, Athens 124 61, Greece. [92] Division of General Internal Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [93] Welch Center for Prevention, Epidemiology, and Clinical Research, Johns Hopkins University, Baltimore, Maryland 21205, USA. [94] Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles 90033, USA. [95] Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [96] University of Leicester, Leicester LE1 7RH, UK. [97] Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, North Carolina 27106, USA. [98] Department of Internal Medicine, Division of Endocrinology, Metabolism, Pathobiochemistry and Clinical Chemistry and Institute of Diabetes Research and Metabolic Diseases, University of Tübingen, Tübingen DE-72076, Germany. [99] German Center for Diabetes Research (DZD), Neuherberg DE-85764, Germany. [100] The Oxford Centre for Diabetes, Endocrinology and Metabolism, University of Oxford, Oxford OX3 7LE, UK. [101] Department of Biostatistics, University of Washington, Seattle, Washington 98195, USA. [102] Science for Life Laboratory, Karolinska Institutet, Stockholm SE-171 77, Sweden. [103] Department of Gastroenterology, Gennimatas General Hospital, Athens 11527, Greece. [104] Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan 48109, USA. [105] Department of Hematology, Long Road, Cambridge CB2 0XY, UK. [106] Department of Internal Medicine, Division of Gastroenterology and Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, Michigan 48109, USA. [107] Division of Psychiatric Genomics, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [108] William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 4NS, UK. [109] First Department of Propaedeutic and Internal Medicine, Athens University Medical School, Laiko General Hospital, Athens 11527, Greece. [110] Diabetes Centre, 2nd Department of Internal Medicine, National University of Athens, Hippokration General Hospital, Athens 11527, Greece. [111] Anogia Medical Centre, Anogia 740 51, Greece. [112] The GeneSTAR Research Program, Division of General Internal Medicine, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. [113] University of Sheffield, Sheffield S10 2TN, UK. [114] Department of Statistics, University of California at Berkeley, Berkeley, California 94720, USA. [115] Faculty of Health Science, University of Copenhagen, Copenhagen 1165, Denmark. [116] Division of Research, Kaiser Permanente, Northern California Region, Oakland, California 94612, USA. [117] Department of Medicine, University of Texas Health Science Center, San Antonio, Texas 77030, USA. [118] Department of Epidemiology, Colorado School of Public Health, University of Colorado Denver, Aurora, Colorado 80204, USA. [119] Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts 02142, USA. [120] Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. [121] College of Medicine, National Defense Medical Center, Taipei 114, Taiwan. [122] Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, North Carolina 27106, USA. [123] Department of Epidemiology, German Institute of Human Nutrition Potsdam Rehbrücke, Nuthetal DE-14558, Germany. [124] Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK. [125] Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah 22254, Saudi Arabia. [126] CNR Institute of Clinical Physiology, Pisa 73100, Italy. [127] Department of Nutrition, Harvard School of Public Health, Boston, Massachusetts 02115, USA. [128] Department of Public Health & Clinical Medicine, Umeå University, Umeå SE-901 87, Sweden. [129] Genetics of Diabetes, University of Exeter Medical School, University of Exeter, Exeter EX1 2LU, UK. [130] Research Unit, Skellefteå SE-931 87, Sweden. [131] Population Sciences Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. [132] Framingham Heart Study, Framingham, Massachusetts 01702, USA. [133] Cardiology Division, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [134] University of Cambridge Metabolic Research Laboratories, MRC Metabolic Diseases Unit and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge CB2 1TN, UK. [135] Institute of Cardiovascular and Medical Sciences, University of Glasgow, Glasgow G12 8TA, UK. [136] Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, Minnesota 55455, USA. [137] Department of Public Health, Faculty of Medicine, University of Split, Split 21000, Croatia. [138] Center for Public Health Genomics, Department of Public Health Sciences, University of Virginia, Charlottesville, Virginia 22908, USA. [139] Division of Cardiology, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts 02115, USA. [140] Centre for Population Health Sciences, Medical School, University of Edinburgh, Edinburgh, Scotland EH8 9YL, UK. [141] Medical Research Institute, University of Dundee, Dundee DD1 9SY, UK. [142] Institute of Cellular Medicine, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK. [143] Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA. [144] Department of Epidemiology, University of Washington, Seattle, Washington 98195, USA. [145] Department of Health Services, University of Washington, Seattle, Washington 98195, USA. [146] Group Health Research Institute, Group Health Cooperative, Seattle, Washington 98195, USA. [147] Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117597, Singapore. [148] Center for Medical Systems Biology, Leiden 2300, The Netherlands. [149] Genetics, PCPS, GlaxoSmithKline, Philadelphia, Pennsylvania 19104, USA. [150] Department of Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA. [151] The Mindich Child Health and Development Institute, The Icahn School of Medicine at Mount Sinai, New York, New York 10029, USA. [152] Department of Physiology and Biophysics, University of Mississippi Medical Center, Jackson, Mississippi 38677, USA. [153] New York Academy of Medicine, New York, New York 10029, USA. [154] Cardiovascular Health Research Unit, Departments of Medicine and Epidemiology, University of Washington, Seattle, Washington 98195, USA. [155] Division of Endocrinology, Diabetes and Metabolism, Cedars-Sinai Medical Center, Los Angeles, California 90048, USA. * These authors contributed equally to this work. § A list of The EPIC-InterAct Consortium members is provided below.

The EPIC-InterAct Consortium

Stephen J. Sharp[6], Nita G. Forouhi[6], Nicola D. Kerrison[6], Debora ME Lucarelli[6], Matt Sims[6], Inês Barroso[33,134], Mark I. McCarthy[48,100,156], Larraitz Arriola[157,158,159], Beverley Balkau[160,161], Aurelio Barricarte[159,162], Carlos Gonzalez[163], Sara Grioni[164], Rudolf Kaaks[165], Timothy J. Key[166], Carmen Navarro[159,167,168], Peter M. Nilsson[50], Kim Overvad[169,170], Domenico Palli[171], Salvatore Panico[172], J. Ramón Quirós[173], Olov Rolandsson[70], Carlotta Sacerdote[174,175], María-José Sánchez[159,176,177], Nadia Slimani[178], Anne Tjonneland[179], Rosario Tumino[180,181], Daphne L. van der A[182], Yvonne T. van der Schouw[183] & Elio Riboli[184]

[156] Oxford NIHR Biomedical Research Centre, Oxford, UK. [157] Public Health Division of Gipuzkoa, San Sebastian, Spain. [158] Instituto BIO-Donostia, Basque Government, San Sebastian, Spain. [159] CIBER Epidemiología y Salud Pública (CIBERESP), Spain. [160] Inserm, CESP, U1018, Villejuif, France. [161] Univ Paris-Sud, UMRS 1018, Villejuif, France. [162] Navarre Public Health Institute (ISPN), Pamplona, Spain. [163] Catalan Institute of Oncology (ICO), Barcelona, Spain. [164] Epidemiology and Prevention Unit, Milan, Italy. [165] German Cancer Research Centre (DKFZ), Heidelberg, Germany. [166] Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. [167] Department of Epidemiology, Murcia Regional Health Council, Murcia, Spain. [168] Unit of Preventive Medicine and Public Health, School of Medicine, University of Murcia, Murcia, Spain. [169] Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark. [170] Aalborg University Hospital, Aalborg, Denmark. [171] Cancer Research and Prevention Institute (ISPO), Florence, Italy. [172] Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy. [173] Public Health Directorate, Asturias, Spain. [174] Unit of Cancer Epidemiology, Citta' della Salute e della Scienza Hospital-University of Turin and Center for Cancer Prevention (CPO), Torino, Italy. [175] Human Genetics Foundation (HuGeF), Torino, Italy. [176] Andalusian School of Public Health, Granada, Spain. [177] Instituto de Investigación Biosanitaria de Granada (Granada.ibs), Granada, Spain. [178] International Agency for Research on Cancer, Lyon, France. [179] Danish Cancer Society Research Center, Copenhagen, Denmark. [180] ASP Ragusa, Italy. [181] Aire Onlus, Ragusa, Italy. [182] National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. [183] University Medical Center Utrecht, Utrecht, Utrecht, the Netherlands. [184] School of Public Health, Imperial College London, London, UK