UCLA
UCLA Previously Published Works

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

Permalink
https://escholarship.org/uc/item/8rr9q67v

Journal
Nature Communications, 6

Authors
Wessel, J
Chu, AY
Willems, SM
et al.

Publication Date
2015

DOI
10.1038/ncomms6897

Peer reviewed
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 ($\beta = -0.09 \pm 0.01$ 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 ($\beta = -0.07 \pm 0.035$ pmol insulin mmol glucose$^{-1}$, $P = 0.048$), but higher 2-h glucose ($\beta = 0.16 \pm 0.05$ mmol l$^{-1}$, $P = 4.3 \times 10^{-4}$). We identify a gene-based association with FG at G6PC2 ($p_{SKAT} = 6.8 \times 10^{-5}$) 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 ($\beta = 0.02 \pm 0.004$ mmol l$^{-1}$, $P = 1.3 \times 10^{-5}$). 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\(^5\)\(^6\). 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\(^1\)\(^1\). Thus, the Illumina HumanExome BeadChip (or exome chip) has been designed to allow the capture of rare (MAF <1\%), low frequency (MAF = 1–5\%) and common (MAF ≥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\(^1\)\(^2\)) T2D-Glycemia Exome Consortium\(^1\)\(^3\). 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, GPRM1, SLC2A2, SLC30A8, RREB1 and COBLL1) and five with T2D (ARAPI, 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.

**Novel association of a GLP1R variant with glycaemic traits.** We identified a novel association of a nonsynonymous SNV (nsSNV) (A316T, rs10305492, MAF = 1.4\%) in the gene encoding the receptor for glucagon-like peptide 1 (GLP1R), with the minor (A) allele associated with lower FG (\(\beta = -0.09 \pm 0.01 \text{mmol L}^{-1}\) (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 (\(\beta\) 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 (\(\beta\) 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-hyperinsulinaemic 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.**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variation type</th>
<th>Chr</th>
<th>Build 37 position</th>
<th>dbSNPID</th>
<th>Alleles</th>
<th>African and European Proportion of trait variance explained</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLP1R</td>
<td>A316T</td>
<td>6</td>
<td>39046794</td>
<td>rs10305492</td>
<td>A</td>
<td>G</td>
</tr>
<tr>
<td>ABO</td>
<td>intergenic</td>
<td>9</td>
<td>136153875</td>
<td>rs651007</td>
<td>A</td>
<td>G</td>
</tr>
</tbody>
</table>

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 mmol L\(^{-1}\).
GLP1R which indicates only one variant in the glycaemic traits or T2D susceptibility, further supported by Fig. 2, variants at this locus may influence FG, including two variants (variant analysis identified association of 12 other SNVs with FG discovery sequence project (Supplementary Table 5). Single-SNVs identified from whole-exome sequencing in up to 14,118 and its association with FG at exome chip showing association with FG.

reductions in BMI, 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 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 \((\beta = 0.02 \pm 0.004 \text{ mmol}^{-1}, \text{MAF} = 20\%, \quad P = 1.3 \times 10^{-8}, \text{variance explained} = 0.02\%, \text{Table 1})\). 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 \((\beta = 0.008 \pm 0.003, \text{P} = 0.02, \text{Supplementary Table 1})\) and T2D risk \((\text{OR [95\%CI]} = 1.05 [1.01–1.08], \text{P} = 0.01, \text{Supplementary Data 3})\). Further, we independently replicated the association at this locus with FG in non-overlapping data from MAGIC using rs579459, a variant in LD with rs651007 and genotyped on the Illumina CardioMetaChip \((\beta = 0.008 \pm 0.003 \text{ mmol}^{-1}, \text{P} = 5.0 \times 10^{-3}; \text{N}_{\text{MAGIC}} = 88,287)\). The FG-associated SNV at ABO was in low LD with the three variants that distinguish between the four major blood groups O, A1, A2 and B \((r^2 = 0.18, \text{rs8176749} \text{r}^2 = 0.01 \text{ and rs8176750} \text{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 \((\beta = 0.025 \pm 0.01 \text{ kg m}^{-2}, \text{P} = 3.4 \times 10^{-4})\) but...
not in men. As previously reported\textsuperscript{24,25}, the FG increasing allele of rs651007 was associated with increased LDL and TC (LDL: $\beta = 2.3 \pm 0.28$ mg dl$^{-1}$, $P = 6.1 \times 10^{-16}$; TC: $\beta = 2.4 \pm 0.35$ 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\textsuperscript{16} (http://www.gtexportal.org/home/) and the ENCODE Consortium resources\textsuperscript{14} in the UCSC Genome Browser\textsuperscript{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 DNase1, 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\textsuperscript{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 rs536534 ($P = 1.1 \times 10^{-4}$) 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\textsuperscript{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%, $p_{\text{SKAT}} = 8.2 \times 10^{-18}$, $p_{\text{WST}} = 4.1 \times 10^{-9}$, Table 2). The combination of 15 rare SNVs remained associated with FG after conditioning on two known common SNVs in LD\textsuperscript{27} with each other (rs560887 in intron 1 of G6PC2 and rs563694 located in the intergenic region between G6PC2 and ABCB11) (conditional $p_{\text{SKAT}} = 5.2 \times 10^{-9}$, $p_{\text{WST}} = 3.1 \times 10^{-5}$, 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\textsuperscript{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\textsuperscript{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 MAF 1-5% and circle symbols indicate variants with MAF < 1%.

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

<table>
<thead>
<tr>
<th>Gene</th>
<th>Chr: Build 37 position</th>
<th>cMAF*</th>
<th>SVNs (n)</th>
<th>Weighted sum test (WST)</th>
<th>Sequence Kernel Association Test (SKAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$p$</td>
<td>$p^*$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$p^2$</td>
<td>$p^3$</td>
</tr>
<tr>
<td>G6PC2</td>
<td>2:169757930-169764491</td>
<td>0.016</td>
<td>15</td>
<td>$4.1 \times 10^{-9}$</td>
<td>$2.6 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

Fasting glucose concentrations were adjusted for sex, age, cohort effects and up to 10 principal components in up to 60,564 non-diabetic individuals.

*MAF—minor allele frequency of all variants included in the analysis.

\textsuperscript{1}SVNs(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.

\textsuperscript{1}P value for gene-based test after conditioning on rs563694.

\textsuperscript{1}P value for gene-based test after conditioning on rs650887.

\textsuperscript{1}P value for gene-based test after conditioning on rs560887.
haplotype association statistics and P values generated from the 15 rare SNVs from gene-based analysis of 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 (r²) indicated by colour scale legend. p-Values scaled to show associations for variant rs560887 (purple dot, MAF = 43%, P = 4.2 × 10⁻⁸⁷). (b) Regional association results (−log10(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 functional variations. (d) Conditional association results (−log10(p)) for fasting glucose conditioned on common SNV (rs560887) 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 × 10⁻¹⁰), S324P (P = 1.4 × 10⁻⁷) or Y207S (P = 1.5 × 10⁻⁶) compared with the most common haplotype. Addition of the known common intronic variant (rs560887) resulted in a stronger global haplotype association test (p_global = 1.5 × 10⁻¹¹), 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; it is also near the transcription start site of the 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. 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_global = 1.1 × 10⁻¹⁷) 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 × 10⁻¹⁰), S324P (P = 1.4 × 10⁻⁷) or Y207S (P = 1.5 × 10⁻⁶) compared with the most common haplotype. Addition of the known common intronic variant (rs560887) resulted in a stronger global haplotype association test (p_global = 1.5 × 10⁻¹¹), 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; 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 four 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%, $P_{SKAT} = 1.4 \times 10^{-3}$, $P_{WST} = 5.4 \times 10^{-4}$, 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 ($P_{SKAT} = 1.3 \times 10^{-3}$, $P_{WST} = 3.2 \times 10^{-3}$ vs $P_{SKAT} = 0.6$, $P_{WST} = 0.04$ 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 Ingenta (http://www.ingenuity.com), no pathways were included in the exome chip gene-based analyses. Analyses with names suggestive of roles in glucose, insulin or broader metabolic pathways, we did not identify any pathways met 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 nonsynonymous 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 dbNSFP31 in genes within 500 kb of known glycaemic loci (Supplementary Table 20). At the DNILZ-GPSM1 locus, a common nsSNV (rs60980157; S391L) in the GPSM1 gene was significantly associated with FG (Bonferroni corrected $P$ value $<3.9 \times 10^{-3}$ = 0.05/1281 SNVs for FG), further suggesting GPSM1 as the causal gene, despite prior functional evidence that GRB14 may represent the causal gene at the locus33 (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 loci1,24 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 ($r^2_{AF} = 0$, $D'_{AF} = 1$; 1000 Genomes). SNV rs1552224 at ARAP1 was recently shown to increase ARAP1 mRNA expression in pancreatic islets35, which further supports ARAP1 as the causal gene underlying the common GWAS signal36. The association for nsSNV rs73317647 in SLC30A8 (ORAF1%CI): 0.45[0.31–0.65], $P_{AF} = 2.4 \times 10^{-5}$, MAF$_{AF} = 0.6%$) is consistent with the recent report that rare or low frequency protein-altering variants at this locus are associated with protection against T2D10. 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 ($P_{FG} <1.5 \times 10^{-3}$ (0.05/34 FG SNVs), $P_{FI} <2.9 \times 10^{-3}$ (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: $p_{FG} = 2.0 \times 10^{-9}$, $p_{FI} = 1.4 \times 10^{-4}$, 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: $P = 3.1 \times 10^{-13}$, 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, GFSM1, SLC2A2, SLC30A8, RREB1 and COBLL1) and five with T2D (ARAPI, GIPR, KCNJ11, SLC30A8 and WFS1). For example, via gene-based analyses, we identified 15 rare variants in G6PC2 (pSKAT = 8.2 × 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). 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. The GCKR variant was found to increase active cytosolic GCK, promoting glycolysis and hepatic glucose uptake while increasing substrate for lipid synthesis.

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). 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(7-36NH₂) 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 response we observed for this variant, further supported by the Glp1r-knockout mouse, which shows lower early insulin response. The association of variation in GLP1R 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, 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 (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, but blunted or lost ligand-dependent signalling.

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 ABCB1/ABCC8/ KCNJ11 (encoding the targets of sulfonylureas) and PPARγ (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 functional characterization, in revealing novel genetic effects on glycaemic regulation 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. The CHARGE T2D-Glycaemia Exome Consortium was formed by cohorts within the CHARGE consortium as well as collaborating non-CHARGE studies to examine 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 analyses were approved by their local institutional review boards and written informed consent was obtained from all study participants.

Quantitative traits and phenotypes. FG (mmol l⁻¹) and FI (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 genetic glycaemic 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 ≥ 7 mmol l⁻¹, random plasma glucose ≥ 11.1 mmol l⁻¹ or haemoglobin A1C ≥ 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-of-function) 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 BeadChip v1.0 (N = 247,870 SNVs) or v1.1 (N = 242,901 SNVs). Illumina’s GenTrain version 2.0 clustering algorithm in GenomeStudio or zCall was used for genotype calling. Details regarding genotyping and QC for each study are summarized in Supplementary Data 1. To improve genotype calling of rare variants 10 studies comprising N ≥ 6,000 participants participated in joint calling centrally, which has been described in detail elsewhere. 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 criteria23 based on call rates, genotyping quality score, reproducibility, heterozygosity 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 xHAP allelic balance 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.

Michigan’s multisample SNP calling pipeline UMAKE was used (H.M. Kang and co-workers, personal communication) to 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.

In ERF, the NARWHAL pipeline51 was used for this purpose as well. In GSK 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 = 1,353) and the GlaxoSmithKline discovery sequence project3 (GSK, N = 6,254) were used. The Exome Sequencing Project (ESP) and the GSK project provided summary level statistics combining data from GEMS, CoLaus and LOLIPOP collections that added additional exome sequence data at GLP1R and ABO variants are shown in Supplementary Fig. 9.

Whole-exome sequencing. For exome sequencing analyses we had data up from 14,118 individuals of European ancestry from seven studies, including four studies contributing exome sequence samples that also participated in the exome chip assays (Athens Ischemic 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 = 1,353) and the GlaxoSmithKline discovery sequence project3 (GSK, N = 6,254). The Exome Sequencing Project (ESP) and the GSK project provided summary level statistics combining data from GEMS, CoLaus and LOLIPOP collections that added additional exome sequence data at GLP1R and ABO variants are shown in Supplementary Fig. 9.

In ERF, the NARWHAL pipeline51 was used for this purpose as well. In GSK pairwise short reads were aligned with SOAP32, GATK32 (http://www.broadinstitute.org/gatk) and Picard (broadinstitute.org) were used to remove systematic biases and to do quality recalibration. In ARIC, CHS and FHS the Atlas24 suite (Atlas-NP and Atlas-indel) was used to call variants and produce a variant call file (VCF file). 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 SOPAmp35. In ARIC, CHS and FHS variants were excluded if SNV posterior probability was <0.05 (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. Sanger sequencing of targeted bases at <90% coverage was performed 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 × 10−5 were removed. After variant-level QC, a quality assessment of the final sequence data was performed in ARIC, CHS and FHS based on mean allele balance 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 homozygote to heterozygote 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.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 37,380 individuals. We collected results for 2-h insulin (C2) and 30 min and 2-h insulin levels for a total sample size of 37,601 individuals. Analyses of 2-h glucose, 2-h insulin and 30 min-insulin were adjusted using three models: (1) age, sex and centile; (2) age, sex, centile and BMI; and (3) age, sex, centile, 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.38 Adjusting for baseline FG assures the effect of a variant on these glycaemic physiologic traits is independent of FG.

We calculated the insulinogenic index using the standard formula: [insulin 30 min insulin baseline]/[glucose 30 min – glucose baseline] (corrected intact 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 trapezoidal method.37 For the analysis of AUCOGTT, (N = 16,126 individuals) we used three models, including above discussed above. For AUCOGTT, (N = 16,015) 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: [AUCOGTT - AUCIVGTT]/AUCOGTT in RISC (N = 208). We used models 1 and 2 (as discussed above) for adjustment (Supplementary Table 3).

Overall 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 study60, the Ely study61 for insulinogenic index and the RISC study62 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 analyses33 (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 model for variants of interest and in ESP discovered single variant tests the significance threshold was set to P < 3 × 10−5 (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 MAF ≥ 20 in cases; N = 111,347) were included in single-variant tests; the significance threshold was set to P < 4.5 × 10−7 (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 Brownwring weights to analyze variants with MAF<1% in genes with a cumulative MAC ≥ 20 for quantitative traits and cumulative MAC ≥ 20 for binary traits. These analyses were performed using the R package GENKRAKEN2 and the stop variant calls were performed using the seqMeta package (ref. 31). We considered a Bonferroni-corrected significance threshold of P ≤ 1.6 × 10−6 (0.05/30,520 tests (15,260 genes × 2 gene-based tests)) in the analyses of FG and FI and P ≤ 1.7 × 10−6 (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 T2D, we performed a conditional test with known T2D loci. For FG and FI we performed a conditional test with known FG and FI loci. For FG and FI we performed a conditional test with known FG and FI loci.

The threshold of significance was set at P = 1.1 × 10−3 (P = 0.05/151 known FG loci; 0.05/151 known FI loci). For FG, FI and T2D functional variant analyses the threshold of significance was computed as P = 1.1 × 10−3 (P = 0.05/151 known FG loci; 0.05/151 known FI loci). For FG, FI and T2D functional variant analyses the threshold of significance was computed as P = 3.5 × 10−3 (P = 0.05/72) and the genotypic test was performed using the formula: [AUCOGTT - AUCIVGTT]/AUCOGTT in RISC (N = 208). We used models 1 and 2 (as discussed above) for adjustment (Supplementary Table 3).

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: [AUCOGTT - AUCIVGTT]/AUCOGTT in RISC (N = 208). We used models 1 and 2 (as discussed above) for adjustment (Supplementary Table 3).

Overall 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 study60, the Ely study61 for insulinogenic index and the RISC study62 for incretin effects to allow for comparison of effect sizes across phenotypes.
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

$$Y = \mu + X_1 + \beta_1 b_1 + \ldots + X_K + \beta_K b_K + e + e_i$$

(1)

Where Y is the trait; X is the covariates matrix; b(m = 2,…, 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 0 is for unrelated samples; e is the random error.

For meta-analysis, we adapted a multiple parameter meta-analysis method to summarize the findings from each cohort. 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 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 kg m⁻²), waist hip ratio (WHR) adjusted for BMI, high-density lipoprotein cholesterol (HDL-C, in mg dl⁻¹), low-density lipoprotein cholesterol (LDL-C, in mg dl⁻¹), triglycerides (TG, natural log transformed, in % change units) and total cholesterol (TC, in mg dl⁻¹). These traits were examined in single-variant exome chip analysis results in collaboration with other groups and conducted using UCSC Genome Browser skatMeta or seqMeta. 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 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 × 10⁻⁵ (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 available genes in the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Reactome, Panther, BioCarta and Ingenuity pathway data bases. 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 were enriched for evaluation with genes in strong 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 × 10⁻³ 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 "glucol," "glucol," "insulin" or "metabo." We ran MAGENTA with FG and FI data sets on these ‘glucolometabolic’ 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 ≤ 1.0 × 10⁻⁴. Similarly, when examining whether associated variant 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 signalling 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. First, the three-dimensional structure was induced in the A316T receptor structure using molecular dynamics (MD). This was followed by introducing the full receptor-membrane-water to the full-length running MD simulation. Then, the receptor structure is inserted back into the relaxed membrane-water system from the WT structure. T316 residue and other residues within 5 Å of itself are minimized using the CHARMM force field in the NAMD molecular dynamics (MD) programme. This is followed by heating the full receptor-membrane-water to 310 K and running MD simulation for 500 ps using the NAMD programme. Electrostatics are treated by Ewald summation and a time step of 0.5 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. The Encyclopedia of DNA Elements (ENCODE) Consortium results 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 to visualize these data sets, along with 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 transcript groups were displayed in Pack or Full mode and the Experimental Matrix for each track was configured to display only direct interactions between these regulatory and transcriptional states with a selection of cell or tissue types comprised of ENCODE T1 and T2 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 with no exonic positional information in signals relevant to T2D or features 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: wgEncodeBroadHistoneH3K27acStadAin.bed files); and all H3K4me1 peaks (all: wgEncodeBroadHistoneH3K4me1StadAin.bed files). The histone mark files were merged and the maximal score was taken at each base over all cell lines. These features were Encyclopaedia of 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 × 10⁻⁸. LD values were obtained using the PLINK program based on the Rotterdam Study for SNPs within 100 kb with an r² 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


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) (SRC2HL102419). 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. Sequence data was carried out at the Baylor Genome Center (US HG030273). 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 Project Consortium. Bedred and Marcus Stoiber were supported by the LDR+ 14-200 (B.B. and M.S.) and 400HG006698-03 (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, Hjartater (the Icelandic Heart Association) and the Althing (the Icelandic Parliament).

ARIC: The Atherosclerosis Risk in Communities (ARIC) Study is conducted as a collaborative study by the National Heart, Lung, and Blood Institute (NHLBI), the National Institute on Aging (NIA), the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). Support for ARIC operations is provided by NHLBI, through the ARIC Coordinating Center, and through contracts from other Institutes of the National Institutes of Health.

We are indebted to the EFSOCH study participants and the people of Curland for their help in data collection.

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

FENLAND: The Fenland Study is funded by the Medical Research Council (MC, U101617941) 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, the Medical and Laboratory team, and the Famil Study team. The EFSOCH study was supported by the Medical Research Council (MC_U101617941) 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. Loos). The original FHS Exome data were generated 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 DK80140 and American Diabetes Association Mentor-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 Daunouy and the Senior Fellowship Study was funded by the Medical Research Council (MC_U101617941) and Wellcome Trust.

The Genetics Epidemiology of Metabolic Syndrome (GEMS) Study: We thank Michael B. Eisenstein, Scott GEMS investigators, Scott Griffin, Jonathan Cohen, Ruth McPherson, Antero Kesaniemi, Robert Mahley, Tom Berlot, Philip Barter and Gerard Waeghe. We gratefully acknowledge the contributions of the study personnel at each of the collaborating sites: John Farrel, Nicholas Nikolopoulos and Maureen Sutton (Boston); Judy Walsehe, Monica Prentice, Anne Whitehouse, Julie Butters and Tori Nicholls (Australia); Heather Doelle, Lynn Lewis and Anna Tomma (Canada); Maria Canas, Seppe Poykko, Elina Mikkola and Minna Vainio (Finland); Claire Hurrel, Diane Morin, Alice Mermod, Myriam Genomed and Roger Darioldi (Switzerland); Guy Pepin, Sibel Tanir, Halen Palaguo, Kereem Ozer, Linda Mahley and Ayse Agacikine (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 NHLBI grants through the National Heart, Lung, and Blood Institute (HL58625-01A, HL59684, HL071025-01A, U01HL72518, HL110264, and HL087698) and the National Institutes of Nursing Research (NR0224103) and by M01-RM000052 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 HHSN2620100037C from the National Heart, Lung, and Blood Institute.

GLACIER: We are indebted to the study participants who dedicated their time and samples to the GLACIER Study as part of the Vasterbotten halsosondoringskor (Vasterbotten Health Survey). We thank John Huitininen and Åsa Agrén (Northern Sweden Research and Development for data collection and Kjirsten Einarsson and Thomas Johansson (Vasterbotten 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 Vasterbotten County Council.

The Genetics Epidemiology of Metabolic Syndrome (GEMS) Study: We thank Michael B. Eisenstein, Scott GEMS investigators, Scott Griffin, Jonathan Cohen, Ruth McPherson, Antero Kesaniemi, Robert Mahley, Tom Berlot, Philip Barter and Gerard Waeghe. We gratefully acknowledge the contributions of the study personnel at each of the collaborating sites: John Farrel, Nicholas Nikolopoulos and Maureen Sutton (Boston); Judy Walsehe, Monica Prentice, Anne Whitehouse, Julie Butters and Tori Nicholls (Australia); Heather Doelle, Lynn Lewis and Anna Tomma (Canada); Maria Canas, Seppe Poykko, Elina Mikkola and Minna Vainio (Finland); Claire Hurrel, Diane Morin, Alice Mermod, Myriam Genomed and Roger Darioldi (Switzerland); Guy Pepin, Sibel Tanir, Halen Palaguo, Kereem Ozer, Linda Mahley and Ayse Agacikine (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 NHLBI grants through the National Heart, Lung, and Blood Institute (HL58625-01A, HL59684, HL071025-01A, U01HL72518, HL110264, and HL087698) and the National Institutes of Nursing Research (NR0224103) and by M01-RM000052 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 HHSN2620100037C 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 Vasterbotten halsosondoringskor (Vasterbotten Health Survey). We thank John Huitininen and Åsa Agrén (Northern Sweden Research and Development for data collection and Kjirsten Einarsson and Thomas Johansson (Vasterbotten County Council) for technical assistance with DNA extraction. This particular project was supported by project grants from the Swedish Heart-Lung Foundation, the Swedish Research Council, Pålshons Foundation, The Swedish Heart Lung Foundation of Northern Sweden Research and Development for data collection and 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, Pålshons Foundation and Novo nordisk (all grants to P. W. Franks).
GOMAP (Genetic Overlap between Metabolic and Psychiatric Disease): This work was funded by the Welcome Trust (080851). We thank all participants for their important contribution. We are grateful to Georgia Markou, Laiko General Hospital Diabetes Centre, Maria Emeisud and Panagiota Fiotopoulou, Hippokratio General Hospital Diabetes Centre, Athina Karabela, Dafni Psychiatric Hospital, Eirini Gezou and Marios Matzioros, Dromokatoio Psychiatric Hospital, Angela Rentari, Harokopio University of Athens, and Danielle Walker, Wellcome Trust Sanger Institute. GOMAP was conducted and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (1HL074787, 1HL074688, 1HL074780 and 1HL47902). Exome chip genotyping and data analyses were conducted and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (1HL074787, 1HL074688, 1HL074780 and 1HL47902). All other genotyping and data analyses were conducted in part by grants DK081350 and HG00712. Exome data were generated in 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 (HL074787, 1HL074688, 1HL074780 and 1HL47902). Exome chip genotyping and data analyses were conducted and supported by grants in part by grants DK081350 and HG00712. Computing resources were provided, in part, by the Wake Forest School of Medicine Center for Public Health Genomics.

JHS: The JHS is supported by grants HHSN268201300004C, HHSN268201300004C, HHSN268201300004C, HHSN268201300004C, HHSN268201300004C from the National Heart, Lung and Blood Institute and the National Institute on Aging and Health Disparities. Genotyping was supported by the NHLBI of the National Institutes of Health under award number R101HL071816 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-principal investigators of the LOLIPOP study: Jaspal Kooner, John Chambers and Paul Elliott. The LOLIPOP study is funded 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).

Data on glycanatic 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-02456 from the National Heart, Lung, and Blood Institute (NHLBI). Funding for MESA SHARe genotyping was provided by NHLBI Contract N02-HL-6-4278. MESA Funded Research is a collaboration with the list of participating MESA investigators; support is provided by grants and contracts R01HL101501, R01HL072125, R01HL072120, 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 R83169701. We thank the participants of the MESA study, the Coordinating Center, MESA investigators, and staff for their valuable contribution. 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 P30DK064541 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.K. acknowledges funding from the Academy of Finland. M.B. and K.M. acknowledge funding from NIH grants DK062370, DK093757, DK072193.

MRC Ely: The Ely Study was funded by the Medical Research Council (MC_U1016174971) 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/04/01), the Swedish Research Council, the Knut and Alice Wallenberg Foundation, the Swedish Heart-Lung Foundation, the Torsten and Ragnar Söderberg Foundation, the Scandinavian Heart Foundation, Mar- olinna University of Health Sciences, the Foundation for Old Servants, 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-Ileal acknowledges funding from Åke-wiberg, Tore Nilsson and Karolinska Institutet Foundations. Mattias Fränkberg 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 QLGI-CT-2001-01252. Additional support for the RISC Study has been provided by AstraZeneca (Sweden). The RISC Study was supported by European Union grant QLGI-CT-2001-01252 and AstraZeneca. Ele Ferranannio acknowledges grant fund 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 (RIDE), the Netherlands Genome Initiative (NGI)/Netherlands Organization for Scientific Research (NWO) project nr. 050-060-810, CHANCES (nr 224244), 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 (OCW), The Netherlands Organization for Scientific Research (ORG) (DG XII) and the Municipality of Rotterdam. Abbas Debghann 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. Matsias Frännberg 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 0767/2003, IRG07nov013, IRG09nov014, NMRC 1176/2008, STAR0003/2008, CC/SGERU2010) 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-funded 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 (20130197). 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-Novigugligure for support and Fiammetta Vigano` for technical help. We also thank Professor Val Borbera for participating in the study, the local administrations and the ASL-Ferrara for support and Mammiglioni 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. This 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 HL88467 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.

The EPIC-InterAct Consortium

Stephen J. Sharp⁶, Nita G. Forouhi⁶, Nicola D. Kerrison⁶, Debora ME Lucarelli⁶, Matt Sims⁶, Inês Barroso³,³³,³,³⁴, Mark I. McCarthy⁴⁸,¹⁰⁰,¹⁵⁶, Larraitz Arriola¹⁵⁷,¹⁵⁸,¹⁵⁹, Beverley Balkau¹⁶⁰,¹⁶¹, Aurelio Barricarte¹⁵⁹,¹⁶², Carlos Gonzalez¹⁶³, Sara Grioni¹⁶⁴, Rudolf Kaaks¹⁶⁵, Timothy J. Key¹⁶⁶, Carmen Navarro¹⁵⁹,¹⁶⁷,¹⁶⁸, Peter M. Nilsson⁵⁰, Kim Overvad¹⁶⁹,¹⁷⁰, Domenico Palli¹⁷¹, Salvatore Panico¹⁷², J. Ramón Quiroº¹⁷³, Olov Rolandsson¹⁷⁰, Carlotta Sacerdote¹⁷⁴,¹⁷⁵, Maria–José Sánchez¹⁵⁹,¹⁷⁶,¹⁷⁷, Nadia Slimani¹⁷⁸, Anne Tjonneland¹⁷⁹, Rosario Tumino¹⁸⁰,¹⁸¹, Daphne L. van der A¹⁸², Yvonne T. van der Schouw¹⁸³ & Elio Riboli¹⁸⁴

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, UMR 1018, Villejuif, France. ¹⁶² Navarre Public Health Institute (ISPAN), Pamplona, Spain. ¹⁶³ Catalan Institute of Oncology (IC.0), 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 chirurgica Federico II University, Naples, Italy. ¹⁷³ Public Health Directorate, Asturias, Spain. ¹⁷⁴ Unit of Cancer Epidemiology, Città' della Salute e della Scienza Hospital-University of Turin and Center for Cancer Prevention (CPO), Torino, Italy. ¹⁷⁵ Human Genetics Foundation (HuGeF), Turin, Italy. ¹⁷⁶ Andalusian School of Public Health, Granada, Spain. ¹⁷⁷ Instituto de Investigación Biosanitaria de Granada (Granada.iks), 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.