

Title: Sequencing of 640,000 exomes identifies *GPR75* variants associated with protection against obesity

Authors: Parsa Akbari^{1,*}, Ankit Gilani^{2,*}, Olukayode Sosina^{1,*}, Jack A. Kosmicki¹, Lori Khrimian¹, Yi-Ya Fang¹, Trikaladarshi Persaud¹, Victor Garcia², Dylan Sun¹, Alexander Li¹, Joelle Mbatchou¹, Adam Locke¹, Christian Benner¹, Niek Verweij¹, Nan Lin³, Sakib Hossain², Kevin Agostinucci², Jonathan V. Pascale², Ercument Dirice², Michael Dunn¹, Regeneron Genetics Center[§], DiscovEHR Collaboration[§], William E. Kraus^{5,6}, Svati H. Shah^{7,8}, Yii-Der I. Chen⁹, Jerome I. Rotter⁹, Daniel J. Rader¹⁰, Olle Melander^{11,12}, Christopher D. Still¹³, Tooraj Mirshahi¹³, David J. Carey¹³, Jaime Berumen-Campos¹⁴, Pablo Kuri-Morales¹⁴, Jesus Alegre-Díaz¹⁴, Jason M. Torres¹⁵, Jonathan R. Emberson¹⁵, Rory Collins¹⁵, Suganthi Balasubramanian¹, Alicia Hawes¹, Marcus Jones¹, Brian Zambrowicz¹, Andrew J. Murphy¹, Charles Paulding¹, Giovanni Coppola¹, John D. Overton¹, Jeffrey G. Reid¹, Alan Shuldiner¹, Michael Cantor¹, Hyun M. Kang¹, Goncalo R. Abecasis¹, Katia Karalis¹, Aris N. Economides¹, Jonathan Marchini¹, George D. Yancopoulos¹, Mark W. Sleeman¹, Judith Altarejos¹, Giusy Della Gatta¹, Roberto Tapia-Conyer^{14,†}, Michal L. Schwartzman^{2,†}, Aris Baras^{1,†,‡}, Manuel A. R. Ferreira^{1,†}, Luca A. Lotta^{1,†,‡}

Affiliations:

¹Affiliations should be preceded by superscript numbers corresponding to the author list, and each affiliation should end with a period.

²Each affiliation should be a separate paragraph.

³For large groups, use the name of the group or consortium and include a full list of the authors in the Supplementary Material.

¹Regeneron Genetics Center, Regeneron Pharmaceuticals Inc., 777 Old Saw Mill River Rd., Tarrytown, NY, 10591, USA.

²Department of Pharmacology and Medicine, New York Medical College School of Medicine, Valhalla, NY, 10595, USA.

³Regeneron Genetics Center, 777 Old Saw Mill River Rd., Tarrytown, NY, 10591, USA,

⁴(No affiliation data provided).

⁵Division of Cardiology, Duke University Medical Center, Durham, NC, USA.

⁶Duke Center for Living, Duke University Medical Center, Durham, NC.

⁷Department of Medicine, Duke University Medical Center, Durham, NC, USA.

⁸Duke Molecular Physiology Institute, Duke University Medical Center, Durham, NC.

⁹Institute for Translational Genomics and Population Sciences, The Lundquist Institute for Biomedical Innovation and Department of Pediatrics, Harbor-UCLA Medical Center, 1124 W. Carson Street, Torrance, CA, 90502, USA.

¹⁰Department of Genetics, Perelman School of Medicine, University of Pennsylvania, PA, USA.

¹¹Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden.

¹²Department of Emergency and Internal Medicine, Skåne University Hospital, Malmö, Sweden.

¹³Geisinger Obesity Institute, Geisinger Health System, Danville, PA, 17882, USA.

¹⁴Faculty of Medicine, National Autonomous University of Mexico, Circuito Escolar 411A, Copilco Universidad, Coyoacán, Ciudad de México, 4360, Mexico.

5 ¹⁵Nuffield Department of Population Health, University of Oxford, Oxford, England, UK.

* These authors contributed equally to the work.

§ A complete list of investigators contributing to the Regeneron Genetics Center and MyCode Consortium is provided in the Supplementary Materials.

† These authors contributed equally to the work.

10 ‡ Corresponding authors. Email: luca.lotta@regeneron.com (L.A.L), aris.baras@regeneron.com (A.B).

One Page Summary (~500 words)

Introduction: Obesity accounts for a substantial and growing burden of disease globally. Body adiposity is highly heritable and human genetic studies can be a pathway to new biological and therapeutic insights.

Rationale: Genetic studies of extreme obesity and genome-wide association studies in large cohorts have partly clarified the inherited basis of human obesity and affirmed the critical importance of central nervous system genes. Whole-exome sequencing of hundreds of thousands of individuals from population or health-system based studies is complementary to those approaches and has the potential to identify large-effect protein-coding variants associated with propensity for or protection against obesity. Here, we sequenced the exomes of 645,626 individuals from the United Kingdom, United States and Mexico, and estimated associations of rare coding variants with body mass index (BMI), a measure of overall adiposity used to define obesity in clinical practice. We complemented exome-sequencing with fine-mapping of common alleles, polygenic score analysis, *in vitro* and *in vivo* modelling work.

Results: We identified 16 genes for which the burden of rare nonsynonymous variants was associated with BMI at the exome-wide level of statistical significance (inverse-variance weighted meta-analysis $p\text{-value} < 3.6 \times 10^{-07}$), including associations at five brain-expressed G-protein coupled receptors (*CALCR*, *MC4R*, *GIPR*, *GPR151*, *GPR75*). A tissue expression analysis revealed an over-representation of genes highly expressed in the hypothalamus, a key center for the neuro-endocrine regulation of energy balance. Protein truncating variants in *GPR75* were found in ~4/10,000 sequenced people, with similar frequency across populations, and were associated with 1.9 kg/m² lower BMI (95% confidence interval [CI], -2.6, -1.3 kg/m²; $p = 1.4 \times 10^{-08}$; corresponding to 5.5 kilograms lower body weight) and with protection against obesity (odds ratio, 0.46; 95% CI, 0.31, 0.67; $p = 6.9 \times 10^{-05}$) in heterozygous carriers. Knock-out of *Gpr75* in mice resulted in resistance to weight gain in a high-fat diet model, which was allele-dose dependent (25% lower weight gain for heterozygous *Gpr75*^{+/-} mice and 44% lower weight gain for knock-out *Gpr75*^{-/-} mice compared with wild-type *Gpr75*^{+/+} mice) and was accompanied by improved glycemic control and insulin sensitivity. Protein-truncating variants in *CALCR* were associated with higher BMI and obesity risk, while protein truncating variants in *GIPR* and two missense alleles (Arg190Gln or Glu288Gly), which we show result in loss-of-function *in vitro*, were associated with lower BMI and obesity risk. Rare protein-truncating variants in *UBR2*, *ANO4* and *PCSK1* were associated with over 2-fold higher odds of obesity in heterozygous carriers, an effect estimate similar to that of protein-truncating or predicted-deleterious missense variants in *MC4R* which are considered the most common cause of monogenic obesity. Among monogenic obesity genes in the leptin-melanocortin pathway, heterozygous carrier status for predicted loss-of-function variants in *MC4R*, *PCSK1*, *POMC* or *LEP* was associated with higher BMI, while heterozygous protein-truncating variants in *LEPR* were not associated with BMI. In Europeans from the general population in the United Kingdom, the penetrance of obesity in individuals carrying protein-truncating variants in *MC4R* varied from less than 30% to more than 60% in people at the bottom versus top quintile for the distribution of a polygenic score capturing predisposition to higher BMI due to over 2 million common genetic variants.

Conclusion: These results suggest that inhibition of GPR75 may be a therapeutic strategy for obesity and illustrate the power of massive-scale exome sequencing for the identification of large-effect coding variant associations and drug targets for complex traits.

Abstract (~125 words)

Large-scale human exome sequencing can identify rare protein-coding variants with large impact on complex traits as body adiposity. We sequenced the exomes of 645,626 individuals from the United Kingdom, United States, and Mexico and estimated associations of rare coding variants with body mass index (BMI). We identified 16 genes with an exome-wide significant association with BMI, including five brain-expressed G-protein coupled receptors (*CALCR*, *MC4R*, *GIPR*, *GPR151*, *GPR75*). Protein-truncating variants in *GPR75* were observed in ~4/10,000 sequenced individuals and were associated with 1.9 kg/m² lower BMI and 54% lower odds of obesity in heterozygous state. Knock-out of *Gpr75* in mice resulted in resistance to weight gain and improved glycemic control in a high-fat diet model. Inhibition of GPR75 may provide a therapeutic strategy for obesity.

Obesity and its health complications account for a substantial and growing burden of global disease. Understanding the genetic and molecular underpinnings of body adiposity can be a pathway to the development of safe and effective therapeutic strategies. Body fat is a highly heritable trait and genetic studies have revealed biological pathways that regulate energy balance. Studies in mouse models (1-3) and human forms of extreme, early-onset obesity (4-10) uncovered the influence of the leptin-melanocortin system on appetite regulation. Additionally, genome wide association studies (GWAS) of body mass index (BMI) have highlighted the polygenic contribution to the inherited basis of adiposity, identifying thousands of common genetic variants each with small effect sizes, and reaffirmed the broad influence of the central nervous system on body mass regulation (11-15).

Studies of rare protein-coding variants have helped translate genetic associations into new biological and therapeutic insights (4, 6-10, 16-24). Analyses of coding variation in human obesity have focused on (a) candidate gene or exome sequencing in pedigrees or case collections with extreme phenotypes or (b) array-genotyping of catalogued variant sites in large cohorts. Whole-exome sequencing of hundreds of thousands of individuals from population or health-system based studies can be a complementary approach with the potential to identify large-effect coding variants influencing the propensity to become obese or protection against obesity (25-28). Here, we report a multi-ethnic exome-sequencing association study for BMI in over 640,000 individuals across 3 distinct cohorts and regions of the world (United Kingdom, United States, and Mexico).

Exome-wide association of rare coding alleles with body mass index

We performed high-coverage whole exome sequencing in 645,626 individuals (29), including 428,719 European ancestry individuals from the United Kingdom Biobank cohort (UKB; **Table S1**) (30), 121,061 European ancestry individuals from the MyCode Community Health Initiative cohort from the USA-based Geisinger Health System (GHS; **Table S1**) (31) and 95,846 Admixed American ancestry individuals from the Mexico City Prospective Study (MCPS; **Table S1**) (32).

In an exome-wide meta-analysis across these three cohorts, there were sixteen genes for which the burden of rare nonsynonymous genetic variants was associated with BMI at the exome-wide level of statistical significance (inverse-variance weighted [IVW] meta-analysis $p < 3.6 \times 10^{-07}$, a

5 Bonferroni correction for 20,000 genes and seven variant selection models (29); **Table 1 and Fig. S1**). These associations were conditionally-independent of BMI-associated common variants identified by fine-mapping genome-wide association signals (29) and were consistent across the constituent datasets of the meta-analysis (**Table S2**).

10 Among the 16 genes, our analysis identified two genes where rare mutations are known to cause monogenic obesity (*MC4R* (8, 9) and *PCSK1* (6)) and two genes where rare coding variants have been associated with BMI (*GPR151* (33) and *GIPR* (16)). For the other 12 genes, our study provides genetic evidence linking rare coding variation to BMI and obesity-related phenotypes. Five of the sixteen genes encode G-protein coupled receptors (GPCRs; the largest class of drug targets in the human genome (34)) expressed in the brain and central nervous system (*ie.* *GPR75*, *CALCR*, *GIPR*, *GPR151* and *MC4R*). A tissue expression analysis for BMI-associated genes in our results revealed an over-representation of genes that are highly and specifically expressed in the hypothalamus, a key center for the neuro-endocrine regulation of energy balance (**Fig. S2**).

15 **Protein-truncating *GPR75* variants associated with leanness and protection against obesity in humans**

20 We explored in depth the new association for rare pLOF variants in the *GPR75* gene as this gene encodes a GPCR highly expressed in the brain across species and this association was the largest effect-size association with *lower* BMI in our exome-wide analysis.

25 Predicted loss-of-function variants in *GPR75* were observed in ~4 out of every 10,000 sequenced people, with similar frequency across populations (**Table S2**), and carrier status was associated with a 0.35 standard deviation lower BMI, corresponding to 1.9 kg/m² lower BMI or approximately 5.5 kg or 12 lbs lower body weight (**Table 1 and Fig. 1A**).

30 The association with lower BMI was directionally consistent and statistically significant in each of the constituent cohorts of our discovery meta-analysis (**Table S2**), as well as within age and sex subgroups (**Table S3**). We further corroborated the association of *GPR75* pLOF variants with lower BMI in a combined analysis including an additional 91,328 individuals not included in the discovery set (per-allele beta in standard deviation [SD] units of BMI in the meta-analysis of discovery and additional cohorts, -0.35, 95% confidence interval [CI], -0.47, -0.23, $p=3.9\times 10^{-9}$; **Fig. 1B**). This strong association with lower BMI was accompanied by a corresponding association with protection against obesity. Heterozygous carriers of *GPR75* pLOF variants had 35 54% lower odds of obesity compared to non-carriers in a meta-analysis of the UKB, GHS and MCPS cohorts (**Table S4**; per-allele odds ratio, 0.46, 95% confidence interval, 0.31, 0.67, $p=6.9\times 10^{-5}$) and their distribution across BMI categories was dramatically shifted towards lower BMI categories (**Fig. 1C**). None of 228 heterozygous carriers of *GPR75* pLOF variants were underweight (**Fig. 1C**). In UKB, *GPR75* pLOF carriers were more likely to self-report a thinner than average comparative body size at age 10 compared to non-carriers (**Table S5**).

45 We examined the genomic context of this new BMI association. The first and smallest exon of *GPR75*, containing untranslated sequence, is included in both *GPR75* and in a putative *GPR75-ASB3* readthrough gene with the nearby Ankyrin Repeat and SOCS Box Containing 3 (*ASB3*; **Fig. S3**). The second and final *GPR75* exon (containing the entire translated region of *GPR75*) is not shared with any other gene or transcript (**Fig. S3**). We conducted a number of analyses to

ensure that the association of pLOF variants could be firmly attributed to the *GPR75* gene. First, 45 of the 46 pLOF variants in *GPR75* that contributed to the association with lower BMI were located in exon 2 (**Table S6**), which is exclusive to the *GPR75* gene (**Fig. S3**). Accordingly, the burden genotypes for pLOF variants in *GPR75* had no linkage disequilibrium (LD; $R^2 < 0.0001$) with the burden genotype for pLOF variants affecting the *GPR75-ASB3* readthrough gene or the *ASB3* gene. Second, we estimated the association with BMI of the burden of rare coding variants in *ASB3* or in the *GPR75-ASB3* readthrough gene in our large exome sequencing meta-analysis. There was no association with BMI for the burden of rare nonsynonymous variants in *ASB3* or *GPR75-ASB3* across multiple statistical models with different variant annotation and allele frequency inclusion criteria (**Table S7**), including a lack of association for pLOF variants in either *ASB3* or *GPR75-ASB3* (**Table S7**). Finally, we estimated the association with BMI for the burden of rare pLOF variants in *GPR75* conditional upon *ASB3* and *GPR75-ASB3* genotypes. The association of *GPR75* pLOF variants with lower BMI was unaffected by adjusting for *ASB3* and *GPR75-ASB3* genotypes (beta in SD units of BMI, -0.35, 95% CI, -0.48 to -0.23, $p = 1.4 \times 10^{-08}$ before adjustment for *ASB3* or *GPR75-ASB3* genotypes; beta in SDs of BMI, -0.35, 95% CI, -0.47 to -0.23, $p = 1.3 \times 10^{-08}$ after adjustment; **Table S8**). Therefore, the association with lower BMI for rare pLOF variants in *GPR75* can be confidently attributed to the *GPR75* gene.

We also explored whether there were common variant associations in the locus. In the 1Mb window surrounding *GPR75* (500kb either side of the gene), there were 26 common variants associated with BMI at the genome-wide level of statistical significance (IVW meta-analysis, $p < 5 \times 10^{-08}$) in our GWAS of imputed common variants in Europeans (**Table S9, Fig. S4**), while there were no genome-wide significant associations in admixed Americans (**Fig. S4**). These 26 variants all fine-mapped to a signal led by rs59428052 (G-allele frequency, 14.7%; posterior probability of causal association, 30.4%; per-allele beta in SD units of BMI, -0.015; 95% confidence interval, -0.020 to -0.010; $p = 1.3 \times 10^{-09}$), which is an intergenic variant nearest to *ASB3* and approximately 200kb downstream of *GPR75*. The rs59428052 variant did not co-localize with any eQTL signal nor were any of the additional 25 variants at the locus in LD ($R^2 > 0.8$) with any sentinel eQTLs in GTEx v8 (**Table S9**). Two of the 26 variants were in LD with a missense variant in *ASB3* and *GPR75-ASB3* (rs36020289), which does not affect the *GPR75* transcript (**Table S9**).

We performed a formal conditional analysis adjusting for the 26 common variants associated with BMI in the region and identified that the association with lower BMI for pLOF variants in *GPR75* remains unchanged (**Table S8**). Therefore, the association with lower BMI for rare pLOF variants in *GPR75* is conditionally independent of any of the 26 common variants associated with BMI at the locus in Europeans.

In summary, our human genetic analysis at the locus indicates that: (a) rare pLOF variants in *GPR75* are associated with lower BMI with a large effect association, (b) the pLOF association is attributed to *GPR75* and not to other nearby transcripts, (c) the signal is independent of BMI-associated common variants in the region and (d) the small-effect intergenic common variant signal found in that region by GWAS fine-mapping in Europeans has no apparent link with *GPR75*.

The association with lower BMI for pLOF variants in *GPR75* was due to multiple independent rare pLOF variants predicted to truncate *GPR75* at different locations (**Fig. 1A and Table S6**). Due to their rarity, none of the 46 rare pLOF variants found by exome sequencing in our analysis were well ascertained by array-genotyping or imputation (**Table S6**). Leave-one-out analyses showed that the burden signal was robust to the exclusion of one pLOF variant at a time (**Table S10**). Out of 46 rare pLOF variants in *GPR75*, six were individually associated with lower BMI at a nominal level of statistical significance (IVW meta-analysis $p < 0.05$; Ala110fs; Ser219fs; Gln234*; Cys400fs, Lys404*, c.-110+1G>A; **Table S11**), while none was associated with higher BMI. When excluding all 6 of these variant sites from analysis, the remaining set of pLOF variants was still associated with lower BMI (**Table S10**).

We expressed *in vitro* the two most frequent (minor allele count ≥ 10) amongst the six pLOF variants individually associated with BMI and show that they result in cellular retention of a truncated receptor likely leading to a complete loss of function (**Fig. 2**).

We predict that the loss of a functional copy (ie. haploinsufficiency) or production of a truncated protein that disrupts receptor multimers (ie. dominant negative effects) may explain the association of *GPR75* truncation with lower BMI. We hypothesized that in the case of haploinsufficiency the earlier N-terminal truncation of *GPR75* would result in greater phenotypic impact than a C-terminal truncation within the last intracellular domain. Genetic variants resulting in truncation of *GPR75* before the final intracellular domain were associated with -2.2 kg/m^2 lower BMI (IVW meta-analysis $p = 1.3 \times 10^{-07}$) as compared with -1.4 kg/m^2 lower BMI (IVW meta-analysis $p = 0.018$) for variants resulting in truncation within the final domain (**Table S12**). This difference was even more pronounced for truncations within the last 100 amino-acids of the final C-terminal domain (**Table S12**).

Body composition analysis with bioimpedance in UKB showed that the association with lower BMI was driven by an association with lower overall body fat mass and lower body fat percentage (**Fig. S5**). In an agnostic phenome-wide analysis of *GPR75* pLOF variants (29), we did not observe statistically-significant associations with common diagnoses or measured continuous traits after correction for the number of statistical tests performed (2,173 phenotypes tested; Bonferroni-corrected p-value threshold, $p < 2.3 \times 10^{-05}$), reflecting the rarity of these variants and the stringent multiple test correction.

A detailed analysis of metabolic traits revealed a nominally-significant association (IVW meta-analysis $p < 0.05$) with higher high-density lipoprotein cholesterol, which is consistent with a favorable metabolic profile (**Table S13**). Carriers of pLOF in *GPR75* had lower odds of type 2 diabetes compared to non-carriers (63,492 cases and 549,961 controls; per-allele odds ratio, 0.92; 95% confidence interval, 0.59, 1.45; $p = 0.73$; **Table S13**), but the difference was not statistically significant. We interrogated exome sequencing association statistics from up to 20,791 type 2 diabetes cases and 24,440 controls included in the T2D Knowledge Portal (<https://t2d.hugeamp.org/>; Accessed January 8th, 2021), and similarly observed numerically lower odds of type 2 diabetes in carriers of *GPR75* pLOF variants (odds ratio for type 2 diabetes, 0.52, 95% CI, 0.14 to 1.97; $p = 0.30$; allele frequency, 0.03%). Due to the rarity of pLOF variants in *GPR75* and given the genetic relationship between BMI and type 2 diabetes, we estimate that millions of people would need to be sequenced to detect an association at $p < 0.05$ (**Table S13**).

An analysis for HbA1c, a continuous biomarker of glycemic levels, led to similar results (**Table S13**).

***Gpr75* deletion confers resistance to high-fat diet induced obesity in mice**

In a mouse model of high-fat diet (HFD) induced obesity, experimental deletion of *Gpr75* protected against weight gain and its associated abnormalities in glucose and insulin metabolism (**Fig. 3**). When placed on HFD for 14 weeks, *Gpr75*^{+/+} mice approximately doubled their weight. Body weight changed from an average (standard deviation) of 20.9 (2.1) to 43.3 (6.5) grams (body weight change, +22.4 grams). In contrast, mice with a genetic deletion of *Gpr75* gained less weight in an allele-dose dependent fashion (body weight change +16.9 grams, difference in weight-change with wild-type -5.5 grams or -25% for *Gpr75*^{+/-} mice; body weight change +12.6 grams, difference in weight-change with wild-type -9.8 grams or -44% for the *Gpr75*^{-/-} mice; **Fig. 3A**). Increases in fasting blood glucose seen with HFD in *Gpr75*^{+/+} mice were reduced in an allele-dose dependent manner in *Gpr75*^{+/-} and *Gpr75*^{-/-} mice (**Fig. 3B**). Mice with a genetic deletion in *Gpr75* were also resistant to HFD-induced impairments in glucose tolerance and insulin sensitivity (**Fig. 3C-D**). At the end of 14 weeks of HFD, plasma leptin levels were lower in *Gpr75*^{-/-} and *Gpr75*^{+/-} mice compared to wild-type mice (**Fig. S6**), whereas adiponectin levels were higher resulting in a 2- and 10-fold lower leptin to adiponectin ratio in *Gpr75*^{+/-} and *Gpr75*^{-/-} mice compared to wild-type (**Fig. S6**).

Genomic insights at other BMI-associated GPCRs and known monogenic obesity genes from exome sequencing

Gene burden associations at other GPCRs illustrate the complementarity of large-scale exome sequencing analyses to common variants GWAS in identifying effector genes, establishing directionality of association (ie. whether LOF in a gene is associated with higher or lower BMI levels) and identifying variants whose functional follow-up can provide biological insights.

In our GWAS fine-mapping analysis, we identified four distinct signals in the 1Mb region around the *CALCR* gene (**Table S14**). While these common variant associations point to *CALCR* as a possible effector gene, they do not inform on their own on whether reduced *CALCR* function would be associated with higher or lower BMI. In the exome analysis, we identified a significant association for the burden of rare (AAF < 0.1%) pLOF and predicted-deleterious missense variants in *CALCR* with 0.08 SDs (~0.5 kg/m²) higher BMI and 20% higher odds of obesity (OR, 1.20; 95% CI, 1.12, 1.20; p=8.9×10⁻⁰⁷; **Tables 1 and S4**). In addition, the burden of *CALCR* pLOF genetic variants alone (ie. excluding missense variants) was associated with higher BMI (**Table S15**), indicating that loss-of-function in *CALCR* is associated with higher BMI and obesity risk in humans.

At *GIPR*, a known BMI locus (16), we identified an exome-wide significant association for the burden of pLOF and predicted deleterious missense variants with lower BMI (**Table 1**). This association remained statistically significant, albeit attenuated, after accounting for Arg190Gln or Glu288Gly (**Table S15**), two rare missense variants with unknown functional consequences previously associated with lower BMI (16).

In our exome sequencing meta-analysis, rare protein truncating variants in *GIPR* were associated with lower BMI, with a similar effect size to that of Arg190Gln or Glu288Gly (**Fig. 4A and**

Table S15), suggesting that these missense variants may result in a loss-of-function. We tested this hypothesis in cell-based expression experiments, showing that both Arg190Gln-GIPR and Glu288Gly-GIPR result in a near-complete loss-of-function with respect to Gs and Gq signaling when agonized with recombinant glucose-dependent insulintropic polypeptide (**Fig. 4B**), compared to wild-type. These results indicate that heterozygous loss-of function in *GIPR* results in lower BMI and obesity risk in humans.

Using exome sequencing, we confirmed the association of pLOF and deleterious missense variants in *MC4R* with higher BMI (**Table 1**) and that of low-frequency gain-of-function variants in *MC4R* (Val103Ile and Ile251Leu) with lower BMI (**Table S15**). In cell-based experiments, we also show that the Val103Ile variant is resistant to the agouti-related peptide mediated inhibition of MC4R signaling and confirm the signaling preference towards β -arrestin pathway of this gain of function variant (17) (**Fig. S7**).

In the MCPS cohort, we identified an *MC4R* signal that was fine-mapped to a single likely-causal missense variant (Ile269Asn, 18:60371544:A:T in **Table S14**). This variant was previously associated with childhood and adult obesity in Mexico (35) and experimentally shown to result in a complete LOF of both cAMP signaling and β -arrestin recruitment (17). The variant is extremely rare in individuals of European ancestry but the allele frequency was of 1% in admixed American individuals from the MCPS cohort. Heterozygous *MC4R* deficiency is considered the most common monogenic form of human obesity (prevalence in cases of severe early-onset obesity, ~6%) (36). In obese individuals of European ancestry from the UKB cohort, the prevalence of pLOF or rare (AAF < 1%) deleterious missense variants in *MC4R* was 0.4%, indicating that MC4R deficiency is a rare genetic contributor to general obesity in that population. In obese admixed American participants from the MCPS cohort, the prevalence of MC4R deficiency due to the Ile269Asn variant was approximately 3%. Therefore, the prevalence of heterozygous MC4R deficiency in general obesity is over 7-fold greater in admixed Americans from Mexico, than in Europeans from the UK.

Homozygous *MC4R* deficiency has been described in only a handful of cases of severe early onset obesity (36) but the penetrance of obesity in this condition is unknown. In MCPS, there were 17 homozygous carriers of Ile269Asn, 12 of whom were obese and 5 overweight. These data suggest that homozygous MC4R deficiency might be incompatible with the maintenance of a healthy weight in adulthood.

Our exome-wide study also identified additional large effect-size associations with higher BMI and obesity risk for rare coding variants in other genes, including associations with over 2-fold higher odds of obesity for protein truncating variants in *PCSK1* (in heterozygous state), *UBR2* or *ANO4* (**Table 1 and S4**). These associations were similar in effect size as those for rare (AAF < 1%) pLOF or deleterious missense variants in *MC4R*, but the frequency of these genotypes was lower than that of *MC4R* pLOF or deleterious missense alleles (**Table 1**). Hence, these rare mutations occur in only a small number of people in the general population. However, they may have a large impact on obesity risk in those small groups of carriers.

We used our large exome sequencing dataset to investigate the association with BMI of pLOF variants in five leptin-melanocortin pathway genes where pLOF variants have been previously

linked to monogenic obesity (*LEP* [leptin deficiency, MIM: 614962], *LEPR* [leptin receptor deficiency, MIM: 614963], *POMC* [proopiomelanocortin deficiency, MIM: 609734], *PCSK1* [proprotein convertase 1/3 deficiency, MIM: 600955] and *MC4R* [melanocortin 4 receptor deficiency, MIM: 618406]). Of these, only *MC4R* deficiency is considered to have autosomal dominant inheritance, while all other deficiencies are considered to be autosomal recessive. Heterozygous carrier status for pLOF variants in *LEP*, *POMC*, *PCSK1* or *MC4R* was associated with higher BMI (IVW meta-analysis $p < 0.05$; **Table S16**), while heterozygous carrier status for pLOF variants in *LEPR* was not associated with BMI suggesting a pure autosomal recessive inheritance for leptin receptor deficiency.

Rare single-variant exome-wide analysis reveals additional signals at the *SOS2* and *SRRM2* genes

In addition to the gene burden association analyses, we also conducted single-variant association analyses, looking for rare nonsynonymous alleles individually associated with BMI at the genome-wide level of statistical significance (IVW meta-analysis $p < 5 \times 10^{-8}$) conditional on GWAS fine-mapped common variants (29). We identified 6 such variants, including 4 occurring in genes identified in the primary gene burden analysis (*GPR151*, *SPARC*, *MC4R* and *ANKRD27*; **Table S17**) and 2 variants in genes not identified in other analyses (Arg2033Pro in *SRRM2* and Pro191Arg in *SOS2*; **Table S17**). The missense variant in *SOS2* was associated with 0.05 SDs (0.28 kg/m²) lower BMI per allele.

Interestingly, mutations in *SOS2* are associated with autosomal dominant forms of Noonan syndrome (MIM: 616559), a condition that has been associated with lower prevalence of overweight (37).

Combined use of common variant GWAS fine-mapping and exome-sequencing gene burden associations to prioritize likely effector genes for BMI

Fine-mapping of GWAS associations in the discovery cohorts of our analysis identified 1,905 independent signals led by sentinel common variants (29) (AAF > 1%; **Table S18**), which had a median 95% credible set size of 36 likely-causal variants (interquartile range, 12 to 119). We used MANTRA (Meta-ANalysis of Transethnic Association studies) (38), a Bayesian trans-ethnic meta-analysis approach, to estimate associations of fine-mapped signals across ancestries and observed strong evidence of association across datasets (median log₁₀ Bayes factor, 7.0; interquartile range, 4.9 to 10.4; **Table S18**).

Out of the 1,905 signals, 13 (0.7%) fine-mapped to a single nonsynonymous sentinel variant with >95% posterior probability of causal association (12 missense variants and 1 splice site variant in 12 genes; **Table S19**). These included a gene identified in our gene burden analysis (*MC4R*) and an additional 11 genes. We investigated whether there were associations for the burden of rare (AAF < 1%) pLOF or pLOF plus predicted deleterious (5/5 *in silico* prediction algorithms) missense variants in these genes prioritized by GWAS fine-mapping. We found evidence of association (IVW meta-analysis $p < 0.05$) for 4 of these 11 genes (36%; **Table S19**). These common and rare variants association pairs included the His48Arg missense variant and the burden of rare pLOF or predicted deleterious missense variants in *ADH1B*, encoding the key ethanol metabolism enzyme alcohol dehydrogenase 1B. His48Arg is known to associate with higher BMI via increased alcohol consumption (39) and our rare variant analysis corroborates

the causal nature of that association and the importance of alcohol consumption in weight regulation.

We used exome sequencing results to investigate whether there were associations for rare nonsynonymous variants in genes at the *FTO* locus, where common variants have been associated with BMI in early GWAS studies (11, 12) and where experimental studies have suggested the distant *IRX3* and *IRX5* as likely effector genes (40). We did not observe an association between rare coding variants in the *FTO*, *IRX3* or *IRX5* with BMI (Table S20). Since pLOF variants in these genes were rare (AAF \leq 0.02%; Table S20), this analysis can only exclude large-effect associations with BMI for heterozygous pLOF variants in those genes. The strongest associations for the burden of rare coding variants in the region was for rare (AAF < 1%) pLOF or predicted deleterious (5/5 *in silico* prediction algorithms) missense variants in the *CHD9* gene (Table S20).

At 9 out of the sixteen 1Mb-regions around the genes identified in our exome-wide gene burden analysis there were common variant (sentinel AAF > 1%) signals identified by GWAS fine-mapping, while at the remaining 7 regions there were no nearby fine-mapped common sentinel variants (Table S14). At the 9 loci with common variant signals, we used physical proximity, common nonsynonymous variants and eQTL co-localization in either European and all-ancestry GTEx v8 datasets to identify likely effector genes for the common variant associations (29) (Table S14 and S21). At 6 of the 9 regions, the gene identified in the exome-wide analysis was also one of the genes that were prioritized by common variants associations (Table S14). However, at 4 of those 6 loci, other genes were also prioritized resulting in an uncertain effector gene attribution based on common variants alone. In these 4 loci, gene burden associations for the gene identified in the exome analysis had an average of 18-orders of magnitude stronger statistical association than any of the other possible effector genes prioritized by common variants (average difference in $-\log_{10}[p]$, 17.9; range, 5.8 to 42.4; Table S14), suggesting that gene burden analyses may dramatically help prioritize effector genes at some GWAS-associated loci. At the remaining three loci (*DPP9*, *UBR2*, *GPR75*), common variant associations prioritized different genes than the ones identified in the exome-wide analysis (Table S14). These results suggest that, at those loci, common variants associations act via different genes than the ones found by exome-sequencing or that the gene prioritization from common variant associations did not identify the correct effector gene for those common variant signals.

Polygenic burden influences the penetrance of obesity in carriers of high-impact rare coding alleles

Both common and rare alleles contribute to the risk of general obesity, but their interplay in shaping obesity risk has been understudied due to the lack of datasets with both common and rare variant ascertainment. Here, we generated a genome-wide polygenic score capturing genetic predisposition to higher BMI due to over 2.5 million common alleles (29) and studied its interplay with rare, large effect-size coding variants in shaping risk for obesity in the population-based UKB cohort. This analysis suggests that polygenic burden influences the penetrance of obesity and the level of BMI in carriers of *GPR75* (large-effect protective association) or *MC4R* (large-effect risk-increasing association) pLOF variants in a linearly additive manner (Fig. 5 and Fig. S8). The penetrance of obesity in individuals carrying protein-truncating variants in *MC4R*

varied from less than 30% to more than 60% in people at the bottom versus top quintile for the distribution of the polygenic score (**Fig. 5**). There was a nearly 60% absolute difference in the prevalence of obesity between extremes of genetic predisposition, ie. *GPR75* pLOF carriers in the bottom quintile of polygenic burden and *MC4R* pLOF carriers in the top quintile (**Fig. 5**).

Discussion

By conducting a large exome-sequencing study on the influences of rare coding variation on body adiposity, we make a number of observations that advance our understanding of the genes and pathways involved in propensity for and resistance to obesity in humans.

We discovered a new association for rare protein-truncating genetic variants in *GPR75* with lower adiposity and substantial protection from obesity. We validated this human genetic association in a high-fat diet model of obesity in mice, where genetic ablation of *Gpr75* was associated with resistance to weight gain, greater insulin sensitivity and improved glycemic control. In our analysis, the association for pLOF variants in *GPR75* showed the largest effect-size genetic association with *lower* adiposity and protection against obesity at the genome-wide level. The estimated effect size for this association appears to be 3- to 4-times larger than the largest effect size associations for common genetic variants at the *FTO* locus (11, 13) or for low-frequency gain-of-function variants in *MC4R* (17). The observation of a consistent association across cohorts from different regions of the world and with different study design, as well as multiple ancestries highlights the generalizability of this association in people with various genetic backgrounds and environmental exposures.

Human genetics validation is a predictor of the likely success of drug development programs (41, 42) where the identification of naturally-occurring protective alleles has catalyzed the translation from genetic association to therapeutic drug development in a growing number of examples (17-22, 43-45). Therefore, our findings suggest that GPR75 inhibition could be a therapeutic approach for obesity.

The expression of *GPR75* in the hypothalamic nuclei and other brain regions, protection from weight gain for *Gpr75* knock-out mice under high-fat diet challenge and previous evidence of the role of brain GPCRs in energy balance regulation suggest that this receptor may be implicated in the brain-mediated regulation of energy balance, providing an important direction for future mechanistic research. It will also be important to clarify whether 20-hydroxyeicosatetraenoic acid (20-HETE), an eicosanoid metabolite of arachidonic acid, or C-C Motif Chemokine Ligand 5 (CCL5), a chemokine, which have been previously proposed as putative ligands for GPR75 (46, 47) or other yet-undiscovered ligands are responsible for the link between GPR75 loss-of-function and body weight regulation.

In addition to GPR75, our agnostic exome-wide analysis identified four other GPCRs expressed in the brain and previously implicated in energy balance regulation, highlighting once again the importance of neurological pathways in obesity risk first shown in family studies of extreme obesity and more recently in genome-wide association analyses of common variants. GIPR is expressed in adipose, brain, bone and other metabolically-active tissues where it acts as the receptor for glucose-dependent insulinotropic polypeptide, an incretin hormone involved in the regulation of insulin secretion, gastric emptying and other metabolic processes (48). CALCR is the receptor of calcitonin and amylin, a peptide hormone secreted by pancreatic beta-cells which has been shown to promote satiety, delayed gastric emptying and weight control in patients with type 2 diabetes (49-51). Ablation of *Calcr*-expressing neurons in the nucleus tractus solitarius has been recently implicated in the disruption of a leptin-independent pathway of appetite regulation in murine models (52). *GPR151* encodes a habenular receptor involved in addictive

behavior and differential food-intake response to nicotine (53). The association effect size for pLOF and predicted deleterious missense variants in our analysis as well as the previously reported association for the Arg95* allele (33) is tiny (-0.3 kg/m² per allele), suggesting that it may be secondary to other behavioral or neurological phenotypes possibly related to addiction. In addition to these GPCRs, our exome-wide analysis revealed associations with body adiposity for rare coding variants in several other genes. While in this study we primarily focused on the associations in *GPR75* and other GPCRs due to their more immediate translational potential, the identification of these additional associations in our large study provides an initial foundation for understanding the role in the regulation of body fat and obesity risk of these genes.

Our study also illustrates the power and versatility of massive-scale exome sequencing in population- or health-system based cohort studies as a genetic discovery approach complementary to common variant GWAS and pedigree-based studies. We show the utility of this approach in (a) discovering large-effect protective or risk-increasing associations for genes and individual rare coding variants, (b) identifying functional variants that can be studied *in vitro* for biological insight, (c) prioritizing effector genes for common variant signals identified by fine-mapping, and (d) understanding the impact of rare variants on complex traits like obesity across populations and their interplay with common variation in shaping disease risk or resistance.

In conclusion, the results from this exome-sequencing analysis in over 640,000 people suggest that inhibition of GPR75 may be a therapeutic strategy for obesity and illustrate the power of massive-scale exome sequencing for the identification of large-effect coding variant associations and drug targets for complex traits.

Methods

Detailed materials and methods are described in the study supplement (29). Briefly, we performed high-coverage whole exome sequencing in 645,626 individuals, including 428,719 European ancestry individuals from the UK Biobank cohort, 121,061 European ancestry individuals from the MyCode Community Health Initiative cohort from the USA-based Geisinger Health System and 95,846 Admixed American ancestry individuals from the Mexico City Prospective Study. The outcome measure was body mass index (BMI), calculated as weight in kilograms divided by the square of standing height in meters. We estimated associations with BMI for the burden of rare nonsynonymous variants in each sequenced gene by fitting mixed-effects regression models accounting for population stratification and relatedness using BOLT-LMM v2.3.4 (54) or REGENIE v1.0 (55). Results across studies were pooled by inverse-variance weighted meta-analysis. Consistent with previous literature (27, 28), for our primary gene-burden association analysis we considered a threshold of exome-wide statistical significance of $p < 3.6 \times 10^{-07}$, a Bonferroni correction for 20,000 genes and seven variant selection models. In parallel, to leverage evidence from genome-wide common variants, we performed genome-wide association analyses (GWAS) of imputed common alleles in the same discovery dataset used for our exome sequencing analysis. We leveraged fine-mapping of common alleles, formal conditional analyses, physical proximity mapping, linkage disequilibrium with common nonsynonymous alleles, eQTL co-localization and polygenic score analysis to highlight the complementarity of evidence from exome-sequencing and GWAS of common variants. To illustrate the translational value of exome-sequencing identified rare coding associations, we

performed targeted *in vitro* and *in vivo* experiments. We performed *in vitro* expression studies for the functional characterization of naturally-occurring variants in *GPR75*, *MC4R* and *GIPR*. We also developed genetically engineered knock-out *Gpr75*^{-/-} mouse strains using the VelociGene® technology (56, 57) and studied the weight gain, glycemic and insulinemic phenotypes of knock-out *Gpr75*^{-/-} and *Gpr75*^{+/-} heterozygous mice compared with *Gpr75*^{+/+} wild-type mice in a high-fat diet model.

References and Notes:

1. Y. Zhang *et al.*, Positional cloning of the mouse obese gene and its human homologue. *Nature* **372**, 425-432 (1994).
2. W. Fan, B. A. Boston, R. A. Kesterson, V. J. Hruby, R. D. Cone, Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165-168 (1997).
3. D. Huszar *et al.*, Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131-141 (1997).
4. C. T. Montague *et al.*, Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature* **387**, 903-908 (1997).
5. S. O'Rahilly *et al.*, Brief report: impaired processing of prohormones associated with abnormalities of glucose homeostasis and adrenal function. *N Engl J Med* **333**, 1386-1390 (1995).
6. R. S. Jackson *et al.*, Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet* **16**, 303-306 (1997).
7. K. Clement *et al.*, A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature* **392**, 398-401 (1998).
8. G. S. Yeo *et al.*, A frameshift mutation in *MC4R* associated with dominantly inherited human obesity. *Nat Genet* **20**, 111-112 (1998).
9. C. Vaisse, K. Clement, B. Guy-Grand, P. Froguel, A frameshift mutation in human *MC4R* is associated with a dominant form of obesity. *Nat Genet* **20**, 113-114 (1998).
10. H. Krude *et al.*, Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by *POMC* mutations in humans. *Nat Genet* **19**, 155-157 (1998).
11. T. M. Frayling *et al.*, A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889-894 (2007).
12. R. J. Loos *et al.*, Common variants near *MC4R* are associated with fat mass, weight and risk of obesity. *Nat Genet* **40**, 768-775 (2008).
13. A. E. Locke *et al.*, Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197-206 (2015).
14. L. Yengo *et al.*, Meta-analysis of genome-wide association studies for height and body mass index in approximately 700000 individuals of European ancestry. *Hum Mol Genet* **27**, 3641-3649 (2018).
15. A. V. Khera *et al.*, Polygenic Prediction of Weight and Obesity Trajectories from Birth to Adulthood. *Cell* **177**, 587-596 e589 (2019).
16. V. Turcot *et al.*, Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat Genet* **50**, 26-41 (2018).
17. L. A. Lotta *et al.*, Human Gain-of-Function *MC4R* Variants Show Signaling Bias and Protect against Obesity. *Cell* **177**, 597-607 e599 (2019).
18. K. Musunuru *et al.*, Exome sequencing, *ANGPTL3* mutations, and familial combined hypolipidemia. *N Engl J Med* **363**, 2220-2227 (2010).
19. F. E. Dewey *et al.*, Genetic and Pharmacologic Inactivation of *ANGPTL3* and Cardiovascular Disease. *N Engl J Med* **377**, 211-221 (2017).
20. N. S. Abul-Husn *et al.*, A Protein-Truncating *HSD17B13* Variant and Protection from Chronic Liver Disease. *N Engl J Med* **378**, 1096-1106 (2018).

21. A. B. Jorgensen, R. Frikke-Schmidt, B. G. Nordestgaard, A. Tybjaerg-Hansen, Loss-of-function mutations in APOC3 and risk of ischemic vascular disease. *N Engl J Med* **371**, 32-41 (2014).
22. Tg *et al.*, Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *N Engl J Med* **371**, 22-31 (2014).
- 5 23. J. Flannick *et al.*, Loss-of-function mutations in SLC30A8 protect against type 2 diabetes. *Nat Genet* **46**, 357-363 (2014).
24. G. Marenne *et al.*, Exome Sequencing Identifies Genes and Gene Sets Contributing to Severe Childhood Obesity, Linking PHIP Variants to Repressed POMC Transcription. *Cell Metab* **31**, 1107-1119 e1112 (2020).
- 10 25. F. E. Dewey *et al.*, Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. *Science* **354**, (2016).
26. C. V. Van Hout *et al.*, Exome sequencing and characterization of 49,960 individuals in the UK Biobank. *Nature* **586**, 749-756 (2020).
27. J. Flannick *et al.*, Exome sequencing of 20,791 cases of type 2 diabetes and 24,440 controls. *Nature* **570**, 71-76 (2019).
- 15 28. R. Do *et al.*, Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102-106 (2015).
29. Materials and methods are available as supplementary materials.
30. C. Sudlow *et al.*, UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* **12**, e1001779 (2015).
- 20 31. D. J. Carey *et al.*, The Geisinger MyCode community health initiative: an electronic health record-linked biobank for precision medicine research. *Genet Med* **18**, 906-913 (2016).
32. R. Tapia-Conyer *et al.*, Cohort profile: the Mexico City Prospective Study. *Int J Epidemiol* **35**, 243-249 (2006).
- 25 33. C. A. Emdin *et al.*, Analysis of predicted loss-of-function variants in UK Biobank identifies variants protective for disease. *Nat Commun* **9**, 1613 (2018).
34. A. S. Hauser, M. M. Attwood, M. Rask-Andersen, H. B. Schioth, D. E. Gloriam, Trends in GPCR drug discovery: new agents, targets and indications. *Nat Rev Drug Discov* **16**, 829-842 (2017).
35. M. Vazquez-Moreno *et al.*, The Melanocortin 4 Receptor p.Ile269Asn Mutation Is Associated with Childhood and Adult Obesity in Mexicans. *J Clin Endocrinol Metab* **105**, (2020).
- 30 36. I. S. Farooqi *et al.*, Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085-1095 (2003).
37. G. Binder *et al.*, Health and quality of life in adults with Noonan syndrome. *J Pediatr* **161**, 501-505 e501 (2012).
- 35 38. A. P. Morris, Transethnic meta-analysis of genomewide association studies. *Genet Epidemiol* **35**, 809-822 (2011).
39. M. V. Holmes *et al.*, Association between alcohol and cardiovascular disease: Mendelian randomisation analysis based on individual participant data. *BMJ* **349**, g4164 (2014).
40. M. Claussnitzer *et al.*, FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895-907 (2015).
- 40 41. D. Cook *et al.*, Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat Rev Drug Discov* **13**, 419-431 (2014).
42. M. R. Nelson *et al.*, The support of human genetic evidence for approved drug indications. *Nat Genet* **47**, 856-860 (2015).
- 45 43. N. O. Stitzel *et al.*, ANGPTL3 Deficiency and Protection Against Coronary Artery Disease. *J Am Coll Cardiol* **69**, 2054-2063 (2017).
44. T. I. Pollin *et al.*, A null mutation in human APOC3 confers a favorable plasma lipid profile and apparent cardioprotection. *Science* **322**, 1702-1705 (2008).
45. J. C. Cohen, E. Boerwinkle, T. H. Mosley, Jr., H. H. Hobbs, Sequence variations in PCSK9, low LDL, and protection against coronary heart disease. *N Engl J Med* **354**, 1264-1272 (2006).
- 50

46. V. Garcia *et al.*, 20-HETE Signals Through G-Protein-Coupled Receptor GPR75 (Gq) to Affect Vascular Function and Trigger Hypertension. *Circ Res* **120**, 1776-1788 (2017).
47. S. Dedoni, L. A. Campbell, B. K. Harvey, V. Avdoshina, I. Mocchetti, The orphan G-protein-coupled receptor 75 signaling is activated by the chemokine CCL5. *J Neurochem* **146**, 526-539 (2018).
48. J. J. Holst, M. M. Rosenkilde, Recent advances of GIP and future horizons. *Peptides* **125**, 170230 (2020).
49. P. A. Hollander *et al.*, Pramlintide as an adjunct to insulin therapy improves long-term glycemic and weight control in patients with type 2 diabetes: a 1-year randomized controlled trial. *Diabetes Care* **26**, 784-790 (2003).
50. P. Hollander *et al.*, Effect of pramlintide on weight in overweight and obese insulin-treated type 2 diabetes patients. *Obes Res* **12**, 661-668 (2004).
51. O. Schmitz, B. Brock, J. Rungby, Amylin agonists: a novel approach in the treatment of diabetes. *Diabetes* **53 Suppl 3**, S233-238 (2004).
52. W. Cheng *et al.*, Calcitonin Receptor Neurons in the Mouse Nucleus Tractus Solitarius Control Energy Balance via the Non-aversive Suppression of Feeding. *Cell Metab* **31**, 301-312 e305 (2020).
53. B. Antolin-Fontes *et al.*, The habenular G-protein-coupled receptor 151 regulates synaptic plasticity and nicotine intake. *Proc Natl Acad Sci U S A* **117**, 5502-5509 (2020).
54. P. R. Loh, G. Kichaev, S. Gazal, A. P. Schoech, A. L. Price, Mixed-model association for biobank-scale datasets. *Nat Genet* **50**, 906-908 (2018).
55. J. Mbatchou *et al.*, Computationally efficient whole genome regression for quantitative and binary traits. *bioRxiv*, 2020.2006.2019.162354 (2020).
56. D. M. Valenzuela *et al.*, High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21**, 652-659 (2003).
57. W. T. Poueymirou *et al.*, F0 generation mice fully derived from gene-targeted embryonic stem cells allowing immediate phenotypic analyses. *Nat Biotechnol* **25**, 91-99 (2007).

Acknowledgments

The authors are grateful to Prof. Jonathan C. Cohen, University of Texas Southwestern, Dallas (USA). **Funding** This research received funding from Regeneron Pharmaceuticals. New York Medical College investigators received funding from the National Institutes of Health grants HL034300 (M.L.S.) and HL139793 (M.L.S.) and the diversity supplement HL139793-1S (V.G). The Mexico City Prospective Study has received funding from the Mexican Health Ministry, the Mexican National Council of Science and Technology, the Wellcome Trust, the UK Medical Research Council Population Health Research Unit at the University of Oxford, and the Nuffield Department of Population Health at the University of Oxford. Y.I.C, J.I.R were supported by the National Center for Advancing Translational Sciences, CTSI grant UL1TR001881, 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. **Author contributions:** [draft to be finalized with co-authors input] All authors contributed to the collection of data, interpretation of results, critical review and editing of the manuscript for important intellectual content, final approval of submission of the manuscript for publication; A.B, M.A.R.F, and L.A.L project conceptualization; P.A, A.G, O.S, J.A.K, L.K, Y.F, T.P, V.G, D. S, A.Li, J.Mb., A.E.L, C.B, N.V, N.L, S.H, K.A, J.V.P, E.D, M.D, S.B, A.H, M.J, B.Z, A.J.M, C.P, G.C, J.D.O, J.G.R, A.R.S, M.C, H.M.K, G.R.A, K.K, A.N.E, J.M, G.D.Y, M.W.S, J.A, G.D, M.L.S, A.B, M.A.R.F, and L.A.L development or application of methods; P.A, O.S, and L.A.L paper drafting; P.A, O.S, J.A.K, D.S, A.Li, J.Mb, A.E.L, C.B, N.V, N.L, S.B, A.H, H.M.K, G.R.A, J.M., M.A.R.F., L.A.L data analysis; L.K, Y.F, T.P, M.W.S, J.A, G.D *in vitro* experiments design, analysis, interpretation; A.G, V.G, S.H, K.A, J.V.P, E.D, M.D, and M.L.S design, analysis, interpretation of mouse models and experiments; J.B, P.K, J.A,

J.M.T, J.R.E, R.C, and R.T design, data and biological sample collection and organization, securing funding, implementation and coordination for the Mexico City Prospective Study; C.D.S, T.M, and D.J.C design, data and biological sample collection and organization, securing funding, implementation and coordination for the Geisinger Health System study; O.M design, data and biological sample collection and organization, securing funding, implementation and coordination for the Malmö Diet and Cancer study; Y.I.C, and J.I.R design, data and biological sample collection and organization, securing funding, implementation and coordination for the Taiwan metabochip consortium; D.J.R design, data and biological sample collection and organization, securing funding, implementation and coordination for the University of Pennsylvania Penn Medicine BioBank; W.E.K, S.H.S, design, data and biological sample collection and organization, securing funding, implementation and coordination for the Duke Catheterization Genetics cohort. **Competing Interests:** [draft to be finalized with co-authors input] Regeneron authors receive salary from and own options and/or stock of the company. Parts of this work have been described in one or more pending provisional patent applications. G.D.Y is the Chief Scientific Officer and member of the Board of Directors at Regeneron Pharmaceuticals; A.J.M is an Executive Officer of Regeneron Pharmaceuticals. **Data and materials availability:** The data reported in this paper are tabulated in Table 1 and Tables S1 to S21. UKB individual-level genotypic and phenotypic data, which accounts for the majority of the data included in this genetic analysis, is available to approved investigators via the UK Biobank study (<https://www.ukbiobank.ac.uk/>). Additional data for reproducing the results described in the article is available upon reasonable request and subject to appropriate data use agreements. eQTL summary statistics were downloaded from the GTEx portal (<https://gtexportal.org/>). T2D Knowledge Portal summary statistics for associations with type 2 diabetes were accessed via the website (<https://t2d.hugeamp.org/>).

Software and code used in this manuscript include:

Coloc R package (<https://github.com/chr1swallace/coloc>)

BOLT-LMM (<http://data.broadinstitute.org/alkesgroup/BOLT-LMM/downloads/>)

REGENIE (<https://github.com/rgcgithub/regenie>)

FINEMAP (<http://www.christianbenner.com/>)

MANTRA (available on request to Prof. Andrew Morris andrew.morris-5@manchester.ac.uk)

LDpred (<https://pypi.org/project/LDpred/>)

METAL (<http://csg.sph.umich.edu/abecasis/metal/download/>)

snpEff software (<https://pcingola.github.io/SnpEff/>)

FlowJo (<https://www.flowjo.com/solutions/flowjo/downloads>)

Illumina sample de-multiplexing software

(https://support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html)

Supplementary Materials:

Regeneron Genetics Center Banner Author List and Contribution Statements

Materials and Methods

Figures S1-S8

Tables S1-S21

References (*[supplement only]*)

Fig. 1. Exome-wide association analysis for body mass index identifies protein-truncating variants in *GPR75* associated with lower body mass index and obesity risk in humans. *A* Linear model of the *GPR75* protein and its domains (top; intra- and extra-cellular domains in yellow, transmembrane domains in orange), the distribution on the *GPR75* protein of 46 predicted loss of function variants found by exome sequencing in UKB, GHS and MCPS (middle) and the distribution of BMI in standardized units among heterozygous carriers of each variant (bottom). In the bottom sub-panel, horizontal blue bars show the mean BMI in non-carriers, while horizontal red bars show the overall covariates-adjusted mean BMI in carriers of any predicted loss-of-function genetic variants in *GPR75*. *B* Meta-analysis of the association with BMI of predicted loss-of-function variants in *GPR75* in discovery and additional cohorts. *C* Distribution of heterozygous carriers of predicted loss of function genetic variants in *GPR75* (top), non-carriers (middle) and heterozygous carriers of predicted loss of function genetic variants in *MC4R* (bottom) in body mass index categories according to the World Health Organization's classification in the UKB, GHS and MCPS cohorts. Abbreviations: RR, reference-reference genotype; RA, reference-alternative heterozygous genotype; AA, alternative-alternative homozygous genotype; CI, confidence interval; DHS, Dallas Heart Study; SINAI, Mount Sinai BioMe cohort; Duke Catheterization Genetics cohort; TAICHI, Taiwanese Chinese persons from the Taiwan MetaboChip Consortium; PMBB, University of Pennsylvania Medicine BioBank; UKB, UK Biobank; MALMO, Malmö Population-Based Cohort; AFR, African ancestry; AMR, American ancestry; EAS, East Asian ancestry; EUR, European ancestry; SAS, South Asian ancestry. MCPS included individuals of Admixed American ancestry.

Fig. 2. *In vitro* expression studies of two predicted loss-of-function genetic variants in *GPR75*.

A Results of quantitative reverse transcription polymerase chain reaction experiments which measured *GPR75* mRNA levels. Expression of *GPR75* was calculated relative to the beta-actin gene. Values represent the mean and standard deviation of 3 technical replicates representative of 1 of 3 biological replicate experiments performed for each condition. *B* Western blotting analysis of *GPR75* protein levels. *GPR75* Ala110fs and Gln234* protein products correspond to the predicted molecular weight of 14 and 25 kDa, respectively. The results are representative of 3 biological replicates. *C* Immunofluorescence staining experiments describing the cellular localization of *GPR75*. The top images show intracellular staining achieved by membrane permeabilization, while the bottom images show plasma membrane localization (non-permeabilized cellular membrane). *D* Flow cytometry analysis of the cell surface expression of *GPR75*. Identified cell populations are presented in percent (%) of live HA-TAG *GPR75* positive cells. Values represent the mean of 4 biological replicates per condition and their standard deviation. All experiments were performed in HEK293 cells that were transfected with green fluorescent protein control plasmids (Control), *GPR75*-wildtype (*GPR75*), *GPR75*-Ala110fs or *GPR75*-Gln234* plasmids. Abbreviations: SSC, side scatter; HA, hemagglutinin tag.

Fig. 3. Weight-gain during high-fat diet and metabolic phenotype in mice with a genetic deletion of *Gpr75*. *A* Weekly body weight gain during a 14-weeks high-fat diet challenge; *B* Changes in fasting blood glucose before and after the high-fat diet challenge; *C* Results of a glucose tolerance test at the end of the 14-week high-fat diet challenge; *D* Plasma insulin at the end of the 14-week high-fat diet challenge. Each panel shows results in *Gpr75*^{+/+} (WT), *Gpr75*^{+/-} (HET), and *Gpr75*^{-/-} (KO) mice. Number of mice included in each group and analysis are in parenthesis. Results are presented as mean \pm standard error. Abbreviations: ns, not statistically-significant; **p*<0.05, ***p*<0.01, ****p*<0.001, *****p*<0.0001 by two-way ANOVA with Tukey's multiple comparisons test.

Fig. 4. Association of loss-of-function genetic variants in *GIPR* with lower body mass index.

A Association with BMI for predicted loss of function genetic variants in *GIPR* across both discovery and additional cohorts. Abbreviations: RR, reference-reference genotype; RA, reference-alternative heterozygous genotype; AA, alternative-alternative homozygous genotype; CI, confidence interval; DHS, Dallas Heart Study; SINAI, Mount Sinai BioMe cohort; Duke Catheterization Genetics cohort; TAICHI, Taiwanese Chinese persons from the Taiwan MetaboChip Consortium; UPENN, University of

Pennsylvania Medicine BioBank; UKB, UK Biobank; MALMO, Malmö Population-Based Cohort; AFR, African ancestry; AMR; American ancestry; EAS, East Asian ancestry; EUR, European ancestry; SAS, South Asian ancestry. *B in vitro* expression results of the Arg190Gln and Glu288Gly missense variants in *GIPR*, which were associated with lower BMI. Both genetic variants showed lower GIP-induced Gs signaling by CRE-reporter assay (top left panel), compared to wild-type. In addition, both missense genetic variants showed reduced response to GIP-induced Gq signaling through a NFAT-dependent assay (top right panel), compared to wild-type. Both missense variants exhibit similar cell surface expression as wild-type in flow cytometry experiments (bottom panel).

Fig. 5. Interplay between a genome-wide polygenic score for higher body mass index with rare, large effect-size genetic mutations in *GPR75* and *MC4R*. This Figure shows the prevalence of obesity (defined as BMI equal or above 30 kg/m²) in heterozygous carriers of predicted loss of function variants in *GPR75*, non-carriers, or heterozygous carriers of predicted loss of function variants in *MC4R* within quintiles of a genome-wide polygenic score for higher BMI. P-values for interaction between the polygenic score and rare pLOF variants on body mass index were 0.09 and 0.85 for *GPR75* and *MC4R*, respectively. Error bars represent 95% confidence intervals.

Table 1. Association with body mass index of the burden of rare protein-coding variants in the exome-wide analysis.

Gene	Genomic coordinates	Genetic exposure, variant type; frequency cut-off in %	Beta (95% CI) per allele in SD units of BMI	p	AAF, fraction of 1	Genotype counts, RR RA AA genotypes	Beta (95% CI) per allele in kg/m ² units of BMI	Beta (95% CI) per allele in kgs of body weight	Beta (95% CI) per allele in lbs of body weight
<i>LRRC41</i>	1: 46278704	pLOF plus missense (5/5); AAF < 1%	0.20 (0.13, 0.28)	2.3×10 ⁻⁰⁷	0.0005	645,041 585 0	1.1 (0.7, 1.5)	3.2 (2.0, 4.4)	7.0 (4.3, 9.6)
<i>UHMK1</i>	1: 162497250	pLOF plus missense (5/5); AAF < 1%	-0.07 (-0.09, -0.04)	5.8×10 ⁻⁰⁸	0.0048	639,499 6,117 10	-0.4 (-0.5, -0.2)	-1.0 (-1.4, -0.6)	-2.2 (-3.0, -1.4)
<i>GPR75</i>	2: 53853133	pLOF; AAF < 1%	-0.35 (-0.48, -0.23)	1.4×10 ⁻⁰⁸	0.0002	645,398 228 0	-1.9 (-2.6, -1.3)	-5.5 (-7.4, -3.6)	-12.2 (-16.4, -8.0)
<i>TRANK1</i>	3: 36828274	pLOF plus any missense; AAF < 1%	0.02 (0.01, 0.03)	3.3×10 ⁻⁰⁷	0.0356	599,766 45,799 61	0.1 (0.1, 0.2)	0.4 (0.2, 0.5)	0.8 (0.5, 1.1)
<i>ROBO1</i>	3: 78598912	pLOF; AAF < 1%	0.25 (0.16, 0.33)	3.6×10 ⁻⁰⁸	0.0003	645,182 444 0	1.3 (0.9, 1.8)	3.8 (2.5, 5.2)	8.5 (5.5, 11.5)
<i>KIAA1109</i>	4: 122170647	pLOF; AAF < 1%	0.13 (0.08, 0.18)	1.9×10 ⁻⁰⁷	0.0011	644,238 1,388 0	0.7 (0.4, 1.0)	2.1 (1.3, 2.8)	4.5 (2.8, 6.2)
<i>PCSK1</i>	5: 96393000	pLOF; AAF < 1%	0.34 (0.22, 0.45)	8.9×10 ⁻⁰⁹	0.0002	645,368 258 0	1.8 (1.2, 2.4)	5.3 (3.5, 7.1)	11.6 (7.7, 15.6)
<i>GPR151</i>	5: 146514853	pLOF plus missense (5/5); AAF < 1%	-0.05 (-0.07, -0.03)	3.8×10 ⁻⁰⁹	0.0097	633,209 12,372 45	-0.3 (-0.4, -0.2)	-0.8 (-1.0, -0.5)	-1.7 (-2.3, -1.1)
<i>SPARC</i>	5: 151663570	pLOF plus missense (1/5); AAF < 1%	0.05 (0.03, 0.06)	1.0×10 ⁻⁰⁹	0.0113	631,094 14,494 38	0.3 (0.2, 0.3)	0.8 (0.5, 1.0)	1.7 (1.1, 2.2)
<i>UBR2</i>	6: 42564319	pLOF; AAF < 1%	0.37 (0.25, 0.48)	2.7×10 ⁻¹⁰	0.0002	645,363 263 0	2.0 (1.4, 2.6)	5.7 (4.0, 7.5)	12.7 (8.7, 16.6)
<i>CALCR</i>	7: 93426355	pLOF plus missense (1/5); AAF < 0.1%	0.08 (0.06, 0.11)	1.9×10 ⁻¹²	0.0048	639,416 6,208 2	0.5 (0.3, 0.6)	1.3 (1.0, 1.7)	2.9 (2.1, 3.7)
<i>ANO4</i>	12: 100901785	pLOF; AAF < 1%	0.25 (0.16, 0.35)	2.2×10 ⁻⁰⁷	0.0003	645,248 378 0	1.4 (0.8, 1.9)	3.9 (2.4, 5.4)	8.7 (5.4, 12.0)
<i>MC4R</i>	18: 60371350	pLOF plus missense (5/5); AAF < 1%	0.29 (0.25, 0.33)	9.7×10 ⁻⁴⁴	0.0016	643,589 2,036 1	1.6 (1.3, 1.8)	4.5 (3.9, 5.2)	10.0 (8.6, 11.4)
<i>DPP9</i>	19: 4676563	pLOF plus missense (5/5); AAF < 1%	-0.06 (-0.08, -0.04)	5.3×10 ⁻⁰⁸	0.0058	638,141 7,469 16	-0.3 (-0.4, -0.2)	-0.9 (-1.3, -0.6)	-2.0 (-2.8, -1.3)
<i>ANKRD27</i>	19: 32598144	pLOF plus missense (1/5); AAF < 1%	0.03 (0.02, 0.05)	5.7×10 ⁻⁰⁸	0.0171	623,572 22,016 38	0.2 (0.1, 0.3)	0.5 (0.3, 0.7)	1.2 (0.8, 1.6)
<i>GIPR</i>	19: 45669520	pLOF plus missense (1/5); AAF < 1%	-0.09 (-0.11, -0.07)	1.1×10 ⁻¹⁸	0.0069	636,680 8,937 9	-0.5 (-0.6, -0.4)	-1.4 (-1.7, -1.1)	-3.0 (-3.7, -2.4)

The Table reports genes for which the gene burden of rare nonsynonymous variants was associated with body mass index at the exome-wide level of statistical significance ($p < 3.6 \times 10^{-7}$). Analyses were performed in 645,626 participants from the UKB, GHS and MCPS studies. Genomic coordinates reflect chromosome and physical position in base pairs according to Genome Reference Consortium Human Build 38.

- 5 Abbreviations: CI, confidence interval; SD, standard deviation; BMI, body mass index; AAF, alternative allele frequency; RR, reference-reference genotype; RA, reference-alternative heterozygous genotype; AA, alternative-alternative homozygous genotype; pLOF, predicted loss of function; Missense (1/5), missense variant predicted to be deleterious by at least 1 out of 5 *in silico* prediction algorithms; Missense (5/5), missense variant predicted to be deleterious by 5 out of 5 *in silico* prediction algorithms.