

SCIENTIFIC REPORTS



OPEN

Rare Variant Analysis of Human and Rodent Obesity Genes in Individuals with Severe Childhood Obesity

Audrey E. Hendricks^{1,2}, Elena G. Bochukova^{3,4}, Gaëlle Marenne¹, Julia M. Keogh³, Neli Atanassova³, Rebecca Bounds³, Eleanor Wheeler¹, Vanisha Mistry³, Elana Henning³, Understanding Society Scientific Group*, Antje Körner^{5,6}, Dawn Muddyman¹, Shane McCarthy¹, Anke Hinney⁷, Johannes Hebebrand⁷, Robert A. Scott⁸, Claudia Langenberg⁸, Nick J. Wareham⁸, Praveen Surendran⁹, Joanna M. Howson⁹, Adam S. Butterworth^{9,10}, John Danesh^{1,9,10}, EPIC-CVD Consortium*, Børge G Nordestgaard^{11,12}, Sune F Nielsen^{11,12}, Shoaib Afzal^{11,12}, Sofia Papadia³, Sofie Ashford³, Sumedha Garg³, Glenn L. Millhauser¹³, Rafael I. Palomino¹³, Alexandra Kwasniewska³, Ioanna Tachmazidou¹, Stephen O'Rahilly³, Eleftheria Zeggini¹, UK10K Consortium*, Inês Barroso^{1,3} & I. Sadaf Farooqi³

Obesity is a genetically heterogeneous disorder. Using targeted and whole-exome sequencing, we studied 32 human and 87 rodent obesity genes in 2,548 severely obese children and 1,117 controls. We identified 52 variants contributing to obesity in 2% of cases including multiple novel variants in *GNAS*, which were sometimes found with accelerated growth rather than short stature as described previously. Nominally significant associations were found for rare functional variants in *BBS1*, *BBS9*, *GNAS*, *MKKS*, *CLOCK* and *ANGPTL6*. The p.S284X variant in *ANGPTL6* drives the association signal (rs201622589, MAF=0.1%, odds ratio = 10.13, p-value = 0.042) and results in complete loss of secretion in cells. Further analysis including additional case-control studies and population controls (N = 260,642) did not support association of this variant with obesity (odds ratio = 2.34, p-value = 2.59×10^{-3}), highlighting the challenges of testing rare variant associations and the need for very large sample sizes. Further validation in cohorts with severe obesity and engineering the variants in model organisms will be needed to explore whether human variants in *ANGPTL6* and other genes that lead to obesity when deleted in mice, do contribute to obesity. Such studies may yield druggable targets for weight loss therapies.

Studies focused on severe early onset obesity alone, or obesity with developmental delay and/or dysmorphic features have identified a number of genes harbouring highly penetrant causal mutations^{1,2}. The further characterisation of rare, highly penetrant variants identified in such individuals can provide insights into the cellular and

¹Wellcome Trust Sanger Institute, Cambridge, UK. ²Department of Mathematical and Statistical Sciences, University of Colorado-Denver, Denver, CO, 80204, USA. ³University of Cambridge Metabolic Research Laboratories and NIHR Cambridge Biomedical Research Centre, Wellcome Trust-MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. ⁴The Blizzard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. ⁵Center for Pediatric Research, University Children's Hospital Leipzig, Leipzig, Germany. ⁶IFB Adiposity Diseases Medical Faculty, University of Leipzig, Leipzig, Germany. ⁷Department of Child and Adolescent Psychiatry, Psychotherapy, and Psychosomatics, University Hospital Essen and University of Duisburg-Essen, Essen, Germany. ⁸MRC Epidemiology Unit, Institute of Metabolic Science, University of Cambridge School of Clinical Medicine, Cambridge, UK. ⁹Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ¹⁰The National Institute for Health Research Blood and Transplant Unit (NIHR BTRU) in Donor Health and Genomics, University of Cambridge, Cambridge, UK. ¹¹Department of Clinical Biochemistry and The Copenhagen General Population Study, Herlev and Gentofte Hospital, Copenhagen University Hospital, Copenhagen, Denmark. ¹²Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. ¹³Department of Chemistry & Biochemistry, University of California Santa Cruz, Santa Cruz, CA, 95064, USA. Audrey E. Hendricks and Elena G. Bochukova contributed equally to this work. *A comprehensive list of consortium members appears at the end of the paper. Correspondence and requests for materials should be addressed to I.B. (email: ib1@sanger.ac.uk) or I.S.F. (email: isf20@cam.ac.uk)

physiological mechanisms involved in energy homeostasis and human obesity, and can identify and/or validate targets for therapeutic intervention. The aim of our study was to examine the prevalence of rare/novel variants in human and mouse obesity genes using high-throughput next-generation sequencing in a large cohort of individuals with severe early onset obesity. Here we describe results from an analysis of 119 candidate genes sequenced in 2,548 individuals with severe, early-onset obesity from the Severe Childhood Onset Obesity Project (SCOOP)³ (UK individuals of European ancestry recruited to the Genetics of Obesity Study, GOOS; BMI standard deviation score (SDS) > 3; onset of obesity before the age of 10 years; Methods) from the UK10K project⁴. Our analysis includes 737 SCOOP individuals with whole-exome sequence, and 1,811 additional SCOOP individuals, not consented for whole-exome analysis, in whom we performed targeted sequencing. As this work was performed as part of a consortium (UK10K project), this data was compared to 1,117 individuals with other disorders (e.g. neurodevelopmental and rare disease) in whom exome sequencing and analysis were performed using the same methods. For the purposes of this analysis, these individuals are designated as “controls” (Methods), although we recognise that there are limitations with this design.

Study Design

SCOOP individuals likely to have congenital leptin deficiency, a treatable cause of severe obesity, were excluded by measurement of serum leptin, and individuals with mutations in the melanocortin 4 receptor gene (*MC4R*) (the most common genetic form of penetrant obesity) were excluded by prior Sanger sequencing.

We focused on six tiers of genes (Methods and Supplementary Table 1): (1) genes known to harbour variants causing human obesity alone (*Obesity Alone*; $n = 6$); (2) genes in which known variants cause human obesity combined with developmental delay and/or dysmorphology (*Obesity and Delay*; $n = 26$); (3) genes in which loss-of-function leads to obesity in mice (*LoF Mice*; $n = 51$); (4) genes in which gain-of-function leads to obesity in mice (*GoF Mice*, $n = 5$); (5) genes encoding anorectic peptides and their receptors (*Anorectic Molecules*, $n = 7$), and (6) genes in which loss-of-function is associated with other metabolic phenotypes in mice (*Complex Metabolic Effects*, $n = 24$). Our aim was to include genes where there was sufficient evidence to indicate disorders with Mendelian inheritance (obesity syndromes) or where complete deletion or overexpression (as opposed to conditional knockouts) causes an obesity phenotype in mice (Methods) (i.e. genes in which inherited loss/gain of function variants might exist). To identify variants more likely to be causally linked to obesity, we focused on *rare* (MAF < 1%) and *novel* (not seen in the data we used for filtering) variants predicted to be functional (i.e. non-sense variants, missense amino acid substitutions, alterations of conserved splice sites or small insertions/deletions (indels) that introduced a frameshift) (Methods). We filtered our data against approximately 8,000 publicly available sequenced samples, as well as 2,097 exomes (sequenced in parallel on the same platform as our samples) and 3,781 whole-genome sequenced samples also from the UK10K project (Methods).

Human Obesity Syndrome Genes. Firstly, we sought to identify rare and novel functional variants in the 32 genes known to cause human obesity, with or without, additional developmental delay and/or dysmorphology features (*Obesity Alone*, or *Obesity and Delay*) (Methods). We identified 11 rare potentially functional variants in ClinVar with pathogenic/likely pathogenic status and 321 novel functional variants, which we confirmed by Sanger sequencing (Fig. 1, Methods). Based on inheritance patterns (where available) and the functional properties of variants that have previously been characterised, variants in these genes may contribute to obesity, sometimes in a non-fully penetrant manner, in 52 (2%) individuals (Supplementary Figure 1 & Supplementary Table 2; Fig. 2, Methods). As variants in *MC4R* account for approximately 5% of severe obesity in this cohort⁵, these findings indicate that > 90% of patients within this cohort do not have their phenotype explained by variants in known human obesity genes. Further analysis exploring the whole-exome in an agnostic manner, with appropriately matched non-disease controls not available in the UK10K project, will be an aim of future investigations.

Notably, we identified thirteen clinically associated *GNAS* variants in fourteen SCOOP individuals (12 of these confirmed on Sanger sequencing; 0.5%) (Supplementary Table 3). *GNAS* is an imprinted gene in which heterozygous loss-of-function variants are associated with obesity, short stature and skeletal abnormalities, and, when maternally inherited, hormone resistance syndromes⁶. As *GNAS* sequencing has traditionally only been performed in individuals with classical clinical features, our findings suggest that the true prevalence in childhood obesity may be underappreciated. Three of the variants found in our study have been described previously in patients with classical features (p.Y163X⁷; p.R258W⁸, and p.R265H⁹). We identified a novel nonsense variant (p.Y169X) predicted to remove the entire Ras-like GTPase domain and several missense variants predicted to affect downstream signalling by affecting the interaction with G-protein coupled receptors (GPCRs), G-protein β - and γ -subunits, or downstream adenylyl cyclase when mapped onto the protein structure of *GNAS* (Fig. 3). We confirmed maternal transmission of variants in three out of the five families where parental samples were available for genotyping. Although four *GNAS* variant carriers exhibited endocrinopathies and nine had developmental delay as anticipated, unexpectedly four individuals had accelerated linear growth in childhood (height SDS > 2) rather than short stature (defined as height sds < 2.0) (Supplementary Table 3). Further molecular and physiological studies will be needed to investigate potential genotype-phenotype correlations. As studies in rodents have shown that *Gnas* is imprinted in the paraventricular nucleus of the hypothalamus¹⁰, the location of the majority of neurons expressing the G-protein coupled receptor *MC4R*, it is plausible to hypothesize that some *GNAS* variants may contribute to obesity and accelerated linear growth by reducing melanocortin signalling.

Next we formally tested for enrichment of rare, or novel, functional variants in case-control analysis¹¹ (Methods). Although no gene attained experiment-wide threshold of p -value < 2×10^{-4} (Methods and Supplementary Tables 4–9), four genes (*BBS1*, *BBS9*, *GNAS*, and *MKKS*) known to cause obesity and developmental delay and/or dysmorphology (*Obesity and Delay*) had nominally significant burden test p -values (p -value < 0.05) and a higher burden of variants in cases than controls (Table 1).

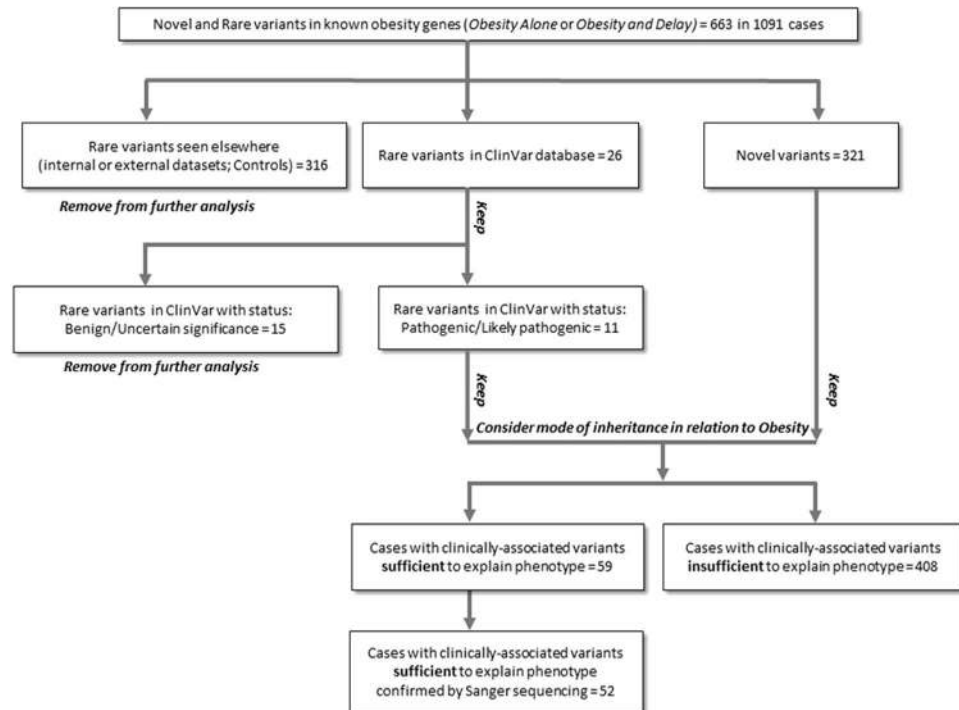


Figure 1. Identification of clinically-associated variants in known human obesity genes. Schematic outlining the analysis strategy.

Mouse Obesity Genes. We found a higher burden of rare variants in cases than controls for two genes known to cause obesity when disrupted in mice - *ANGPTL6* and *CLOCK* (Table 1). Testing the functional consequences of *CLOCK* variants in cells is not straightforward as the molecular mechanisms that lead to obesity are not fully understood. As such, we focused on verifying the accuracy of the computationally predicted function of all twelve rare *ANGPTL6* variants (Methods, Fig. 4, Supplementary Table 10). All variants found in cases (but not the one variant found exclusively in controls, V143L) were predicted to affect the fibrinogen-like domain of the protein (Fig. 4a). *ANGPTL6* is predominantly expressed in the liver but is also expressed in white adipose tissue and kidney. To evaluate the effect of these variants, HEK293 cells were transiently transfected with constructs encoding wild type and variant forms of *ANGPTL6*. Whereas wild-type *ANGPTL6* was readily detected in the medium, several mutants reduced, and the S286X mutant abolished, secretion of *ANGPTL6* (Fig. 4b). Incorporation of this functional information did not qualitatively change the results of the *ANGPTL6* gene-based test (Supplementary Table 10). However, limiting the analysis to the single variant that completely abolished protein secretion (p.S286X, rs201622589) led to an adjusted OR = 10.13 and p-value = 0.028 (Fisher's p-value = 0.041) suggesting that the association signal at this gene is primarily driven by this variant (Supplementary Table 10). To increase power, and given that there were no additional cases of European descent from our obesity cohort, we analysed existing exome-chip data from an additional 253,587 unrelated European, non-Finnish population controls (Methods, Supplementary Table 11). In an updated analysis with the 2,548 cases and 253,587 controls, we observed an odds ratio (OR) = 2.90 and a Fisher's p-value = 0.0022, which did not reach experiment or exome-wide significance. Further analysis of 1,436 non-overlapping obesity cases and 1,954 non-overlapping controls from two studies did not lend additional support, and when combined with the original data, yielded an overall OR = 2.34 and chi-squared p-value = 0.0060 (total 3,984 cases vs 256,658 controls, Supplementary Table 11). The results are similar when limiting to non-obese controls (Supplementary Table 12). These findings highlight the difficulty in studying very rare variants in complex diseases, and potentially the effect of winner's curse, reinforcing the need for extremely large sample sizes¹². Of interest, mice with targeted deletion of *Angptl6* that survive to birth (20%) develop marked obesity, have increased food intake, reduced energy expenditure, exhibit lipid accumulation in liver and muscle, and develop insulin resistance¹³. Conversely, mice with targeted overexpression of *Angptl6* are lean, insulin sensitive and are protected from diet-induced obesity¹³. As *ANGPTL6* is a liver-derived circulating peptide, and thus could potentially be manipulated for therapeutic purposes¹⁴, further genetic studies in larger cohorts and experimental studies in mice and humans are necessary to explore its role in obesity and potential utility as an anti-obesity drug target.

Power and Gene Set Analysis. Assuming an experiment-wide significance level of 2×10^{-4} and that 30% of rare (MAF < 1%) variants within a given gene are causal we have very limited power (~20%) to detect an association to a gene region using our sample of 2,548 cases and 1,117 controls (Supplementary Figure 2, Online Methods, Supplementary Note 4)¹². As there is evidence that looking across group of genes with stronger priors may increase power^{15–17}, we tested for association of rare or novel functional variants within each of the six

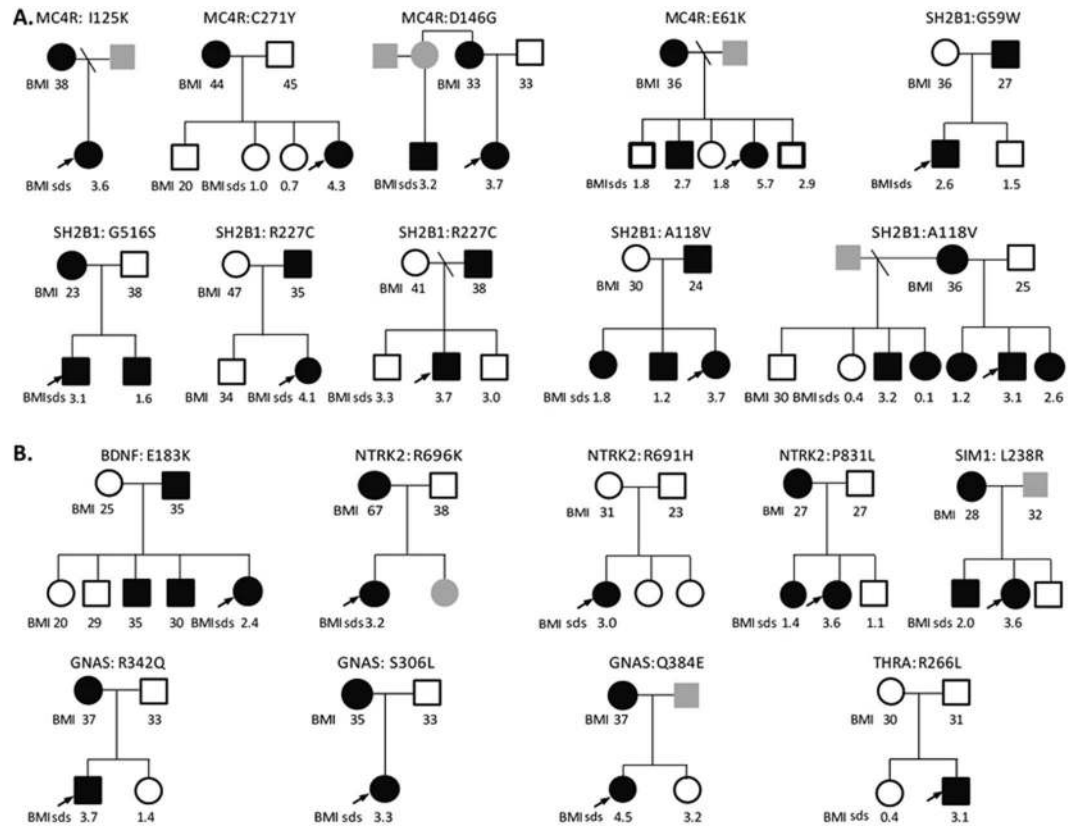


Figure 2. Pedigrees of patients with clinically associated variants contributing to obesity. Co-segregation of variants is shown where male (squares) and female (circles) family members consented to genotyping. Heterozygous (filled) and wild-type (empty) mutation carriers are indicated; in some cases, individuals were not available for genotyping (grey). Body mass index (BMI) ($>27\text{ kg/m}^2$ = overweight; $>30\text{ kg/m}^2$ = obesity) for adults and BMI standard deviation scores (BMI sds) for children are shown where data was available. (A) Obesity alone; (B) Obesity and Delay. *MC4R* mutations were excluded by prior Sanger sequencing of all individuals undergoing whole exome sequencing and the majority (1744/1811) of individuals undergoing targeted resequencing. The finding of four *MC4R* mutations in the remaining 82 individuals is in keeping with the prevalence of heterozygous *MC4R* mutations in this cohort as reported previously⁵.

candidate gene sets. We further filtered variants with respect to their likelihood of being deleterious and performed six nested tests for each gene set (Methods). Although none of the analyses of tiered gene sets reached experiment-wide significance (Supplementary Table 13, Fig. 5), we see an increase in OR as we focus on rarer variants with more evidence of predicted deleterious effect within the *Obesity Alone* gene set. This suggests that focusing on extremely rare variants with strong evidence of being deleterious may be a good strategy for diseases with a complex genetic architecture, which likely includes rare, highly penetrant alleles. Our power analyses also suggests that restricting the MAF threshold to be near the very rare allele frequency of that expected for the causal variants, rather than a more lenient MAF threshold of 1%, increases power (Supplementary Figure 3). Finally, there are nominally significant signals ($p\text{-value} < 0.05$) within the *LoF Mice* candidate gene set for the novel, functional variant tests suggesting that within this rather large gene set ($n = 51$) there may be additional genes associated with human early onset obesity (Fig. 5).

Discussion

In summary, analysis of sequence data in 119 genes across 2,548 severely obese children and 1,117 other disease controls from the UK10K project identified variants in known obesity genes, contributing to obesity in 2% of children within this cohort, in whom mutations in *MC4R* and leptin had been excluded. It also highlighted that *GNAS* mutations may be more prevalent in human obesity than previously thought, and that the spectrum of phenotypic consequences may be broader and more heterogeneous than previously described (6). Additional detailed genotype-phenotype studies will be required to further elucidate the molecular underpinnings of the physiological consequences of these variants.

Taking the information generated in this study together with the published literature, we suggest that molecular genetic investigations, including testing known obesity genes, or agnostic exome-wide or genome-wide approaches (as these become more established at the point of care), should become part of the assessment for a child presenting with severe obesity (BMI SDS > 3) in the absence of other syndromic features. Genetic studies can allow the identification of congenital leptin deficiency which is entirely treatable with recombinant leptin therapy¹⁸. Also, early reports suggest that patients with mutations that disrupt POMC signalling may be effectively

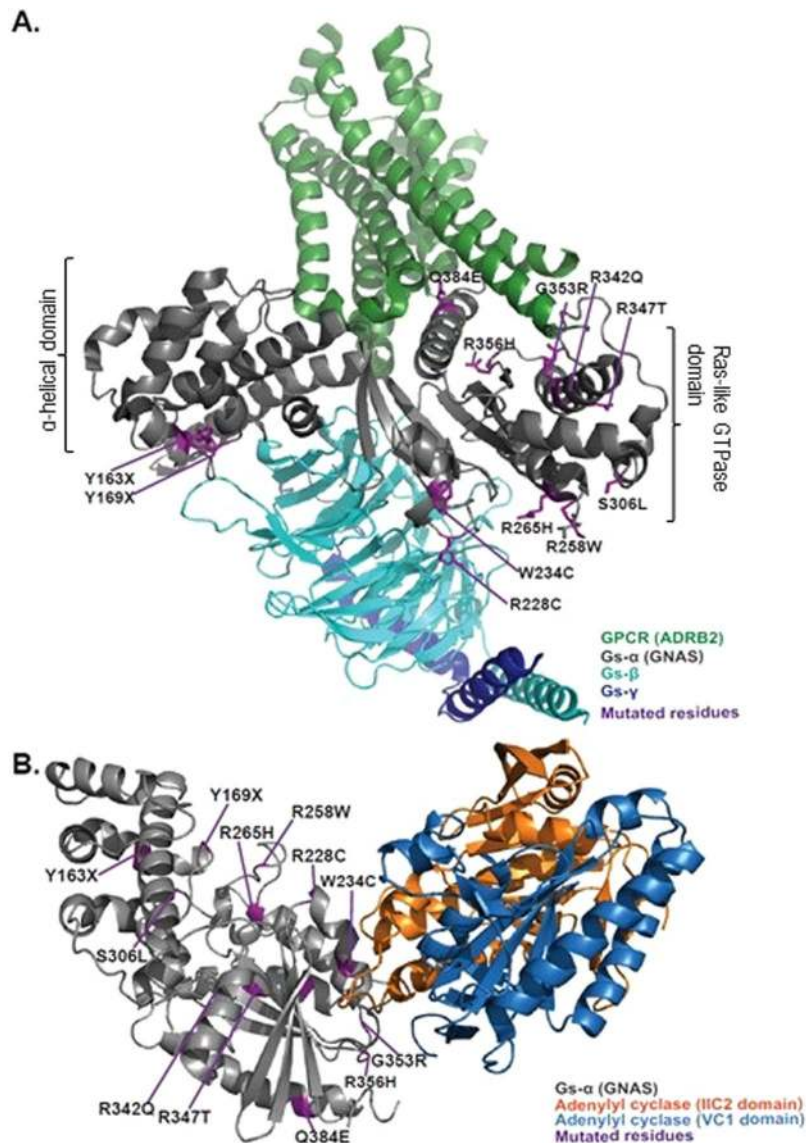


Figure 3. Structural model of variants identified in GNAS. (A) Structure of the active state ternary protein complex of G-protein coupled receptor (GPCR) beta 2-adrenergic receptor (ADRB2), and the nucleotide-free Gs heterotrimer, composed of Gs- α (GNAS), Gs- β , and Gs- γ subunits (based on Rasmussen *et al.*²⁰, pdb file: 3sn6). The two major domains of GNAS are noted, α -helical domain and Ras-like GTPase domain. The principal interactions between GNAS and GPCRs involve the amino- and carboxy-terminal α -helices. The Ras-like GTPase domain contains most of the catalytic residues necessary for GTP hydrolysis, as well as the G $\beta\gamma$ and effector binding regions (switch regions I, II and III), which change confirmation upon binding to GTP or GDP). (B) Structure of the complex of Gs-alpha with the catalytic domains of mammalian adenylyl cyclase (based on Tesmer *et al.*²¹, pdb file: 1azs). Variant residues (purple), and the different components of the complex are highlighted.

treated with Setmelanotide (a MC4R agonist)¹⁹. Finally, our data suggest that focusing on very rare, deleterious variants may increase power to find genetic association. Further detailed investigation of genes that lead to obesity when deleted in mice may identify new pathophysiological mechanisms involved in human obesity that can be targeted for drug discovery.

Methods

Sample Sets. All studies were approved by the Cambridge Local Research Ethics Committee and all participants and their parents gave written informed consent. All methods were performed in accordance with the relevant laboratory/clinical guidelines and regulations.

The SCOOP cohort³ was studied as part of the UK10K consortium⁴. Data was compared to subsets within the neurodevelopmental and rare disease groups that were consented for use as controls. Details and further

Gene	Tier	# Variants	Rare alleles in SCOOP TS (N = 1,811)	Rare alleles in SCOOP WES (N = 737)	Rare alleles in Control WES (N = 1,117)	OR	Adjusted OR	p-value SKATO	p-value BURDEN
Rare Functional									
<i>BBS1</i>	Obesity and Delay	18	36	18	10	2.368		0.022	0.018
<i>BBS9</i>	Obesity and Delay	20	24	5	5	2.543		0.081	0.045
<i>GNAS</i>	Obesity and Delay	14	12	4	0	Inf	14.505	0.015	0.008
<i>MKKS</i>	Obesity and Delay	19	90	25	28	1.801		0.006	0.007
<i>ANGPTL6</i>	Mouse LOF	12	18	10	4	3.073		0.043	0.027
Novel Functional									
<i>GNAS</i>	Obesity and Delay	13	11	4	0	Inf	13.629	0.019	0.010
<i>CLOCK</i>	Mouse LOF	9	8	2	0	Inf	9.208	0.066	0.036

Table 1. Summary of case-control results. Summary of case-control results in genes with nominally significant (p-value < 0.05, bold) burden of *Rare* (top) or *Novel* (bottom) functional variants in SCOOP cases (N = 1,811 with targeted sequence and N = 737 with whole-exome sequence), compared to 1,117 controls with whole-exome sequence data from the UK10K project. The number of variants (# variants) per gene are shown, as well as the number of alleles in cases and controls, odds ratios (OR) and p-values from SKAT-O (p-value SKATO) and from burden tests (p-value BURDEN). For variants not detected in controls an adjusted odds ratio (Adjusted OR) was calculated by adding 0.5 to the number of alleles in each cell of the two-by-two table.

information about the UK10K project can be found at <http://www.uk10k.org/> and in the UK10K consortium paper from 2015⁴.

We gathered the genotypes for the variant S284X (rs201622589) in 202,981 unrelated population controls with European non-Finnish ancestry from existing cohorts (UKHLS²², Fenland, EPIC Norfolk²³, CCHS^{24–26}, CGPS^{24–26}, CIHDS^{24–26}, EPIC-CVD²⁷, UK Biobank²⁸ and ExAC²⁹) and in two additional studies of obese children and healthy or thin controls^{30,31} including one from the Leipzig Childhood Obesity Cohort³². (Supplementary Note 1)

Sequencing, Variant Calling, and Quality Control. Targeted Sequencing (TS) and Whole Exome Sequencing (WES) was performed as described elsewhere (refs 33 and 4 respectively). All UK10K WES samples from all arms of the exome study available at the time of the variant calling and SCOOP TS samples (N = 5,233 and 2,819 respectively) were called together on the non-redundant targets +/− 100 bp using multi-sample calling. Using SAMtools³⁴, a BCF file was created and the site genotype likelihoods were calculated. Variants (SNPs and Indels) were called using BCFtools. Indels were left-aligned using ‘vcf norm’ from the htlib package. Variants were filtered for sequencing and genotype quality at the site and genotype level using *vcf-annotate*³⁵. Additional details in Supplementary Note 2.

Variant Annotation. All variant annotation was applied using the GRCh37 human reference. Variants were annotated with rsIDs from dbSNP 137, and allele frequencies from the final 1000Genomes Phase 1 integrated (v3) callset³⁶, the NHLBI Exome Sequencing Project (ESP) v2³⁷, and the UK10K WGS sample set⁴. The Ensembl Variant Effect Predictor (<http://www.ensembl.org/info/docs/variation/vep/index.html>)³⁸ v2.8 with Ensembl 66 was used to add variant consequence annotations including the predicted deleteriousness of each missense variant as predicted by SIFT^{39,40}, PolyPhen^{41,42}, and Condel⁴³. For the 32 known human obesity genes, only the variant consequences on the most clinically relevant transcript were considered for further analysis. For the other 87 candidate obesity genes, the most severe consequence on any transcript was retained for each variant (Supplementary Table 1). Using Sequence Ontology terms^{44,45}, variant consequences were defined as functional (i.e. essential splice site, stop gained, stop lost, complex indel, frameshift coding, non synonymous coding, within mature miRNA, partial codon).

To identify *rare* variants, we used the thirteen UK10K WES sample sets that were not obese sample sets and were not included as controls in this study (Supplementary Note 1). These sample sets were processed in parallel with our cases and controls and thus provide a similar sequencing coverage and depth across the called regions. We removed variants that had a MAF > 1% across all or a MAF > 10% in any of the thirteen sample sets. We then removed variants with a MAF > 1% in any of seven additional sample sets: 1. UK10K WGS sample set⁴, 2–3. European and African American NHLBI ESP v2 sample sets³⁷, and 4–7, the four continent sample sets (AFR, AMR, ASN, EUR) from 1000 Genomes Phase 1 integrated v3 call set³⁶. *Novel* variants were identified as sites not seen in any of the internal or external datasets used for MAF annotation, i.e. novel at the time of this study.

Sample quality control. Samples were identified as contaminated using a combination of two methods: VerifyBamID v1.0⁴⁶ and “fraction skewed hets”⁴. For the TS samples, the FREEMIX value was estimated using 11,250 high quality, autosomal, biallelic SNPs with an alternate AF ≥ 0.01, and a call rate ≥ 0.5 in both the TS sample set and the 1000Genomes Project Phase 1 v3³⁶. Of the 2,819 TS SCOOP samples, the 784 WES SCOOP samples, and the 1427 WES samples used as controls 37, one, and nine were excluded due to contamination

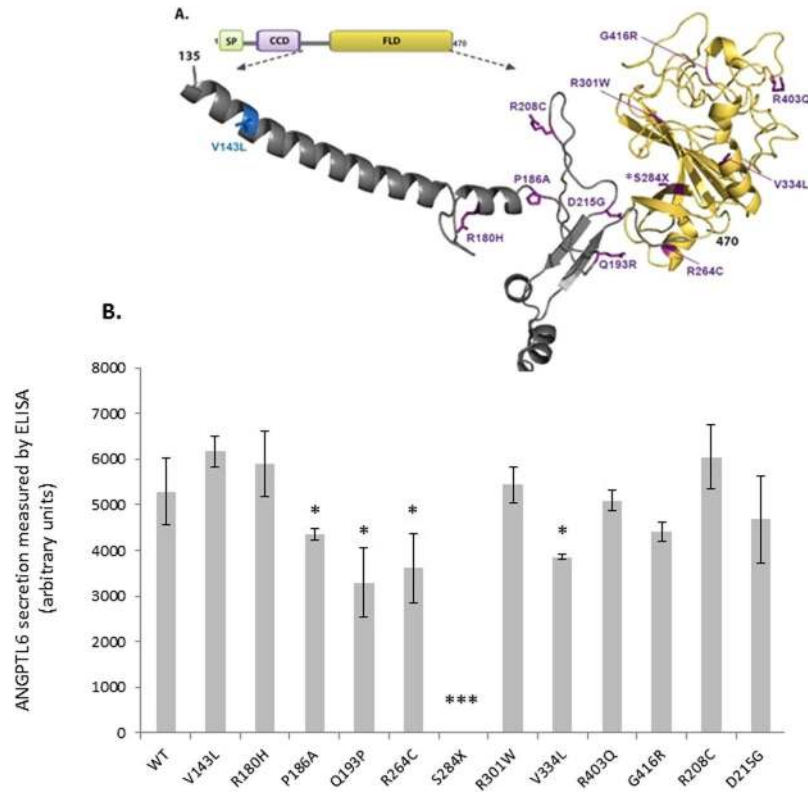


Figure 4. Functional characterisation of ANGPTL6 variants location and functional effect (A) Human ANGPTL6 protein and location of the genetic variants. Schematic of the human ANGPTL6 protein with the three recognisable domains: signal peptide (SP) in green, coiled-coil domain (CCD) in purple, and fibrinogen-like domain (FLD) in yellow. As all genetic variants identified are located predominantly in the FLD, the partial tertiary structure of the C-terminal protein portion (amino acids 135–470), based on previously solved fibrinogen structure (1lwu) is presented. The single variant seen in controls only is shown in blue, variants identified in cases are in purple, the S284X variant is starred. (B) Functional characterisation of ANGPTL6 variants. Cells were transiently transfected with constructs encoding wild-type (WT) or mutant ANGPTL6; levels of protein secretion into the media were measured by ELISA; means \pm standard deviation (SD) for experiments performed in triplicate are shown (results were confirmed by Western blotting; data not shown). Statistical significance was measured using unpaired T-test with Welch's correction using the GraphPad Prism software. * $p < 0.05$; *** $p < 0.001$.

respectively. Sixteen non-contaminated TS samples were excluded based on a mean sample read-depth 3 SD below the average for all samples (i.e. mean sample read-depth < 12.09). Genotype concordance for 436 WES SCOOP samples and 1,035 TS SCOOP samples for which we had both sequence and GWAS data³ was also calculated. Four WES samples and six TS samples with a concordance rate below 90% were identified and excluded. A set of highly polymorphic markers (MAF > 0.3) was genotyped and compared to the sequencing calls from each sample. Non-concordant WES samples were removed prior to variant calling and are not included in the original sequencing numbers. Eighty-one TS samples with low concordance were removed after variant calling. Three TS samples were removed due to having an extremely high genotype missing rate of $> 50\%$.

To identify non-European samples, we calculated principal components (PCs) from the 1000Genomes Phase I integrated call set³⁶ using either EIGENSTRAT v4.2⁴⁷ or LASER 2.0⁴⁸ for the WES and TS samples respectively (Supplementary Note 3). Of the 2,676 TS samples, 837 were classified as non-European due to genetic ancestry and three were excluded due to a reported non-European ancestry resulting in 1,836 good quality TS samples of European ancestry. Of the 779 WES SCOOP samples, 37 samples were classified as non-European due to genetic ancestry resulting in 742 good quality WES SCOOP samples of European ancestry. Of the 1,418 control samples, 146 were identified as being non-European leaving 1272 high quality WES controls.

Known and cryptic relatedness was identified by estimating pairwise identity by descent using PLINK v1.07⁴⁹. Genetic relationships within the WES sample set is described in detail elsewhere⁴. We removed individuals sequentially by: (1) largest number of relationships, (2) diseased controls, non-diseased controls, cases, (3) lower mean depth of sequencing. Five SCOOP cases and 155 controls were removed resulting in 737 SCOOP WES cases and 1,117 WES controls all of high quality, unrelated, and European ancestry. A similar process was used to identify genetic relationships within the TS sample set and between the TS and WES sample sets (Supplementary Note 3). This resulted in 25 additional exclusions (18 TS cases related to WES cases and seven related within the TS sample set) for a total of 1,811 high quality, unrelated, TS samples of European ancestry.

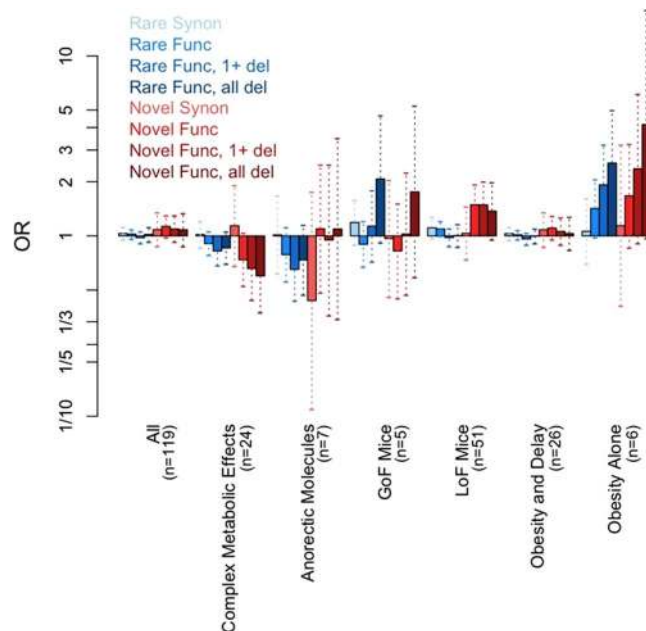


Figure 5. Tiered analysis of obesity candidate genes sets. ORs (bars) and 95% (dotted vertical lines) confidence intervals for each of the six gene tiers and across all 119 genes combined each for eight different filtering scenarios by MAF (*Rare* or *Novel*) and functional prediction (synonymous - *Synon*, functional - *Func*, functional with at least 1 deleterious consequence prediction - *Func, 1 + del*, and functional with all deleterious consequence predictions - *Func, all del*).

Candidate Gene Sets. We performed database searches using keywords ‘obesity’, ‘growth’, ‘size’, ‘adipose tissue’, (details below) and manually curated the results to arrive at a set of six tiers of candidate gene with definite or likely links to obesity. The Online Mendelian Inheritance of Man (OMIM database, <http://www.omim.org/>) accessed in February, 2013⁵⁰ was used to identify 32 genes (Supplementary Table 1) which directly lead to human obesity. These were further categorised into two groups based on consistent clinical features: genes characterised by “Obesity alone” (*Obesity Alone*; $n = 6$), and “Obesity and developmental delay and/or dysmorphology” (*Obesity and Delay*; $n = 26$). Any putative list of candidate genes has limitations and other genes that contribute to energy homeostasis/obesity in animal models deserve interrogation in the future.

Genes that cause obesity when disrupted in rodents were identified using the Mouse Genome Informatics Database (MGI, <http://www.informatics.jax.org/>)⁵¹ and the Rat Genome Database (RGD, <http://rgd.mcw.edu>)⁵² both accessed in February 2013, as well as published research (www.ncbi.nlm.gov/Pubmed). Eighty-seven genes were associated with a growth/size phenotype when perturbed in rodent animal models. We evaluated the strength of evidence for a role of each gene in obesity phenotype and divided them into four further categories, specifically: Loss of function associated with obesity in mice (*LoF Mice*; $n = 51$); gain of function associated with obesity in mice (*GoF Mice*, $n = 5$); anorectic peptides and receptors (*Anorectic Molecules*, $n = 7$); loss of function associated with other metabolic phenotypes in mice (*Complex Metabolic Effects*, $n = 24$) (Supplementary Table 1).

Sequence validation of rare functional variants. Novel, functional variants in all 119 candidate genes were taken forward for validation by conventional Sanger sequencing (90% validation rate). Briefly, customized PCR primers were designed ± 250 bp surrounding the variant; and sequencing was performed using BigDye Terminator v3 kit (Applied Biosystems) and analysed by capillary electrophoresis on an ABI3730 DNA Analyzer platform (Applied Biosystems), according to the manufacturers’ instructions. Familial segregation analysis of variants was performed where family samples were available.

Identification of clinically-associated variants in known human obesity genes. The strategy for identification of clinically-associated variants in known human obesity genes is outlined in Fig. 1. Briefly, novel and rare variants in the known human obesity genes were filtered using the ClinVar database (<http://www.ncbi.nlm.nih.gov/clinvar/>). We retained novel variants and those with a ClinVar status of Pathogenic/Likely pathogenic ($N = 332$).

Structural Analysis of GNAS Variants. For structural analysis of the genetic variants and prediction of their impact on protein function, we modelled their location in relation to previously solved protein crystal structures of GNAS in complex with either G-protein coupled receptor (ADRB2) or downstream effector (adenylyl cyclase). Specifically, the first model represented in Fig. 3a is based on crystal structure of the active state ternary protein complex of GPCR beta-2-adrenergic receptor (ADRB2), and the nucleotide-free Gs heterotrimer, composed of Gs- α (GNAS), Gs- β , and Gs- γ subunits (ref. 20, PDB number: 3SN6, www.rcsb.org). The second model, represented in Fig. 3b, is based on complex of GNAS with the catalytic domains of mammalian adenylyl cyclase (ref. 21, PDB

number: 1AZS, www.rcsb.org). The structural representations using ribbon-depicted models were generated using the Open-Source PyMOL Molecular Graphics System, Version 1.7.x Schrödinger, LLC (<http://pymol.org>).

Association Analyses. We implemented the optimal Sequence Kernel Association Test SKAT-O^{11,53} with the SKAT R package v1.1.2⁵⁴ using options *method* = "optimal.adj" for SKAT-O and *r.corr* = 1 for burden. For each gene and candidate gene set, we performed two primary tests using the burden test: 1. rare functional (*Rare Func*) or (2) novel functional (*Novel Func*). We performed secondary tests restricting either to variants that were predicted to be deleterious by at least one of the three algorithms (i.e. SIFT, PolyPhen, and Condel) (*1 + del*) or to variants that were predicted to be deleterious by all three algorithms (*all del*). For SIFT and Condel, variants were classified as deleterious if they were labelled as "deleterious". For PolyPhen, variants were classified as deleterious if they were labelled as "probably damaging" or "damaging". Our conservative Bonferroni adjusted significance level of the primary analyses for 119 genes and 6 gene sets was $0.05/(2 \times 125) = 2.0E-4$. Once considering our additional ten secondary tests, our Bonferroni adjusted significance level was $0.05/(12 \times 125) = 3.3E-5$. We repeated all primary and secondary tests using SKAT-O (Table 1; Supplementary Table 4–9, 13).

Functional studies of variants in ANGPTL6. To predict the impact of variants on the ANGPTL6 protein, we modelled their location using the crystal structure of a related protein fibrinogen in a complex with a peptide Gly-His-Pro-amide (described in Yang *et al.*⁵⁵, PDB number: 1LWU, www.rcsb.org). The structural representations using ribbon-depicted models were generated using the Open-Source PyMOL Molecular Graphics System, Version 1.7.x Schrödinger, LLC (<http://pymol.org>). N-terminal triple Flag tag was added to human ANGPTL6 cDNA cloned into a pEZ-M14 mammalian expression vector (Capital Bioscience). Mutations were introduced into this construct using QuikChange (Agilent technologies), and confirmed by Sanger sequencing. ANGPTL6 protein expression was studied by transient transfection in HEK293 cells. Media was collected 48 h – post transfection, the cell medium was centrifuged for 5 min (5,000 g at 4 °C) and the supernatants were collected. Cells were harvested prior to centrifugation for 15 min (15,000 g at 4 °C). Aliquots from the medium and cells were subjected to SDS-PAGE and immunoblot analysis. Primary antibodies (monoclonal Flag M2 antibody (Sigma) and a polyclonal antibody to Calnexin (Cell signalling)) were used at 1:1000, and secondary Horseradish peroxidase-conjugated anti-mouse or anti-rabbit IgG (Dako) antibody at 1:2000 dilution. After staining with ECL West Dura Substrate kit (Thermo Scientific), visualisation was performed on Chemidoc Digital Imager (Bio-Rad) (Supplementary Information). The relative protein secretion of ANGPTL6 for each variant was standardized to wild-type protein secretion. Experiments were performed in triplicate and analysed using an unpaired T-test with Welch's correction.

ANGPTL6 statistical follow-up analyses. We ran gene-based analyses using SKAT-O as described above on two subsets of ANGPTL6 variants (Supplementary Table 10): (1) the one variant that resulted in a complete loss of protein secretion (S284X; rs201622589), and (2) excluding S284X. For (1), we also used a Fisher's Exact Test. We repeated the Fisher's Exact Test of the S284X variant including additional population controls with pre-existing genotype data and replication samples (Online Methods, [Sample Sets](#); Supplementary Tables 11 and 12).

Power Analysis. We calculated the power to detect association to a gene region using the burden test with the *Power_Logistic_R* function and the haplotype dataset within the SKAT R-package⁵⁴. We performed power calculations using 500 simulations on a random 2Kb sub region. The effect sizes of the causal variants are equal to $\log_{10}(\text{MAF})$ with a maximum effect size of 1.6 (MAF = 0.0001) and all have the same direction of effect. We limited the MAF threshold for causal variants to $\leq 0.01, 0.001, 0.0005$ and varied the percentage of causal variants in the region to be between 10–90%. (More details in Supplementary Note 4).

References

1. El-Sayed Moustafa, J. S. & Froguel, P. From obesity genetics to the future of personalized obesity therapy. *Nat Rev Endocrinol* **9**, 402–13 (2013).
2. van der Klaauw, A. A. & Farooqi, I. S. The hunger genes: pathways to obesity. *Cell* **161**, 119–32 (2015).
3. Wheeler, E. *et al.* Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet* **45**, 513–7 (2013).
4. Consortium, U. K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90 (2015).
5. Farooqi, I. S. *et al.* Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N Engl J Med* **348**, 1085–95 (2003).
6. Lemos, M. C. & Thakker, R. V. GNAS mutations in Pseudohypoparathyroidism type 1a and related disorders. *Hum Mutat* **36**, 11–9 (2015).
7. Aldred, M. A. & Trembath, R. C. Activating and inactivating mutations in the human GNAS1 gene. *Hum Mutat* **16**, 183–9 (2000).
8. Warner, D. R., Weng, G., Yu, S., Matalon, R. & Weinstein, L. S. A novel mutation in the switch 3 region of Gsalpha in a patient with Albright hereditary osteodystrophy impairs GDP binding and receptor activation. *J Biol Chem* **273**, 23976–83 (1998).
9. Bastida Eizaguirre, M. *et al.* [Albright hereditary osteodystrophy: identification of a novel mutation in a family]. *An Esp Pediatr* **54**, 598–600 (2001).
10. Chen, M. *et al.* Central nervous system imprinting of the G protein G(s)alpha and its role in metabolic regulation. *Cell Metab* **9**, 548–55 (2009).
11. Lee, S., Wu, M. C. & Lin, X. Optimal tests for rare variant effects in sequencing association studies. *Biostatistics* **13**, 762–75 (2012).
12. Moutsianas, L. *et al.* The power of gene-based rare variant methods to detect disease-associated variation and test hypotheses about complex disease. *PLoS Genet* **11**, e1005165 (2015).
13. Oike, Y. *et al.* Angiopoietin-related growth factor antagonizes obesity and insulin resistance. *Nat Med* **11**, 400–8 (2005).
14. Kadomatsu, T., Tabata, M. & Oike, Y. Angiopoietin-like proteins: emerging targets for treatment of obesity and related metabolic diseases. *FEBS J* **278**, 559–64 (2011).
15. Purcell, S. M. *et al.* A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–90 (2014).

16. Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet* **94**, 677–94 (2014).
17. Krumm, N. *et al.* Excess of rare, inherited truncating mutations in autism. *Nat Genet* **47**, 582–8 (2015).
18. Farooqi, I. S. *et al.* Effects of recombinant leptin therapy in a child with congenital leptin deficiency. *N Engl J Med* **341**, 879–84 (1999).
19. Kuhnen, P. *et al.* Proopiomelanocortin Deficiency Treated with a Melanocortin-4 Receptor Agonist. *N Engl J Med* **375**, 240–6 (2016).
20. Rasmussen, S. G. *et al.* Crystal structure of the beta2 adrenergic receptor-Gs protein complex. *Nature* **477**, 549–55 (2011).
21. Tesmer, J. J., Sunahara, R. K., Gilman, A. G. & Sprang, S. R. Crystal structure of the catalytic domains of adenylyl cyclase in a complex with G α .GTP γ S. *Science* **278**, 1907–16 (1997).
22. Lynn, P. Sample design for Understanding Society. *Understanding Society Working Paper Series* **2009–01**(2009).
23. Day, N. *et al.* EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer* **80**(Suppl 1), 95–103 (1999).
24. Kamstrup, P. R., Tybjaerg-Hansen, A., Steffensen, R. & Nordestgaard, B. G. Genetically elevated lipoprotein(a) and increased risk of myocardial infarction. *JAMA* **301**, 2331–9 (2009).
25. Nordestgaard, B. G., Benn, M., Schnohr, P. & Tybjaerg-Hansen, A. Nonfasting triglycerides and risk of myocardial infarction, ischemic heart disease, and death in men and women. *JAMA* **298**, 299–308 (2007).
26. Varbo, A. *et al.* Remnant cholesterol as a causal risk factor for ischemic heart disease. *J Am Coll Cardiol* **61**, 427–36 (2013).
27. Danesh, J. *et al.* EPIC-Heart: the cardiovascular component of a prospective study of nutritional, lifestyle and biological factors in 520,000 middle-aged participants from 10 European countries. *Eur J Epidemiol* **22**, 129–41 (2007).
28. Sudlow, C. *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).
29. Exome Aggregation Consortium (ExAC), Cambridge, MA ([URL: http://exac.broadinstitute.org](http://exac.broadinstitute.org)) ([September, 2015]).
30. Hinney, A. *et al.* Genome Wide Association (GWA) Study for Early Onset Extreme Obesity Supports the Role of Fat Mass and Obesity Associated Gene (FTO) Variants. *PLoS ONE* **2**, e1361 (2007).
31. Quante, M. *et al.* The LIFE child study: a life course approach to disease and health. *BMC Public Health* **12**, 1021 (2012).
32. Korner, A., Berndt, J., Stumvoll, M., Kiess, W. & Kovacs, P. TCF7L2 gene polymorphisms confer an increased risk for early impairment of glucose metabolism and increased height in obese children. *J Clin Endocrinol Metab* **92**, 1956–60 (2007).
33. Grozeva, D. *et al.* De novo loss-of-function mutations in SETD5, encoding a methyltransferase in a 3p25 microdeletion syndrome critical region, cause intellectual disability. *Am J Hum Genet* **94**, 618–24 (2014).
34. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–9 (2009).
35. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–8 (2011).
36. Genomes Project, C. *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–73 (2010).
37. Tennessen, J. A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–9 (2012).
38. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–70 (2010).
39. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res* **31**, 3812–4 (2003).
40. Ng, P. C. & Henikoff, S. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet* **7**, 61–80 (2006).
41. Sunyaev, S., Ramensky, V. & Bork, P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* **16**, 198–200 (2000).
42. Sunyaev, S. *et al.* Prediction of deleterious human alleles. *Hum Mol Genet* **10**, 591–7 (2001).
43. Gonzalez-Perez, A. & Lopez-Bigas, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet* **88**, 440–9 (2011).
44. Eilbeck, K. *et al.* The Sequence Ontology: a tool for the unification of genome annotations. *Genome Biol* **6**, R44 (2005).
45. Mungall, C. J., Batchelor, C. & Eilbeck, K. Evolution of the Sequence Ontology terms and relationships. *J Biomed Inform* **44**, 87–93 (2011).
46. Jun, G. *et al.* Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* **91**, 839–48 (2012).
47. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–9 (2006).
48. Wang, C., Zhan, X., Liang, L., Abecasis, G. R. & Lin, X. Improved ancestry estimation for both genotyping and sequencing data using projection procrustes analysis and genotype imputation. *Am J Hum Genet* **96**, 926–37 (2015).
49. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–75 (2007).
50. Online Mendelian Inheritance in Man, OMIM[®]. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD), (<http://omim.org/>) (February, 2013).
51. The Mouse Genome Database (MGD), Mouse Genome Informatics. *The Jackson Laboratory, Bar Harbor, Maine*. <http://www.informatics.jax.org> (February, 2013).
52. The Rat Genome Browser (RGD), Rat Genome Database Web Site. *Medical College of Wisconsin, Milwaukee, Wisconsin* <http://rgd.mcw.edu/> (February, 2013).
53. Wu, M. C. *et al.* Rare-variant association testing for sequencing data with the sequence kernel association test. *Am J Hum Genet* **89**, 82–93 (2011).
54. Seunggeun Lee, with contributions from Larisa Miropolsky and Michael Wu. SKAT: SNP-Set(Sequence) Kernel Association Test.. R package version 1.1.2. <http://CRAN.R-project.org/package=SKAT> (2015).
55. Yang, Z. *et al.* Crystal structure of fragment D from lamprey fibrinogen complexed with the peptide Gly-His-Arg-Pro-amide. *Biochemistry* **41**, 10218–24 (2002).

Acknowledgements

We are indebted to the patients and their families for their participation and to the physicians involved in the Genetics of Obesity Study (GOOS). This work was supported by the Wellcome Trust (ISF, IB) (098497/Z/12/Z; WT098051), Medical Research Council (ISF, SOR) (MRC_MC_UU_12012/5), NIHR Cambridge Biomedical Research Centre (ISF, IB, SOR), Bernard Wolfe Health Neuroscience Endowment (ISF), European Research Council (ISF) and NIH grant DK064265 (GLM), the European Community's Seventh Framework Programme (FP7/2007–2013) project Beta-JUDO n°279153 (ISF, AK). This study comprises one arm of the UK10K Consortium (WT091310). The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at the University of Essex and funded by the Economic and Social Research Council. The survey was conducted by NatCen and the genome-wide scan data were analysed and deposited by the Wellcome Trust

Sanger Institute. Information on how to access the data can be found on the Understanding Society website <https://www.understandingsociety.ac.uk/>. AH and JH were funded by the German Ministry for Education and Research (National Genome Research Net-Plus 01GS0820), the German Research Foundation (DFG; HI865/2-1), the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreements n°245009 and n°262055. We thank participants and staff of the Copenhagen City Heart Study, Copenhagen Ischemic Heart Disease Study, and the Copenhagen General Population Study for their important contributions (CCHS, CGPS, CIHDS). CHD case ascertainment and validation, genotyping, and clinical chemistry assays in EPIC-CVD were supported by grants awarded to the University of Cambridge from the EU Framework Programme 7 (HEALTH-F2-2012-279233), the UK Medical Research Council (G0800270) and British Heart Foundation (SP/09/002), the European Research Council (268834), the UK National Institute for Health Research Cambridge Biomedical Research Centre, Merck and Pfizer. We thank all EPIC participants and staff for their contribution to the study, the laboratory teams at the Medical Research Council Epidemiology Unit for sample management and Cambridge Genomic Services for genotyping, Sarah Spackman for data management, and the team at the EPIC-CVD Coordinating Centre for study coordination and administration (EPIC-CVD). JD is supported by a British Heart Foundation Professorship, NIHR Senior Investigator, European Research Council Senior Investigator. Leipzig LIFE Child as part of the Leipzig Childhood Obesity cohort was funded by the European Union, by the European Regional Development Fund (ERFD) by means of the Free State of Saxony within the framework of the excellence initiative. This research has been conducted using the UK Biobank Resource.

Author Contributions

I.B., I.S.F., A.E. and E.G.B. designed the study, analysed the data and wrote the paper. J.M.K., R.B., S.A., E.H. contributed to recruitment and consenting of the cohort; N.A., V.M., A.K., S.P., S.G. to validation of genetic findings in family studies and functional studies; G.M., E.W., I.T., E.Z. contributed to data analysis and G.L.M., R.I.P., S.O.R., to data collection and/or analysis. A.K., J.K. contributed data in Leipzig children. A.H., J.H., R.A.S., C.L., N.J.W., P.S., J.M.M.H., A.S.B., J.D., B.G.N., S.F.N., S.A., Understanding Society Scientific Group, EPIC-CVD Consortium contributed data for the *ANGPTL6* (p.S284X) variant. Targeted and whole-exome sequencing was performed as part of the UK10K consortium. All authors reviewed and contributed to the final version of the paper.

Additional Information

Supplementary information accompanies this paper at doi:[10.1038/s41598-017-03054-8](https://doi.org/10.1038/s41598-017-03054-8)

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017

Consortia Understanding Society Scientific Group

Michaela Benzeval¹⁴, Jonathan Burton¹⁴, Nicholas Buck¹⁴, Annette Jäckle¹⁴, Meena Kumari¹⁴, Heather Laurie¹⁴, Peter Lynn¹⁴, Stephen Pudney¹⁴, Birgitta Rabe¹⁴ & Dieter Wolke¹⁵

¹⁴Institute for Social and Economic Research, University of Essex, Colchester, UK. ¹⁵University of Warwick, Warwick, UK.

UK10K Consortium

Saeed Al Turki¹⁶, Carl A. Anderson¹, Richard Anney¹⁷, Dinu Antony¹⁸, María Soler Artigas¹⁹, Muhammad Ayub²⁰, Senduran Bala¹, Jeffrey C. Barrett¹, Phil Beales¹⁸, Jamie Bentham²¹, Shoumo Bhattacharya²¹, Ewan Birney²², Douglas Blackwood²³, Martin Bobrow²⁴, Patrick F. Bolton^{25,26,27}, Chris Boustred²⁸, Gerome Breen^{26,27}, Mattia Calissano²⁹, Keren Carss¹, Ruth Charlton³⁰, Krishna Chatterjee³, Lu Chen³¹, Antonio Ciampi³², Sebahattin Cirak^{29,33}, Peter Clapham¹, Gail Clement³⁴, Guy Coates¹, Massimiliano Cocca^{35,36}, David A. Collier^{27,37}, Catherine Cosgrove²¹, Tony Cox¹, Nick Craddock³⁸, Lucy Crooks³⁹, Sarah Curran^{25,40,41}, David Curtis⁴², Allan Daly¹, Petr Danecek¹, Ian N. M. Day⁴³, Aaron Day-Williams⁴⁴, Anna Dominiczak⁴⁵, Thomas Down⁴⁶, Yuanping Du⁴⁷, Ian Dunham²², Richard Durbin¹, Sarah Edkins¹, Rosemary Ekong⁴⁸, Peter Ellis¹, David M. Evans^{49,50}, David R. Fitzpatrick⁵¹, Paul Flicek²², James Floyd⁵², A. Reghan Foley²⁹, Christopher S. Franklin¹, Marta Futema⁵³, Louise Gallagher¹⁷, Tom R. Gaunt⁴⁹, Matthias Geijs¹, Daniel Geschwind⁵⁴, Celia M. T. Greenwood^{32,55,56,57}, Heather Griffin⁵⁸, Detelina Grozeva²⁴, Xiaosen Guo^{47,59}, Xueqin Guo⁴⁷, Hugh Gurling⁶⁰, Deborah Hart³⁴, Peter Holmans³⁸, Bryan Howie⁶¹, Jie Huang¹, Liren Huang⁴⁷, Tim Hubbard^{1,46}, Steve E. Humphries⁵³, Matthew E. Hurles¹, Pirro Hysi³⁴, Valentina Iotchkova^{1,22}, David K. Jackson¹, Yalda Jamshidi⁶², Chris Joyce¹, Konrad J. Karczewski^{63,64}, Jane Kaye⁵⁸, Thomas Keane¹, John P. Kemp^{49,50}, Karen Kennedy^{1,65}, Alastair Kent⁶⁶, Farrah Khawaja⁶⁷, Margriet van Kogelenberg¹, Anja Kolb-Kokocinski¹, Genevieve Lachance³⁴, Cordelia Langford¹, Daniel Lawson⁶⁸, Irene Lee⁶⁹, Monkol Lek⁶³, Rui Li^{55,56,70}, Yingrui Li⁴⁷, Jieqin Liang⁴⁷, Hong Lin⁴⁷, Ryan Liu⁷¹, Jouko Lönnqvist⁷², Luis R. Lopes^{73,74}, Margarida Lopes^{1,75,76}, Daniel G. MacArthur^{63,64}, Massimo Mangino^{34,77}, Jonathan Marchini^{75,78}, John Maslen¹, Iain Mathieson⁷⁹, Peter McGuffin²⁷, Andrew M. McIntosh²³, Andrew G. McKechnie^{23,80}, Andrew McQuillin⁶⁰, Yasin Memari¹, Sarah Metrustry³⁴, Nicola Migone⁸¹, Josine L. Min⁴⁹, Hannah M. Mitchison¹⁸, Alireza Moayyeri^{34,52}, Andrew Morris⁸³, James Morris¹, Francesco Muntoni²⁹, Kate Northstone⁴⁹, Michael C. O'Donovan³⁸, Alexandros Onoufriadis⁴⁶, Karim Oualkacha⁸⁴, Michael J. Owen³⁸, Aarno Palotie^{1,85,86}, Kalliope Panoutsopoulou¹, Victoria Parker³, Jeremy R. Parr⁸⁷, Lavinia Paternoster⁴⁹, Tiina Paunio^{72,88}, Felicity Payne¹, Stewart J. Payne⁸⁹, John R. B. Perry^{8,34}, Olli Pietilainen^{72,85}, Vincent Plagnol⁹⁰, Rebecca C. Pollitt⁹¹, David J. Porteous⁹², Sue Povey⁴⁸, Michael A. Quail¹, Lydia Quayle³⁴, F. Lucy Raymond²⁴, Karola Rehnström¹, J. Brent Richards^{32,34,55,56,70}, Cheryl K. Ridout⁹³, Susan Ring⁹⁴, Graham R. S. Ritchie²², Nicola Roberts²⁴, Rachel L. Robinson³⁰, David B. Savage³, Peter Scambler¹⁸, Stephan Schiffels¹, Miriam Schmidts^{18,95}, Nadia Schoenmakers³, Richard H. Scott^{18,96}, Robert K. Semple³, Eva Serra¹, Sally I. Sharp⁶⁰, Adam Shaw⁹⁷, Hashem A. Shihab⁴⁹, So-Youn Shin^{1,49}, David Skuse⁶⁹, Kerrin S. Small³⁴, Carol Smee¹, Blair H. Smith⁹⁸, George Davey Smith⁴⁹, Nicole Soranzo^{1,31}, Lorraine Southam^{1,75}, Olivera Spasic-Boskovic²⁴, Timothy D. Spector³⁴, David St Clair⁹⁹, Beate St Pourcain^{49,100,101}, Jim Stalker¹, Elizabeth Stevens²⁹, Jianping Sun^{32,55}, Gabriela Surdulescu³⁴, Jaana Suvisaari⁷², Petros Syrris⁷³, Rohan Taylor⁶⁷, Jing Tian⁴⁷, Nicholas J. Timpson⁴⁹, Martin D. Tobin^{19,102}, Ana M. Valdes³⁴, Anthony M. Vandersteent¹⁰³, Parthiban Vijayarangakannan¹, Peter M. Visscher^{50,104}, Louise V. Wain¹⁹, Klaudia Walter¹, James T. R. Walters³⁸, Guangbiao Wang⁴⁷, Jun Wang^{47,59,105,106,107}, Yu Wang⁴⁷, Kirsten Ward³⁴, Tamiaka Whyte²⁹, Hywel J. Williams^{38,108}, Kathleen A. Williamson⁵¹, Crispian Wilson²⁴, Scott G. Wilson^{24,109,110}, Kim Wong¹, ChangJiang Xu^{32,55}, JianYang^{50,104}, Feng Zhang³⁴, Pingbo Zhang⁴⁷ & Hou-Feng Zheng^{55,56,70}

¹The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton CB10 1HH, Cambridge, UK.

¹⁶Department of Pathology, King Abdulaziz Medical City, P.O. Box 22490, Riyadh 11426, Saudi Arabia. ¹⁷Department of Psychiatry, Trinity Centre for Health Sciences, St James Hospital, James's Street, Dublin 8, Ireland. ¹⁸Genetics and Genomic Medicine and Birth Defects Research Centre, UCL Institute of Child Health, London WC1N 1EH, UK.

¹⁹Departments of Health Sciences and Genetics, University of Leicester, Leicester LE1 7RH, UK. ²⁰Division of Developmental Disabilities, Department of Psychiatry, Queen's University, Kingston, Ontario N6C 0A7, Canada.

²¹Department of Cardiovascular Medicine and Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford

OX3 7BN, UK. ²²European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. ²³Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh EH10 5HF, UK. ²⁴Academic Laboratory of Medical Genetics, Box 238, Lv 6 Addenbrooke's Treatment Centre, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK. ²⁵Department of Child Psychiatry, Institute of Psychiatry, Psychology and Neuroscience, King's College London, 16 De Crespigny Park, London SE5 8AF, UK. ²⁶NIHR BRC for Mental Health, Institute of Psychiatry, Psychology and Neuroscience and SLAM NHS Trust, King's College London, London, UK. 16 De Crespigny Park, London SE5 8AF, UK. ²⁷MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, Denmark Hill, London SE5 8AF, UK. ²⁸North East Thames Regional Genetics Service, Great Ormond Street Hospital NHS Foundation Trust, London WC1N 3JH, UK. ²⁹Dubowitz Neuromuscular Centre, UCL Institute of Child Health & Great Ormond Street Hospital, London WC1N 1EH, UK. ³⁰Leeds Genetics Laboratory, St James University Hospital, Beckett Street, Leeds LS9 7TF, UK. ³¹Department of Haematology, University of Cambridge, Long Road, Cambridge CB2 0PT, UK. ³²Department of Epidemiology, Biostatistics and Occupational Health, McGill University, Montreal, Quebec H3A 1A2, Canada. ³³Institut für Humangenetik, Uniklinik Köln, Kerpener Strasse 34, 50931 Köln, Germany. ³⁴The Department of Twin Research & Genetic Epidemiology, King's College London, St Thomas' Campus, Lambeth Palace Road, London SE1 7EH, UK. ³⁵Medical Genetics, Institute for Maternal and Child Health IRCCS "Burlo Garofolo", 34100 Trieste, Italy. ³⁶Department of Medical, Surgical and Health Sciences, University of Trieste, 34100 Trieste, Italy. ³⁷Lilly Research Laboratories, Eli Lilly & Co. Ltd., Erl Wood Manor, Sunninghill Road, Windlesham GU20 6PH, UK. ³⁸MRC Centre for Neuropsychiatric Genetics & Genomics, Institute of Psychological Medicine & Clinical Neurosciences, School of Medicine, Cardiff University, Cardiff CF24 4HQ, UK. ³⁹Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK. ⁴⁰University of Sussex, Brighton BN1 9RH, UK. ⁴¹Sussex Partnership NHS Foundation Trust, Swandean, Arundel Road, Worthing BN13 3EP, UK. ⁴²University College London (UCL), UCL Genetics Institute, Darwin Building, Gower Street, London WC1E 6BT, UK. ⁴³Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK. ⁴⁴Computational Biology & Genomics, Biogen Idec, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁴⁵Institute of Cardiovascular and Medical Sciences, University of Glasgow, Wolfson Medical School Building, University Avenue, Glasgow G12 8QQ, UK. ⁴⁶Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Hospital, London SE1 9RT, UK. ⁴⁷BGI-Shenzhen, Shenzhen 518083, China. ⁴⁸University College London (UCL) Department of Genetics, Evolution & Environment (GEE), Gower Street, London WC1E 6BT, UK. ⁴⁹MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK. ⁵⁰University of Queensland Diamantina Institute, Translational Research Institute, Brisbane, Queensland 4102, Australia. ⁵¹MRC Human Genetics Unit, MRC Institute of Genetics and Molecular Medicine, at the University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK. ⁵²The Genome Centre, John Vane Science Centre, Queen Mary, University of London, Charterhouse Square, London EC1M 6BQ, UK. ⁵³Cardiovascular Genetics, BHF Laboratories, Rayne Building, Institute of Cardiovascular Sciences, University College London, London WC1E 6JJ, UK. ⁵⁴UCLA David Geffen School of Medicine, Los Angeles, California 90095, USA. ⁵⁵Lady Davis Institute, Jewish General Hospital, Montreal, Quebec H3T 1E2, Canada. ⁵⁶Department of Human Genetics, McGill University, Montreal, Quebec H3A 1B1, Canada. ⁵⁷Department of Oncology, McGill University, Montreal, Quebec H2W 1S6, Canada. ⁵⁸HeLEX – Centre for Health, Law and Emerging Technologies, Nuffield Department of Population Health, University of Oxford, Old Road Campus, Oxford OX3 7LF, UK. ⁵⁹Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, DK-2200 Copenhagen, Denmark. ⁶⁰University College London (UCL), Molecular Psychiatry Laboratory, Division of Psychiatry, Gower Street, London WC1E 6BT, UK. ⁶¹Adaptive Biotechnologies Corporation, Seattle, Washington 98102, USA. ⁶²Human Genetics Research Centre, St George's University of London, London SW17 0RE, UK. ⁶³Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁶⁴Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ⁶⁵National Cancer Research Institute, Angel Building, 407 St John Street, London EC1V 4AD, UK. ⁶⁶Genetic Alliance UK, 4D Leroy House, 436 Essex Road, London N1 3QP, UK. ⁶⁷SW Thames Regional Genetics Lab, St George's University, Cranmer Terrace, London SW17 0RE, UK. ⁶⁸Schools of Mathematics and Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK. ⁶⁹Behavioural and Brain Sciences Unit, UCL Institute of Child Health, London WC1N 1EH, UK. ⁷⁰Department of Medicine, Jewish General Hospital, McGill University, Montreal, Quebec H3A 1B1, Canada. ⁷¹BGI-Europe, London EC2M 4YE, UK. ⁷²National Institute for Health and Welfare (THL), Helsinki FI-00271, Finland. ⁷³Institute of Cardiovascular Science, University College London, Gower Street, London WC1E 6BT, UK. ⁷⁴Cardiovascular Centre of the University of Lisbon, Faculty of Medicine, University of Lisbon, Avenida Professor Egas Moniz, 1649-028 Lisbon, Portugal. ⁷⁵Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Oxford OX3 7BN, UK. ⁷⁶Illumina Cambridge Ltd, Chesterford Research Park, Cambridge CB10 1XL, UK. ⁷⁷National Institute for Health Research (NIHR) Biomedical Research Centre at Guy's and St Thomas' Foundation Trust, London SE1 9RT, UK. ⁷⁸Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK. ⁷⁹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸⁰The Patrick Wild Centre, The University of Edinburgh, Edinburgh EH10 5HF, UK. ⁸¹Department of Medical Sciences, University of Torino, 10124 Torino, Italy. ⁸²Institute of Health Informatics, Farr Institute of Health Informatics Research, University College London (UCL), 222 Euston Road, London NW1 2DA, UK. ⁸³Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, 9 Little France Road, Edinburgh EH16 4UX, UK. ⁸⁴Department of Mathematics, Université de Québec À Montréal, Montréal, Québec H3C 3P8, Canada. ⁸⁵Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki FI-00014, Finland. ⁸⁶Program in Medical and Population Genetics and Genetic Analysis Platform, The Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02132, USA. ⁸⁷Institute of Neuroscience, Henry Wellcome Building for Neuroecology, Newcastle University, Framlington Place, Newcastle upon Tyne NE2 4HH, UK. ⁸⁸University of Helsinki, Department of Psychiatry, Helsinki FI-00014, Finland. ⁸⁹North West Thames Regional

Genetics Service, Kennedy-Galton Centre, Northwick Park Hospital, Watford Road, Harrow HA1 3UJ, UK. ⁹⁰University College London (UCL) Genetics Institute (UGI) Gower Street, London WC1E 6BT, UK. ⁹¹Connective Tissue Disorders Service, Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK. ⁹²Centre for Genomic and Experimental Medicine, Institute of Genetics and Experimental Medicine, University of Edinburgh, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. ⁹³Molecular Genetics, Viapath at Guy's Hospital, London SE1 9RT, UK. ⁹⁴ALSPAC & School of Social and Community Medicine, University of Bristol, Oakfield House, Oakfield Grove, Clifton, Bristol BS8 2BN, UK. ⁹⁵Human Genetics Department, Radboudumc and Radboud Institute for Molecular Life Sciences (RIMLS), Geert Grooteplein 25, Nijmegen 6525 HP, The Netherlands. ⁹⁶Department of Clinical Genetics, Great Ormond Street Hospital, London, WC1N 3JH, UK. ⁹⁷Clinical Genetics, Guy's & St Thomas' NHS Foundation Trust, London SE1 9RT, UK. ⁹⁸Mackenzie Building, Kirsty Semple Way, Ninewells Hospital and Medical School, Dundee DD2 4RB, UK. ⁹⁹Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, UK. ¹⁰⁰School of Oral and Dental Sciences, University of Bristol, Lower Maudlin Street, Bristol BS1 2LY, UK. ¹⁰¹School of Experimental Psychology, University of Bristol, 12a Priors Road, Bristol BS8 1TU, UK. ¹⁰²National Institute for Health Research (NIHR) Leicester Respiratory Biomedical Research Unit, Glenfield Hospital, Leicester LE3 9QP, UK. ¹⁰³Maritime Medical Genetics Service, 5850/5980 University Avenue, PO Box 9700, Halifax, Nova Scotia B3K 6R8, Canada. ¹⁰⁴Queensland Brain Institute, University of Queensland, Brisbane, Queensland 4072, Australia. ¹⁰⁵Princess Al Jawhara Albrahim Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, P.O. Box 80200, Jeddah 21589, Saudi Arabia. ¹⁰⁶Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China. ¹⁰⁷Department of Medicine and State Key Laboratory of Pharmaceutical Biotechnology, University of Hong Kong, 21 Sassoon Road, Hong Kong. ¹⁰⁸The Centre for Translational Omics – GOSgene, UCL Institute of Child Health, London WC1N 1EH, UK. ¹⁰⁹School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia 6009, Australia. ¹¹⁰Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia 6009, Australia.

EPIC-CVD consortium

Kim Overvad^{111,112}, Anne Tjønneland¹¹³, Françoise Clavel-Chapelon¹¹⁴, Rudolf Kaaks¹¹⁵, Heiner Boeing¹¹⁶, Antonia Trichopoulos^{117,118}, Pietro Ferrari¹¹⁹, Domenico Palli¹²⁰, Vittorio Krogh¹²¹, Salvatore Panico¹²², Rosario Tumino¹²³, Giuseppe Matullo^{124,125}, Jolanda Boer¹²⁶, Yvonne van der Schouw¹²⁷, Elisabete Weiderpass^{128,129,130,131}, J. Ramon Quiros¹³², María-José Sánchez^{133,134}, Carmen Navarro^{135,136}, Conchi Moreno-Iribas^{113,134}, Larraitz Arriola¹³⁷, Olle Melander¹³⁸, Patrik Wennberg¹³⁹, Timothy J. Key¹⁴⁰ & Elio Riboli¹⁴¹

¹¹¹Department of Public Health, Section for Epidemiology, Aarhus University, Aarhus, Denmark. ¹¹²Department of Cardiology, Aalborg University Hospital, Aalborg, Denmark. ¹¹³Diet, Genes and Environment, Danish Cancer Society Research Center, Copenhagen, Denmark. ¹¹⁴INSERM, Centre for Research in Epidemiology and Population Health (CESP), U1018, Nutrition, Hormones, and Women's Health Team, Institut Gustave Roussy, Villejuif, France. ¹¹⁵Division of Cancer Genetic Epidemiology, German Cancer Research Centre (DKFZ), im Neuenheimer Feld 581, 69121 Heidelberg, Germany. ¹¹⁶Department of Epidemiology, German Institute of Human Nutrition (DIFE), Potsdam-Rehbrücke, Germany. ¹¹⁷WHO Collaborating Center for Nutrition and Health, Unit of Nutritional Epidemiology and Nutrition in Public Health, Department of Hygiene, Epidemiology and Medical Statistics, University of Athens Medical School, Athens, Greece. ¹¹⁸Hellenic Health Foundation, Athens, Greece. ¹¹⁹IARC, Lyon, France. ¹²⁰Molecular and Nutritional Epidemiology Unit, Centro per lo Studio e la Prevenzione Oncologica-Scientific Institute of Tuscany, Florence, Italy. ¹²¹Epidemiology and Prevention Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy. ¹²²Dipartimento di Medicina Clinica e Chirurgia, Federico II University, Naples, Italy. ¹²³Cancer Registry and Histopathology Unit, Civic- M.P. Arezzo Hospital, ASP Ragusa, Italy. ¹²⁴Human Genetics Foundation, Turin, Italy. ¹²⁵Department of Medical Sciences, University of Turin, Italy. ¹²⁶Centre for Nutrition, Prevention and Health Services, National Institute for Public Health and the Environment (RIVM), Bilthoven, The Netherlands. ¹²⁷Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. ¹²⁸Department of Community Medicine, Faculty of Health Sciences, University of Tromsø, The Arctic University of Norway, Tromsø, Norway. ¹²⁹Department of Research, Cancer Registry of Norway, Institute of Population-Based Cancer Research, Oslo, Norway. ¹³⁰Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. ¹³¹Genetic Epidemiology Group, Folkhälsan Research Center, Helsinki, Finland. ¹³²Public Health Directorate, Asturias, Spain. ¹³³Public Health Institute of Navarra, Pamplona, Spain. ¹³⁴Red de Investigación en Servicios de Salud en Enfermedades Crónicas, Madrid, Spain. ¹³⁵Epidemiology Department, Murcia Health Authority, Murcia, Spain. ¹³⁶Centro de Investigación Biomédica en red en Epidemiología y Salud Pública, Barcelona, Spain. ¹³⁷Public Health Division of Gipuzkoa, Instituto Bio-Donostia, Basque Government, CIBERESP, Spain. ¹³⁸CRR, Lund University, 20502 Malmö, Sweden. ¹³⁹Department of Public Health and Clinical Medicine, Family Medicine, Umeå University, Umeå, Sweden. ¹⁴⁰Cancer Epidemiology Unit, Nuffield Department of Population Health, University of Oxford, Oxford, UK. ¹⁴¹German Cancer Research Center (DKFZ), Division of Cancer Epidemiology, Heidelberg, Germany.