ARTICLES
https://doi.org/10.1038/s41588-017-0011-x
nature genetics

Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity

Genome-wide association studies (GWAS) have identified >250 loci for body mass index (BMI), implicating pathways related to neuronal biology. Most GWAS loci represent clusters of common, noncoding variants from which pinpointing causal genes remains challenging. Here we combined data from 718,734 individuals to discover rare and low-frequency (minor allele frequency (MAF) < 5%) coding variants associated with BMI. We identified 14 coding variants in 13 genes, of which 8 variants were in genes (*ZBTB7B*, *ACHE*, *RAPGEF3*, *RAB21*, *ZFHX3*, *ENTPD6*, *ZFR2* and *ZNF169*) newly implicated in human obesity, 2 variants were in genes (*MC4R* and *KSR2*) previously observed to be mutated in extreme obesity and 2 variants were in *GIPR*. The effect sizes of rare variants are ~10 times larger than those of common variants, with the largest effect observed in carriers of an *MC4R* mutation introducing a stop codon (p.Tyr35Ter, MAF = 0.01%), who weighed ~7 kg more than non-carriers. Pathway analyses based on the variants associated with BMI confirm enrichment of neuronal genes and provide new evidence for adipocyte and energy expenditure biology, widening the potential of genetically supported therapeutic targets in obesity.

Obesity is a heritable disease and represents a major unmet public-health problem with only a few safe and long-term effective therapies[1] and intervention strategies[2]. In efforts to understand the genetic basis of obesity and identify potential targets for new therapies, GWAS for BMI and obesity risk have identified >250 common variants over the past decade[3–7]. Consistent with single-gene disorders of obesity[8], tissue expression and gene set enrichment analyses for genes implicated in BMI-associated loci have shown that the central nervous system (CNS) has a critical role in body weight regulation[5]. While the numerous GWAS loci have provided insight into broad biological mechanisms underlying body weight regulation, pinpointing the causal gene(s) and variant(s) remains a major challenge[9], as GWAS-identified variants are typically noncoding and may affect genes at long distance. The association of intronic *FTO* variants with BMI illustrates the challenges of identifying causal regulatory effects. The proposed causal variant in this locus was found to regulate the expression of nearby *RPGRIP1L* in some studies[10–12], whereas others found that it regulated the distant *IRX3* and *IRX5* genes in specific cell types[13,14].

To expedite mapping of obesity-related genes, we performed an exome-wide search for low-frequency (MAF = 1–5%) and rare (MAF < 1%) single-nucleotide variants (SNVs) associated with BMI using exome-targeted genotyping arrays. A total of 125 studies ($N_{individuals}$ = 718,734) performed single-variant association between up to 246,328 SNVs and BMI. In addition, we performed gene-based meta-analyses to aggregate rare and low-frequency coding SNVs across 14,541 genes. Using genetic, functional and computational follow-up analyses, we gained insights into the function of BMI-implicated genes and the biological pathways through which they may influence body weight.

Results

Fourteen rare and low-frequency coding variants in 13 genes. Our study comprises a discovery stage and a follow-up stage (Methods, Supplementary Fig. 1 and Supplementary Tables 1–3). In our primary analysis, the discovery stage included data from 123 studies (N_{max} = 526,508) across five ancestry groups, with individuals predominantly of European ancestry (~85%). Each study performed single-variant association analyses of coding variants present on the exome array, including up to 13,786 common (MAF > 5%) and 215,917 rare or low-frequency coding SNVs (affecting exons and splicing sites). Summary statistics were combined using fixed-effect meta-analyses. SNV associations for rare or low-frequency variants that reached a suggestive level of significance ($P < 2.0 \times 10^{-6}$) were taken forward for follow-up in two European cohorts, deCODE (N_{max} = 72,613) and UK Biobank (N_{max} = 119,613 (interim release)). Overall significance was assessed after combining the results of the discovery and follow-up studies into a final meta-analysis (all-ancestry, sex-combined additive model, N_{max} = 718,734); SNV associations that reached $P < 2 \times 10^{-7}$ were considered array-wide significant[15,16] (Table 1, Supplementary Figs. 2–4 and Supplementary Table 4). In secondary analyses, we performed sex-specific analyses, analyses limited to individuals of European ancestry and analyses using a recessive model of inheritance.

In our primary analysis of rare and low-frequency variants, we identified five rare SNVs in three genes (one in *KSR2*, two in *MC4R* and two in *GIPR*) and nine low-frequency SNVs in eight genes (one in *ZBTB7B*, two in *ACHE*, and one each in *RAPGEF3*, *PRKAG1*, *RAB21*, *HIP1R*, *ZFHX3* and *ENTPD6*) associated with BMI (Table 1, Box 1, Supplementary Fig. 3a and Supplementary Table 5). In secondary analyses, we identified two additional low-frequency SNVs: one in the all-ancestry, women-only analysis (*ZFR2*) and one in the European-ancestry-only analysis (*ZNF169*) (Table 1, Supplementary Fig. 3b,c and Supplementary Tables 6–8). Of these 16 SNVs, located in 13 genes, the 2 SNVs in *MC4R* ($r^2 = 1$; $D' = 1$) and the 2 SNVs in *ACHE* ($r^2 = 0.98$; $D' = 0.99$) were in high linkage disequilibrium (LD), whereas the two SNVs in *GIPR* ($r^2 = 0$; $D' = 0.16$) were independent of each other. Hence, the 16 SNVs represent 14 independent variants (4 rare and 10 low frequency), of which 8 are located in genes not previously implicated in BMI

Full list of authors and affiliations appear at the end of the paper.

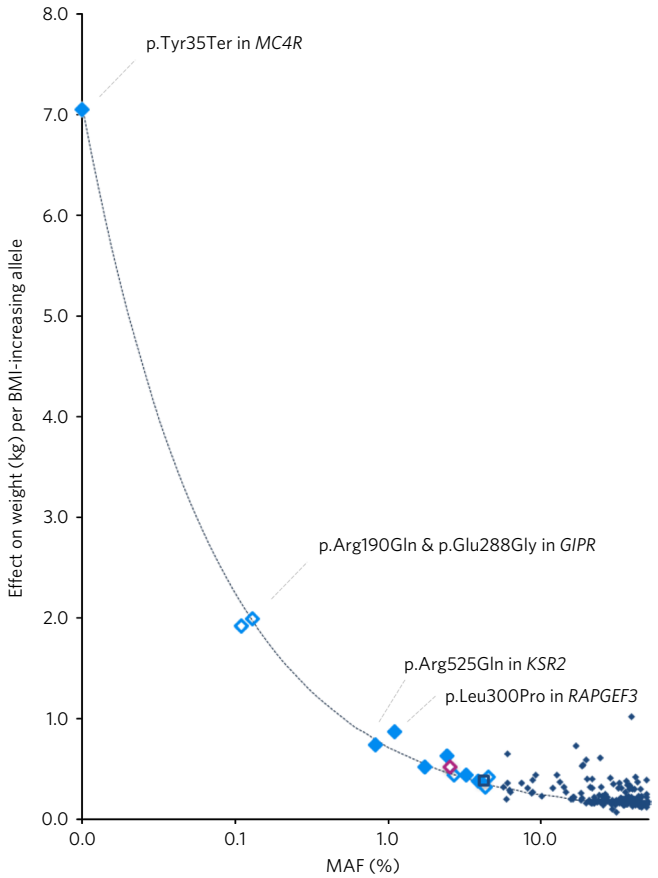

Fig. 1 | Effect sizes of the 14 BMI-associated rare and low frequency coding variants by variant minor allele frequency. Effect sizes are expressed in body weight (kg) per allele, assuming an s.d. of 4.5 kg and an average-sized person of 1.7 m in height. Filled markers indicate that the minor allele is associated with higher BMI, and unfilled markers indicate that the minor allele is associated with lower BMI. Variants were identified in all-ancestry analyses (light blue diamond), European-ancestry-only analyses (dark blue square) and women-only analyses (pink diamond). Effect sizes for previously identified GWAS loci are shown by navy blue diamonds. The dotted line represents 80% power, assuming $\alpha = 2 \times 10^{-7}$ and $N = 525{,}000$ (discovery sample size).

(*ZBTB7B*, *ACHE*, *RAPGEF3*, *RAB21*, *ZFHX3*, *ENTPD6*, *ZFR2* and *ZNF169*) and 6 are located in five loci that were previously identified by GWAS (*PRKAG1–BCDIN3D*, *HIP1R–CLIP1*, *MC4R* and *GIPR–QPCTL*)[5] and/or through sequencing of severe, early-onset obesity cases (*MC4R* and *KSR2*)[17–19] (Supplementary Fig. 5). Conditional analyses established that the coding SNVs in *PRKAG1*, *MC4R* and *GIPR* were independent of the common lead variants in GWAS loci (rs7138803, rs17782313 and rs2287019, respectively), whereas the SNVs in the *HIP1R* and GWAS locus near *CLIP1* (rs11057405) represent the same signal (Methods, Supplementary Fig. 5 and Supplementary Tables 9 and 10).

Next, we performed gene-based association tests (sequence kernel association test (SKAT) and variable threshold (VT); broad, strict) in up to 14,541 genes[20] to examine whether these aggregated analyses would yield new evidence of multiple rare or low-frequency coding SNVs in the same gene affecting BMI (Methods). Using broad SNV inclusion criteria, associations for 13 genes reached array-wide significance ($P < 2.5 \times 10^{-6}$)[15,16], 4 of which had not been highlighted in single-variant analyses (Table 2 and Supplementary Table 11). Conditional analyses showed that only for *GIPR* was the gene-based association driven by multiple SNVs (Table 2 and Supplementary Table 12). For all other genes, the associations were driven by a single SNV only, but these SNVs had not reached array-wide significance in single-variant analyses.

Taking these findings together, we identified 14 rare or low-frequency coding SNVs in 13 genes that are independently associated with BMI: 4 rare SNVs in 3 genes and 10 low-frequency SNVs in 10 genes. One SNV (*ZFR2*) showed a sex-specific effect, whereas no ancestry-specific effects were observed (Supplementary Fig. 6, Supplementary Tables 6–8 and Supplementary Note). Eight (*ACHE*, *ENTPD6*, *RAB21*, *RAPGEF3*, *ZBTB7B*, *ZFHX3*, *ZFR2* and *ZNF169*) of these 13 genes have not been previously implicated in body weight regulation (Table 1).

New common coding variants associated with BMI. Although the main focus of our study was rare and low-frequency coding SNVs, we also identified 92 common coding variants associated with BMI ($P < 2.0 \times 10^{-7}$; Supplementary Figs. 4 and 7, and Supplementary Table 4), of which 41 were new (Supplementary Table 9 and Supplementary Note). These new common loci were identified in our study but not in previous GWAS efforts because our current sample size is more than two times as large as that of the most recent GWAS meta-analysis[5] and also because some SNVs were not tested before, as they were not present on the HapMap reference panel and/or are on the X chromosome, which was not analyzed. Because of the increased sample size, the effect sizes of the 41 new common loci were smaller (on average, 0.014 s.d./allele; range, 0.010–0.024) than for previously established common loci (0.021 s.d./allele; 0.010–0.050) (Supplementary Fig. 7).

Impact of rare and low-frequency SNVs on BMI and obesity risk. The minor allele for half of the 14 rare and low-frequency SNVs was associated with lower BMI (Fig. 1 and Table 1). The effects of low-frequency SNVs ranged between 0.024 and 0.066 s.d./allele, equivalent to ~0.11 to 0.30 kg/m^2 in BMI or ~0.315 to 0.864 kg in body weight for a 1.7-m-tall person. The effects of rare SNVs ranged between 0.06 and 0.54 s.d./allele, equivalent to 0.26 to 2.44 kg/m^2 in BMI or 0.74 to 7.05 kg per allele (Fig. 1 and Table 1). By comparison, these rare SNV effect sizes were on average ten times larger than those for previously identified GWAS loci ($\text{effect}_{\text{mean}} = 0.019$ s.d./allele, equivalent to ~0.086 kg/m^2 or ~0.247 kg per allele), for which the largest effect was seen for the *FTO* locus (0.08 s.d./allele, equivalent to ~0.35 kg/m^2 or 1 kg per allele), and those for other GWAS loci ranged between 0.010 and 0.056 s.d./allele (~0.045 to 0.25 kg/m^2 or 0.130 to 0.728 kg)[5].

Effect sizes increased as MAF decreased, in particular for SNVs with MAF <0.5% (~1 heterozygote carrier in 100 people), consistent with the statistical power of our sample (Fig. 1). For example, the nonsense *MC4R* SNV encoding p.Tyr35Ter (rs13447324, MAF = 0.01%) was present in ~1 in 5,000 individuals and resulted in a ~7-kg increase in body weight for a 1.7-m-tall person. The two *GIPR* SNVs contributed independently to lower body weight: carriers (1 in ~455 individuals) of p.Arg190Gln (rs139215588) weighed ~1.92 kg (–0.148 s.d. BMI) less than non-carriers and carriers (1 in ~385 individuals) of p.Glu288Gly (rs143430880) weighed ~1.99 kg (–0.153 s.d. BMI) less. Among 115,611 individuals from the UK Biobank, one apparently healthy 61-year-old woman, with no reported illnesses, carried both rare *GIPR* alleles and weighed ~11.2 kg less (equivalent to –0.86 s.d. BMI or –3.87 kg/m^2) than the average non-carrier of the same height (Supplementary Fig. 8). The possible synergistic effect for the two *GIPR* alleles needs confirmation in additional individuals who carry both variants.

Even though the effect sizes of low-frequency and, in particular, rare SNVs tended to be larger than those of common GWAS-identified loci[5], the 14 SNVs combined explained <0.1% of BMI variation because of their low population frequencies (Table 1 and Methods). Also, although the effects of the four rare SNVs (one each

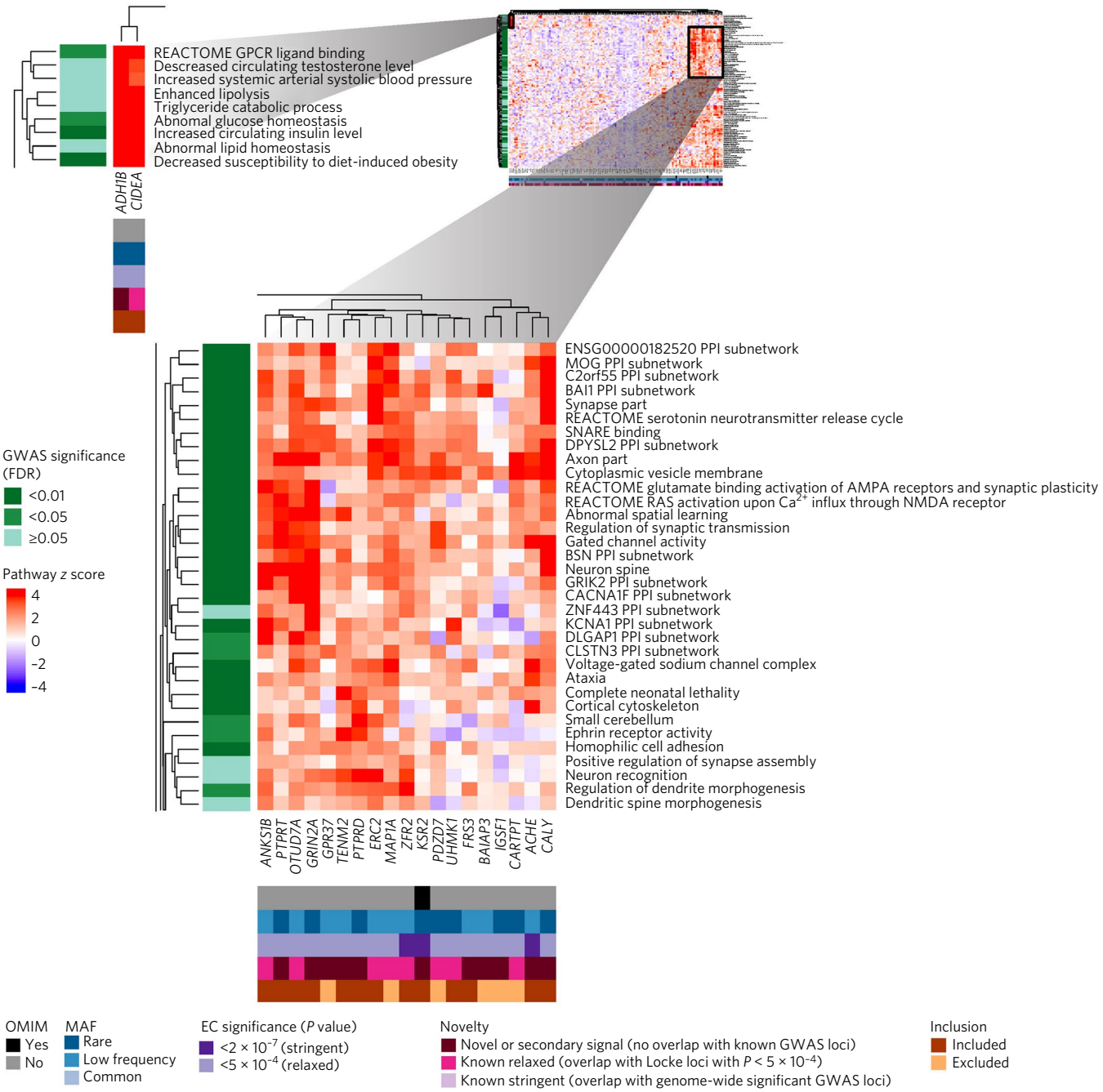

Fig. 2 | Heat map showing DEPICT gene set enrichment results for rare and low-frequency coding SNVs with suggestive and significant evidence of association with BMI. For any given cell, the color indicates how strongly the corresponding gene (*x* axis) is predicted to belong to the reconstituted gene set (*y* axis), based on the gene's *z* score for gene set inclusion in DEPICT's reconstituted gene sets (red indicates a higher *z* score, and blue indicates a lower *z* score). To visually reduce redundancy and increase clarity, we chose one representative 'meta-gene set' for each group of highly correlated gene sets based on affinity propagation clustering (Methods and Supplementary Note). Heat map intensity and DEPICT *P* values (Supplementary Table 17) correspond to the most significantly enriched gene set within the meta-gene set. Annotations for genes indicate (i) whether the gene has an Online Mendelian Inheritance in Man (OMIM) annotation as underlying a monogenic obesity disorder (black/gray), (ii) the MAF of the significant ExomeChip variant (blue), (iii) whether the variant's *P* value reached array-wide significance ($<2 \times 10^{-7}$) or suggestive significance ($<5 \times 10^{-4}$) (purple), (iv) whether the variant was new, overlapped 'relaxed' GWAS signals from Locke et al.[5] (GWAS $P < 5 \times 10^{-4}$) or overlapped 'stringent' GWAS hits (GWAS $P < 5 \times 10^{-8}$) (pink), and (v) whether the gene was included in the gene set enrichment analysis or excluded by filters (orange/brown) (Methods and Supplementary Note). Annotations for gene sets indicate whether the meta-gene set was significant (green; false discovery rate (FDR)) in the DEPICT analysis of GWAS results[5]. Here two regions of particularly strong gene set membership are shown (see full heat map in Supplementary Fig. 10a).

in *KSR2* and *MC4R* and 2 in *GIPR*) were large by GWAS standards, penetrance for obesity is still expected to be low. Indeed, using data from the UK Biobank ($N_{max} = 119{,}781$), we compared the prevalence of normal weight ($18.5\,kg/m^2 \leq BMI < 25\,kg/m^2$) and obesity (BMI $\geq 30\,kg/m^2$) between carriers and non-carriers of these variants (Methods and Supplementary Table 13). For the variants in *GIPR* (p.Arg190Gln and p.Glu288Gly), both of which are BMI-decreasing SNVs, carriers tended ($P < 0.05$) to have lower obesity prevalence

Table 1 | Rare and low-frequency coding variants significantly associated with BMI

Chr:position	Variant	Coding locus	Allele		Amino acid change	EAF (%)	β (s.d./allele)	SE	P value	N	Explained variance (%)
			Effect	Other							
All ancestry additive											
1:154987704	rs141845046	*ZBTB7B*[a]	T	C	p.Pro190Ser	2.44	0.048	0.006	7.73×10^{-18}	718,628	0.011
7:100490797	rs1799805	*ACHE*[a]	T	G	p.His353Asn	3.90	0.029	0.005	2.82×10^{-10}	707,448	0.006
12:48143315	rs145878042	*RAPGEF3*[a]	G	A	p.Leu300Pro	1.10	0.066	0.008	1.56×10^{-15}	700,852	0.010
12:49399132	rs1126930	*PRKAG1*	C	G	p.Thr38Ser	3.22	0.034	0.005	3.98×10^{-12}	712,354	0.007
12:72179446	rs61754230	*RAB21*[a]	T	C	p.Ser224Phe	1.74	0.040	0.007	1.33×10^{-9}	693,373	0.005
12:117977550	rs56214831	*KSR2*	T	C	p.Arg525Gln	0.82	0.057	0.010	1.08×10^{-8}	655,049	0.005
12:123345509	rs34149579	*HIP1R*	T	G	p.Cys938Phe	4.54	−0.032	0.004	2.00×10^{-14}	716,253	0.009
16:72830539	rs62051555	*ZFHX3*[a]	G	C	p.Gln1100His	4.34	−0.024	0.004	4.01×10^{-8}	690,637	0.005
18:58039478	rs13447324	*MC4R*	T	G	p.Tyr35Ter	0.01	0.542	0.086	2.26×10^{-10}	631,683	0.006
19:46178020	rs139215588	*GIPR*	A	G	p.Arg190Gln	0.11	−0.148	0.028	1.25×10^{-7}	695,800	0.005
19:46180976	rs143430880	*GIPR*	G	A	p.Glu288Gly	0.13	−0.153	0.028	2.96×10^{-8}	599,574	0.006
20:25195509	rs6050446	*ENTPD6*[a]	A	G	p.Lys185Glu	2.71	−0.034	0.005	2.40×10^{-10}	717,084	0.006
All ancestry sex specific additive (women only)											
19:3813906	rs45465594	*ZFR2*[a]	C	A	p.Ile718Met	2.55	−0.040	0.008	1.94×10^{-7}	373,848	0.008
European ancestry only additive											
9:97062981	rs12236219	*ZNF169*[a]	T	C	p.Arg381Cys	4.23	−0.029	0.005	8.78×10^{-10}	612,396	0.007

Array-wide significance is defined as $P < 2 \times 10^{-7}$. Variant positions are reported according to Build 37, and alleles are coded according to the positive strand. Alleles (effect/other), effect allele frequency (EAF), β value, standard error (SE) and P value are based on the meta-analysis of discovery-stage (GIANT) and validation-stage (deCODE, UK Biobank) studies. The effect allele is always the minor allele. Effects (β) are expressed in s.d., assuming mean = 0 and s.d. = 1. The amino acid change from the most abundant coding transcript is shown in this table (see Supplementary Table 24 for more details on protein annotation based on the VEP tool and transcript abundance from the GTEx database).[a] New gene (not previously implicated in human obesity).

(21.2% and 20.1%, respectively) than non-carriers (25.1% and 25%). For the *MC4R* p.Tyr35Ter and *KSR2* p.Arg525Gln variants, the prevalence of obesity for carriers (30% and 25.7%, respectively) and non-carriers (25.1% and 25.3%) was not significantly different.

We examined whether rare and low-frequency SNVs affected obesity risk early in life by combining data from three case–control studies of childhood obesity ($N_{cases} = 4{,}395$; $N_{controls} = 13{,}072$) (Methods and Supplementary Table 14). Associations for 10 of the 13 SNVs were directionally consistent with those observed for BMI in adults (77%, $P_{binomial} = 0.046$), with 3 of the variants (*ZBTB7B*, *PRKAG1* and *RAB21*) reaching nominal significance ($P < 0.05$). Although no carriers of the *MC4R* mutations were available for analyses, the role of *MC4R* in body weight regulation in childhood was established almost two decades ago[17,19,21].

Impact of rare and low-frequency SNVs on cardiometabolic and other traits. To examine whether the identified SNVs affect other traits, we obtained results from multiple large-scale genetic consortia (GIANT[15], MAGIC, GoT2D/T2D-GENES[16], GLGC, ICBP[22] and REPROGEN[23]) (Supplementary Fig. 9 and Supplementary Table 15) and performed phenome-wide association study (PheWAS) analyses using electronic medical record (EMR) data from BioVu and UK Biobank (Methods and Supplementary Table 16). The BMI-increasing allele of the *ZBTB7B* SNV encoding p.Pro190Ser was associated with greater height, and those for the *PRKAG1*, *ACHE* and *RAPGEF3* SNVs were associated with shorter height, but association with other traits differed. Specifically, carriers of the serine-encoding allele for *PRKAG1* p.Thr38Ser appeared heavier and shorter and had lower HDL cholesterol levels, earlier age at menarche (reported before[23]) and higher systolic blood pressure, in agreement with PheWAS analyses showing an increased risk of 'malignant essential hypertension' and 'hypertension' (Supplementary Table 16). While carriers of the proline-encoding allele for *RAPGEF3* p.Leu300Pro were also heavier and shorter, they had a lower BMI-adjusted waist-to-hip ratio (WHR_{adjBMI}) and lower fasting insulin levels (Supplementary Table 15), consistent with PheWAS results showing lower odds of 'secondary diabetes mellitus' (Supplementary Table 16). Thus, while all SNVs were associated with BMI, their patterns of association with other traits suggest that they may affect different physiological pathways.

Gene set enrichment analyses. To test whether the BMI-associated rare and low-frequency variants implicate biological pathways, we performed gene set enrichment analyses. Similarly to our previous analysis of GWAS for BMI[5], we analyzed coding variants that reached $P < 5 \times 10^{-4}$, using a DEPICT version adapted for exome array analysis[15] (Methods and Supplementary Note). We used 50 rare and low-frequency coding variants as input (all $P < 5 \times 10^{-4}$; Methods) and observed significant enrichment (Fig. 2, Supplementary Fig. 10a and Supplementary Table 17). Many of the enriched gene sets were related to neuronal processes, such as neurotransmitter release and synaptic function (for example, glutamate receptor activity, regulation of neurotransmitter levels and synapse part), consistent with previous findings from GWAS[5]. When we excluded variants near (±1 Mb with respect to) previously identified GWAS loci, we still observed 29 significantly enriched gene sets (in 12 meta-gene sets) (Supplementary Fig. 10b and Supplementary Table 18), thereby providing independent confirmation of the GWAS gene set enrichment results. In addition to neuronal-related gene sets, the analyses with rare and low-frequency coding variants newly identified a cluster of metabolic pathways related to insulin action and adipocyte/lipid metabolism (for example, enhanced lipolysis, abnormal lipid homeostasis and increased circulating insulin level; Fig. 2). Finally, we observed that rare and low-frequency BMI-associated coding variants were more effective at identifying enriched gene sets than common coding variants. Specifically, adding 192 common

Box 1 | Brief description of the 13 genes (in alphabetical order) identified

***ACHE* (acetylcholinesterase)**. In GTEx, *ACHE* is mainly expressed in brain and muscle[74]. Its encoded protein hydrolyzes acetylcholine (Ach) at brain cholinergic synapses and neuromuscular junctions and thus terminates signal transmission[65]. Knockout mice show a reduction in expression of muscarinic Ach receptors in brain regions associated with learning and memory and show lower ability to initiate the signaling cascade[75]. This gene has fewer missense variants than expected and is highly intolerant to loss-of-function mutations[51].

***ENTPD6* (ectonucleoside triphosphate diphosphohydrolase 6)**. Previously known as IL-6 signal transducer-2, this gene encodes a protein similar to the E-type nucleotidases that participate in purine and pyrimidine metabolism, calcium ion binding, hydrolase activity, magnesium ion binding and nucleoside diphosphatase activity[76]. In GTEx, it is widely expressed in many different tissues, in particular in the brain[74].

***GIPR* (gastric inhibitory polypeptide receptor)**. *GIPR* encodes a G-protein-coupled receptor for gastric inhibitory polypeptide, which is secreted by intestinal K cells after food ingestion[58]. GIPR activation stimulates insulin secretion from pancreatic beta cells and mediates fat deposition by increasing lipoprotein lipase activity, lipogenesis, and fatty acid and glucose uptake in adipocytes. In GTEx, GIPR is mostly expressed in Epstein–Barr virus (EBV)-transformed lymphocytes, stomach and visceral adipose tissue[74].

***HIP1R* (huntingtin-interacting protein 1 related)**. HIP1R is a multidomain protein that promotes actin binding and cell survival and interacts with CLTB and HIP1 (GeneCards). HIP1 and HIP1R appear to have central roles in clathrin-coated vesicle formation and intracellular membrane trafficking by promoting transient interaction between actin filaments and the endocytic machinery[77,78]. In GTEx, HIP1R is most expressed in stomach tissue, brain (substantia nigra, spinal cord, hippocampus) and sun-exposed skin[74].

***KSR2* (kinase suppressor of ras 2)**. KSR2 is an intracellular protein that functions as a molecular scaffold to regulate the MAP kinases ERK1 and ERK2 and determine cell fate. KSR2 also regulates AMPK activity controlling cellular thermogenesis, fat oxidation and glucose metabolism[18,60,61]. Knockout mouse models and human mutations have demonstrated a link with obesity risk[61]. KSR2 is almost exclusively expressed in the brain. It has fewer missense variants than expected and is highly intolerant to loss-of-function mutations[51].

***MC4R* (melanocortin 4 receptor)**. MC4R is a seven-transmembrane-domain G-protein-coupled receptor, predominantly expressed in the brain[74]. MC4R has been known to have a key role in body weight regulation for more than 20 years. Activation of MC4R by α-MSH, a POMC-derived peptide, suppresses food intake; MC4R antagonists increase food intake, and *MC4R* deficiency in human and rodent models results in hyperphagia and severe and early-onset obesity[79]. More than 150 *MC4R* mutations have been identified in individuals with severe, early-onset obesity[79], many of which lead to complete or partial loss of function[50,80]. Up to 6% of individuals with severe, early-onset obesity carry pathogenic mutations in *MC4R*, making *MC4R* deficiency the most common form of monogenic obesity[50,80,81].

***PRKAG1* (protein kinase AMP–activated non-catalytic subunit γ1)**. The protein encoded by *PRKAG1* is one of the γ regulatory subunits of AMPK, which is an important energy-sensing enzyme that monitors cellular energy status[33]. AMPK and PRKAG1 are ubiquitously expressed[74]. In the hypothalamus, AMPK influences food intake, energy expenditure and glucose homeostasis[35]. Muscle-specific overexpression of the AMPK γ1 subunit in mice results in increased food intake but does not affect body weight, presumably through a compensatory increase in energy expenditure[82].

***RAB21* (member *RAS* oncogene family)**. RAB21 belongs to the Rab family of monomeric GTPases involved in the control of cellular membrane trafficking. The encoded protein is widely expressed[74], has a role in the targeted trafficking of integrins and is involved in the regulation of cell adhesion and migration[83]. *RAB21* is thought to be intolerant to loss-of-function mutations[51].

***RAPGEF3* (rap guanine nucleotide exchange factor 3; also known as *EPAC1*)**. *RAPGEF3* encodes the exchange protein directly activated by cAMP isoform 1 (EPAC1), one of two cAMP sensors that are involved in numerous intracellular-cAMP-mediated functions[84]. EPAC1 is ubiquitously expressed[74], and insights from mouse knockout models suggest a role in energy homeostasis and the development of obesity and diabetes through the regulation of leptin and insulin signaling[30,84].

***ZFR2* (zinc-finger RNA-binding protein 2)**. The biological function of the gene product is as yet undetermined. Gene Ontology (GO) annotations related to this gene include 'nucleic acid binding'. It may have a role in dendritic branching and axon guidance[85,86]. ZFR2 is predominantly expressed in the brain[74].

***ZBTB7B* (zinc-finger- and BTB-domain-containing 7B; also known as *ThPOK*)**. ZBTB7B is a transcription factor regulating T cell fate in the thymus, particularly as the master regulator of $CD4^+$ lineage commitment[87]. It is a repressor of type 1 collagen gene expression[88]. This gene is mainly expressed in T cell lineages, skin and gastrointestinal tissues. *ZBTB7B* is thought to be intolerant to loss-of-function mutations[51].

***ZFHX3* (zinc-finger homeobox 3)**. *ZFHX3* encodes a transcription factor with multiple homeodomains and zinc-finger motifs and has a role in the cell cycle and in myogenic and neuronal differentiation. This gene is a tumor suppressor[89] that influences circadian rhythms[90,91] and sleep[91]. It may also contribute to the genesis of atrial fibrillation[92]. In GTEx, *ZFHX3* is highly expressed in arterial tissue as well as other tissues[74]. The *ZFHX3* gene is highly intolerant to loss-of-function mutations[51].

***ZNF169* (zinc-finger protein 169)**. The biological function of the gene product is as yet unclear. GO annotations suggest that *ZNF169* is involved in 'nucleic acid binding' and 'transcriptional regulation'. This gene is ubiquitously expressed[74].

More details and references are given in Supplementary Table 23.

coding SNVs (all $P < 5 \times 10^{-4}$) to the analysis decreased the number of enriched gene sets from the 471 (106 meta-gene sets) seen with rare and low-frequency coding SNVs to 62 (24 meta-gene sets) (Supplementary Fig. 10c and Supplementary Table 19). We observed fewer significant gene sets with the combined common and rare/low-frequency analysis, despite including more total coding variants and a higher fraction of array-wide significant coding variants. One possible explanation for this is that rare and low-frequency coding variants may fall in the causal gene more often than common coding variants, which suggests that the rare and low-frequency variants are more likely to be causal rather than simply in LD with causal variants.

We also used gene set enrichment analysis to prioritize candidate genes. Among the genes with rare and low-frequency coding

Table 2 | Genes significantly associated with BMI in gene-based meta-analyses, aggregating rare and low-frequency coding SNVs

Gene	Location of longest coding transcript	Test	*N* variants	*P* value	Conditioned *P* value[a]	Single variant	
						Top variant	*P* value
All ancestry, sex combined							
SLC6A17	1:110693132-110744823	SKAT	13	2.73×10^{-7}	0.13	rs41313405	4.45×10^{-7}
RAPGEF3	12:48128453-48152889	SKAT	19	8.91×10^{-15}	0.20	rs145878042	5.16×10^{-14}
PRKAG1	12:49396055-49412629	SKAT	4	2.75×10^{-12}	0.53	rs1126930	2.63×10^{-12}
RAB21	12:72148643-72187256	SKAT	5	4.81×10^{-8}	0.27	rs61754230	4.96×10^{-8}
KSR2	12:117890817-118406028	SKAT	7	7.15×10^{-9}	0.19	rs56214831	4.59×10^{-8}
MAP1A	15:43809806-43823818	SKAT	25	9.42×10^{-7}	0.16	rs55707100	1.01×10^{-6}
MC4R	18:58038564-58040001	VT	4	3.72×10^{-9}	0.01	rs13447325	2.97×10^{-11}
GIPR	19:46171502-46186982	VT	10	8.24×10^{-9}	1.12×10^{-4}	rs143430880	5.76×10^{-6}
All ancestry, sex specific							
ALDH3A1 (men only)	17:19641298-19651746	SKAT	15	3.24×10^{-7}	0.003	rs142078447	8.62×10^{-6}
ZFR2 (women only)	19:3804022-3869027	SKAT	19	1.81×10^{-7}	0.82	rs45465594	3.64×10^{-7}
European, sex combined							
ACHE	7:100487615-100493592	SKAT	6	3.30×10^{-10}	0.12	rs386545548	7.22×10^{-10}
European, sex specific							
ANGPTL7 (men only)	1:11249346-11256038	VT	3	2.50×10^{-6}	0.008	rs202182115	2.56×10^{-5}
ZNF169 (women only)	9:97021548-97064111	SKAT	9	1.89×10^{-7}	0.24	rs12236219	1.06×10^{-6}

Array-wide significant gene-based association is defined as $P < 2.5\times10^{-6}$. *P* values are based on the meta-analysis of discovery-stage studies.Gene-based analyses were performed with SKAT and VT; results shown are from the test (SKAT or VT) for which the significance exceeded $P < 2.5\times10^{-6}$. Only results using the broad SNV inclusion criteria reached array-wide significance. Transcript positions are reported according to Build 37 for the longest coding transcript supported by RefSeq (as displayed in the USCS Genome Browser). [a]*P* value after conditioning on the most significant (top) single variant aggregated in the gene-based test.

variants associated with BMI at $P < 5\times10^{-4}$, a subset of genes were prominently represented in the CNS-related enriched gene sets (Fig. 2) and are proposed to influence neurotransmission and/or synaptic organization, function and plasticity. These included genes in regions with suggestive evidence of association from GWAS (*CARTPT*, *MAP1A* and *ERC2*) and genes in regions not previously implicated by GWAS (*CALY*, *ACHE*, *PTPRD* and *GRIN2A*). The non-neuronal metabolic gene sets implicated two genes (*CIDEA* and *ADH1B*) that are markers of brown or 'beige' adipose tissue[24,25], providing new supporting evidence for a causal role in body weight regulation of this aspect of adipocyte biology.

***Drosophila* results.** To test for potential adiposity-driving effects of gene regulation, we performed tissue-specific RNA interference (RNAi) knockdown experiments in *Drosophila melanogaster*. We generated adipose-tissue-specific (*cg*-Gal4) or neuronal-specific (*elav*-Gal4) RNA interference (RNAi) knockdown crosses for 9 of the 13 candidate genes for which fly orthologs exist (Supplementary Table 20) and performed whole-body triglyceride analysis in male young adult flies. Triglycerides, the major lipid storage form in animals, were chosen as a direct measure of fly adiposity. Both neuronal and fat-body knockdown of *zfh2*, the ortholog of *ZFHX3*, resulted in significantly increased triglyceride levels. Adipose-tissue-specific, but not neuronal, knockdown of *epac* (*RAPGEF3*) was lethal. Tissue-specific loss of function of the other seven genes tested did not affect triglyceride levels.

Rare and low-frequency coding SNVs in genes implicated in monogenic and syndromic obesity. We identified 39 genes in the literature that have been convincingly implicated in monogenic obesity or syndromes for which obesity is one of the main features (Supplementary Fig. 11 and Supplementary Tables 21 and 22). Of the 652 rare and low-frequency SNVs in these 39 genes in monogenic and/or syndromic obesity, 5 SNVs were significantly associated with BMI (Bonferroni-corrected $P = 7.7\times10^{-5}$ (= 0.05/652)). Besides the SNVs in *MC4R* (p.Tyr35Ter and p.Asp37Val) and *KSR2* (p.Arg525Gln) already highlighted in the single-variant analyses, we identified an additional SNV in *MC4R* (p.Ile251Leu) and one in *BDNF* (p.Glu6Lys) associated with BMI. *MC4R* p.Ile251Leu has previously been shown to protect against obesity[26], whereas *BDNF* p.Glu6Lys, which is independent of SNVs previously identified by GWAS ($r^2 = 0.01$; $D' = 1.0$)[5], has not been implicated in body weight regulation before. We examined whether the 652 rare and low-frequency SNVs were enriched for association with BMI as compared to rare and low-frequency coding SNVs in all other genes, but found no evidence to support this.

Discussion

In this meta-analysis of exome-targeted genotyping data, we identified 14 rare and low-frequency coding variants in 13 genes associated with BMI. Eight of these genes (*ACHE*, *ENTPD6*, *RAB21*, *RAPGEF3*, *ZBTB7B*, *ZFHX3*, *ZFR2* and *ZNF169*) have not previously been implicated in human obesity, but evidence from animal studies provides support for a role in energy metabolism for some of these, such as *ACHE*[27,28], *RAPGEF3*[29–32] and *PRKAG1*[33–38]. Others fall into established BMI GWAS loci (*PRKAG1–BCDIN3D*, *HIP1R–CLIP1*, *MC4R* and *GIPR–QPCTL*)[5] and/or were previously implicated in severe, early-onset obesity (*MC4R* and *KSR2*)[17–19] and, using this exome-targeted approach, we pinpoint rare and low-frequency variants in these loci that have a role in obesity in

the general population. Pathway analyses confirm a key role for neuronal processes and newly implicate adipocyte and energy expenditure biology.

Consistent with other polygenic traits[15,23,39–42], we show that large sample sizes are needed to identify associated rare and low-frequency variants. The observed effect sizes reflect the statistical power of our sample size and are particularly large for SNVs with MAF <0.05%. The existence of rare alleles with larger effects on BMI than have been observed for common alleles might reflect negative or stabilizing selection on the extremes of BMI. However, rare variants with smaller effects almost certainly exist, but larger samples will be needed to uncover these. Our study was limited to coding variants on the exome array; large-scale sequencing studies will be needed to test for variants not covered by exome arrays.

The strongest association was observed for a mutation in *MC4R* introducing a stop codon (p.Tyr35Ter, rs13447324, MAF = 0.01%), with carriers weighing on average 7 kg more than non-carriers. *MC4R* is widely expressed in the CNS and is an established key player in energy balance regulation[43,44]. Mouse and human studies already showed two decades ago that MC4R deficiency results in extreme obesity, mainly through increased food intake[45–48]. *MC4R* p.Tyr35Ter, which results in MC4R deficiency[49], was one of the first *MC4R* mutations discovered in individuals with monogenic obesity[17,19], in whom the mutation is >20 times more prevalent than in the general population[17,50–52]. Here we show that *MC4R* p.Tyr35Ter has a role outside the setting of early-onset and extreme obesity. Despite its large effect, penetrance is low, and this variant does not fit the model of a fully penetrant Mendelian variant.

While significantly associated rare and low-frequency coding variants are strong candidates for being causal, the strongest implication of causal genes is provided by association with multiple independent coding variants, as we demonstrate for *GIPR*. We identified two rare variants in *GIPR* (p.Arg190Gln, rs139215588, MAF = 0.11%; p.Glu288Gly, rs143430880, MAF = 0.13%) independently associated with lower BMI; carriers of either variant weigh ~2 kg less than non-carriers. Common variants in or near *GIPR* have been found to associate with lower BMI[53] and delayed glucose and insulin response to an oral glucose challenge[54]. However, the two rare variants influence BMI independently of these common ones and were not associated with type 2 diabetes or the glycemic traits tested. Rodent models have provided strong evidence of a role for GIPR in body weight regulation. *Gipr*-deficient mice are protected from diet-induced obesity[55] and have an increased resting metabolic rate[56]. Blocking GIP signaling using a vaccination approach in mice on a high-fat diet reduces weight gain, mainly through reduced fat accumulation, mediated by increased energy expenditure[57]. Manipulation of incretins (GIP and GLP1) and their receptors has complex effects on obesity and insulin secretion and action that may differ between humans and mice[58]. The human genetic data suggest that inhibition of GIPR signaling might represent a therapeutic approach for the treatment of obesity[59].

A fourth rare variant, in *KSR2* (p.Arg525Gln, rs56214831, MAF = 0.82%), increases body weight by ~740 g/allele. *KSR2* is another gene previously implicated in energy metabolism and obesity[18,60,61]. In a recent study, mutation carriers were hyperphagic, had a reduced basal metabolic rate and showed severe insulin resistance[18]. Consistent with human data, *Ksr2*$^{-/-}$ mice are obese and hyperphagic, and they have reduced energy expenditure[18,60–62]. KSR2 is almost exclusively expressed in the brain and interacts with multiple proteins[63], including AMP-activated protein kinase (AMPK), a key regulator of energy homeostasis[60,61]. Notably, *KSR2* is one of the first genes implicated in severe, early-onset obesity in which mutations affect not only food intake but also basal metabolic rate, and it is thought to act via neuronal effects[18] (Fig. 2).

Despite convincing evidence of association for these four rare variants in *MC4R*, *GIPR* and *KSR2*, the penetrance of these variants for obesity is low (Supplementary Table 13). This is consistent with the polygenic and multifactorial nature of obesity, where variants across a range of frequencies and effect sizes contribute to the phenotype in any one person. Despite the low predictive power of the variants, it remains possible that particular variants in any one person may contribute to different balances of underlying physiology and, hence, different responses to treatments. This was illustrated in two patients with monogenic obesity due to *POMC* mutations; these patients lacked the main activator of MC4R and were effectively treated with an MC4R agonist[64].

Of the coding variants in the newly identified genes, some have well-known connections with obesity. For example, *PRKAG1* encodes the γ1 subunit of AMPK, a critical cellular energy sensor[33]. In the hypothalamus, AMPK integrates hormonal and nutritional signals with neuronal networks to regulate food intake and whole-body energy metabolism[34–36]. Furthermore, hypothalamic AMPK is a key regulator of brown adipose tissue in mice[35,37,38]. The BMI-decreasing allele at the associated *PRKAG1* variant (p.Thr38Ser, rs1126930, MAF = 3.22%) has additional beneficial effects on blood pressure, providing additional genetic support for modulation of AMPK as an ongoing therapeutic avenue for treatment.

ACHE, in which p.His353Asn (rs1799805, MAF = 3.9%) is associated with increased BMI, is another candidate gene related to neuronal biology, involved in the signaling of acetylcholine at neuromuscular junction and brain cholinergic synapses[65,66]. Inhibitors of ACHE, used to treat moderate-to-severe Alzheimer's disease[67], result in weight loss in humans, and *Ache*-deficient mice have delayed weight gain[27,28]. However, these phenotypes may be indirect consequences of adverse gastrointestinal and neuromuscular effects, respectively[27,28,68,69].

Another low-frequency coding variant (p.Leu300Pro, rs145878042, MAF = 1.1%) is located in *RAPGEF3* and has strong effects on multiple other phenotypes. The BMI-increasing allele encoding Pro300 is associated with shorter height, lower WHR_{adjBMI} and lower insulin levels, suggesting that this variant has multiple physiological consequences. Data from animal models also suggest complex effects of RAPGEF3 on adipocyte biology, energy balance and glucose metabolism[29–32]. For example, in one study, mice with global deletion of *Rapgef3* on a high-fat diet were resistant to obesity owing to reduced food intake and had increased glucose tolerance[30]. However, in a similar study, *Rapgef3*$^{-/-}$ mice developed severe obesity, increased respiratory exchange ratio and impaired glucose tolerance[32]. Mice with adipose-tissue-specific knockout of *Rapgef3* on a high-fat diet are also more prone to obesity and show increased food intake, reduced energy expenditure, impaired glucose tolerance and reduced circulating leptin levels[70]. More research is needed to understand the consequences of *RAPGEF3* manipulation.

The remaining genes with significant associations—*ENTPD6*, *HIP1R*, *RAB21*, *ZFR2*, *ZBTB7* and *ZFHX3*—have no clear prior evidence for a role in energy homeostasis, and in-depth functional follow-up is needed to gain insight into how they affect body weight. Here we performed gene set enrichment analyses to better understand the biology implicated by our genetic data and confirm the importance of neuronal processes, in particular synaptic function and neurotransmitter release, providing an independent validation of previous GWAS findings[5]. The combination of gene set enrichment analyses and association analyses of coding variants also enables us to highlight candidate genes that are both within these gene sets and show association with BMI at rare and low-frequency coding variants. These include genes reaching array-wide significance (for example, *ACHE* and *ZFR2*) and others with clear prior evidence for a role in body weight regulation (for example, *CARTPT*[71]) but that had not been highlighted in our single-variant or gene-based association analyses. Of note, the enrichment signals were stronger with rare and low-frequency

coding variants only than with all coding variants, suggesting that these variants are more likely to be causal and may more often point directly to relevant genes, whereas common coding variants may more often be proxies for common noncoding variants that affect nearby genes.

In addition, our gene set enrichment analyses now provide supporting evidence for a role of non-neuronal mechanisms as well. Specifically, *CIDEA* and *ADH1B* are both strongly predicted to be members of enriched gene sets related to insulin action and adipocyte biology, and both are markers that distinguish brown from white fat depots in mice[24] and humans[25]. CIDEA is predominantly expressed in adipose tissue and is known as a key regulator of energy metabolism[24]. *Cidea*-deficient mice are resistant to diet-induced obesity with increased lipolysis and mitochondrial uncoupling[24]. The connection of *ADH1B* with obesity is less clear, but the gene is highly expressed in human adipocytes, has been implicated by gene expression analyses in obesity and insulin resistance, and functions early in a potentially relevant metabolic pathway (retinoid biosynthesis)[24,25,72,73]. Similar pathways were implicated by recent work dissecting the signal near *FTO*[13]. However, because SNV association signals at *ADH1B* and *CIDEA* did not reach array-wide significance, additional genetic analysis of the role of these genes in obesity would be warranted.

In summary, we performed association analyses of rare and low-frequency variants with BMI in >700,000 individuals and identified 14 associated variants in 13 genes, including 5 known and 8 new genes. Although each variant contributes little to BMI variation in the general population, the variants may have substantial impact on body weight at an individual level. Furthermore, prior literature for these genes and unbiased gene set enrichment analysis indicate a strong role for neurobiology in body weight regulation and also provide new support for a causal role for aspects of adipocyte biology. The identified genes provide potential targets that may lead to new and more precise approaches for the treatment of obesity, where there has been minimal innovation over the past 30 years[1].

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0011-x.

Received: 14 August 2017; Accepted: 15 November 2017;
Published online: 22 December 2017

References

1. Bray, G. A. & Ryan, D. H. Update on obesity pharmacotherapy. *Ann. NY Acad. Sci.* **1311**, 1–13 (2014).
2. Bray, G. A., Frühbeck, G., Ryan, D. H. & Wilding, J. P. Management of obesity. *Lancet* **387**, 1947–1956 (2016).
3. Monda, K. L. et al. A meta-analysis identifies new loci associated with body mass index in individuals of African ancestry. *Nat. Genet.* **45**, 690–696 (2013).
4. Wen, W. et al. Meta-analysis of genome-wide association studies in East Asian-ancestry populations identifies four new loci for body mass index. *Hum. Mol. Genet.* **23**, 5492–5504 (2014).
5. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).
6. Winkler, T. W. et al. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS Genet.* **11**, e1005378 (2015).
7. Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. *Nat. Genet.* **49**, 1458–1467 (2017).
8. van der Klaauw, A. A. & Farooqi, I. S. The hunger genes: pathways to obesity. *Cell* **161**, 119–132 (2015).
9. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
10. Stratigopoulos, G. et al. Hypomorphism of *Fto* and *Rpgrip1l* causes obesity in mice. *J. Clin. Invest.* **126**, 1897–1910 (2016).
11. Stratigopoulos, G., LeDuc, C. A., Cremona, M. L., Chung, W. K. & Leibel, R. L. Cut-like homeobox 1 (CUX1) regulates expression of the fat mass and obesity-associated and retinitis pigmentosa GTPase regulator–interacting protein-1-like (*RPGRIP1L*) genes and coordinates leptin receptor signaling. *J. Biol. Chem.* **286**, 2155–2170 (2011).
12. Stratigopoulos, G. et al. Hypomorphism for *RPGRIP1L*, a ciliary gene vicinal to the *FTO* locus, causes increased adiposity in mice. *Cell Metab.* **19**, 767–779 (2014).
13. Claussnitzer, M. et al. *FTO* obesity variant circuitry and adipocyte browning in humans. *N. Engl. J. Med.* **373**, 895–907 (2015).
14. Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).
15. Marouli, E. et al. Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).
16. Fuchsberger, C. et al. The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47 (2016).
17. Sina, M. et al. Phenotypes in three pedigrees with autosomal dominant obesity caused by haploinsufficiency mutations in the melanocortin-4 receptor gene. *Am. J. Hum. Genet.* **65**, 1501–1507 (1999).
18. Pearce, L. R. et al. *KSR2* mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell* **155**, 765–777 (2013).
19. Hinney, A. et al. Several mutations in the melanocortin-4 receptor gene including a nonsense and a frameshift mutation associated with dominantly inherited obesity in humans. *J. Clin. Endocrinol. Metab.* **84**, 1483–1486 (1999).
20. Purcell, S. M. et al. A polygenic burden of rare disruptive mutations in schizophrenia. *Nature* **506**, 185–190 (2014).
21. van den Berg, L. et al. Melanocortin-4 receptor gene mutations in a Dutch cohort of obese children. *Obesity* **19**, 604–611 (2011).
22. Surendran, P. et al. Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
23. Lunetta, K. L. et al. Rare coding variants and X-linked loci associated with age at menarche. *Nat. Commun.* **6**, 7756 (2015).
24. Zhou, Z. et al. *Cidea*-deficient mice have lean phenotype and are resistant to obesity. *Nat. Genet.* **35**, 49–56 (2003).
25. Tews, D. et al. Comparative gene array analysis of progenitor cells from human paired deep neck and subcutaneous adipose tissue. *Mol. Cell. Endocrinol.* **395**, 41–50 (2014).
26. Stutzmann, F. et al. Non-synonymous polymorphisms in melanocortin-4 receptor protect against obesity: the two facets of a Janus obesity gene. *Hum. Mol. Genet.* **16**, 1837–1844 (2007).
27. Lin, H. Q., Wang, Y., Chan, K. L., Ip, T. M. & Wan, C. C. Differential regulation of lipid metabolism genes in the brain of acetylcholinesterase knockout mice. *J. Mol. Neurosci.* **53**, 397–408 (2014).
28. Vignaud, A. et al. Genetic ablation of acetylcholinesterase alters muscle function in mice. *Chem. Biol. Interact.* **175**, 129–130 (2008).
29. Ji, Z., Mei, F. C. & Cheng, X. Epac, not PKA catalytic subunit, is required for 3T3-L1 preadipocyte differentiation. *Front. Biosci.* **2**, 392–398 (2010).
30. Yan, J. et al. Enhanced leptin sensitivity, reduced adiposity, and improved glucose homeostasis in mice lacking exchange protein directly activated by cyclic AMP isoform 1. *Mol. Cell. Biol.* **33**, 918–926 (2013).
31. Almahariq, M., Mei, F. C. & Cheng, X. Cyclic AMP sensor EPAC proteins and energy homeostasis. *Trends Endocrinol. Metab.* **25**, 60–71 (2014).
32. Kai, A. K. et al. Exchange protein activated by cAMP 1 (*Epac1*)-deficient mice develop β-cell dysfunction and metabolic syndrome. *FASEB J.* **27**, 4122–4135 (2013).
33. Hardie, D. G., Ross, F. A. & Hawley, S. A. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat. Rev. Mol. Cell Biol.* **13**, 251–262 (2012).
34. Hardie, D. G. & Ashford, M. L. AMPK: regulating energy balance at the cellular and whole body levels. *Physiology* **29**, 99–107 (2014).
35. López, M., Nogueiras, R., Tena-Sempere, M. & Diéguez, C. Hypothalamic AMPK: a canonical regulator of whole-body energy balance. *Nat. Rev. Endocrinol.* **12**, 421–432 (2016).
36. Minokoshi, Y. et al. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569–574 (2004).
37. Viollet, B. et al. The AMP-activated protein kinase α2 catalytic subunit controls whole-body insulin sensitivity. *J. Clin. Invest.* **111**, 91–98 (2003).
38. Xue, B. et al. Neuronal protein tyrosine phosphatase 1B deficiency results in inhibition of hypothalamic AMPK and isoform-specific activation of AMPK in peripheral tissues. *Mol. Cell. Biol.* **29**, 4563–4573 (2009).
39. Warren, H. R. et al. Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nat. Genet.* **49**, 403–415 (2017).
40. Chami, N. et al. Exome genotyping identifies pleiotropic variants associated with red blood cell traits. *Am. J. Hum. Genet.* **99**, 8–21 (2016).

41. Li, M. et al. *SOS2* and *ACP1* loci identified through large-scale exome chip analysis regulate kidney development and function. *J. Am. Soc. Nephrol.* **28**, 981–994 (2017).
42. Liu, C. et al. Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci. *Nat. Genet.* **48**, 1162–1170 (2016).
43. Schwartz, M. W., Woods, S. C., Porte, D. Jr, Seeley, R. J. & Baskin, D. G. Central nervous system control of food intake. *Nature* **404**, 661–671 (2000).
44. Garfield, A. S. et al. A neural basis for melanocortin-4 receptor–regulated appetite. *Nat. Neurosci.* **18**, 863–871 (2015).
45. Huszar, D. et al. Targeted disruption of the melanocortin-4 receptor results in obesity in mice. *Cell* **88**, 131–141 (1997).
46. Fan, W., Boston, B. A., Kesterson, R. A., Hruby, V. J. & Cone, R. D. Role of melanocortinergic neurons in feeding and the agouti obesity syndrome. *Nature* **385**, 165–168 (1997).
47. Yeo, G. S. H. et al. A frameshift mutation in *MC4R* associated with dominantly inherited human obesity. *Nat. Genet.* **20**, 111–112 (1998).
48. Vaisse, C., Clement, K., Guy-Grand, B. & Froguel, P. A frameshift mutation in human *MC4R* is associated with a dominant form of obesity. *Nat. Genet.* **20**, 113–114 (1998).
49. Lubrano-Berthelier, C. et al. Melanocortin 4 receptor mutations in a large cohort of severely obese adults: prevalence, functional classification, genotype–phenotype relationship, and lack of association with binge eating. *J. Clin. Endocrinol. Metab.* **91**, 1811–1818 (2006).
50. Farooqi, I. S. et al. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. *N. Engl. J. Med.* **348**, 1085–1095 (2003).
51. Lek, M. et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
52. Hinney, A. et al. Melanocortin-4 receptor gene: case–control study and transmission disequilibrium test confirm that functionally relevant mutations are compatible with a major gene effect for extreme obesity. *J. Clin. Endocrinol. Metab.* **88**, 4258–4267 (2003).
53. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).
54. Saxena, R. et al. Genetic variation in *GIPR* influences the glucose and insulin responses to an oral glucose challenge. *Nat. Genet.* **42**, 142–148 (2010).
55. Miyawaki, K. et al. Inhibition of gastric inhibitory polypeptide signaling prevents obesity. *Nat. Med.* **8**, 738–742 (2002).
56. Hansotia, T. et al. Extrapancreatic incretin receptors modulate glucose homeostasis, body weight, and energy expenditure. *J. Clin. Invest.* **117**, 143–152 (2007).
57. Fulurija, A. et al. Vaccination against GIP for the treatment of obesity. *PLoS One* **3**, e3163 (2008).
58. Finan, B. et al. Reappraisal of GIP pharmacology for metabolic diseases. *Trends Mol. Med.* **22**, 359–376 (2016).
59. Irwin, N. & Flatt, P. R. Evidence for beneficial effects of compromised gastric inhibitory polypeptide action in obesity-related diabetes and possible therapeutic implications. *Diabetologia* **52**, 1724–1731 (2009).
60. Revelli, J. P. et al. Profound obesity secondary to hyperphagia in mice lacking kinase suppressor of ras 2. *Obesity* **19**, 1010–1018 (2011).
61. Costanzo-Garvey, D. L. et al. KSR2 is an essential regulator of AMP kinase, energy expenditure, and insulin sensitivity. *Cell Metab.* **10**, 366–378 (2009).
62. Brommage, R. et al. High-throughput screening of mouse knockout lines identifies true lean and obese phenotypes. *Obesity* **16**, 2362–2367 (2008).
63. Liu, L. et al. Proteomic characterization of the dynamic KSR-2 interactome, a signaling scaffold complex in MAPK pathway. *Biochim. Biophys. Acta* **1794**, 1485–1495 (2009).
64. Kühnen, P. et al. Proopiomelanocortin deficiency treated with a melanocortin-4 receptor agonist. *N. Engl. J. Med.* **375**, 240–246 (2016).
65. Xiang, Y. Y., Dong, H., Yang, B. B., Macdonald, J. F. & Lu, W. Y. Interaction of acetylcholinesterase with neurexin-1β regulates glutamatergic synaptic stability in hippocampal neurons. *Mol. Brain* **7**, 15 (2014).
66. Bartels, C. F., Zelinski, T. & Lockridge, O. Mutation at codon 322 in the human acetylcholinesterase (*ACHE*) gene accounts for YT blood group polymorphism. *Am. J. Hum. Genet.* **52**, 928–936 (1993).
67. Farlow, M. R. et al. Effectiveness and tolerability of high-dose (23 mg/d) versus standard-dose (10 mg/d) donepezil in moderate to severe Alzheimer's disease: a 24-week, randomized, double-blind study. *Clin. Ther.* **32**, 1234–1251 (2010).
68. Farlow, M. et al. Safety and tolerability of donepezil 23 mg in moderate to severe Alzheimer's disease. *BMC Neurol.* **11**, 57 (2011).
69. Tariot, P., Salloway, S., Yardley, J., Mackell, J. & Moline, M. Long-term safety and tolerability of donepezil 23 mg in patients with moderate to severe Alzheimer's disease. *BMC Res. Notes* **5**, 283 (2012).
70. Hu, Y. et al. Role of exchange protein directly activated by cyclic AMP isoform 1 in energy homeostasis: regulation of leptin expression and secretion in white adipose tissue. *Mol. Cell. Biol.* **36**, 2440–2450 (2016).
71. Altarejos, J. Y. et al. The Creb1 coactivator Crtc1 is required for energy balance and fertility. *Nat. Med.* **14**, 1112–1117 (2008).
72. Winnier, D. A. et al. Transcriptomic identification of *ADH1B* as a novel candidate gene for obesity and insulin resistance in human adipose tissue in Mexican Americans from the Veterans Administration Genetic Epidemiology Study (VAGES). *PLoS One* **10**, e0119941 (2015).
73. Molotkov, A., Deltour, L., Foglio, M. H., Cuenca, A. E. & Duester, G. Distinct retinoid metabolic functions for alcohol dehydrogenase genes *Adh1* and *Adh4* in protection against vitamin A toxicity or deficiency revealed in double null mutant mice. *J. Biol. Chem.* **277**, 13804–13811 (2002).
74. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
75. Volpicelli-Daley, L. A., Duysen, E. G., Lockridge, O. & Levey, A. I. Altered hippocampal muscarinic receptors in acetylcholinesterase-deficient mice. *Ann. Neurol.* **53**, 788–796 (2003).
76. Ivanenkov, V. V., Murphy-Piedmonte, D. M. & Kirley, T. L. Bacterial expression, characterization, and disulfide bond determination of soluble human NTPDase6 (CD39L2) nucleotidase: implications for structure and function. *Biochemistry* **42**, 11726–11735 (2003).
77. Jain, R. N. et al. Hip1r is expressed in gastric parietal cells and is required for tubulovesicle formation and cell survival in mice. *J. Clin. Invest.* **118**, 2459–2470 (2008).
78. Engqvist-Goldstein, A. E. et al. RNAi-mediated Hip1R silencing results in stable association between the endocytic machinery and the actin assembly machinery. *Mol. Biol. Cell* **15**, 1666–1679 (2004).
79. Tao, Y. X. The melanocortin-4 receptor: physiology, pharmacology, and pathophysiology. *Endocr. Rev.* **31**, 506–543 (2010).
80. Stutzmann, F. et al. Prevalence of melanocortin-4 receptor deficiency in Europeans and their age-dependent penetrance in multigenerational pedigrees. *Diabetes* **57**, 2511–2518 (2008).
81. Vaisse, C. et al. Melanocortin-4 receptor mutations are a frequent and heterogeneous cause of morbid obesity. *J. Clin. Invest.* **106**, 253–262 (2000).
82. Schönke, M., Myers, M. G. Jr., Zierath, J. R. & Björnholm, M. Skeletal muscle AMP-activated protein kinase $\gamma1^{H151R}$ overexpression enhances whole body energy homeostasis and insulin sensitivity. *Am. J. Physiol. Endocrinol. Metab.* **309**, E679–E690 (2015).
83. Pellinen, T. et al. Small GTPase Rab21 regulates cell adhesion and controls endosomal traffic of β1-integrins. *J. Cell Biol.* **173**, 767–780 (2006).
84. Banerjee, U. & Cheng, X. Exchange protein directly activated by cAMP encoded by the mammalian *rapgef3* gene: structure, function and therapeutics. *Gene* **570**, 157–167 (2015).
85. Rippey, C. et al. Formation of chimeric genes by copy-number variation as a mutational mechanism in schizophrenia. *Am. J. Hum. Genet.* **93**, 697–710 (2013).
86. Schmitz, C., Kinge, P. & Hutter, H. Axon guidance genes identified in a large-scale RNAi screen using the RNAi-hypersensitive *Caenorhabditis elegans* strain *nre-1(hd20) lin-15b(hd126)*. *Proc. Natl. Acad. Sci. USA* **104**, 834–839 (2007).
87. Setoguchi, R. et al. Repression of the transcription factor Th-POK by Runx complexes in cytotoxic T cell development. *Science* **319**, 822–825 (2008).
88. Widom, R. L., Culic, I., Lee, J. Y. & Korn, J. H. Cloning and characterization of hcKrox, a transcriptional regulator of extracellular matrix gene expression. *Gene* **198**, 407–420 (1997).
89. Sun, X. et al. Deletion of Atbf1/Zfhx3 in mouse prostate causes neoplastic lesions, likely by attenuation of membrane and secretory proteins and multiple signaling pathways. *Neoplasia* **16**, 377–389 (2014).
90. Parsons, M. J. et al. The regulatory factor ZFHX3 modifies circadian function in SCN via an AT motif–driven axis. *Cell* **162**, 607–621 (2015).
91. Balzani, E. et al. The Zfhx3-mediated axis regulates sleep and interval timing in mice. *Cell Rep.* **16**, 615–621 (2016).
92. Kao, Y. H. et al. ZFHX3 knockdown increases arrhythmogenesis and dysregulates calcium homeostasis in HL-1 atrial myocytes. *Int. J. Cardiol.* **210**, 85–92 (2016).

Acknowledgements

A.P.R. was supported by R01DK089256. A.W.H. is supported by an NHMRC Practitioner Fellowship (APP1103329). A.K.M. received funding from NIH/NIDDK K01DK107836. A.T.H. is a Wellcome Trust Senior Investigator (WT098395) and an NIH Research Senior Investigator. A.P.M. is a Wellcome Trust Senior Fellow in Basic Biomedical Science (WT098017). A.R.W. is supported by the European Research Council (SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC). A.U.J. is supported by the American Heart Association (13POST16500011) and the NIH (R01DK089256, R01DK101855, K99HL130580). B.K. and E.K.S. were supported by the Doris Duke Medical Foundation, the NIH (R01DK106621), the University of Michigan Internal Medicine Department, Division of Gastroenterology, the University of Michigan Biological Sciences Scholars Program and the Central Society for Clinical Research. C.J.W. is supported by the NIH (HL094535, HL109946). D.J.L. is supported by R01HG008983 and R21DA040177. D.R.W. is supported by the Danish Diabetes Academy, which is funded by the Novo

Nordisk Foundation. V. Salomaa has been supported by the Finnish Foundation for Cardiovascular Research. F.W.A. is supported by Dekker scholarship–Junior Staff Member 2014T001 Netherlands Heart Foundation and the UCL Hospitals NIHR Biomedical Research Centre. F.D. is supported by the UK MRC (MC_UU_12013/1-9). G.C.-P. received scholarship support from the University of Queensland and QIMR Berghofer. G.L. is funded by the Montreal Heart Institute Foundation and the Canada Research Chair program. H.Y. and T.M.F. are supported by the European Research Council (323195; SZ-245 50371-GLUCOSEGENES-FP7-IDEAS-ERC). I.M.H. is supported by BMBF (01ER1206) and BMBF (01ER1507m), the NIH and the Max Planck Society. J. Haessler was supported by NHLBI R21HL121422. J.N.H. is supported by NIH R01DK075787. K.E.N. was supported by the NIH (R01DK089256, R01HD057194, U01HG007416, R01DK101855) and the American Heart Association (13GRNT16490017). M.A.R. is supported by the Nuffield Department of Clinical Medicine Award, Clarendon Scholarship. M.I.M. is a Wellcome Trust Senior Investigator (WT098381) and an NIH Research Senior Investigator. M.D. is supported by the NCI (R25CA94880, P30CA008748). P.R.N. is supported by the European Research Council (AdG; 293574), the Research Council of Norway, the University of Bergen, the KG Jebsen Foundation and the Helse Vest, Norwegian Diabetes Association. P.T.E. is supported by the NIH (1R01HL092577, R01HL128914, K24HL105780), by an Established Investigator Award from the American Heart Association (13EIA14220013) and by the Foundation Leducq (14CVD01). P.L.A. was supported by NHLBI R21HL121422 and R01DK089256. P.L.H. is supported by the NIH (NS33335, HL57818). R.S.F. is supported by the NIH (T32GM096911). R.J.F.L. is supported by the NIH (R01DK110113, U01HG007417, R01DK101855, R01DK107786). S.A.L. is supported by the NIH (K23HL114724) and a Doris Duke Charitable Foundation Clinical Scientist Development Award. T.D.S. holds an ERC Advanced Principal Investigator award. T.A.M. is supported by an NHMRC Fellowship (APP1042255). T.H.P. received Lundbeck Foundation and Benzon Foundation support. V.T. is supported by a postdoctoral fellowship from the Canadian Institutes of Health Research (CIHR). Z.K. is supported by the Leenaards Foundation, the Swiss National Science Foundation (31003A-143914) and SystemsX.ch (51RTP0_151019). Part of this work was conducted using the UK Biobank resource (project numbers 1251 and 9072). A full list of acknowledgments appears in the Supplementary Note.

Author contributions

Writing group (wrote and edited the manuscript): A.E.J., A.E.L., C.M.L., C.S., G.L., H.M.H., J.N.H., K.E.N., K.L.Y., M.F.F., M. Graff, P.D., R.J.F.L., T.M.F., V.T., Y. Lu. **Data preparation group** (program development and quality control of data from contributing cohorts for meta-analyses): A.E.J., A.E.L., C.M.-G., C.S., D.J.L., E.M., H.M.H., I.B.B., K.L.Y., K.E.S., K.S.L., M.A.R., M.C.Y.N., M. Graff, N.G.D.M., P. Mudgal, R.J.F.L., S. Feng, S.M.W., S.S., S.V., T.A., T.E., T. Karaderi, T.W.W., V.T., X.Z., Y. Lu. **BMI meta-analyses** (discovery and follow-up, single variant and gene based): A.E.J., C.S., C.-T.L., D.J.L., H.M.H., I.B.B., J.N.H., K.L.Y., R.J.F.L., T.M.F., V.T., Y. Lu. **Childhood data** (analyses and interpretation): A.E.H., G.M., H.H., I. Barroso, I.S.F., J.P.B., S.F.A.G., V.M. Pleiotropy working group: A.M., C.J.W., C.M.L., D.J.L., E.M., F.D., G.A., G.M., G.M.P., H. Kitajima, H.M.H., J.C.F., J.P.C., J.R.B.P., J.W., K.S.R., M. Boehnke, M.I.M., P.B.M., P.D., R.J.F.L., S. Kathiresan, S.M.W., S.W., T.F.V., X.S. Phenome-wide association studies: A.G., A.M., J.C.D., L. Bastarache, M.I.M., T.L.E. Gene set enrichment analyses: D.L., J.N.H., R.S.F., S.B., T.H.P., Z.K. Monogenic and syndromic gene enrichment analyses: A.K.M., H.M.H. Fly obesity screen: A. Lempradl, J.A. Pospisilik. Oversight of contributing studies: A. Linneberg, A. Peters, A. Tönjes, A.C.H., A.D.M., A.G.U., A.I.d.H., A.J.L., A.M.D., A.P.M., A.P.R., A.S.B., A.T.H., A.W.H., B.B., B.G.N., C.A.B., C.C., C.E.P., C.H., C.J.W., C.L., C.M.L., C.N.A.P., D.F.E., D.F.R., D.I.C., D.M.W., D.O.M.-K., D.R.N., D.R.W., D.S., D.W.B., E.B., E.B.L., E.D.A., E.F., E.I., E.K.S., E.P.B., E.Z., F. Karpe, F. Kee, F. Renström, F.W.A., G. Dedoussis, G. Tromp, G.B., G.B.J., G.K.H., G.L., G.P., G.P.J., G.W.M., H. Kuivaniemi, H.B., H.D.W., H.H., H.-J.G., H.M.d.R., H.R.W., I. Barroso, I. Brandslund, I.B.B., I.F., I.J.D., I.M.H., I.R., I.S.F., J. Kaprio, J.C.C., J.-C.T., J.D., J.D.R., J.F., J.G.W., J.I.R., J.M.M.H., J.M.S., J.R.O'C., J.S.K., J.T., K. Stefansson, K. Strauch, K.E.N., K.K., K.K.A., K.L.M., K.M.H., K.N., K.R.O., K.T.Z., L.E.W., L.L., L.W., M. Blüher, M. Kähönen, M.A.I., M.A.R., M.B.S., M.C.H.d.G., M.d.H., M.E.J., M.F., M.H.B., M.I.M., M.L.O'D., M.M., M.P., M.-P.D., M.S., M.U., M.V., M.W., N.D.P., N.J.S., N.J.W., N.S., N.v.L., O. Pedersen, O. Polasek, O.T.R., P. Kovacs, P.A., P.A.P., P.B.M., P.D., P.E., P.G., P.G.-L., P.J.S., P.L.A., P.L.H., P.L.P., P.M.R., P.R.N., P.T.C., P.W.F., R.A.O., R.A.S., R.C., R.E.S., R.J.F.L., R.V., S. Fauser, S. Kathiresan, S.E.M., S.F.A.G., S.J., S.L.R.K., S.M., S.P., S.H.S., T.A.M., T.B.H., T.D.S., T.E., T.H., T.J., T.L., T.L.E., T.M.F., U.T., V. Gudnason, V. Salomaa, V.V., W.H.-H.S., X.G., X.L., Y. Liu. Genotyping of contributing studies: A. Loukola, A.T.-H., A.S.B., A.D'E., A.G.U., A.I.d.H., A.J.L., A.L.M., A.M., A.M.D., A.P.G., A.P.M., A.P.R., A.R.H., A.S.B., A.V., A.W.H., A.Y.C., B.G.N., C.A.B., C.E.P., C.H., C.J.P., C.K., C.L., C.M., C.M.-G., C.M.L., C.M.v.D., C.N.A.P., C.S., D.F.R., D.I.C., D.J.C., D.J.R., D.M.W., D.R.N., E.D.A., E.E., E.I., E.K., E.M.L., E.P.B., E.W.D., F. Karpe, F. Rivadeneira, F.S.C., G. Davies, G. Tromp, G.P.J., G.W.M., H. Kuivaniemi, H.H., H.L.G., H. Li, H.V., I.G., J. Kuusisto, J.A. Perry, J.B.-J., J.C.C., J.D., J.D.F., J.G.D., J.I.R., J.M.M.H., J.S.K., J.T., K. Strauch, K.D.T., K.E.S., K.M., L. Milani, L. Southam, L.A.L., L.A.K., L.M.Y.-A., L.-P.L., M. Benn, M. Boehnke, M. Gorski, M. Kähönen, M.B.S., M.C.H.d.G., M.F., M.H.B., M.I.M., M.L., M.L.B., M.L.G., M.L.O'D., M.M.-N., M.-P.D., M.P.S.-L., N.D.P., N.G., N.J.S., N.J.W., N.v.L., Ø.H., P.B.M., P.G.-L., P.I.W.d.B., P.T.C., P.W.F., R.A.O., R.A.S., R.E.S., R.F.-S., R.J.F.L., R.L.-G., R.R., R.Y., S. Kanoni, S. Kathiresan, S.C., S.F.A.G., S.F.N., S.H.V., S.L.R.K., S.S., S.W.v.d.L., T.A.L., T.B.H., T.E., T.H., T.L., U.V., V.V., Wei Zhao, X.L., Y. Lu, Y.-D.I.C., Y.H., Y. Liu, Y. Wang. Phenotyping of contributing studies: A.A.B., A.-E.F., A.F., A.J. Swift, A. Pattie, A. Peters, A.R.H., A. Robino, A.S.B., A.T.H., A. Tönjes, A.T.-H., A.U.J., A.V., A.W.H., B.B., B.G.N., B.H.T., B.K., C.A.B., C.E.L., C.E.P., C.H., C.J.P., C.K., C.M., C.M.-G., C.M.L., C.N.A.P., C.S., D.E., D.F.R., D.I.C., D.J.R., D.R.N., D.R.V.E., D.R.W., E.C., E.D.A., E.E., E.F., E.I., E.P.B., E.R.B.P., E.T., E.W.D., F. Karpe, F. Kee, F. Renström, F. Rivadeneira, F.W., G.B.J., G.L., G.P.J., G. Tromp, G.W.M., H.B., H.D.W., H.H., H.L.G., H. Li, I.J.D., I.R., J.C., J.C.C., J.D., J.D.F., J.F., J.-H.J., J. Kaprio, J. Lindström, J.M.M.H., J.M.S., J.P.B., J.S.K., K.E.N., K.-H.L., K.K., K.K.A., K.M.H., K.N., K.R.O., K.S.L., K.S.S., K.T.Z., L.A.K., L.E.B., L.E.W., L.L., L.M.Y.-A., L. Southam, L. Sun, L.W., M.A., M.A.I., M. Blüher, M. Brumat, M.C.H.d.G., M.F.F., M.I.M., M.J.C., M. Kähönen, M. Karaleftheri, M.L.B., M.M., M.M.-N., M.N., M.R., M.S., N.D.P., N.F., N.G., N.J.S., N.J.W., N.N., N.R.L., Ø.H., O.H.F., O.L.H., O. Polasek, O.R., O.T.R., P.A., P.G.-L., P. Komulainen, P. Kovacs, P.L.P., P.M.R., P. Mitchell, P.R.K., P.R.N., P.T.C., P.T.E., R.d.M., R.E.S., R.F.-S., R.M.-C., R.R., R.S.K., R.V., R.Y., S.A.L., S.E.M., S.F.A.G., S. Fauser, S.H.S., S.H.V., S.L.R.K., S.M., S.S., S.T., T.A.L., T.A.M., T.B.H., T.D.S., T.E.G., T.J., T.J.P., T.L., T.L.E., T.N.P., V. Giedraitis, V. Salomaa, V.T., W.H.-H.S., X.L., X.S., Y. Liu, Y. Lu. Data analysis of contributing studies: A.E.H., A.E.J., A.E.L., A.G., A.J.C., A.J. Slater, A. Lophatananon, A.M., A.P.M., A.P.P., A. Pirie, A.R.W., A. Rasheed, A. Robino, A.S.B., A. Teumer, A.V.S., A.Y.C., B.K., C.A.B., C.A.W., C.H., C.M.-G., C.P.N., C.S., C.T.H., C.-T.L., D.F.R., D.I.C., D.J.T., D.M.W., D.S.A., D.S.C., D.V., E.B.W., E.E., F. Rivadeneira, G.C.-P., G. Davies, G.L., G.M., G. Thorleifsson, G. Tromp, G.V., H. Li, H. Lin, H.M.S., H.P., H.R.W., H.T., H.Y., I.F., I.G., J.A. Perry, J.B.-J., J.C.C., J.C.G., J.E.H., J.G., J.G.D., J.H.Z., J. Haessler, J. Hernesniemi, J.I.R., J. Kuusisto, J. Li, J. Luan, J.M.M.H., J.P.B., J.P.T., J.R.O'C., J.S.K., J.v.S., J.W.J., J.Y., K.E.N., K.E.S., K.E.T., K.L.Y., K.M., K.S.L., L.-A.L., L.A.L., L. Broer, L.F.B., L.H., L.M.O.L., L.M.Y.-A., L. Moilanen, L.-P.L., L. Southam, M.A.N., M.C., M.C.H.d.G., M.C.Y.N., M.D., M.F., M.F.F., M. Gorski, M. Graff, M.L., M.P.S.-L., M.R., M.U., M.V., N.D.P., N.G., N.G.D.M., N.J.S., N.L.H.-C., N.R.R., N.v.L., Ø.H., P.L.A., P. Mudgal, P.S., P.Y., R.A.S., R.C., R.L.-G., R.U., R.Y., S.A.L., S.E.A., S.G., S.J., S.M., S.M.W., S.P., S.S., S.V., S.W.v.d.L., T.E., T. Karaderi, T. Korhonen, T.L.E., T.M.F., T.V.V., T.W.W., V.M., V. Steinthorsdottir, V.T., V.V., W.G., W. Zhang, W. Zhou, Wei Zhao, X.G., X.L., X.S., Y.H., Y.J., Y. Lu, Y.S., Y. Wu.

Competing interests

G. Thorleifsson, V. Steinthorsdottir, U.T. and K. Stefansson are employed by deCODE Genetics/Amgen, Inc.I. Barroso and spouse own stock in GlaxoSmithKline and Incyte, Ltd.S. Kathiresan has received grant support from Bayer Healthcare and Amarin; holds equity in San Therapeutics and Catabasis; and has received personal fees for participation in scientific advisory boards for Bayer Healthcare, Catabasis, Regeneron Genetics Center, Merck, Celera, Genomics PLC, Novartis, Sanofi, AstraZeneca, Alnylam, Eli Lilly Company, Leerink Partners, Noble Insights and Ionis Pharmaceuticals. A.Y.C., D.F.R. and T.F.V. are employees of Merck Sharp Dohme Corp., New Jersey, USA. D.I.C. receives genotyping and collaborative scientific support from Amgen and receives support for genetic analysis from AstraZeneca. P.M.R. receives genotyping and collaborative scientific support from Amgen and receives support for genetic analysis from AstraZeneca. M.J.C. is Chief Scientist for Genomics England, a UK government company.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-017-0011-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to J.N.H. or R.J.F.L.

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

Valérie Turcot[1], Yingchang Lu[2,3,4], Heather M. Highland[5,6], Claudia Schurmann[3,4], Anne E. Justice[5], Rebecca S. Fine[7,8,9], Jonathan P. Bradfield[10,11], Tõnu Esko[7,9,12], Ayush Giri[13], Mariaelisa Graff[5], Xiuqing Guo[14], Audrey E. Hendricks[15,16], Tugce Karaderi[17,18], Adelheid Lempradl[19], Adam E. Locke[20,21], Anubha Mahajan[17], Eirini Marouli[22], Suthesh Sivapalaratnam[23,24,25], Kristin L. Young[5], Tamuno Alfred[3], Mary F. Feitosa[26], Nicholas G. D. Masca[27,28], Alisa K. Manning[7,24,29,30], Carolina Medina-Gomez[31,32], Poorva Mudgal[33], Maggie C. Y. Ng[33,34], Alex P. Reiner[35,36], Sailaja Vedantam[7,8,9], Sara M. Willems[37], Thomas W. Winkler[38], Gonçalo Abecasis[20], Katja K. Aben[39,40], Dewan S. Alam[41], Sameer E. Alharthi[22,42], Matthew Allison[43], Philippe Amouyel[44,45,46], Folkert W. Asselbergs[47,48,49], Paul L. Auer[50], Beverley Balkau[51], Lia E. Bang[52], Inês Barroso[15,53,54], Lisa Bastarache[55], Marianne Benn[56,57], Sven Bergmann[58,59], Lawrence F. Bielak[60], Matthias Blüher[61,62], Michael Boehnke[20], Heiner Boeing[63], Eric Boerwinkle[64,65], Carsten A. Böger[66], Jette Bork-Jensen[67], Michiel L. Bots[68], Erwin P. Bottinger[3], Donald W. Bowden[33,34,69], Ivan Brandslund[70,71], Gerome Breen[72,73], Murray H. Brilliant[74], Linda Broer[32], Marco Brumat[75], Amber A. Burt[76], Adam S. Butterworth[77,78], Peter T. Campbell[79], Stefania Cappellani[80], David J. Carey[81], Eulalia Catamo[80], Mark J. Caulfield[22,82], John C. Chambers[83,84,85], Daniel I. Chasman[7,86,87,88], Yii-Der I. Chen[14], Rajiv Chowdhury[77], Cramer Christensen[89], Audrey Y. Chu[90,91], Massimiliano Cocca[92], Francis S. Collins[93], James P. Cook[94], Janie Corley[95,96], Jordi Corominas Galbany[97,98], Amanda J. Cox[33,34,99], David S. Crosslin[100], Gabriel Cuellar-Partida[101,102], Angela D'Eustacchio[80], John Danesh[15,77,78,103], Gail Davies[95,96], Paul I. W. Bakker[68,104], Mark C. H. Groot[105,106], Renée Mutsert[107], Ian J. Deary[95,96], George Dedoussis[108], Ellen W. Demerath[109], Martin Heijer[110], Anneke I. Hollander[98], Hester M. Ruijter[111], Joe G. Dennis[112], Josh C. Denny[55], Emanuele Angelantonio[77,78], Fotios Drenos[113,114], Mengmeng Du[36,115], Marie-Pierre Dubé[1,116], Alison M. Dunning[117], Douglas F. Easton[112,117], Todd L. Edwards[13], David Ellinghaus[118], Patrick T. Ellinor[7,24,30], Paul Elliott[119], Evangelos Evangelou[84,120], Aliki-Eleni Farmaki[108,121], I. Sadaf Farooqi[53,54], Jessica D. Faul[122], Sascha Fauser[123], Shuang Feng[20], Ele Ferrannini[124,125], Jean Ferrieres[126], Jose C. Florez[7,24,29,30], Ian Ford[127], Myriam Fornage[128], Oscar H. Franco[31], Andre Franke[118], Paul W. Franks[129,130,131], Nele Friedrich[132], Ruth Frikke-Schmidt[57,133], Tessel E. Galesloot[40], Wei Gan[17], Ilaria Gandin[134], Paolo Gasparini[75,80], Jane Gibson[135], Vilmantas Giedraitis[136], Anette P. Gjesing[67], Penny Gordon-Larsen[137,138], Mathias Gorski[38,66], Hans-Jörgen Grabe[139,140], Struan F. A. Grant[10,141,142], Niels Grarup[67], Helen L. Griffiths[143], Megan L. Grove[64], Vilmundur Gudnason[144,145], Stefan Gustafsson[146], Jeff Haessler[36], Hakon Hakonarson[10,141], Anke R. Hammerschlag[147], Torben Hansen[67], Kathleen Mullan Harris[137,148], Tamara B. Harris[149], Andrew T. Hattersley[150], Christian T. Have[67], Caroline Hayward[151], Liang He[152,153], Nancy L. Heard-Costa[90,154], Andrew C. Heath[155], Iris M. Heid[38,156], Øyvind Helgeland[157,158], Jussi Hernesniemi[159,160,161], Alex W. Hewitt[162,163,164], Oddgeir L. Holmen[165], G. Kees Hovingh[23], Joanna M. M. Howson[77], Yao Hu[166], Paul L. Huang[24], Jennifer E. Huffman[151], M. Arfan Ikram[31,167,168], Erik Ingelsson[146,169], Anne U. Jackson[20], Jan-Håkan Jansson[170,171], Gail P. Jarvik[76,172], Gorm B. Jensen[173], Yucheng Jia[14], Stefan Johansson[158,174], Marit E. Jørgensen[175,176], Torben Jørgensen[57,177,178], J. Wouter Jukema[179,180], Bratati Kahali[181,182,183,184], René S. Kahn[185], Mika Kähönen[186,187], Pia R. Kamstrup[56], Stavroula Kanoni[22], Jaakko Kaprio[153,188,189], Maria Karaleftheri[190], Sharon L. R. Kardia[60], Fredrik Karpe[191,192], Sekar Kathiresan[7,24,88], Frank Kee[193], Lambertus A. Kiemeney[40], Eric Kim[14], Hidetoshi Kitajima[17], Pirjo Komulainen[194,195,196], Jaspal S. Kooner[83,85,197,198], Charles Kooperberg[36], Tellervo Korhonen[153,189,199], Peter Kovacs[61], Helena Kuivaniemi[81,200,201], Zoltán Kutalik[59,202], Kari Kuulasmaa[189], Johanna Kuusisto[203], Markku Laakso[203], Timo A. Lakka[194,195,196], David Lamparter[58,59], Ethan M. Lange[204], Leslie A. Lange[204], Claudia Langenberg[37], Eric B. Larson[76,205,206], Nanette R. Lee[207,208], Terho Lehtimäki[160,161], Cora E. Lewis[209],

Huaixing Li[166], Jin Li[169], Ruifang Li-Gao[107], Honghuang Lin[210], Keng-Hung Lin[211], Li-An Lin[128], Xu Lin[166], Lars Lind[212], Jaana Lindström[189], Allan Linneberg[178,213,214], Ching-Ti Liu[215], Dajiang J. Liu[216], Yongmei Liu[217], Ken S. Lo[1], Artitaya Lophatananon[218], Andrew J. Lotery[143], Anu Loukola[153,188], Jian'an Luan[37], Steven A. Lubitz[7,24,30], Leo-Pekka Lyytikäinen[160,161], Satu Männistö[189], Gaëlle Marenne[15], Angela L. Mazul[5], Mark I. McCarthy[17,191,192], Roberta McKean-Cowdin[219], Sarah E. Medland[102], Karina Meidtner[220,221], Lili Milani[12], Vanisha Mistry[53,54], Paul Mitchell[222], Karen L. Mohlke[204], Leena Moilanen[223], Marie Moitry[224,225], Grant W. Montgomery[102,226], Dennis O. Mook-Kanamori[107,227], Carmel Moore[78,228], Trevor A. Mori[229], Andrew D. Morris[230], Andrew P. Morris[17,94], Martina Müller-Nurasyid[156,231,232], Patricia B. Munroe[22,82], Mike A. Nalls[233,234], Narisu Narisu[93], Christopher P. Nelson[27,28], Matt Neville[191,192], Sune F. Nielsen[56,57], Kjell Nikus[159], Pål R. Njølstad[157,158], Børge G. Nordestgaard[56,57], Dale R. Nyholt[102,235], Jeffrey R. O'Connel[236], Michelle L. O'Donoghue[237], Loes M. Olde Loohuis[238], Roel A. Ophoff[185,238], Katharine R. Owen[191,192], Chris J. Packard[127], Sandosh Padmanabhan[127], Colin N. A. Palmer[239], Nicholette D. Palmer[69], Gerard Pasterkamp[111,240], Aniruddh P. Patel[7,24,88], Alison Pattie[96], Oluf Pedersen[67], Peggy L. Peissig[74], Gina M. Peloso[215], Craig E. Pennell[241], Markus Perola[12,188,189,242], James A. Perry[236], John R. B. Perry[37], Tune H. Pers[67,243], Thomas N. Person[74], Annette Peters[221,232,244], Eva R. B. Petersen[245], Patricia A. Peyser[60], Ailith Pirie[117], Ozren Polasek[230,246], Tinca J. Polderman[147], Hannu Puolijoki[247], Olli T. Raitakari[248,249], Asif Rasheed[250], Rainer Rauramaa[194,195], Dermot F. Reilly[91], Frida Renström[129,251], Myriam Rheinberger[66], Paul M. Ridker[87,88,237], John D. Rioux[1,116], Manuel A. Rivas[7,252], David J. Roberts[78,253,254], Neil R. Robertson[17,191], Antonietta Robino[80], Olov Rolandsson[170,255], Igor Rudan[230], Katherine S. Ruth[256], Danish Saleheen[250,257], Veikko Salomaa[189], Nilesh J. Samani[27,28], Yadav Sapkota[102], Naveed Sattar[127], Robert E. Schoen[258], Pamela J. Schreiner[109], Matthias B. Schulze[220,221], Robert A. Scott[37], Marcelo P. Segura-Lepe[84], Svati H. Shah[259], Wayne H.-H. Sheu[260,261,262], Xueling Sim[20,263], Andrew J. Slater[264,265], Kerrin S. Small[266], Albert V. Smith[144,145], Lorraine Southam[15,17], Timothy D. Spector[266], Elizabeth K. Speliotes[181,182,183], John M. Starr[95,267], Kari Stefansson[144,268], Valgerdur Steinthorsdottir[268], Kathleen E. Stirrups[22,25], Konstantin Strauch[156,269], Heather M. Stringham[20], Michael Stumvoll[61,62], Liang Sun[152,153], Praveen Surendran[77], Amy J. Swift[93], Hayato Tada[237,270], Katherine E. Tansey[114,271], Jean-Claude Tardif[1], Kent D. Taylor[14], Alexander Teumer[272], Deborah J. Thompson[112], Gudmar Thorleifsson[268], Unnur Thorsteinsdottir[144,268], Betina H. Thuesen[178], Anke Tönjes[273], Gerard Tromp[81,201], Stella Trompet[179,274], Emmanouil Tsafantakis[275], Jaakko Tuomilehto[189,276,277,278], Anne Tybjaerg-Hansen[57,133], Jonathan P. Tyrer[117], Rudolf Uher[279], André G. Uitterlinden[31,32], Matti Uusitupa[280], Sander W. Laan[111], Cornelia M. Duijn[31], Nienke Leeuwen[281,282], Jessica van Setten[47], Mauno Vanhala[283,284], Anette Varbo[56,57], Tibor V. Varga[129], Rohit Varma[285], Digna R. Velez Edwards[286], Sita H. Vermeulen[40], Giovanni Veronesi[287], Henrik Vestergaard[67,176], Veronique Vitart[151], Thomas F. Vogt[288], Uwe Völker[289,290], Dragana Vuckovic[75,80], Lynne E. Wagenknecht[217], Mark Walker[291], Lars Wallentin[292], Feijie Wang[166], Carol A. Wang[241], Shuai Wang[215], Yiqin Wang[166], Erin B. Ware[60,122], Nicholas J. Wareham[37], Helen R. Warren[22,82], Dawn M. Waterworth[293], Jennifer Wessel[294], Harvey D. White[295], Cristen J. Willer[181,182,296], James G. Wilson[297], Daniel R. Witte[298,299], Andrew R. Wood[256], Ying Wu[204], Hanieh Yaghootkar[256], Jie Yao[14], Pang Yao[166], Laura M. Yerges-Armstrong[236,300], Robin Young[77,127], Eleftheria Zeggini[15], Xiaowei Zhan[301], Weihua Zhang[83,84], Jing Hua Zhao[37], Wei Zhao[60,257], Wei Zhou[181,182], Krina T Zondervan[17,302], CHD Exome+ Consortium[303], EPIC-CVD Consortium[303], ExomeBP Consortium[303], Global Lipids Genetic Consortium[303], GoT2D Genes Consortium[303], EPIC InterAct Consortium[303], INTERVAL Study[303], ReproGen Consortium[303], T2D-Genes Consortium[303],

The MAGIC Investigators[303], Understanding Society Scientific Group[303], Jerome I. Rotter[14], John A. Pospisilik[19], Fernando Rivadeneira[31,32], Ingrid B. Borecki[26], Panos Deloukas[22,42], Timothy M. Frayling[256], Guillaume Lettre[1,116], Kari E. North[304], Cecilia M. Lindgren[17,305], Joel N. Hirschhorn[7,8,9,306]* and Ruth J. F. Loos[3,4,307]*

[1]Montreal Heart Institute, Université de Montréal, Montreal, Quebec, Canada. [2]Division of Epidemiology, Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt Epidemiology Center, Vanderbilt University School of Medicine, Nashville, TN, USA. [3]Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [4]Genetics of Obesity and Related Metabolic Traits Program, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [5]Department of Epidemiology, University of North Carolina, Chapel Hill, NC, USA. [6]Human Genetics Center, University of Texas School of Public Health, University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, University of Texas Health Science Center at Houston, Houston, TX, USA. [7]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [8]Department of Genetics, Harvard Medical School, Boston, MA, USA. [9]Division of Endocrinology and Center for Basic and Translational Obesity Research, Boston Children's Hospital, Boston, MA, USA. [10]Center for Applied Genomics, Division of Human Genetics, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [11]Quantinuum Research, LLC, San Diego, CA, USA. [12]Estonian Genome Center, University of Tartu, Tartu, Estonia. [13]Division of Epidemiology, Department of Medicine, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA. [14]Institute for Translational Genomics and Population Sciences, LABioMed at Harbor-UCLA Medical Center, Torrance, CA, USA. [15]Wellcome Trust Sanger Institute, Hinxton, UK. [16]Department of Mathematical and Statistical Sciences, University of Colorado, Denver, CO, USA. [17]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [18]Department of Biological Sciences, Faculty of Arts and Sciences, Eastern Mediterranean University, Famagusta, Cyprus. [19]Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany. [20]Department of Biostatistics and Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA. [21]McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO, USA. [22]William Harvey Research Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [23]Department of Vascular Medicine, AMC, Amsterdam, The Netherlands. [24]Massachusetts General Hospital, Boston, MA, USA. [25]Department of Haematology, University of Cambridge, Cambridge, UK. [26]Division of Statistical Genomics, Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA. [27]Department of Cardiovascular Sciences, University of Leicester, Glenfield Hospital, Leicester, UK. [28]NIHR Leicester Cardiovascular Biomedical Research Unit, Glenfield Hospital, Leicester, UK. [29]Department of Medicine, Harvard Medical School, Boston, MA, USA. [30]Medical and Population Genetics Program, Broad Institute, Cambridge, MA, USA. [31]Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands. [32]Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands. [33]Center for Diabetes Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [34]Center for Genomics and Personalized Medicine Research, Wake Forest School of Medicine, Winston-Salem, NC, USA. [35]Department of Epidemiology, University of Washington, Seattle, WA, USA. [36]Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [37]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge, UK. [38]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany. [39]Netherlands Comprehensive Cancer Organisation, Utrecht, The Netherlands. [40]Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands. [41]School of Kinesiology and Health Science, Faculty of Health, York University, Toronto, Ontario, Canada. [42]Princess Al-Jawhara Al-Brahim Centre of Excellence in Research of Hereditary Disorders (PACER-HD), King Abdulaziz University, Jeddah, Saudi Arabia. [43]Department of Family Medicine and Public Health, University of California, San Diego, La Jolla, CA, USA. [44]INSERM U1167, Lille, France. [45]Institut Pasteur de Lille, U1167 Lille, France. [46]Université de Lille, U1167, RID-AGE, Risk Factors and Molecular Determinants of Aging-Related Diseases, Lille, France. [47]Department of Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. [48]Durrer Center for Cardiogenetic Research, ICIN-Netherlands Heart Institute, Utrecht, The Netherlands. [49]Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK. [50]Zilber School of Public Health, University of Wisconsin–Milwaukee, Milwaukee, WI, USA. [51]INSERM U1018, Centre de Recherche en Épidémiologie et Santé des Populations (CESP), Villejuif, France. [52]Department of Cardiology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [53]Metabolic Research Laboratories, University of Cambridge, Cambridge, UK. [54]NIHR Cambridge Biomedical Research Centre, Wellcome Trust–MRC Institute of Metabolic Science, Addenbrooke's Hospital, Cambridge, UK. [55]Department of Biomedical Informatics, Vanderbilt University, Nashville, TN, USA. [56]Department of Clinical Biochemistry, Herlev and Gentofte Hospital, Copenhagen University Hospital, Herlev, Denmark. [57]Department of Public Health, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [58]Department of Computational Biology, University of Lausanne, Lausanne, Switzerland. [59]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [60]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, MI, USA. [61]IFB Adiposity Diseases, University of Leipzig, Leipzig, Germany. [62]Department of Medicine, University of Leipzig, Leipzig, Germany. [63]Department of Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), Nuthetal, Germany. [64]School of Public Health, Human Genetics Center, University of Texas Health Science Center at Houston, Houston, TX, USA. [65]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX, USA. [66]Department of Nephrology, University Hospital Regensburg, Regensburg, Germany. [67]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [68]Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands. [69]Department of Biochemistry, Wake Forest School of Medicine, Winston-Salem, NC, USA. [70]Department of Clinical Biochemistry, Lillebaelt Hospital, Vejle, Denmark. [71]Institute of Regional Health Research, University of Southern Denmark, Odense, Denmark. [72]MRC Social Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. [73]NIHR Biomedical Research Centre for Mental Health, South London and Maudsley Hospital, London, UK. [74]Marshfield Clinic Research Institute, Marshfield, WI, USA. [75]Department of Medical Sciences, University of Trieste, Trieste, Italy. [76]Department of Medicine, University of Washington, Seattle, WA, USA. [77]MRC/BHF Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [78]NIHR Blood and Transplant Research Unit in Donor Health and Genomics, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [79]Epidemiology Research Program, American Cancer Society, Atlanta, GA, USA. [80]Institute for Maternal and Child Health, IRCCS 'Burlo Garofolo', Trieste, Italy. [81]Weis Center for Research, Geisinger Health System, Danville, PA, USA. [82]NIHR Barts Cardiovascular Research Unit, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK. [83]Department of Cardiology, London North West Healthcare NHS Trust, Ealing Hospital, Middlesex, UK. [84]Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, UK. [85]Imperial College Healthcare NHS Trust, London, UK. [86]Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [87]Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [88]Harvard Medical School, Boston, MA, USA. [89]Medical Department, Lillebaelt Hospital, Vejle, Denmark. [90]NHLBI Framingham Heart Study, Framingham, MA, USA. [91]Genetics and Pharmacogenomics, Merck, Sharp & Dohme, Boston, MA, USA. [92]Department of Medical, Surgical and Health Sciences, University of

Trieste, Trieste, Italy. [93]Medical Genomics and Metabolic Genetics Branch, National Human Genome Research Institute, US National Institutes of Health, Bethesda, MD, USA. [94]Department of Biostatistics, University of Liverpool, Liverpool, UK. [95]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [96]Department of Psychology, University of Edinburgh, Edinburgh, UK. [97]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. [98]Department of Ophthalmology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, The Netherlands. [99]Menzies Health Institute Queensland, Griffith University, Southport, Queensland, Australia. [100]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, USA. [101]Diamantina Institute, University of Queensland, Brisbane, Queensland, Australia. [102]QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [103]British Heart Foundation Cambridge Centre of Excellence, Department of Medicine, University of Cambridge, Cambridge, UK. [104]Department of Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands. [105]Department of Clinical Chemistry and Haematology, Division of Laboratory and Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands. [106]Utrecht Institute for Pharmaceutical Sciences, Division Pharmacoepidemiology and Clinical Pharmacology, Utrecht University, Utrecht, The Netherlands. [107]Department of Clinical Epidemiology, Leiden University Medical Center, Leiden, The Netherlands. [108]Department of Nutrition and Dietetics, School of Health Science and Education, Harokopio University, Athens, Greece. [109]Division of Epidemiology and Community Health, School of Public Health, University of Minnesota, Minneapolis, MN, USA. [110]Department of Internal Medicine, VU University Medical Center, Amsterdam, The Netherlands. [111]Laboratory of Experimental Cardiology, Division Heart and Lungs, University Medical Center Utrecht, Utrecht, The Netherlands. [112]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [113]Institute of Cardiovascular Science, University College London, London, UK. [114]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol, UK. [115]Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [116]Department of Medicine, Faculty of Medicine, Université de Montréal, Montreal, Quebec, Canada. [117]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, UK. [118]Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany. [119]Department of Epidemiology and Biostatistics, MRC-PHE Centre for Environment and Health, School of Public Health, Imperial College London, London, UK. [120]Department of Hygiene and Epidemiology, University of Ioannina Medical School, Ioannina, Greece. [121]Department of Health Sciences, University of Leicester, Leicester, UK. [122]Survey Research Center, Institute for Social Research, University of Michigan, Ann Arbor, MI, USA. [123]Department of Ophthalmology, University of Cologne, Cologne, Germany. [124]CNR Institute of Clinical Physiology, Pisa, Italy. [125]Department of Clinical and Experimental Medicine, University of Pisa, Pisa, Italy. [126]Toulouse University School of Medicine, Toulouse, France. [127]University of Glasgow, Glasgow, UK. [128]Institute of Molecular Medicine, University of Texas Health Science Center at Houston, Houston, TX, USA. [129]Department of Clinical Sciences, Genetic and Molecular Epidemiology Unit, Lund University, Malmo, Sweden. [130]Department of Nutrition, Harvard School of Public Health, Boston, MA, USA. [131]Department of Public Health and Clinical Medicine, Unit of Medicine, Umeå University, Umeå, Sweden. [132]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [133]Department of Clinical Biochemistry, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark. [134]Research Unit, AREA Science Park, Trieste, Italy. [135]Centre for Biological Sciences, Faculty of Natural and Environmental Sciences, University of Southampton, Southampton, UK. [136]Geriatrics, Department of Public Health, Uppsala University, Uppsala, Sweden. [137]Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA. [138]Department of Nutrition, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA. [139]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [140]German Center for Neurodegenerative Diseases (DZNE), Rostock/Greifswald, Greifswald, Germany. [141]Department of Pediatrics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [142]Division of Endocrinology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [143]Vision Sciences, Clinical Neurosciences Research Group, Faculty of Medicine, University of Southampton, Southampton, UK. [144]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [145]Icelandic Heart Association, Kopavogur, Iceland. [146]Department of Medical Sciences, Molecular Epidemiology and Science for Life Laboratory, Uppsala University, Uppsala, Sweden. [147]Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, VU University Amsterdam, Amsterdam, The Netherlands. [148]Department of Sociology, University of North Carolina, Chapel Hill, NC, USA. [149]Laboratory of Epidemiology and Population Sciences, National Institute on Aging, Intramural Research Program, US National Institutes of Health, Bethesda, MD, USA. [150]University of Exeter Medical School, University of Exeter, Exeter, UK. [151]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [152]Biodemography of Aging Research Unit, Social Science Research Institute, Duke University, Durham, NC, USA. [153]Department of Public Health, University of Helsinki, Helsinki, Finland. [154]Department of Neurology, Boston University School of Medicine, Boston, MA, USA. [155]Department of Psychiatry, Washington University, St. Louis, MO, USA. [156]Institute of Genetic Epidemiology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [157]Department of Pediatrics, Haukeland University Hospital, Bergen, Norway. [158]KG Jebsen Center for Diabetes Research, Department of Clinical Science, University of Bergen, Bergen, Norway. [159]Department of Cardiology, Heart Center, Tampere University Hospital and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [160]Department of Clinical Chemistry, Fimlab Laboratories, Tampere, Finland. [161]Department of Clinical Chemistry, Finnish Cardiovascular Research Center–Tampere, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [162]Centre for Eye Research Australia, Royal Victorian Eye and Ear Hospital, University of Melbourne, Melbourne, Victoria, Australia. [163]Centre for Ophthalmology and Vision Science, Lions Eye Institute, University of Western Australia, Perth, Western Australia, Australia. [164]Menzies Research Institute Tasmania, University of Tasmania, Hobart, Tasmania, Australia. [165]KG Jebsen Center for Genetic Epidemiology, Department of Public Health, NTNU, Norwegian University of Science and Technology, Trondheim, Norway. [166]Key Laboratory of Nutrition and Metabolism, Institute for Nutritional Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, University of the Chinese Academy of Sciences, Shanghai, China. [167]Department of Neurology, Erasmus Medical Center, Rotterdam, The Netherlands. [168]Department of Radiology, Erasmus Medical Center, Rotterdam, The Netherlands. [169]Division of Cardiovascular Medicine, Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA. [170]Department of Public Health and Clinical Medicine, Umeå University, Umeå, Sweden. [171]Research Unit Skellefteå, Skellefteå, Sweden. [172]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [173]Copenhagen City Heart Study, Frederiksberg Hospital, Frederiksberg, Denmark. [174]Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital, Bergen, Norway. [175]National Institute of Public Health, University of Southern Denmark, Copenhagen, Denmark. [176]Steno Diabetes Center Copenhagen, Gentofte, Denmark. [177]Faculty of Medicine, Aalborg University, Aalborg, Denmark. [178]Research Center for Prevention and Health, Capital Region of Denmark, Glostrup, Denmark. [179]Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands. [180]Interuniversity Cardiology Institute of the Netherlands, Utrecht, The Netherlands. [181]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [182]Department of Internal Medicine, University of Michigan, Ann Arbor, MI, USA. [183]Division of Gastroenterology, University of Michigan, Ann Arbor, MI, USA. [184]Centre for Brain Research, Indian Institute of Science, Bangalore, India. [185]Department of Psychiatry, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands. [186]Department of Clinical Physiology, Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. [187]Department of Clinical Physiology, Tampere University Hospital, Tampere, Finland. [188]Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland. [189]National Institute for Health and Welfare, Helsinki, Finland. [190]Echinos Medical Centre, Echinos, Greece. [191]Oxford Centre for Diabetes, Endocrinology and Metabolism, Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [192]Oxford NIHR Biomedical Research Centre, Oxford University Hospitals Trust, Oxford, UK. [193]UKCRC Centre of Excellence for Public Health Research, Queens University Belfast, Belfast, UK. [194]Foundation for Research in Health Exercise and

Nutrition, Kuopio Research Institute of Exercise Medicine, Kuopio, Finland. [195]Institute of Biomedicine, School of Medicine, University of Eastern Finland, Kuopio Campus, Kuopio, Finland. [196]Department of Clinical Physiology and Nuclear Medicine, Kuopio University Hospital, Kuopio, Finland. [197]National Heart and Lung Institute, Imperial College London, Hammersmith Hospital Campus, London, UK. [198]MRC-PHE Centre for Environment and Health, Imperial College London, London, UK. [199]Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [200]Department of Psychiatry, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa. [201]Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, South Africa. [202]Institute of Social and Preventive Medicine, Lausanne University Hospital, Lausanne, Switzerland. [203]Institute of Clinical Medicine, Internal Medicine, University of Eastern Finland and Kuopio University Hospital, Kuopio, Finland. [204]Department of Genetics, University of North Carolina, Chapel Hill, NC, USA. [205]Kaiser Permanente Washington Health Research Institute, Seattle, WA, USA. [206]Department of Health Services, University of Washington, Seattle, WA, USA. [207]Department of Anthropology, Sociology and History, University of San Carlos, Cebu City, Philippines. [208]USC–Office of Population Studies Foundation, Inc., University of San Carlos, Cebu City, Philippines. [209]Division of Preventive Medicine, University of Alabama at Birmingham, Birmingham, AL, USA. [210]Department of Medicine, Boston University School of Medicine, Boston, MA, USA. [211]Department of Ophthalmology, Taichung Veterans General Hospital, Taichung, Taiwan. [212]Department of Medical Sciences, Cardiovascular Epidemiology, Uppsala University, Uppsala, Sweden. [213]Department of Clinical Experimental Research, Rigshospitalet, Copenhagen, Denmark. [214]Department of Clinical Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark. [215]Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA. [216]Department of Public Health Sciences, Institute for Personalized Medicine, Pennsylvania State University College of Medicine, Hershey, PA, USA. [217]Division of Public Health Sciences, Wake Forest School of Medicine, Winston-Salem, NC, USA. [218]Division of Health Sciences, Warwick Medical School, Warwick University, Coventry, UK. [219]Department of Preventive Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. [220]Department of Molecular Epidemiology, German Institute of Human Nutrition Potsdam-Rehbrücke (DIfE), Nuthetal, Germany. [221]German Center for Diabetes Research, Neuherberg, Germany. [222]Westmead Millennium Institute of Medical Research, Centre for Vision Research and Department of Ophthalmology, University of Sydney, Sydney, New South Wales, Australia. [223]Department of Medicine, Kuopio University Hospital, Kuopio, Finland. [224]Department of Epidemiology and Public Health, University of Strasbourg, Strasbourg, France. [225]Department of Public Health, University Hospital of Strasbourg, Strasbourg, France. [226]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [227]Department of Public Health and Primary Care, Leiden University Medical Center, Leiden, The Netherlands. [228]INTERVAL Coordinating Centre, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. [229]School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia. [230]Centre for Global Health Research, Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK. [231]Department of Medicine I, University Hospital Großhadern, Ludwig-Maximilians-Universitat, Munich, Germany. [232]German Centre for Cardiovascular Research (DZHK), partner site Munich Heart Alliance, Munich, Germany. [233]Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, MD, USA. [234]Data Tecnica International, Glen Echo, MD, USA. [235]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia. [236]Program for Personalized and Genomic Medicine, Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA. [237]Division of Cardiovascular Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA. [238]Center for Neurobehavioral Genetics, University of California, Los Angeles, Los Angeles, CA, USA. [239]Pat Macpherson Centre for Pharmacogenetics and Pharmacogenomics, Medical Research Institute, Ninewells Hospital and Medical School, Dundee, UK. [240]Laboratory of Clinical Chemistry and Hematology, Division of Laboratories and Pharmacy, University Medical Center Utrecht, Utrecht, The Netherlands. [241]School of Women's and Infants' Health, University of Western Australia, Perth, Western Australia, Australia. [242]Diabetes and Obesity Research Program, University of Helsinki, Helsinki, Finland. [243]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark. [244]Institute of Epidemiology II, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [245]Department of Clinical Immunology and Biochemistry, Lillebaelt Hospital, Vejle, Denmark. [246]School of Medicine, University of Split, Split, Croatia. [247]Central Hospital of Southern Ostrobothnia, Seinäjoki, Finland. [248]Department of Clinical Physiology and Nuclear Medicine, Turku University Hospital, Turku, Finland. [249]Research Centre of Applied and Preventive Cardiovascular Medicine, University of Turku, Turku, Finland. [250]Centre for Non-Communicable Diseases, Karachi, Pakistan. [251]Department of Biobank Research, Umeå University, Umeå, Sweden. [252]Wellcome Trust Centre for Human Genetics, Nuffield Department of Clinical Medicine, University of Oxford, Oxford, UK. [253]NHS Blood and Transplant–Oxford Centre, Oxford, UK. [254]BRC Haematology Theme and Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [255]Department of Public Health and Clinical Medicine, Unit of Family Medicine, Umeå University, Umeå, Sweden. [256]Genetics of Complex Traits, University of Exeter Medical School, University of Exeter, Exeter, UK. [257]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [258]Departments of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. [259]Duke Molecular Physiology Institute, Duke University, Durham, NC, USA. [260]Division of Endocrinology and Metabolism, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung, Taiwan. [261]School of Medicine, National Defense Medical Center, Taipei, Taiwan. [262]School of Medicine, National Yang-Ming University, Taipei, Taiwan. [263]Saw Swee Hock School of Public Health, National University Health System, National University of Singapore, Singapore, Singapore. [264]Genetics, Target Sciences, GlaxoSmithKline, Research Triangle Park, NC, USA. [265]OmicSoft at Qiagen Company, Cary, NC, USA. [266]Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. [267]Alzheimer Scotland Dementia Research Centre, University of Edinburgh, Edinburgh, UK. [269]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [269]Chair of Genetic Epidemiology, IBE, Faculty of Medicine, LMU Munich, Munich, Germany. [270]Kanazawa University, Kanazawa, Japan. [271]College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. [272]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [273]Center for Pediatric Research, Department for Women's and Child Health, University of Leipzig, Leipzig, Germany. [274]Department of Gerontology and Geriatrics, Leiden University Medical Center, Leiden, The Netherlands. [275]Anogia Medical Centre, Anogia, Greece. [276]Centre for Vascular Prevention, Danube University Krems, Krems, Austria. [277]Dasman Diabetes Institute, Dasman, Kuwait. [278]Diabetes Research Group, King Abdulaziz University, Jeddah, Saudi Arabia. [279]Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [280]Department of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland. [281]Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands. [282]Department of Molecular Cell Biology, Leiden University Medical Center, Leiden, The Netherlands. [283]Primary Health Care Unit, Central Hospital of Central Finland, Jyväskylä, Finland. [284]Primary Health Care Unit, Kuopio University Hospital, Kuopio, Finland. [285]USC Roski Eye Institute, Department of Ophthalmology, Keck School of Medicine of the University of Southern California, Los Angeles, CA, USA. [286]Department of Obstetrics and Gynecology, Institute for Medicine and Public Health, Vanderbilt Genetics Institute, Vanderbilt University, Nashville, TN, USA. [287]Research Center on Epidemiology and Preventive Medicine, Department of Medicine and Surgery, University of Insubria, Varese, Italy. [288]Cardiometabolic Disease, Merck, Sharp & Dohme, Kenilworth, NJ, USA. [289]German Centre for Cardiovascular Research (DZHK), partner site Greifswald, Greifswald, Germany. [290]Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. [291]Institute of Cellular Medicine, Medical School, Newcastle University, Newcastle, UK. [292]Department of Medical Sciences, Cardiology, Uppsala Clinical Research Center, Uppsala University, Uppsala, Sweden. [293]Genetics, Target Sciences, GlaxoSmithKline, King of Prussia, PA, USA. [294]Departments of Epidemiology and Medicine, Diabetes Translational Research Center, Fairbanks School of Public Health and School of Medicine, Indiana University, Indianapolis, IN, USA. [295]Green Lane Cardiovascular Service, Auckland City Hospital and University of Auckland, Auckland, New Zealand. [296]Department of Human Genetics, University of Michigan, Ann Arbor, MI, USA. [297]Department of Physiology and Biophysics, University of Mississippi

Medical Center, Jackson, MS, USA. [298]Danish Diabetes Academy, Odense, Denmark. [299]Department of Public Health, Aarhus University, Aarhus, Denmark. [300]GlaxoSmithKline, King of Prussia, PA, USA. [301]Department of Clinical Sciences, Quantitative Biomedical Research Center, Center for the Genetics of Host Defense, University of Texas Southwestern Medical Center, Dallas, TX, USA. [302]Endometriosis CaRe Centre, Nuffield Department of Obstetrics and Gynaecology, University of Oxford, Oxford, UK. [303]A list of members and affiliations appears in the Supplementary Note. [304]Department of Epidemiology and Carolina Center of Genome Sciences, Chapel Hill, NC, USA. [305]Li Ka Shing Centre for Health Information and Discovery, Big Data Institute, University of Oxford, Oxford, UK. [306]Department of Pediatrics, Harvard Medical School, Boston, MA, USA. [307]Mindich Child Health and Development Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. Valérie Turcot, Yingchang Lu, Heather M. Highland, Claudia Schurmann, Anne E. Justice and Rebecca S. Fine contributed equally to this work. Guillaume Lettre, Kari E. North, Cecilia M. Lindgren, Joel N. Hirschhorn and Ruth J. F. Loos jointly directed this work.
*e-mail: joelh@broadinstitute.org; ruth.loos@mssm.edu

Methods

Study design and participants. The discovery cohort consisted of 123 studies (163 datasets) comprising 526,508 adult (≥18 years) individuals of the following ancestries (Supplementary Fig. 1): (i) European ($N = 449,889$), (ii) South Asian ($N = 29,398$), (iii) African ($N = 27,610$), (iv) East Asian ($N = 8,839$) and (v) Hispanic ($N = 10,772$). All participating institutions and coordinating centers approved this project, and informed consent was obtained from all study participants. Discovery meta-analyses were carried out in each ancestry group separately and in an all-ancestry combined group, for both sex-specific and sex-combined analyses. SNVs for which associations reached suggestive significance ($P < 2.0 \times 10^{-6}$) in the discovery analyses were taken forward for follow-up in 192,226 individuals of European ancestry from the UK Biobank and deCODE. Conditional analyses were conducted in the all-ancestry and European-descent groups. Study-specific design, sample quality control and descriptive statistics are provided in Supplementary Tables 1–3.

Phenotype. BMI (weight (in kg)/height (in m)^2) was corrected for age, age^2 and genomic principal components (derived from GWAS data, the variants with MAF >1% on ExomeChip or ancestry-informative markers available on the ExomeChip), as well as any additional study-specific covariates (for example, recruiting center) in a linear regression model. For studies with non-related individuals, residuals were calculated separately by sex, whereas for family-based studies sex was included as a covariate in the model. Additionally, residuals for case–control studies were calculated separately. Finally, residuals were subjected to inverse normal transformation[93].

Genotype calling. The majority of studies followed a standardized protocol and performed genotype calling using the designated manufacturer's software, which was then followed by zCall[94]. For ten studies, participating in the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) Consortium, the raw intensity data for the samples from seven genotyping centers were assembled into a single project for joint calling[95]. Study-specific quality control measures of the genotyped variants were implemented before association analysis (Supplementary Table 3).

Statistical analyses. *Study-level association analyses.* Individual cohorts were analyzed separately for each ancestry group, in sex-combined and sex-specific groups, with either RAREMETALWORKER (see URLs) or RVTEST[96] (Supplementary Table 2), to associate inverse-normal-transformed BMI with genotype, accounting for potential cryptic relatedness (kinship matrix), in a linear mixed model. These software tools are designed to perform score-statistics-based rare variant association analyses, can accommodate both unrelated and related individuals, and provide single-variant results and variance–covariance matrices. The covariance matrix captures LD relationships between markers within 1 Mb, which is used for gene-level meta-analyses and conditional analyses[97]. Single-variant analyses were performed for both additive and recessive models.

Centralized quality control. A centralized quality control procedure, implemented in EasyQC[98], was applied to individual cohort association summary statistics to identify cohort-specific problems: (i) assessment of possible problems in BMI transformation, (ii) comparison of allele frequency alignment against 1000 Genomes Project phase 1 reference data to pinpoint any potential strand issues and (iii) examination of quantile–quantile plots per study to identify any problems arising from population stratification, cryptic relatedness and genotype biases.

Meta-analyses. Meta-analyses were carried out by two different analysts at two sites in parallel. We excluded variants with a call rate <95%, Hardy–Weinberg equilibrium P value $<1 \times 10^{-7}$ or large allele frequency deviations from reference populations (>0.6 for all-ancestry analyses and >0.3 for ancestry-specific population analyses). Significance for single-variant analyses was defined at the array-wide level (a Bonferroni-corrected threshold of $P < 2 \times 10^{-7}$ for ~250,000 SNVs). To test for sex differences of the significant variants ($P < 2 \times 10^{-7}$), we calculated the P-diff for each SNP, which tests for differences between women-specific and men-specific β estimates using EasyStrata[99]. For gene-based analyses, we applied the SKAT[100] and VT[101] gene-based methods using two different sets of criteria (broad and strict) to select predicted damaging rare and low-frequency variants with MAF <5%, based on coding variant annotation from five prediction algorithms (PolyPhen-2 HumDiv and HumVar, LRT, MutationTaster and SIFT)[20]. Our 'broad' gene-based tests included nonsense, stop-loss, splice-site and missense variants that were annotated as damaging by at least one algorithm mentioned above. Our 'strict' gene-based tests included only nonsense, stop-loss, splice-site and missense variants annotated as damaging by all five algorithms. Statistical significance for gene-based tests was set at a Bonferroni-corrected threshold of $P < 2.5 \times 10^{-6}$ for about 20,000 genes[16,102]. Singe-variant and gene-based meta-analyses were both performed using the RareMETALS R package[103]. As our secondary analyses were nested and/or highly correlated with our primary analysis, we chose the same, already stringent Bonferroni-corrected significance threshold for both analyses.

Genomic inflation. Although the overall λ_{GC} value was in the normal range for all coding variants ($\lambda_{GC} = 1.1$; Supplementary Table 23), we observed a marked genomic inflation of the test statistics even after adequate control for population stratification (linear mixed model) arising from common markers ($\lambda_{GC} = 1.99$; Supplementary Fig. 2a and Supplementary Table 25). Such inflation is expected for a highly polygenic trait like BMI, as was previously confirmed for height[15], and is consistent with our very large sample size[5,104]. Furthermore, some of the inflation may be due to the design of the ExomeChip, which besides rare and low-frequency coding SNVs also contains (common and noncoding) SNVs that include previously identified GWAS loci for all traits, including for BMI and BMI-related traits, reported in the GWAS catalog at the time of its design.

After removing established loci (±1 Mb), the excess of significant associations was markedly reduced and inflation was reduced (Supplementary Fig. 2c,d).

Furthermore, to exclude the possibility that some of the observed associations between BMI and rare/low-frequency SNVs could be due to allele calling problems in the smaller studies, we performed a sensitivity meta-analysis with primarily European-ancestry studies totaling >5,000 participants. We found very concordant effect sizes, suggesting that smaller studies do not bias our results (Supplementary Fig. 12).

Follow-up analysis. We sought additional evidence for association of the top signals ($P < 2.0 \times 10^{-6}$) identified in the discovery meta-analysis using two independent studies from the UK (UK Biobank, interim release, $N = 119,613$) and Iceland (deCODE, $N = 72,613$) (Supplementary Tables 1–3). We used the same quality control and analytical methodology as described above. We used inverse-variance-weighted fixed-effects meta-analysis in METAL[105] to combine the discovery and follow-up association results. Significant associations were defined at $P < 2 \times 10^{-7}$ in the combined meta-analysis of discovery, UK Biobank and deCODE results.

Effect of study design. To investigate the potential effect of study design of the participating studies, we tested for heterogeneity between population-based, all case–control studies; T2D case–control studies (Supplementary Table 26). None of these comparisons showed significant evidence of heterogeneity ($P < 7.4 \times 10^{-5}$, correcting for multiple testing).

Conditional analyses. The RareMETALS R package[103] was used to identify independent BMI-associated signals across the all-ancestry meta-analysis results in the discovery phase. RareMETALS performs conditional analyses by using covariance matrices from each individual cohort to distinguish true signals from the shadows of adjacent significant variants in LD. The conditional associations of all the variants within 1 Mb of each rare or low-frequency coding variant were analyzed (i) to identify nearby secondary signals and (ii) to determine independence from nearby noncoding variants or previously identified GWAS loci (previously defined as a window of 1 Mb surrounding the lead SNP). Gene-based conditional analyses were also performed in RareMETALS.

Owing to the selective coverage of variants on the ExomeChip, we also conducted the respective conditional analyses in the UK Biobank dataset that included 847,441 genome-wide genotyped markers and 72,355,667 variants imputed against the UK10K haplotype reference panel, merged with the 1000 Genomes Phase 3 reference panel. Where available, directly genotyped variants were used for conditional analyses. Otherwise, imputed variants with good imputation quality (IMPUTE2 info score > 0.6) were used. We used QCTOOL to extract variants of interest from the original imputed dataset. Subsequently, GTOOL was used to convert to PLINK format (genotype calling threshold 0.99), and results were merged with the directly genotyped variants for conditional analyses in PLINK v1.90b3.35 64 bit (25 March 2016).

Conversion of effect size and explained variance. We assumed that 1 s.d. = 4.5 kg/m^2 BMI units, on the basis of population-based data, and used 1.7 m as the average height of a person to convert effect sizes in s.d. units into body weight. The variance explained by each variant was calculated using the effect allele frequency (f) and beta (β) from the meta-analyses with the formula[106] of explained variance $= 2f(1-f)\beta^2$.

Penetrance analysis. We examined the penetrance for the four rare SNVs—p.Arg525Gln (rs56214831) in *KSR2*, p.Tyr35Ter (rs13447324) in *MC4R*, and p.Arg190Gln (rs139215588) and p.Glu288Gly (rs143430880) in *GIPR*—in European-ancestry data from the UK Biobank (N up to 120,000). For each variant, we compared the prevalence of underweight (BMI < 18.5 kg/m^2), normal weight (18.5 kg/m^2 ≤ BMI < 25 kg/m^2), overweight (25 kg/m^2 ≤ BMI < 30 kg/m^2) and obesity (BMI ≥ 30 kg/m^2) of carriers with non-carriers. We used a Pearson χ^2 test to test for differences between distributions and a χ^2 for linear trend to test whether distributions of carriers were shifted in comparison to non-carriers. For p.Arg525Gln in *KSR2* and p.Tyr35Ter in *MC4R*, we hypothesized that obesity prevalence was higher in carriers than in non-carriers, whereas for the two GIPR variants we hypothesized that the prevalence of normal weight was higher in carriers than non-carriers.

Associations with obesity for the coding rare and low-frequency loci in children. For each of the 14 rare and low-frequency SNVs identified, we tested for association with childhood obesity in the CHOP cohort (Childhood Obesity: Early Programming by Infant Nutrition), the Severe Childhood-Onset Obesity Project (SCOOP), the UK Household Longitudinal Study (UKHLS) and the INTERVAL Study (INTERVAL). Summary statistics across the studies were combined using a fixed-effects inverse-variance meta-analysis with METAL[105].

In the CHOP study, cases (1,358 boys, 1,060 girls) were defined as having a BMI greater than the 95th percentile at any point in their childhood. Controls (1,412 boys, 1,143 girls) were defined as having BMI below the 50th percentile consistently throughout childhood. The BMI percentiles are based on the CDC 2000 Growth Charts. All children were classified on the basis of their BMI measurements between the ages of 2 and 18. All individuals were of European ancestry and were collected at the Children's Hospital of Philadelphia. Informed consent was obtained from all study participants, and study protocols were approved by the local ethics committees. Genotypes were obtained using the HumanHap550v1, HumanHap550v3 and Human610-Quad high-density SNP arrays from Illumina. The intersection of all SNPs on the arrays was used in all subsequent preimputation analyses. Before imputation, we excluded SNPs with a Hardy–Weinberg equilibrium P value $<1.0 \times 10^{-6}$, call rate of <95% or MAF of < 1%. The genotypes were then prephased using SHAPEIT2 and imputed using the 1000 Genomes phase 1 integrated variant set with IMPUTE2. After imputation, SNPs were excluded if the INFO score was <0.4. Boys and girls were analyzed separately using a logistic regression of case and control status, adjusting for three eigenvectors, and summary statistics were combined using a fixed-effects inverse-variance meta-analysis with METAL[105].

SCOOP is a subcohort of the Genetics of Obesity Study (GOOS) cohort. It includes >1,500 UK European-ancestry individuals with severe, early-onset obesity (BMI s.d. score > 3 and obesity onset before the age of 10 years), in whom known monogenic causes of obesity have been excluded (cases with *MC4R* mutations were excluded). Two case–control analyses with SCOOP cases were performed: (i) SCOOP versus UKHLS, for which array (Illumina HumanCoreExome) data were available, and (ii) SCOOP versus INTERVAL, for which whole-exome sequencing data were available.

For the array-based analyses, UKHLS controls were genotyped on the Illumina HumanCoreExome-12v1-0 BeadChip. SCOOP cases and 48 UKHLS controls were genotyped on the Illumina HumanCoreExome-12v1-1 BeadChip. The 48 overlapping UKHLS samples were used for quality control to ensure that there were no systematic differences and bias between the two versions of the chip. SCOOP and UKHLS samples were phased with SHAPEITv2, and imputed with IMPUTE2 using the combined UK10K–1000 Genomes phase 3 reference panel. For the whole-exome sequencing analyses, SCOOP versus INTERVAL controls were whole-exome sequenced within the UK10K-EXOME project (Agilent v3) and the INTERVAL project (Agilent v5), respectively, and were then jointly called and subjected to quality control on the union of the sequencing baits. Individuals overlapping or related between the array-based and whole-exome sequencing studies were removed.

After quality control, 1,456 SCOOP and 6,460 UKHLS (BMI range 19–30) individuals and 521 SCOOP and 4,057 INTERVAL individuals were available for the two analyses; all were unrelated, of high quality and of European ancestry. For both analyses (i.e., SCOOP versus UKHLS and SCOOP versus INTERVAL), a maximum-likelihood frequentist association test with the additive genetic model was implemented in SNPTEST v2.5. In the SCOOP versus UKHLS analysis, sex and the first six principal components were included as covariates and variants with a SNPTEST INFO score <0.4 and Hardy–Weinberg equilibrium P value $<1 \times 10^{-6}$ were removed. For the SCOOP versus INTERVAL analysis, we performed an unadjusted analysis (adjustment for principal components did not sufficiently change the results) and variants were limited to those covered at ≥7× in at least 80% of each sequencing cohort, meeting the VQSR threshold of −2.52, missingness <80%, Hardy–Weinberg equilibrium P value $<1 \times 10^{-8}$ and GQ ≥30.

Cross-trait analyses. We evaluated each of the 14 rare and low-frequency SNVs for association with other relevant obesity-related traits and conditions. We performed lookups in ExomeChip meta-analysis results from other consortia, including our own, GIANT Consortium (height[15], WHR adjusted for BMI[24]), MAGIC (HbA1c, fasting insulin, fasting glucose, 2-h glucose), GLGC (HDL cholesterol (HDL-C), LDL cholesterol (LDL-C), triglycerides and total cholesterol), IBPC[39] (systolic and diastolic blood pressure), REPROGEN[23] (age at menarche and menopause) and GoT2D/T2D-GENES[16] (type 2 diabetes). Associations were considered significant at $P < 2.0 \times 10^{-5}$, accounting for multiple testing.

Phenome-wide association analysis. To evaluate the potential for pleiotropic effects for SNPs discovered from primary analyses, we performed PheWAS using genotype and phenotype data from two independent sources of EHRs: the Vanderbilt University Medical Center Biorepository (BioVU) and the UK Biobank. The phenotype selection and analysis strategies were synchronized across sites. A total of 1,502 hierarchical phenotype codes from EHRs were curated by grouping International Classification of Disease, Ninth Revision (ICD-9) clinical/billing codes as previously described[107]. Phenotype codes with 20 or more cases and with minor allele count of 5 or greater in cases and controls were eligible for analysis. Series of logistic regression analyses were then performed in individuals of European ancestry for each eligible genotype–phenotype combination while adjusting for five genetic ancestry principal components. Odds ratios from genotype–phenotype combinations present in both BioVU and the UK Biobank were then aggregated using inverse-variance-weighted fixed-effects meta-analysis. Associations with P values corresponding to a FDR cutoff of less than 10% were considered statistically significant.

Gene set enrichment analysis. We adapted DEPICT, a gene set enrichment analysis method for GWAS data, for use with the ExomeChip (EC-DEPICT). DEPICT's primary innovation is the use of 'reconstituted' gene sets, where many different types of gene sets (e.g., canonical pathways, protein–protein interaction networks and mouse phenotypes) were extended through the use of large-scale microarray data (see ref. [108] for details). EC-DEPICT computes P values on the basis of Swedish ExomeChip data (Malmö Diet and Cancer (MDC), All New Diabetics in Scania (ANDIS) and Scania Diabetes Registry (SDR) cohorts, $N = 11,899$) and, unlike DEPICT, takes as input only coding variants and only the genes directly containing these variants, rather than all genes within a specified amount of LD (Supplementary Note).

Four analyses were performed for the BMI EC variants: (i) all coding variants with $P < 5 \times 10^{-4}$, (ii) all coding variants with $P < 5 \times 10^{-4}$ independent of known GWAS variants[5], (iii) all coding rare and low-frequency variants with $P < 5 \times 10^{-4}$ and (iv) all coding rare and low-frequency variants with $P < 5 \times 10^{-4}$ independent of known GWAS variants. Affinity propagation clustering[3] was used to group highly correlated gene sets into meta-gene sets. For each meta-gene set, the member gene set with the best P value was used as representative for the purposes of visualization (Supplementary Note). DEPICT for ExomeChip was written using the Python programming language (see URLs for the link to the code on GitHub).

Drosophila RNAi knockdown experiments. For each of the 13 genes in which rare and low-frequency coding variants were associated with BMI, we searched for the corresponding orthologs in *Drosophila* in the Ensembl ortholog database. Orthologs were available for nine genes but missing for *ZBTB7B*, *MC4R*, *GIPR* and *ZNF169*. For each of the nine genes, we generated crosses with RNAi knockdown specifically in adipose tissue (cg-Gal4) and neurons (elav-Gal4), leveraging upstream activation sequence (UAS)-inducible short hairpin knockdown lines, available through the Vienna *Drosophila* Resource Center (VDRC). We crossed male UAS-RNAi flies and elav-GAL4 or CG-GAL4 virgin female flies. All fly experiments were carried out at 25 °C. Five- to 7-d-old males were sorted into groups of 20, weighed and homogenized in PBS with 0.05% Tween with Lysing Matrix D in a beadshaker. The homogenate was heat inactivated for 10 min in a thermocycler at 70 °C. 10 µl of the homogenate was subsequently used in a triglyceride assay (Sigma, Serum Triglyceride Determination Kit) that was carried out in duplicate according to the protocol with one alteration: samples were cleared of residual particulate debris by centrifugation before absorbance reading. Resulting triglyceride values were normalized to fly weight and larval/population density. We used the non-parametric Kruskal–Wallis test to compare wild-type with knockdown lines.

Enrichment analysis genes associated with monogenic obesity. We identified 39 genes with strong evidence that disruption causes monogenic or syndromic forms of obesity (Supplementary Table 21). To test whether these genes are enriched for rare or low-frequency coding variant associations with BMI, we conducted simulations by matching each of the 39 genes with other genes on the basis of gene length and number of variants tested, to create a matched set of genes. We generated 1,000 matched gene sets from our data and assessed how often the number of rare and low-frequency coding variants that exceeded the given significance thresholds was greater in our monogenic/syndromic obesity gene set than in the matched gene sets.

URLs. CDC 2000 Growth Charts, http://www.cdc.gov/growthcharts/cdc_charts.htm; CHOP cohort, http://www.metabolic-programming.org/obesity/; EC-DEPICT code, https://github.com/RebeccaFine/obesity-ec-depict; ENSEMBL, http://www.ensembl.org/; EasyQC, http://www.genepi-regensburg.de/easyqc; EasyStrata, http://www.genepi-regensburg.de/easystrata; ExAC, http://exac.broadinstitute.org/; GCTA, http://cnsgenomics.com/software/gcta/; GTEx, http://www.gtexportal.org/home/; GTOOL, http://www.well.ox.ac.uk/~cfreeman/software/gwas/gtool.html; IMPUTE2, https://mathgen.stats.ox.ac.uk/impute/impute_v2.html; INTERVAL Study, http://www.intervalstudy.org.uk/; PLINK v1.90, https://www.cog-genomics.org/plink2; QCTOOL, http://www.well.ox.ac.uk/~gav/qctool/#overview; RAREMETALWORKER, http://genome.sph.umich.edu/wiki/RAREMETALWORKER; RareMETALS, http://genome.sph.umich.edu/wiki/RareMETALS; RVTEST, https://github.com/zhanxw/rvtests; Shapeit2, https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html; UKHLS, https://www.understandingsociety.ac.uk/; UK10K Obesity Sample Sets–SCOOP, http://www.uk10k.org/studies/obesity.html; 1000 Genomes Phase 1, http://www.1000genomes.org/category/phase-1/.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. Summary statistics can be downloaded from http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium.

References

93. Auer, P. L., Reiner, A. P. & Leal, S. M. The effect of phenotypic outliers and non-normality on rare-variant association testing. *Eur. J. Hum. Genet.* **24**, 1188–1194 (2016).
94. Goldstein, J. I. et al. zCall: a rare variant caller for array-based genotyping: genetics and population analysis. *Bioinformatics* **28**, 2543–2545 (2012).
95. Grove, M. L. et al. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PLoS One* **8**, e68095 (2013).
96. Zhan, X., Hu, Y., Li, B., Abecasis, G. R. & Liu, D. J. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* **32**, 1423–1426 (2016).
97. Liu, D. J. et al. Meta-analysis of gene-level tests for rare variant association. *Nat. Genet.* **46**, 200–204 (2014).
98. Winkler, T. W. et al. Quality control and conduct of genome-wide association meta-analyses. *Nat. Protoc.* **9**, 1192–1212 (2014).
99. Winkler, T. W. et al. EasyStrata: evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics* **31**, 259–261 (2015).
100. 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).
101. Price, A. L. et al. Pooled association tests for rare variants in exon-resequencing studies. *Am. J. Hum. Genet.* **86**, 832–838 (2010).
102. Kiezun, A. et al. Exome sequencing and the genetic basis of complex traits. *Nat. Genet.* **44**, 623–630 (2012).
103. Feng, S., Liu, D., Zhan, X., Wing, M. K. & Abecasis, G. R. RAREMETAL: fast and powerful meta-analysis for rare variants. *Bioinformatics* **30**, 2828–2829 (2014).
104. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
105. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
106. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).
107. Denny, J. C. et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics* **26**, 1205–1210 (2010).
108. Pers, T. H. et al. Biological interpretation of genome-wide association studies using predicted gene functions. *Nat. Commun.* **6**, 5890 (2015).

Corresponding Author: Ruth Loos

Date: October 25 2017

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

   Describe how sample size was determined.

   We aimed to bring together the largest possible sample size (N>720,000) possible.
   We started with the collection of ExomeChip summary statistics from 123 studies and performed a meta-analyses. Subsequently, data of two large-scale studies became available (UK Biobank and deCODE), which were added to the meta-analyses.
   See Page 14 Par 3 - Page 15 Par 1 and Page 51 Par 1.

2. Data exclusions

   Describe any data exclusions.

   We excluded individuals <18yrs. See also Page 51 Par 1.

3. Replication

   Describe whether the experimental findings were reliably reproduced.

   We designed our study in two stages, a discovery stage of 526,508 individuals in which we identified variants associated at P<10-6. The follow-up stage consisted of two large-sale studies (N=192,226) in which we test association of variants identified in the discovery stage. Direction of association in follow up stage was consistent for all 14 rare and low frequency variants and all 92 common variants.

4. Randomization

   Describe how samples/organisms/participants were allocated into experimental groups.

   NA

5. Blinding

   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

   NA

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
☒	☐	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *p* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

The software used has been described in the Online Methods section, from Page 51 to Page 59. URLs and reference are provided when available.
Softwares are:
zCall, RAREMETALWORKER, RVTEST, EasyQC, EasyStrata, RareMETALS , METAL, IMPUTE2, QCTOOL, PLINK, SHAPEITv2, EC-DEPICT
In addition, study-specific software, used by each study to perform analyses is listed in Supplementary Table 2.

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The *Nature Methods* guidance for providing algorithms and software for publication may be useful for any submission.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

NA

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

NA

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

NA

b. Describe the method of cell line authentication used.

NA

c. Report whether the cell lines were tested for mycoplasma contamination.

NA

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

NA

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	NA

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	We provide a description of outcome and covariates used at Page 51 Par 2. Study specific descriptive are provided in Supplementary Table 3.