Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index

Obesity is globally prevalent and highly heritable, but its underlying genetic factors remain largely elusive. To identify genetic loci for obesity susceptibility, we examined associations between body mass index and ~2.8 million SNPs in up to 123,865 individuals with targeted follow up of 42 SNPs in up to 125,931 additional individuals. We confirmed 14 known obesity susceptibility loci and identified 18 new loci associated with body mass index (P < 5 \times 10^{-8}), one of which includes a copy number variant near GPRC5B. Some loci (at MC4R, POMC, SH2B1 and BDNF) map near key hypothalamic regulators of energy balance, and one of these loci is near GIPR, an incretin receptor. Furthermore, genes in other newly associated loci may provide new insights into human body weight regulation.

Stage 2 follow up identifies additional new loci for BMI

To identify additional BMI-associated loci and to validate the loci that reached genome-wide significance in the stage 1 analyses, we examined SNPs representing 42 independent loci (including the 19 genome-wide significant loci) having a stage 1 P < 5 \times 10^{-6}. Variants were considered to be independent if the pair-wise linkage disequilibrium (LD, \textit{r}2) was less than 0.1 and if they were separated by at least 1 Mb. In stage 2, we examined these 42 SNPs in up to 125,931 additional individuals (79,561 newly genotyped individuals from 16 different studies and 46,370 individuals from 18 additional studies for which genome-wide association data were available; Table 1, Supplementary Note and Online Methods). In a joint analysis of stage 1 and stage 2 results, 32 of the 42 SNPs reached P < 5 \times 10^{-8} (Table 1, Supplementary Table 1 and Supplementary Figs. 1 and 2). Even after excluding SNPs within the 32 confirmed BMI loci, we still observed an excess of small P values compared to the distribution expected under the null hypothesis (Fig. 1b and Supplementary Fig. 3), suggesting that more BMI loci remain to be uncovered.

The 32 confirmed associations included all 19 loci with P < 5 \times 10^{-8} at stage 1, 12 additional new loci near RBP1-ADCY3-POMC, QPCTL-GIPR, SLC39A8, TMEM160, FANCI, CADM2, LRP1B, PTBP2, MTIF3-GTF3A, ZNF608, RPL27A-TUB and NUDT3-HMGCR and one locus (in NRXN3) previously associated with waist circumference (near TFAP2B). The remaining six loci, near GPRC5B, MAP2K5-LBXCOR1, TNNI3K, LRRN6C, FLJ35779-HMGCR and PRKD1, have not previously been associated with BMI or other obesity-related traits.

RESULTS

Stage 1 GWAS identifies new loci associated with BMI

We first conducted a meta-analysis of GWAS of BMI and ~2.8 million imputed or genotyped SNPs using data from 46 studies including up to 123,865 individuals (Online Methods, Supplementary Fig. 1 and Supplementary Note). This stage 1 analysis revealed 19 loci associated with BMI at P < 5 \times 10^{-8} (Table 1, Fig. 1a and Supplementary Table 1). These 19 loci included all ten loci from previous GWAS of BMI6–10, two loci previously associated with body weight10 (at FAIM2 and SEC16B) and one locus previously associated with waist circumference (near TFAP2B). The remaining six loci, near GPRC5B, MAP2K5-LBXCOR1, TNNI3K, LRRN6C, FLJ35779-HMGCR and PRKD1, have not previously been associated with BMI or other obesity-related traits.

Stage 2 follow up identifies additional new loci for BMI

To identify additional BMI-associated loci and to validate the loci that reached genome-wide significance in the stage 1 analyses, we examined SNPs representing 42 independent loci (including the 19 genome-wide significant loci) having a stage 1 P < 5 \times 10^{-6}. Variants were considered to be independent if the pair-wise linkage disequilibrium (LD, \textit{r}2) was less than 0.1 and if they were separated by at least 1 Mb. In stage 2, we examined these 42 SNPs in up to 125,931 additional individuals (79,561 newly genotyped individuals from 16 different studies and 46,370 individuals from 18 additional studies for which genome-wide association data were available; Table 1, Supplementary Note and Online Methods). In a joint analysis of stage 1 and stage 2 results, 32 of the 42 SNPs reached P < 5 \times 10^{-8} (Table 1, Supplementary Table 1 and Supplementary Figs. 1 and 2). Even after excluding SNPs within the 32 confirmed BMI loci, we still observed an excess of small P values compared to the distribution expected under the null hypothesis (Fig. 1b and Supplementary Fig. 3), suggesting that more BMI loci remain to be uncovered.

The 32 confirmed associations included all 19 loci with P < 5 \times 10^{-8} at stage 1, 12 additional new loci near RBP1-ADCY3-POMC, QPCTL-GIPR, SLC39A8, TMEM160, FANCI, CADM2, LRP1B, PTBP2, MTIF3-GTF3A, ZNF608, RPL27A-TUB and NUDT3-HMGCR and one locus (in NRXN3) previously associated with waist circumference (near TFAP2B). The remaining six loci, near GPRC5B, MAP2K5-LBXCOR1, TNNI3K, LRRN6C, FLJ35779-HMGCR and PRKD1, have not previously been associated with BMI or other obesity-related traits.
with body weight\textsuperscript{10} or waist circumference\textsuperscript{14,15}, whereas 18 new loci had not previously associated with any obesity-related trait in the general population. Although we confirmed all loci previously established by large-scale GWAS for BMI\textsuperscript{6-10} and waist circumference\textsuperscript{14,15}, four loci previously identified in GWAS for early-onset or adult morbid obesity\textsuperscript{16,17} (at \textit{NPC1}, rs1805081, \(P = 0.0025\); \textit{MAF}, rs1424233, \(P = 0.25\); \textit{PTER}, rs10508503, \(P = 0.64\); and TNKS-MSRA, rs473034, \(P = 0.23\)) showed limited or no evidence of association with BMI in our study.

### Table 1 Stage 1 and stage 2 results of the 32 SNPs that were associated with BMI at genome-wide significant \((P < 5 \times 10^{-8})\) levels

<table>
<thead>
<tr>
<th>SNP</th>
<th>Nearest gene</th>
<th>Other nearby genes\textsuperscript{a}</th>
<th>Chr.</th>
<th>Position\textsuperscript{b} (bp)</th>
<th>Alleles\textsuperscript{b}</th>
<th>Effect Other</th>
<th>Frequency effect allele</th>
<th>Per allele change in BMI ((\beta \text{ (s.e.m.))})</th>
<th>Explained variance (%)</th>
<th>Stage 1 (P)</th>
<th>Stage 2 (P)</th>
<th>Stage 1 + 2 (n)</th>
<th>(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1558902</td>
<td>FTO</td>
<td>APOB48R (Q,M), SULT1A2 (Q,M), AC138894.2 (M), ATXN2L (M), TUFM (Q)</td>
<td>16</td>
<td>52,361,075</td>
<td>A T</td>
<td>0.42</td>
<td>0.39 (0.02)</td>
<td>0.34%</td>
<td>2.05 (\times 10^{-6})</td>
<td>1.01 (\times 10^{-6})</td>
<td>192,344</td>
<td>4.8 (\times 10^{-10})</td>
<td></td>
</tr>
<tr>
<td>rs2867125</td>
<td>TMEM18</td>
<td>NDUF53 (Q), CUGBP1 (Q)</td>
<td>3</td>
<td>187,317,193</td>
<td>T A</td>
<td>0.82</td>
<td>0.14 (0.03)</td>
<td>0.03%</td>
<td>7.61 (\times 10^{-14})</td>
<td>1.15 (\times 10^{-6})</td>
<td>96,221</td>
<td>1.69 (\times 10^{-18})</td>
<td></td>
</tr>
<tr>
<td>rs571312</td>
<td>MC4R (B)</td>
<td>SH2B1 (Q,B,M)</td>
<td>11</td>
<td>27,682,562</td>
<td>A T</td>
<td>0.78</td>
<td>0.19 (0.03)</td>
<td>0.07%</td>
<td>5.53 (\times 10^{-13})</td>
<td>1.17 (\times 10^{-14})</td>
<td>204,158</td>
<td>4.69 (\times 10^{-26})</td>
<td></td>
</tr>
<tr>
<td>rs10938397</td>
<td>GNPDA2</td>
<td>KCTD15</td>
<td>19</td>
<td>3,091,372</td>
<td>G A</td>
<td>0.67</td>
<td>0.06 (0.02)</td>
<td>0.00%</td>
<td>1.31 (\times 10^{-9})</td>
<td>2.40 (\times 10^{-2})</td>
<td>192,872</td>
<td>3.01 (\times 10^{-9})</td>
<td></td>
</tr>
</tbody>
</table>

\textsuperscript{a}Genes within ± 500 kb of the lead SNP. \textsuperscript{b}Positions according to Build 36 and allele coding based on the positive strand. \textsuperscript{c}Effect sizes in kgs\textsuperscript{2} obtained from stage 2 cohorts only.

Chr., chromosome; Q, association and eQTL data converge to affect gene expression; B, biological candidate; M, BMI-associated variant is in strong LD \((r^2 \geq 0.75)\) with a missense variant in the indicated gene; C, CNV.
As could be expected, the effect sizes of the 18 newly discovered loci are slightly smaller, for a given minor allele frequency, than those of the previously identified variants (Table 1 and Fig. 1c). The increased sample size used here also brought out more signals with low minor allele frequency. The BMI-increasing allele frequencies for the 18 newly identified variants ranged from 4% to 87%, covering more of the allele frequency spectrum than previous, smaller GWAS of BMI (24%–83%) (Table 1 and Fig. 1c).

We tested for evidence of non-additive (dominant or recessive) effects, SNP \times SNP interaction effects and heterogeneity by sex or study among the 32 BMI-associated SNPs (Online Methods). We found no evidence for any such effects (all \( P > 0.001 \) and no significant results were seen after correcting for multiple testing) (Supplementary Table 1 and Supplementary Note).

**Impact of the 32 confirmed loci on BMI, obesity, body size and other metabolic traits**

Together, the 32 confirmed BMI loci explained 1.45% of the inter-individual variation in BMI in the stage 2 samples, with the FTO SNP accounting for the largest proportion of the variance (0.34%) (Table 1). To estimate the cumulative effect of the 32 variants on BMI, we constructed a genetic susceptibility score that summed the number (Table 1 and Supplementary Note).

For each unit increase in the genetic-susceptibility score, which is approximately equivalent to having one additional risk allele, BMI increased by 0.17 kg/m\(^2\), the equivalent of a 6.99–8.85 kg body weight difference in adults of 160–180 cm in height (Fig. 2a). Still, we note that the predictive value for obesity risk and BMI of the 32 variants combined was modest, although it was statistically significant (Fig. 2b and Supplementary Fig. 4). The area under the receiver-operating characteristic (ROC) curve for prediction of risk of obesity (BMI \( \geq 30 \) kg/m\(^2\)) using a model including age, sex and any previously known BMI loci was 0.515 (\( P = 0.023 \)) compared to the area under the curve (AUC) of 0.50, which increased to 0.575 (\( P < 10 \)\(^{-5} \)) when the 32 confirmed SNPs were also included in the model (Fig. 2b). The area under the ROC curve for the model including the 32 SNPs only was 0.574 (\( P < 10 \)\(^{-5} \)).

All 32 confirmed BMI-increasing alleles showed directionally consistent effects on the risk of being overweight (BMI \( \geq 25 \) kg/m\(^2\)) or obese (BMI \( \geq 30 \) kg/m\(^2\)) in the stage 2 samples, with 30 of 32 variants achieving at least nominally significant associations. The BMI-increasing alleles increased the odds of being overweight by 1.013- to 1.38-fold and the odds of being obese by 1.016- to 1.203-fold (Supplementary Table 2). In addition, 30 of the 32 loci also showed directionally consistent effects on the risk of extreme and early-onset obesity in a meta-analysis of seven case-control studies of adults and children (binomial sign test \( P = 1.3 \times 10 \)\(^{-7} \)) (Supplementary Table 3). The BMI-increasing allele observed in adults also increased the BMI in children and adolescents with directionally consistent effects observed for 23 of the 32 SNPs (binomial sign test \( P = 0.01 \)). Furthermore, in family-based studies, the BMI-increasing allele was over-transmitted to the obese offspring for 24 of the 32 SNPs (binomial sign test \( P = 0.004 \)) (Supplementary Table 3). As these studies in extreme obesity cases, children and families were relatively small (with \( n \) ranging from 354 to 15,251 individuals) compared to the overall meta-analyses, their power was likely insufficient to confirm association for 32 loci. Nevertheless, these results show that the effects are unlikely to reflect population stratification and that they extend to BMI differences throughout the life course.
All BMI-increasing alleles were associated with increased weight, as could be expected from the correlation between BMI and body weight (Supplementary Table 2). To confirm an effect of the loci on adiposity rather than general body size, we tested for association with body fat percentage, for which data was available in a subset of the stage 2 replication samples (n = 5,359 to n = 28,425) (Supplementary Table 2). The BMI-increasing allele showed directionally consistent effects on body fat percentage at 31 of the 32 confirmed loci (binomial sign test $P = 1.54 	imes 10^{-5}$) (Supplementary Table 2).

We also examined the association of the BMI loci with metabolic traits (type 2 diabetes, fasting glucose, fasting insulin, indices of β-cell function (HOMA-B) and insulin resistance (HOMA-IR), and blood lipid levels) and with height (Supplementary Tables 2 and 4). Although many nominal associations were expected because of known correlations between BMI and most of these traits, and because of overlap in samples, several associations stood out as possible examples of pleiotropic effects of the BMI-associated variants. Particularly interesting is the variant in the GIPR locus, where the BMI-increasing allele is also associated with increased fasting glucose levels and lower 2-h glucose levels (Supplementary Table 4). The direction of the effect is opposite to what would be expected due to the correlation between obesity and glucose intolerance but is consistent with the suggested roles of GIPR in glucose and energy metabolism (see below). Three loci showed strong associations ($P < 10^{-8}$) with height (MCAR, RBP1-ADCY3-POMC and MTCH2-NDUF53). Because BMI is weakly correlated with height (and indeed, the BMI-associated variants as a group show no consistent effect on height), these associations are also suggestive of pleiotropy. Notably, analogous to the effects of severe mutations in POMC and MCAR on height and weight, the BMI-increasing alleles of the variants near these genes were associated with decreased (POMC) and increased (MCAR) height, respectively (Supplementary Table 2).

**Potential functional roles and pathway analyses**

Although associated variants typically implicate genomic regions rather than individual genes, we note that some of the 32 loci include candidate genes with established connections to obesity. Several of the ten previously identified loci are located in or near genes that encode neuronal regulators of appetite or energy balance, including MCAR, BDNF and SH2B1. Each of these genes has been tied to obesity, not only in animal models, but also by rare human variants that disrupt each of these genes and lead to severe obesity. Using the automated literature search program Snipper (Online Methods), we identified various genes within the newly discovered loci with potential biological links to obesity susceptibility (Supplementary Note). Among the new loci, the location of rs713586 near POMC provides further support for a role of neuroendocrine circuits that regulate energy balance in susceptibility to obesity. POMC encodes several polypeptides, including α-MSH, a ligand of the MCAR gene product, and rare mutations in POMC also cause obesity in humans.

In contrast, the locus near GIPR, which encodes a receptor of gastric inhibitory polypeptide (GIP), suggests a role for peripheral biology in obesity. GIP, which is expressed in the K cell of the duodenum and intestine, is an incretin hormone that mediates incremental insulin secretion in response to oral intake of glucose. The variant associated with BMI is in strong LD ($r^2 = 0.83$) with a missense SNP in GIPR (rs1800437, p.Glu354Gln) that has recently been shown to influence glucose and insulin response to an oral glucose challenge. Although no human phenotype is known to be caused by mutations in GIPR, mice with disruption of Gipr are resistant to diet-induced obesity. The association of a variant in GIPR with BMI suggests that there may be a link between incretins, insulin secretion and body weight regulation in humans as well.

To systematically identify biological connections among the genes located near the 32 confirmed SNPs and to potentially identify new pathways associated with BMI, we performed pathway-based analyses using MAGENATAX. Specifically, we tested for enrichment of genetic associations to BMI in biological processes or molecular functions that contain at least one gene from the 32 confirmed BMI loci (Online Methods). Using annotations from the Kyoto Encyclopedia of Genes and Genomes (KEGG), Ingenuity, Protein Analysis Through Evolutionary Relationships (PANTHER) and Gene Ontology databases, we found evidence of enrichment for pathways involved in platelet-derived growth factor (PDGF) signaling (PANTHER, $P = 0.0008$, false discovery rate (FDR) = 0.0061), translation elongation (PANTHER, $P = 0.0008$, FDR = 0.0066), hormone or nuclear-hormone receptor binding (Gene Ontology, $P < 0.0005$, FDR < 0.0085), homeobox transcription (PANTHER, $P = 0.0001$, FDR = 0.011), regulation of cellular metabolism (Gene Ontology, $P = 0.0002$, FDR = 0.031), neurogenesis and neuron differentiation (Gene Ontology, $P < 0.0002$, FDR < 0.034), protein phosphorylation (PANTHER, $P = 0.0001$, FDR = 0.045).
and numerous other pathways related to growth, metabolism, immune and neuronal processes (Gene Ontology, \( P < 0.002, \text{FDR} < 0.046 \)) (Supplementary Table 5).

### Identifying possible functional variants

We used data from the 1000 Genomes Project and the HapMap Consortium to explore whether the 32 confirmed BMI SNPs were in LD (\( r^2 \geq 0.75 \)) with common missense SNPs or copy number variants (CNVs) (Online Methods). Non-synonymous variants in LD with our signals were present in \( BDNF, \) \( SLC39A8, \) \( FLJ35779-HMGCR, \) \( QPCTL-GIPR, \) \( MTCH2, \) \( ADCY3 \) and \( LBXCOR1. \) In addition, the rs7359397 signal was in LD with coding variants in several genes including \( SH2B1, \) \( ATNXL2, \) \( APOB48R, \) \( SULT1A2 \) and \( ACI38894.2 \) (Table 1, Fig. 3, Supplementary Table 6 and Supplementary Fig. 2). Furthermore, two SNPs tagged common CNVs. The first CNV has been previously identified\(^9\) and is a 45-kb deletion near \( NEGR1. \) The second CNV is a 21-kb deletion that lies 50 kb upstream of \( GPRCSB; \) the deletion allele is tagged by the T allele of rs12444979 (\( r^2 = 1.0 \)) that comprise the best haplotype associating with BMI. Plots were generated using LocusZoom (see URLs).

Evidence for the existence of additional associated variants

Because the variants identified by this large study explain only 1.45% of the variance in BMI (2%–4% of genetic variance based on an estimated heritability of 40%–70%), we considered how much the explained phenotypic variance could be increased by including more SNPs at various degrees of significance in a polygene model using an independent validation set (Online Methods)\(^{37,38}\). We found that including SNPs associated with BMI at lower significance levels (up to \( P > 0.05 \)) increased the explained phenotypic variance in BMI to 2.5%, or 4%–6% of the genetic variance (Fig. 4a). In a separate analysis, we estimated the total number of independent BMI-associated variants that are likely to exist with similar effect sizes as the 32 confirmed here (Online Methods)\(^{38}\). Based on the effect size and allele frequencies of the 32 replicated loci observed in stage 2 and the power to detect association in stage 1 and stage 2 combined, we estimated that there are 284 (95% CI 132–510) loci with similar effect sizes as those currently observed, which together would account for 4.5% (95% CI 3.1%–6.8%) of the phenotypic variation or 6%–11% of the genetic variation in BMI (based on an estimated heritability of 40%–70%) (Supplementary Table 8). In order to detect 95% of these loci, a sample size of approximately 730,000 subjects would be needed (Fig. 4b). This method does not account for the potential of loci of smaller effect than those identified here to explain even more of the variance and thus provides an estimated lower bound of explained variance. These two analyses strongly suggest that larger GWAS will...
Figure 4  Phenotypic variance explained by common variants. (a) The variance explained is higher when SNPs not reaching genome-wide significance are included in the prediction model. The y axis represents the proportion of variance explained at different $P$ value thresholds from the stage 1 meta-analysis. Results are given for three studies (Rotterdam Study II (RSII), Rotterdam Study III (RSIII), Queens Institute of Medical Research (QIMR)) which were not included in the meta-analysis, after exclusion of all samples from The Netherlands (for RSII and RSIII) and the United Kingdom (for QIMR) from the discovery analysis for this sub-analysis. The dotted line represents the weighted average of the explained variance of three validation sets. (b) Cumulative number of susceptibility loci expected to be discovered, including those we have already identified and others that have yet to be detected, by the expected percentage of phenotypic variation explained and the sample size required for a one-stage GWAS assuming a genomic control correction is used. The projections are based on loci that achieved a significance level of $P < 5 \times 10^{-8}$ in the joint analysis of stage 1 and stage 2 and the distribution of their effect sizes in stage 2. The dotted red line corresponds to the expected phenotypic variance explained by the 22 loci that are expected to be discovered in a one-stage GWAS using the sample size of stage 1 of this study.

We examined whether selecting only a single variant from each locus for follow up led us to underestimate the fraction of phenotypic variation explained by the associated loci. To search for additional independent loci at each of the 32 associated BMI loci, we repeated our genome-wide association meta-analysis conditioning on the 32 confirmed SNPs. Using a significance threshold of $P = 5 \times 10^{-8}$ for SNPs at known loci, we identified one apparently independent signal at the $MC4R$ locus; rs7227255 was associated with BMI ($P = 6.56 \times 10^{-7}$) even after conditioning for the most strongly associated variant near $MC4R$ (rs571312) (Fig. 5). Notably, rs7227255 is in perfect LD ($r^2 = 1$) with a relatively rare $MC4R$ missense variant (rs2229616, p.Val103Ile, minor allele frequency = 1.7%) that has been associated with BMI in two independent meta-analyses.

Furthermore, mutations at the $MC4R$ locus are known to influence early-onset obesity$^{24,41}$, supporting the notion that allelic heterogeneity may be a frequent phenomenon in the genetic architecture of obesity.

DISCUSSION

Using a two-stage genome-wide association meta-analysis of up to 249,796 individuals of European descent, we identified 18 additional loci that are associated with BMI at genome-wide significance, bringing the total number of such loci to 32. We estimate that more than 250 common variant loci (that is, 284 predicted loci minus 32 confirmed loci) with effects on BMI similar to those described here remain to be discovered and that even larger numbers of loci with smaller effects remain to be identified. A substantial proportion of these loci should be identifiable through larger GWAS and/or by targeted follow up of the top signals selected from our stage 1 analysis. The latter approach is already being implemented through large-scale genotyping of samples informative for BMI using a custom array (the Metaochip) designed to support follow up of thousands of promising variants in hundreds of thousands of individuals.

The combined effect on BMI of the associated variants at the 32 loci is modest, and even when we try to account for as yet undiscovered variants with similar properties, we estimate that these common variant signals account for only 6%–11% of the genetic variation in BMI. There is a strong expectation that additional variance and biology will be explained using complementary approaches that capture variants not examined in the current study, such as lower frequency variants and short insertion–deletion polymorphisms. There is good reason to believe (based on our findings at $MC4R$ and other loci, such as those at $POMC$, $BDNF$ and $SH2B1$, which feature both common and rare variant associations) that a proportion of such low-frequency and rare causation will map to the loci already identified by GWAS.

A primary goal of human genetic discovery is to improve understanding of the biology

Figure 5  A second signal at the $MC4R$ locus contributing to BMI. SNPs are plotted by position in a 1-Mb window of chromosome 18 against association with BMI ($-\log_{10} P$). (a) Plot highlighting the most significant SNP in the stage 1 meta-analysis. (b) Plot highlighting the most significant SNP after conditional analysis, where the model included the most strongly associated SNP as a covariate. Estimated recombination rates (from HapMap) are plotted in cyan to reflect the local LD structure. The SNPs surrounding the most significant SNP are color coded to reflect their LD with this SNP (taken from pairwise $r^2$ values from the HapMap CEU database). Genes, exons and the direction of transcription from the UCSC genome browser are noted. Hashmarks at the top of the figure represent the positions of SNPs in the meta-analysis. Regional plots were generated using LocusZoom.
of conditions such as obesity\cite{2}. One particularly noteworthy finding in this regard is the association between BMI and common variants near \textit{GIPR}, which may indicate a causal contribution of variation in postprandial insulin secretion in the development of obesity. In most instances, the loci identified by the present study harbor few, if any, annotated genes with clear connections to the biology of weight regulation. This reflects our still limited understanding of the biology of BMI and obesity-related traits and is in striking contrast with the results from equivalent studies of certain other traits (such as autoimmune diseases or lipid levels). Thus, these results suggest that much of the biology that underlies obesity remains to be uncovered and that GWAS may provide an important entry point for investigation.

In conclusion, we performed GWAS in large samples to identify numerous genetic loci associated with variation in BMI, a common measure of obesity. Because current lifestyle interventions are largely ineffective in addressing the challenges of growing obesity\cite{3,4}, new insights into the biology of obesity are critically needed to guide the development and application of future therapies and interventions.


**METHODS**

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

**ACKNOWLEDGMENTS**

A full list of acknowledgments appears in the Supplementary Note.

Funding was provided by Academy of Finland (10404, 77299, 104781, 114382, 117797, 120315, 121584, 124243, 126775, 129825, 127437, 129269, 129306, 129494, 129640, 136326, 201902, 201595, 213225, 213506 and 216374); ADA (014-93-015; RIDE2); Robert Dawson Evans Endowment; Royal Society (UK); Michael Donovan Charitable Foundation; Perklén Foundation; Petrus and Augusta Hultén Foundation; Marie Curie Intra-European Fellowship; Medical Research Council (UK) (G0000649, G0000934, G0521010D, G0505039, G0600331 and G0601261, PrevMetSyn); Ministry of Cultural Affairs and Social Ministry of the Federal State of Mecklenburg-West Pomerania; Ministry for Health, Welfare and Sports (The Netherlands); Ministry of Education (Finland); Ministry of Education, Culture and Science (The Netherlands); Ministry of Internal Affairs and Health (Denmark); Ministry of Science, Education and Sport of the Republic of Croatia (216-1080315-0302); Ministry of Science, Research and the Arts Baden-Württemberg; Montreal Heart Institute Foundation; Municipal Health Care Center and Hospital in Jakobstad; Municipality of Rotterdam; Närpes Health Care Foundation; National Cancer Institute; National Health and Medical Research Council of Australia; National Institute for Health Research Oxford Biomedical Research Centre; National Institute for Health Research comprehensive Biomedical Research Centre; US National Institutes of Health of the paper at http://www.nature.com/naturegenetics/.

1. note: Supplementary information is available on the Nature Genetics website.

**REFERENCES**

A full list of references appears in the Supplementary Note.
Neuronal SH2B1 is essential for controlling energy and glucose homeostasis
A common variant in the
Non-synonymous polymorphisms in melanocortin-4 receptor
© 2010 Nature America, Inc. All rights reserved.

ARTICLES


Cardiovascular Genetics team, Centre Investigation Clinique (CIC) 9501, Nancy, France. 206Steno Diabetes Center, Gentofte, Denmark. 207Center for Human Genomics, Wake Forest University, Winston-Salem, North Carolina, USA. 208School of Pathology and Laboratory Medicine, University of Western Australia, Nedlands, Western Australia, Australia. 209School of Medicine and Pharmacology, University of Western Australia, Perth, Western Australia, Australia. 210Department of Clinical Physiology, University of Tampere and Tampere University Hospital, Tampere, Finland. 211Stanford University School of Medicine, Stanford, California, USA. 212Department of Psychiatry and Human Behavior, University of California, Irvine (UCI), Irvine, California, USA. 213Leipziger Interdisziplinärer Forschungs-komplex zu molekularen Ursachen umwelt- und lebensstilassoziierter Erkrankungen (LIFE) Study Centre, University of Leipzig, Leipzig, Germany. 214Service of Medical Genetics, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, Lausanne, Switzerland. 215Human Genetics Center and Institute of Molecular Medicine, University of Texas Health Science Center, Houston, Texas, USA. 216Faculty of Health Science, University of Southern Denmark, Odense, Denmark. 217New York University Medical Center, New York, New York, USA. 218National Institute for Health and Welfare, Department of Mental Health and Substance Abuse Services, Unit for Child and Adolescent Mental Health, Helsinki, Finland. 219NIHR Oxford Biomedical Research Centre, Churchill Hospital, Oxford, UK. 220Department of Urology, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands. 221Institute of Biomedical Sciences, University of Copenhagen, Copenhagen, Denmark. 222Faculty of Health Science, University of Aarhus, Aarhus, Denmark. 223Department of Psychiatry, Leiden University Medical Centre, Leiden, The Netherlands. 224Department of Psychiatry, University Medical Centre Groningen, Groningen, The Netherlands. 225Department of Neurology, University of Lübeck, Lübeck, Germany. 226Institute for Paediatric Nutrition Medicine, Vestische Hospital for Children and Adolescents, University of Witten-Herdecke, Datteln, Germany. 227Department of Medicine III, Prevention and Care of Diabetes, University of Dresden, Dresden, Germany. 228Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, Maryland, USA. 229Hjelt Institute, Department of Public Health, University of Helsinki, Helsinki, Finland. 230South Ostrobothnia Central Hospital, Seinajoki, Finland. 231Department of Internal Medicine, Centre Hospitalier Universitaire Vaudois (CHUV) University Hospital, Lausanne, Switzerland. 232The Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, USA. 233Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA. 234Pacific Biosciences, Menlo Park, California, USA. 235Sage Bionetworks, Seattle, Washington, USA. 236Division of Biostatistics, Washington University School of Medicine, St. Louis, Missouri, USA. 237Division of Intramural Research, National Heart, Lung, and Blood Institute, Framingham Heart Study, Framingham, Massachusetts, USA. 238Department of Epidemiology and Population Health, Albert Einstein College of Medicine, New York, New York, USA. 239Department of Genetics, University of North Carolina, Chapel Hill, North Carolina, USA. 240Department of Medical Genetics, University of Helsinki, Helsinki, Finland. 241Laboratory of Genetics, National Institute on Aging, Baltimore, Maryland, USA. 242Division of Community Health Sciences, St. George’s, University of London, London, UK. 243Klinikum Grosshadern, Munich, Germany. 244Faculty of Medicine, University of Iceland, Reykjavik, Iceland. 245University of Cambridge Metabolic Research Labs, Institute of Metabolic Science Addenbrooke’s Hospital, Cambridge, UK. 246Carolina Center for Genome Sciences, School of Public Health, University of North Carolina Chapel Hill, Chapel Hill, North Carolina, USA. 247Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. 248These authors contributed equally to this work. Correspondence should be addressed to M.B. (boehnke@umich.edu), K. Stefansson (kstefans@decode.is), K.E.N. (kari_north@unc.edu), M.I.M. (mark.mccarthy@drl.ox.ac.uk), J.N.H. (joelh@broadinstitute.org), E.I. (erik.ingelsson@ki.se) or R.J.F.L. (ruth.loos@mrc-epid.cam.ac.uk).
ONLINE METHODS

Study design. We designed a multistage study (Supplementary Fig. 1) comprising a genome-wide association meta-analysis (stage 1) of data on up to 123,865 genotyped individuals from 46 studies and selected 42 SNPs with \( P < 5 \times 10^{-6} \) for follow up in stage 2. Stage 2 comprised up to 125,931 additional genotyped individuals from 42 studies. Meta-analysis of stage 1 and stage 2 summary statistics identified 32 SNPs that reached genome-wide significance (\( P < 5 \times 10^{-8} \)).

Stage 1 genome-wide association meta-analysis. Samples and genotyping. The GIANT consortium currently encompasses 46 studies with up to 123,865 genotyped adult individuals of European ancestry with data on BMI (Supplementary Note). The samples from 46 studies, including between 276 and 26,799 individuals each, were genotyped using Affymetrix and Illumina whole genome genotyping arrays (Supplementary Note). To allow for meta-analysis across different marker sets, imputation of polymorphic HapMap European CEU SNPs (Supplementary Note) was performed using MACH, IMPUTE or BimBam.

Association analysis with BMI. Each study performed single marker association analyses with BMI using an additive genetic model implemented in MACH2QTL (Y. Li, C.J.W., P.S. Ding and G.R.A., unpublished data), Merlin, SNPTOOLKIT, ProbABEL, GenABEL, LME in R or PLINK. BMI was adjusted for age, age² and other appropriate covariates (for example, principal components) and inverse normally transformed to a mean of 0 and a standard deviation of 1. Analyses were stratified by sex and case status (for samples ascertainment for other diseases) (Supplementary Note). To allow for relatedness in the SardiNIA, Framingham Heart, Amish HAPI Heart and Family Heart studies, regression coefficients were estimated in the context of a variance component model that modeled relatedness in men and women combined with sex as a covariate. Before meta-analyzing the genome-wide association data for the 46 studies, SNPs with poor imputation quality scores \( (r^2 < 0.3 \text{ in MACH, observed/expected dosage} < 0.3 \text{ in BimBam or proper_info < 0.4 in IMPUTE}) \) and those with a minor allele count (MAC = 2 × N × minor allele frequency) of < 6 in each sex- and case-specific stratum were excluded for each study. All individual GWAS were genomic control corrected before meta-analysis. Individual study-specific genomic control values ranged from 0.983 to 1.104 (Supplementary Note).

Meta-analysis of stage 1 association results. Next, we performed the stage 1 meta-analysis using the inverse variance method, which is based on \( \beta \) values and standard errors from each individual GWAS. To ensure consistency of results, we also performed the stage 1 meta-analysis using the weighted \( z \)-score method, which is based on the direction of association and \( P \) values of each of the individual studies. Both meta-analyses were performed using METAL (see URLs), and the correlation between the resulting \(-\log_{10} P\) values was high (\( r > 0.99 \)). For the discovery of replicating variants, the results of the inverse variance meta-analysis were used followed by a final genomic control correction of the meta-analyzed results. The genomic control value for the meta-analyzed results before genomic meta-control correction was 1.318.

Selection of SNPs for follow up. Forty-two lead SNPs, representing the forty-two most significant \( (P < 5 \times 10^{-6}) \) independent loci, were selected for replication analyses (stage 2) (Supplementary Table 1). Loci were considered independent when separated by at least 1 Mb. For some loci, the SNP with the strongest association could not be genotyped for technical reasons and was substituted by a proxy SNP that was in high LD with it \( (r^2 > 0.8) \) according to the HapMap CEU data (Supplementary Table 1). We tested the association of these 42 SNPs in 16 de novo and 18 in silico replication studies in stage 2.

Stage 2 follow up. Samples and genotyping. Directly genotyped data for the 42 SNPs was available from a total of 79,561 adults of European ancestry from 16 studies using Sequenom iPLEX or TaqMan assays (Supplementary Note). Samples and SNPs that did not meet the quality control criteria defined by each individual study were excluded. Minimum genotyping quality control criteria were defined as Hardy-Weinberg Equilibrium \( P > 10^{-6} \), call rate > 90% and concordance > 99% in duplicate samples in each of the follow-up studies. Association results were also obtained for the 42 most significant SNPs from 46,370 individuals of European ancestry from 18 GWAS that had not been included in the stage 1 analyses (Supplementary Note). Studies included between 345 and 22,888 individuals and were genotyped using Affymetrix and Illumina genome-wide genotyping arrays. Autosomal HapMap SNPs were imputed using either MACH or IMPUTE with poor imputation quality scores from the in silico studies \( (r^2 < 0.3 \text{ in MACH or proper_info < 0.4 in IMPUTE}) \), and SNPs with a MAC < 6 in each sex- and case-specific stratum were excluded.

Association analyses and meta-analysis. We tested the association between the 42 SNPs and BMI in each in silico and de novo stage 2 study separately as described for the stage 1 studies. We subsequently meta-analyzed \( \beta \) values and standard errors from the stage 2 studies using the inverse-variance method. The meta-analysis using a weighted \( z \)-score method was similar (the \( r \) between \( P \) values was >0.99) and included up to 249,796 individuals. Data was available for at least 179,000 individuals for 41 of the 42 SNPs. For one SNP (rs6955651 in KIAA1305), data was only available for 125,672 individuals due to technical challenges relating to the genotyping and imputation of this SNP. Next, we meta-analyzed the summary statistics of the stage 1 and stage 2 meta-analyses using the inverse-variance method in METAL.

Assessment of population stratification. To assess for possible inflation of test statistics by population stratification, we performed a family-based analysis, which is immune to stratification, in 5,507 individuals with pedigree information from the Framingham Heart Study using that the QFAM–within procedure in PLINK. Effect sizes and directions in the Framingham Heart Study data are the \( \beta \) statistics reported by PLINK from the within-family analysis, and the \( P \) values are empirical and are based on permutation testing. For imputed SNPs, only those with \( r^2 > 0.3 \) in MACH were analyzed using the best-guess genotypes from dosages reports by MACH. For the 32 loci in general and the 18 new loci in particular, the estimated effect sizes on BMI were essentially identical in the overall meta-analysis and in the Framingham Heart Study sample (Supplementary Note), and, as expected in the absence of substantial stratification, about half of the loci (18 out of 32 loci total and 10 out of 18 new loci) had a larger effect size in the family-based sample. These results indicate that the genome-wide significant associations in our meta-analysis are not substantially confounded by stratification.

In addition, we estimated the fixation index \( (Fst) \) for all SNPs to test whether the 32 confirmed BMI SNPs might be false-positive results due to population stratification. We selected five diverse European populations with relatively large sample sizes (Northern Finland Birth Cohort (NFBC), British 1958 Birth Cohort, SardiNIA, CoLaus and DeCODE) for this analysis. The mean \( Fst \) value for the 32 confirmed BMI SNPs was not significantly different from the mean \( Fst \) for 2.1 million non-BMI associated SNPs \( (t = 0.28) \), suggesting that the SNPs that are associated with BMI do not appear to have strong allele frequency differences across the European samples examined.

Follow-up analyses. Subsequently, we performed an extensive series of follow-up analyses to estimate the impact of the 32 confirmed BMI loci in adults and children and to explore their potential functional roles. These follow-up analyses are described in detail in the Supplementary Note.

In brief, we estimated the cumulative effect of the 32 loci combined on BMI and assessed their predictive ability in obesity and BMI in the ARIC study. Association between the 32 confirmed BMI variants and overweight or obese status was assessed in stage 2 samples, and association with BMI in children and adolescents was examined in four population-based studies. Furthermore, we tested for association between the 32 SNPs and extreme or early-onset obesity in seven case-control studies of extremely obese adults and extremely obese children or adolescents. Data on the association between the 32 SNPs and height and weight were obtained from the stage 2 replication samples, and data on the association with related traits were extracted from previously reported genome-wide association meta-analyses for type 2 diabetes (Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium), lipid levels (the Global Lipids
Genetics Consortium 20) and glycemic traits (Meta-Analyses of Glucose and Insulin-related traits Consortium (MAGIC) 19, 21).

To discover potentially new pathways associated with BMI, we tested whether predefined biological processes or molecular functions that contain at least one gene within 300 kb of the 32 confirmed BMI SNPs were enriched for multiple modest BMI associations using MAGENTA 3. We identified SNPs having $r^2 \geq 0.75$ with the lead SNP that were likely non-synonymous, nonsense or which occurred within 5 bp of the exon-intron boundary and also evaluated whether any of the 32 confirmed BMI SNPs tagged common CNVs. We examined the cis associations between each of the 32 confirmed BMI SNPs and expression of nearby genes in adipose tissue 34, 52, whole blood 34, lymphocytes 36, 52 and brain 35.

We evaluated the amount of phenotypic variance explained by the 32 BMI loci using a method proposed by the International Schizophrenia Consortium 37 and estimated the number of susceptibility loci that are likely to exist using a new method 48 based on the distribution of effect sizes and minor allele frequencies observed for the established BMI loci and the power to detect those effects in the combined stage 1 and stage 2 analysis.

We performed a conditional genome-wide association analysis to examine whether any of the 32 confirmed BMI loci harbored additional independent signals, and we also examined gene-by-gene and gene-by-sex interactions among the BMI loci. Dominant and recessive analyses were performed for the 32 confirmed BMI SNPs to test for non-additive effects.