

Common body mass index-associated variants confer risk of extreme obesity

Chris Cotsapas^{1,2,3,†}, Elizabeth K. Speliotes^{3,4,5,†}, Ida J. Hatoum^{5,6}, Danielle M. Greenawalt⁷, Radu Dobrin⁷, Pek Y. Lum⁷, Christine Suver⁷, Eugene Chudin⁷, Daniel Kemp⁸, Marc Reitman⁸, Benjamin F. Voight^{1,2,3}, Benjamin M. Neale^{1,2,3}, Eric E. Schadt⁷, Joel N. Hirschhorn^{3,4,9}, Lee M. Kaplan^{5,6,*}, Mark J. Daly^{1,2,3,*} and the GIANT Consortium[‡]

¹Center for Human Genetic Research and ²Gastrointestinal Unit, Massachusetts General Hospital, Boston, MA, USA, ³Department of Medicine, Harvard Medical School, Boston, MA, USA, ⁴Medical and Population Genetics Program and ⁵Metabolism Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA, ⁶MGH Weight Center, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA, ⁷Genetics Rosetta Inpharmatics LLC, a wholly owned subsidiary of Merck & Co., Inc., Seattle, WA 98109, USA, ⁸Department of Metabolic Disorders, Merck Research Laboratories, Rahway, NJ 07065, USA and ⁹Program in Genomics and Divisions of Endocrinology and Genetics, Children's Hospital, Boston, MA, USA

Received February 23, 2009; Revised June 10, 2009; Accepted June 17, 2009

To investigate the genetic architecture of severe obesity, we performed a genome-wide association study of 775 cases and 3197 unascertained controls at ~550 000 markers across the autosomal genome. We found convincing association to the previously described locus including the *FTO* gene. We also found evidence of association at a further six of 12 other loci previously reported to influence body mass index (BMI) in the general population and one of three associations to severe childhood and adult obesity and that cases have a higher proportion of risk-conferring alleles than controls. We found no evidence of homozygosity at any locus due to identity-by-descent associating with phenotype which would be indicative of rare, penetrant alleles, nor was there excess genome-wide homozygosity in cases relative to controls. Our results suggest that variants influencing BMI also contribute to severe obesity, a condition at the extreme of the phenotypic spectrum rather than a distinct condition.

INTRODUCTION

The prevalence of extreme obesity, defined as a body mass index (BMI) above 40, is increasingly high in developed countries: the most recent estimates of prevalence in the USA are between 2.2 and 4.4%, with the 20- to 39-year-old age group showing the greatest increase from previous studies (1). Rare mutations in genes encoding the melanocortin 4 receptor, the leptin receptor, small nuclear ribonucleoprotein polypeptide N/necdin and other genes (2,3) have been shown to underlie extreme obesity in family studies, but are not sufficiently frequent to explain the high prevalence in

the general population. To search for additional, more frequent genetic variants that could contribute to extreme obesity, we carried out a genome-wide association analysis of 775 bariatric surgery patients (mean BMI = 50.6) compared with 3197 publicly available controls from the general population. We find that variants near the gene *FTO*, previously associated with increased BMI, are associated with increased risk of extreme obesity with genome-wide statistical significance; six of 12 other loci previously shown to affect BMI in population samples also have nominal associations to extreme obesity, suggesting an overlap between the genetic contribu-

*To whom correspondence should be addressed at: Massachusetts General Hospital, Blake 4, 55 Fruit St., Boston, MA 02114, USA. Tel: +1 617 726-3768; Fax: +1 617 724-6832; Email: kaplan@helix.mgh.harvard.edu (L.M.K.) and Center for Human Genetic Research, Massachusetts General Hospital, CPZN 6818, 185 Cambridge Street, Boston, MA 02114, USA. Tel: +1 6176433290; Fax: +1 6176433293; Email: mjdaly@chgr.mgh.harvard.edu (M.J.D.)

[†]The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

[‡]See appendix for list of participants.

tors to BMI in the general population and the variants that predispose to extreme obesity.

We performed genome-wide genotyping of 655 130 single nucleotide polymorphisms (SNPs) of 972 individuals, including 841 self-identified Caucasians, with BMI > 33.3 (mean 50.61, 722 Caucasians with BMI > 40; see Supplementary Material, Fig. S1) recruited at the Massachusetts General Hospital Weight Center. We obtained publicly available genotypes for 3294 anonymous Caucasian samples genotyped on a comparable Illumina platform (see Materials and Methods). After removing markers and individuals with low-quality data, individuals with appreciable non-European ancestry and correcting for residual stratification, we performed an association study on 775 Caucasian cases and 3197 unascertained controls using 457 251 SNPs. We observed only a modest inflation of association statistics with $\lambda_{GC} = 1.05$ (4), equivalent to 1.04 in a symmetric design totaling 2000 samples (5) suggesting little systematic bias remains in our analysis and that our cases and controls were well matched by ancestry. We calculate that our case-control design of extreme cases, and unascertained controls is equivalent to quantitative trait analysis in a population sample of ~4300 (see Materials and Methods).

RESULTS

We found significant association of rs9941349 to extreme obesity in the previously described (6,7) locus intronic to the *FTO* gene on chromosome 16 ($P = 6.09 \times 10^{-12}$, odds ratio OR = 1.48; see Table 1 for top results). This SNP is in strong linkage disequilibrium with the originally reported rs9939609 ($r^2 = 0.87$ in HapMap/CEU) (8). A second SNP, rs8050136, is in complete LD with the previously reported marker and also shows very strong association ($P < 2.83 \times 10^{-11}$, OR = 1.46). These odds ratios are consistent with the proportion of variance explained in the common population (0.34% in Willer *et al.* 9) given our selection of extreme phenotypes as cases in a threshold model (see Supplementary Material, Fig. S1 and Materials and Methods for details) and in other reports of this association in severe obesity (10–12). No other loci were associated at a genome-wide significance level, suggesting that the remaining common variants interrogated by this genotyping platform are unlikely to have effects of comparable magnitude.

This result, and the recent identification of common alleles that increase BMI in the general population, led us to ask whether our data could distinguish between two extreme models for the genetic basis of extreme obesity. In one case, extreme obesity might simply represent the tail of the population-wide BMI distribution and therefore these individuals would have a greater dose of common BMI increasing alleles than more modestly obese individuals. In the other, extreme obesity might constitute a distinct, Mendelian-like condition and the most extremely obese individuals might then have a reduced burden of these common alleles. Thus we asked whether previously reported BMI-associated variants also modulate risk in our cohort (9,13). We are able to capture all 12 BMI variants (including the *FTO* locus) either directly or through highly correlated markers, and we find nominal evidence of association for six of these 12 with extreme obesity

($P < 0.05$; Table 2) Strikingly, 10 alleles previously shown to increase BMI in the general population were more frequent in extremely obese individuals than in controls, suggesting that as a group, these variants influence both common and extreme obesity (Table 2). More compellingly, we find that cases carry significantly more BMI increasing alleles of these SNPs than controls (Welch's one-sided $P < 1 \times 10^{-14}$). Further, we find some evidence that the more extreme 50% of cases also carry more risk alleles than the remainder (Welch's one-sided $P = 0.038$); this observation is recapitulated when regressing BMI phenotype from BMI risk allele count (ANOVA $P = 0.033$). We therefore suggest that a model of risk allele 'burden' may influence BMI even with the extreme tail of the phenotype.

To further search for rare, penetrant recessive alleles that could underlie extreme obesity, we asked whether our cases tended to have more homozygosity-by-descent than our control population (see Materials and Methods). Overall, our cases did not appear more homozygous than controls, and there was no association between homozygosity at any location and extreme obesity (data not shown). Our case cohort thus appears to represent a phenotypic extreme of the normal population rather than a collection of familial disease patients.

DISCUSSION

Our results suggest that most loci known to influence BMI in the general population also contribute to extreme obesity; this predicts that severely obese individuals should carry more BMI increasing alleles than expected by chance across the population, which we find is true in this cohort. These observations support the hypothesis that extreme obesity in the general population is part of the BMI continuum rather than a distinct condition. We fully expect that studies of rarer variation should uncover higher penetrance alleles that are strongly enriched in the extremely obese population; however, rather than distinct Mendelian subtypes, these alleles will likely behave similarly as risk factors across the spectrum of BMI but, like the common alleles, show increasing enrichment with more extreme phenotype. Dissecting the biological processes affected by all of these genetic variants should increase our understanding of the genetics underlying obesity, one of the largest sources of ill-health in the developed world.

MATERIALS AND METHODS

Cohort description

The cohort studied includes men and women, aged 18–75 years, who were recruited from the population of patients undergoing RYGB surgery at the Massachusetts General Hospital Weight Center. Patients at this center routinely undergo extensive preoperative clinical evaluation and phenotyping and are followed closely during the post-operative period. DNA was obtained from 1008 patients who underwent RYGB between 2000 and 2007 and consented to participate (consent rate 97%). Each operation was performed by one of four surgeons using a standardized operative technique for

Table 1. Independent SNPs most associated with extreme obesity

| Nearby gene ^a | SNP | Chromosome ^b | Position ^b | Effect allele ^c | Other allele | Frequency ^d | OR ^e | 95% CI ^f | P-value |
|--------------------------|------------|-------------------------|-----------------------|----------------------------|--------------|------------------------|-----------------|---------------------|----------|
| <i>FTO</i> | rs9941349 | 16 | 52382989 | T | C | 0.43 | 1.48 | 1.33–1.66 | 6.09E–12 |
| <i>ZNF248</i> | rs7474896 | 10 | 38022103 | T | C | 0.14 | 1.46 | 1.25–1.69 | 8.87E–07 |
| <i>RTN4</i> | rs6726292 | 2 | 55010134 | A | G | 0.27 | 0.72 | 0.63–0.82 | 1.11E–06 |
| <i>ITPR3</i> | rs999943 | 6 | 33732711 | C | T | 0.28 | 0.73 | 0.64–0.83 | 1.14E–06 |
| <i>TRAM1L1</i> | rs10433903 | 4 | 118312585 | C | T | 0.34 | 0.75 | 0.66–0.84 | 2.98E–06 |
| <i>MUC15</i> | rs12295638 | 11 | 26561907 | C | T | 0.10 | 1.49 | 1.26–1.77 | 3.94E–06 |
| <i>FBN2</i> | rs374748 | 5 | 127727274 | G | A | 0.11 | 1.47 | 1.25–1.73 | 3.97E–06 |
| <i>RARB</i> | rs1435703 | 3 | 25535235 | T | G | 0.06 | 1.66 | 1.34–2.06 | 4.41E–06 |
| <i>RAFTLIN</i> | rs12635698 | 3 | 16383493 | C | T | 0.15 | 1.41 | 1.21–1.63 | 5.20E–06 |
| <i>PRF1</i> | rs10999409 | 10 | 72002446 | T | C | 0.45 | 1.30 | 1.16–1.45 | 5.32E–06 |
| <i>C20orf133</i> | rs6110577 | 20 | 15283754 | C | T | 0.16 | 1.40 | 1.21–1.61 | 5.33E–06 |
| <i>MLN</i> | rs2274459 | 6 | 33870220 | A | G | 0.17 | 1.38 | 1.20–1.58 | 5.92E–06 |
| <i>RAFTLIN</i> | rs12492816 | 3 | 16434072 | C | A | 0.11 | 1.46 | 1.24–1.72 | 6.94E–06 |
| <i>FLJ20309</i> | rs7603514 | 2 | 206544857 | A | G | 0.20 | 1.36 | 1.19–1.56 | 7.74E–06 |
| <i>MLN</i> | rs9366829 | 6 | 33881177 | G | A | 0.49 | 1.29 | 1.15–1.44 | 8.90E–06 |

Our results contain 15 independent loci associating at $P < 1 \times 10^{-6}$; here we present the most significant SNP in each locus. We note that only the *FTO* locus contains SNPs at genome-wide significance.

^aGene nearest analyzed SNP.

^bGenomic coordinates from NCBI build 35.

^cAllele for which effect is reported.

^dAllele for which OR reported.

^eOdds ratio.

^f95% Confidence intervals on OR.

open (41%) or laparoscopic (59%) RYGB. Liver biopsies were performed at the time of surgery, and samples were rapidly frozen and stored at -80°C .

Sample genotyping and data curation

Nine hundred and seventy-two case samples, including 841 self-identifying as ‘white’, were genotyped on the Illumina HumanHap 650Y product at Rosetta Inpharmatics. We converted data to the PLINK format and used this software for subsequent QC analyses unless otherwise indicated (15).

We first removed poorly performing samples and markers. One individual had more than 10% missing data and was excluded from further analysis. Of the 655 130 SNPs, genotyped in the remaining samples 6402 had more than 10% missing data and 33 570 had less than 1% minor allele frequency (MAF). A further 15 769 sex chromosome and mitochondrial SNPs were also excluded to give an initial dataset of 600 173 autosomal SNPs genotyped in 971 individuals.

We assessed cryptic relatedness in our samples by calculating identity-by-descent (IBD) coefficients for all pairwise combinations of individuals. We were able to detect seven isolated pairs and five extended pedigrees of more than two individuals who appeared to be at least first cousins from their IBD patterns. We also identified pairs with an absolute proportion of IBD $\pi_{\text{hat}} > 0.15$ (equivalent to second cousins in an outbred population). In each case, we retained the sample with the least missing data and discarded the others, eventually removing 50 individuals from our data.

We next addressed population structure in our data: of the 921 samples remaining in our analysis, 806 self-identified as ‘white’ and were kept in the analysis. Recognizing that self-reported ethnicity may still mask considerable population heterogeneity, we calculated principal components of ancestry

using EIGENSTRAT (16). Through this analysis, we identified and removed 25 outlying samples (six standard deviations).

We performed two further data integrity checks using PLINK analyses: we assessed the extent of heterozygosity in these samples and found four samples with either high rates (suggesting either admixture or sample contamination) or low (suggesting inbreeding). We then performed an identity-by-missingness rate (15), where samples are clustered on the basis of missing genotypes in common. Two samples had suspiciously high rates of concordance for missing data, indicating a possible technical artifact in the genotyping process.

Finally, we filtered SNPs more aggressively in these samples: we excluded 3020 markers failing Hardy–Weinberg equilibrium testing (HWE $P < 0.001$); 8040 markers with more than 5% missing data; 61 633 markers with MAF $< 5\%$ and 3972 markers with missing data biased by genotype ($P < 1 \times 10^{-6}$). Our final dataset thus comprises 775 samples genotyped at 525 054 markers across the autosomal genome.

Case/control association using publicly available controls

We obtained publicly available genotypes for 3294 population control samples from the repository created by Illumina for this purpose (see <http://www.illumina.com/pages.ilmn?ID?231>). All samples were of self-reported caucasian ancestry and genotyped on the Illumina HumanHap 550Y product. After filtering as described above (more than 5% MAF, less than 5% missing data per SNP and individual, HWE $P > 0.001$, and excluding related samples at IBD $\pi_{\text{hat}} > 0.15$), we merged these data with our bariatric case data, retaining only the 457 251 SNPs that passed QC in both sets independently.

Table 2. Effects of BMI-associated SNPs on extreme obesity

| Nearby gene | GIANT SNP | Chromosome | Position | Bariatric sample | | Position | R_2 | D' | EA | OA | Frequency | OR | 95% CI | P -value* | Direction | Power $P < 0.05$ |
|----------------------------|------------|------------|-----------|---|------------|----------|-------|------|----|----|-----------|------|-----------|-------------|-----------|---------------------|
| | | | | Proxy | Chromosome | | | | | | | | | | | |
| <i>FTO</i> | rs9939609 | 16 | 52378028 | rs8050136 | 16 | 52373776 | 1 | 1 | A | C | 0.42 | 1.46 | 1.311.64 | 1.42E-11 | Same | 0.97 |
| <i>MC4R</i> | rs17782313 | 18 | 56002077 | rs10871777 | 18 | 56002743 | 1 | 1 | G | A | 0.24 | 1.02 | 0.901.17 | 3.60E-01 | Same | 0.55 |
| <i>TMEM18</i> | rs6548238 | 2 | 624905 | rs4854344 | 2 | 628144 | 0.85 | 0.92 | G | T | 0.18 | 0.85 | 0.730.99 | 1.58E-02 | Same | 0.66 |
| <i>GNPDA2</i> | rs10938397 | 4 | 45023455 | rs12641981 | 4 | 44874640 | 1 | 1 | T | C | 0.43 | 1.18 | 1.061.32 | 1.67E-03 | Same | 0.66 |
| <i>MTCH2</i> | rs10838738 | 11 | 47619625 | Same SNP | — | — | — | — | G | A | 0.35 | 1.09 | 0.971.23 | 6.34E-02 | Same | 0.15 |
| <i>KCTD15</i> | rs11084753 | 19 | 39013977 | rs29941 | 19 | 39001372 | 0.65 | 0.87 | T | C | 0.31 | 0.87 | 0.770.98 | 1.27E-02 | Same | 0.10 |
| <i>NEGR1</i> | rs2815752 | 1 | 72524461 | rs2568958 | 1 | 72537704 | 1 | 1 | G | A | 0.35 | 0.90 | 0.801.01 | 3.45E-02 | Same | 0.21 |
| <i>STK33</i> | rs10769908 | 11 | 8440665 | rs725502 | 11 | 8460319 | 1 | 1 | G | A | 0.49 | 1.09 | 0.981.22 | 5.47E-02 | Same | NA |
| <i>SH2B1</i> | rs7498665 | 16 | 28790742 | Same SNP | — | — | — | — | G | A | 0.37 | 1.12 | 1.001.26 | 2.24E-02 | Same | 0.46 |
| Thorleifsson <i>et al.</i> | | | | | | | | | | | | | | | | |
| <i>SEC16B</i> | rs10913469 | 1 | 176180142 | Same SNP | — | — | — | — | C | T | 0.18 | 1.07 | 0.931.25 | 3.18E-01 | Same | 0.75 |
| <i>BDNF</i> | rs925946 | 11 | 27623778 | Same SNP | — | — | — | — | T | G | 0.28 | 1.13 | 0.9-1.28 | 6.42E-02 | Same | 0.77 |
| <i>ETV5</i> | rs7647305 | 3 | 187316984 | Same SNP | — | — | — | — | T | C | 0.21 | 0.86 | 0.740.99 | 3.75E-02 | Opposite | 0.85 |
| Meyre <i>et al.</i> | | | | | | | | | | | | | | | | |
| <i>NPC1</i> | rs1805081 | 18 | 19394430 | Same SNP | — | — | — | — | G | A | 0.38 | 0.86 | 0.770.97 | 1.65E-02 | Opposite | 1.00 |
| <i>MAF</i> | rs1424233 | 16 | 78240252 | Same SNP | — | — | — | — | A | G | 0.49 | 1.02 | 0.911.15 | 7.03E-01 | Same | 1.00 |
| <i>PTER</i> | rs10508503 | 10 | 16339957 | No proxy with $r^2 > 0.5$ exists on the Illumina 550K product | | | | | | | | | | | | |
| <i>PRL</i> | rs4712652 | 6 | 22186594 | Same SNP | — | — | — | — | A | G | 0.44 | 1.05 | 0.94-1.18 | 3.83E-01 | Opposite | 0.91 |

Twelve loci have been reported to influence BMI in the general population. These associations are captured either directly or by highly correlated markers on our genotyping platform; six of these loci show nominal association evidence (uncorrected $P < 0.05$) with a further three showing weaker association ($P = 0.06, 0.054$ and 0.06). We note that in 11 of 12 cases, the direction of effect is the same (i.e. the BMI increasing allele confers risk of extreme obesity), irrespective of statistical significance (combinatorial probability $P = 0.013$). We also report the power of our cohort to detect the reported associations at $P < 0.05$. For comparison, we also include four variants recently reported as influencing severe childhood and adult obesity (marked Meyre *et al.* 14). EA, effect allele; OA, other allele.

We investigated population stratification in the merged dataset with EIGENSTRAT: we identified 80 controls as outliers, which we removed prior to calculating the top ten principal components of population variation. We then used logistic regression to calculate association using these principal components as covariates. This approach allows us to estimate odds ratios and their confidence intervals.

One way to gauge residual stratification is by measuring the overall inflation in association statistics. The ratio of observed to expected median, λ_{GC} should be 1 (4). Without correcting for stratification, we observed $\lambda_{GC} = 1.125$ (an inflation of 12.5%); after correcting for the top ten axes of variation with EIGENSTRAT, we observed $\lambda_{GC} = 1.05$, a value in line with other published reports. We then felt comfortable adjusting for this residual inflation by dividing each association statistic by λ_{GC} to give our final association test statistics. We note that λ_{GC} scales with sample size, so that our final value of 1.05 in 775 cases and 3204 controls is equivalent to 1.04 in a dataset of 1000 cases and 1000 controls.

Any SNPs either within a 250 kb window around the most associated (or index) SNP or in linkage disequilibrium with the index SNP ($r^2 > 0.5$) was considered not to be independent of the original SNP and was not reported.

Odds ratio comparisons and power calculations

To determine the relative power of the bariatric case samples, we examined the observed χ^2 from the nine known and replicable associations for BMI. Assuming these effects as positive controls, we compared the observed non-centrality parameter (NCP) with the theoretical expectation for the NCP of 1000 population-based quantitative sample for each effect size from the GIANT meta-analysis. The median of the nine observed/expected ratios give us an approximation of the effective quantitative population size equivalent (in thousands) to our case/control cohort.

Homozygosity analysis

Our goal is to estimate an individual's autozygosity by means of maximum likelihood, across the autosomes by adapting a previously described model (17). We assume that homozygosity, especially in long stretches, is likely to be identical-by-descent, and autozygous, rather than identical-by-state. The process of identity by descent for each genomic region, k , has two states: either homozygous with two alleles identical by descent ($X_k = 1$) or not ($X_k = 0$). We approximate this process along the genome by a first-order Markov chain, as this approximation has been shown to hold for a variety of inbred relationships (18). Because the homozygous state is unknown, we invoke a hidden Markov model, using the SNP data and marker map to infer the unknown identity by descent state at each site along the genome, as well as to estimate by numerical maximization across the genome underlying inbreeding coefficient (F) and identity by descent switch parameter (a) which govern the coverage of autozygosity genome-wide. To complete the model, we need to enumerate the identity by descent probability at a specific locus and transition probabilities between loci. In order to describe these quantities, we turn to an explanation on how the large

density of SNP data is managed. To minimize marker linkage disequilibrium, we partitioned the SNP map into segments flanked by recombination hotspots, requiring a minimum of 20 SNPs per segment (as this was shown to improve parameter estimation, data not shown). This has the effect of maintaining tight linkage within segments, but allows for transitions to occur at hotspots, where most recombination occurs (19). Each segment has an associated identity by descent probability, based on the frequency of the phased haplotypes in the region. Transition probabilities in either case are based on these probabilities as well as the fine-scale recombination rate (8). This work is described in more detail in a manuscript in preparation by B.F.V. and others.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* online.

ACKNOWLEDGEMENTS

We are indebted to the study subjects for their participation, without which this research would be impossible.

Conflict of Interest statement. None declared.

FUNDING

This work was supported by grants to L.M.K. from Merck Research Laboratories and the National Institutes of Health [grant numbers DK046200, DK043351 and DK057478]. C.C. is supported by an MGH ECOR/FMD fellowship. E.K.S. was supported by the National Institutes of Health [grant numbers DK07191-32, DK079466-01 and DK080145-01].

REFERENCES

- Centers for Disease Control and Prevention (CDC) (1996) The Third National Health and Nutrition Examination Survey (NHANES III 1988-94) Reference Manuals and Reports (CD-ROM). Centers for Disease Control and Prevention: Bethesda, MD.
- O'Rahilly, S. and Farooqi, I.S. (2006) Genetics of obesity. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **361**, 1095-1105.
- Farooqi, I.S. (2006) Genetic aspects of severe childhood obesity. *Pediatr. Endocrinol. Rev.*, **3**, (Suppl. 4), 528-536.
- Devlin, B. and Roeder, K. (1999) Genomic control for association studies. *Biometrics*, **55**, 997-1004.
- Freedman, M.L., Reich, D., Penney, K.L., McDonald, G.J., Mignault, A.A., Patterson, N., Gabriel, S.B., Topol, E.J., Smoller, J.W., Pato, C.N. *et al.* (2004) Assessing the impact of population stratification on genetic association studies. *Nat. Genet.*, **36**, 388-393.
- Frayling, T.M., Timpson, N.J., Weedon, M.N., Zeggini, E., Freathy, R.M., Lindgren, C.M., Perry, J.R., Elliott, K.S., Lango, H., Rayner, N.W. *et al.* (2007) A common variant in the *FTO* gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*, **316**, 889-894.
- Dina, C., Meyre, D., Gallina, S., Durand, E., Körner, A., Jacobson, P., Carlsson, L.M., Kiess, W., Vatin, V., Lecoecur, C. *et al.* (2007) Variation in *FTO* contributes to childhood obesity and severe adult obesity. *Nat. Genet.*, **39**, 724-726.
- International HapMap (2007) Consortium A second generation human haplotype map of over 3.1 million SNPs. *Nature*, **449**, 851-861.
- Willer, C., Speliotes, E.K., Loos, R.J., Li, S., Lindgren, C.M., Heid, I.M., Berndt, S.I., Elliott, A.L., Jackson, A.U., Lamina, C. *et al.* (2008) Six new

- loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.*, **41**, 25–34.
10. Chu, X., Erdman, R., Susek, M., Gerst, H., Derr, K., Al-Agha, M., Wood, G.C., Hartman, C., Yeager, S., Blosky, M.A. *et al.* (2008) Association of morbid obesity with FTO and INSIG2 allelic variants. *Arch. Surg.*, **143**, 235–240.
 11. Price, R.A., Li, W.D. and Zhao, H. (2008) FTO gene SNPs associated with extreme obesity in cases, controls and extremely discordant sister pairs. *BMC Med. Genet.*, **9**, 4–4.
 12. Hinney, A., Nguyen, T.T., Scherag, A., Friedel, S., Brönnner, G., Müller, T.D., Grallert, H., Illig, T., Wichmann, H.E., Rief, W. *et al.* (2007) Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS ONE*, **2**, e1361.
 13. Thorleifsson, G., Walters, G.B., Gudbjartsson, D.F., Steinthorsdottir, V., Sulem, P., Helgadóttir, A., Styrkarsdóttir, U., Gretarsdóttir, S., Thorlacius, S., Jonsdóttir, I. *et al.* (2009) Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.*, **41**, 18–24.
 14. Meyre, D., Delplanque, J., Chèvre, J.C., Lecoœur, C., Lobbens, S., Gallina, S., Durand, E., Vatin, V., Degraeve, F., Proença, C. *et al.* (2009) Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. *Nat. Genet.*, **41**, 157–159.
 15. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. and Sham, P.C. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
 16. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N.A. and Reich, D. (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.*, **38**, 904–909.
 17. Leutenegger, A.L., Prum, B., Génin, E., Verny, C., Lemainque, A., Clerget-Darpoux, F. and Thompson, E.A. (2003) Estimation of the inbreeding coefficient through use of genomic data. *Am. J. Hum. Genet.*, **73**, 516–523.
 18. Thompson, E.A. (1994) Monte Carlo likelihood in the genetic mapping of complex traits. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **344**, 345–350.
 19. Myers, S., Bottolo, L., Freeman, C., McVean, G. and Donnelly, P. (2005) A fine-scale map of recombination rates and hotspots across the human genome. *Science*, **310**, 321–324.