Exome sequencing followed by genotyping suggests SYPL2 as a susceptibility gene for morbid obesity

Hong Jiao*,1, Peter Arner2, Paul Gerdm3,4, Rona J Strawbridge5, Erik Näslund6, Anders Thorell6,7, Anders Hamsten5, Juha Kere1 and Ingrid Dahlman*,2

Recently developed high-throughput sequencing technology shows power to detect low-frequency disease-causing variants by deep sequencing of all known exons. We used exome sequencing to identify variants associated with morbid obesity. DNA from 100 morbidly obese adult subjects and 100 controls were pooled (n = 10/pool), subjected to exome capture, and subsequent sequencing. At least 100 million sequencing reads were obtained from each pool. After several filtering steps and comparisons of observed frequencies of variants between obese and non-obese control pools, we systematically selected 144 obesity-enriched non-synonymous, splicing site or 5’ upstream single-nucleotide variants for validation. We first genotyped 494 adult subjects with morbid obesity and 496 controls. Five obesity-associated variants (nominal P-value < 0.05) were subsequently genotyped in 1425 morbidly obese and 782 controls. Of the five variants, only rs62623713:A → G (NM_001040709:c.A296G:p.E99G) was confirmed. rs62623713 showed strong association with body mass index (beta = 2.13 (1.09, 3.18), P = 6.28 x 10^-5) in a joint analysis of all 3197 genotyped subjects and had an odds ratio of 1.32 for obesity association. rs62623713 is a low-frequency (2.9% minor allele frequency) non-synonymous variant in exon 4 of the synaptophysin-like 2 (SYPL2) gene. rs62623713 was not covered by Illumina or Affymetrix genotyping arrays used in previous genome-wide association studies. Mice lacking Sypl2 have been reported to display reduced body weight. In conclusion, using exome sequencing we identified a low-frequency coding variant in the SYPL2 gene that was associated with morbid obesity. This gene may be involved in the development of excess body fat.

European Journal of Human Genetics (2015) 23, 1216–1222; doi:10.1038/ejhg.2014.255; published online 19 November 2014

INTRODUCTION

Obesity is a major contributor to the global burden of chronic disease, such as type 2 diabetes.1 A strong genetic impact on obesity has been established. However, the identification of specific genes involved in common forms of human obesity has proven elusive, despite great efforts made through different approaches, including candidate gene studies, genome-wide linkage and genome-wide association (GWA) studies.2,3 So far, GWA studies have identified around 50 obesity-associated genetic loci. However, these loci together explain only about 1–2% of body mass index (BMI) variations in the studied populations.4 Therefore a major question is how the remaining so called ‘missing’ heritability can be explained.5 There are two principal concerns about the interpretation of GWA results. First, the approach is based on the common disease, common variant hypothesis. The commercial single-nucleotide polymorphism (SNP) arrays used in GWA studies were designed to cover the common genetic variants (minor allele frequency (MAF) >5% in populations). Therefore rare variants could not be directly evaluated using such arrays. Second, imperfect statistical models, which do not account for gene–gene and gene–environment interactions, were used in GWA data analyses.6

Recently developed high-throughput sequencing technology complements SNP arrays and shows power to detect rare disease-causing variants by deep sequencing of all known exons.7–9 This revolution of technology holds promise to elucidate complex diseases by allowing a genome-wide search for low frequency and rare and putatively functional variants. Exome sequencing might be most efficient when applied to patients with more extreme forms of common diseases, such as morbid obesity; first, it increases the chances to obtain significant association for rare variants with high penetrance; second, genetic variants in the protein coding sequence might be more likely to have a strong impact on the phenotype than variants in intergenic regions.

In this study, we employed exome sequencing to detect low frequency and rare gene variants enriched in adult subjects with morbid obesity and validated them against never-obese elderly adults.

SUBJECTS AND METHODS

Sample selection

The cohorts for genetic studies are described in Table 1. Subjects were recruited by local advertisement for the purpose of studying genes regulating body weight.

Hong Jiao*1, Peter Arner2, Paul Gerdem3,4, Rona J Strawbridge5, Erik Näslund6, Anders Thorell6,7, Anders Hamsten5, Juha Kere1 and Ingrid Dahlman*,2

1Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, Stockholm, Sweden; 2Department of Medicine (H7), Karolinska Institutet, Stockholm, Sweden; 3Department of Orthopaedics, Karolinska University Hospital, Stockholm, Sweden; 4CLINTEC, Karolinska Institutet, Stockholm, Sweden; 5Cardiovascular Genetics and Genomics Group, Atherosclerosis Research Unit, Department of Medicine, Karolinska Institutet, Stockholm, Sweden; 6Department of Clinical Sciences, Danderyd Hospital, Karolinska Institutet, Stockholm, Sweden; 7Department of Surgery, Ensta Hospital, Karolinska Institutet, Stockholm, Sweden.

*Correspondence: Dr H Jiao, Department of Biosciences and Nutrition, Science for Life Laboratory, Karolinska Institutet, Stockholm 141 57, Sweden. Tel: +46 8 524 811 51; E-mail: hong.jiao@ki.se

or Dr I Dahlman, Department of Medicine, Karolinska Institutet, Stockholm 141 86, Sweden. Tel: +46 8 5888 0000; E-mail: ingrid.dahlman@ki.se

Received 17 July 2014; revised 26 September 2014; accepted 10 October 2014; published online 19 November 2014
Table 1 Clinical information on samples used for exome sequencing and variant validation

|                        | Nationality | Female/male | BMI (kg/m²) | Age (years) | Nationality | Female/male | BMI (kg/m²) | Age (years) |
|------------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Discovery by exome sequencing | Swedish     | 69/31       | 52.2 ± 3.8  | 41.5 ± 11.5 | Swedish     | 76/24       | 21.9 ± 4.3  | 24.5 ± 12.8 |
| First validation        | Swedish     | 357/137     | 51.1 ± 3.7  | 40.9 ± 11.6 | Swedish     | 351/145     | 24.4 ± 2.7  | 55.1 ± 5.8  |
| Second confirmation     | Swedish     | 1068/357    | 43.1 ± 2.5  | 42.3 ± 11.6 | Swedish     | 696/86      | 23.8 ± 2.9  | 44.3 ± 4.4  |
| Total                   |             | 1494/525    |             | 1123/255    |             |             |             |             |

Values for BMI and age are mean ± SD. Controls used in the first validation and second confirmation were non-obese.

or in association with planned visits to our medical or surgical units for morbid obesity or scoliosis, respectively. Most of them have been described.\textsuperscript{26,27} From the obese subset, we selected, among the 10% with the highest BMI, 100 subjects with extreme obesity for exome sequencing (69 women, 31 men, age 41.5 ± 11.5 years, BMI 52.2 ± 3.8 kg/m²). All selected subjects had morbid obesity defined as BMI >40 kg/m². As controls for the exome sequencing, we used subjects already collected for an ongoing genetic study on idiopathic scoliosis (76 women, 24 men, age 24.3 ± 12.8 years, BMI 21.9 ± 4.3 kg/m²). For the first genotyping validation, we selected the 494 morbidly obese subjects with the highest BMI in the large sample collection described above, including the 100 subjects who had been used for exome sequencing. As controls, we used 496 never-obese subjects with age >40 years and BMI <30 kg/m². In the second confirmation, we genotyped remaining 1425 morbidly obese subjects and 782 never-obese controls with age >40 years and BMI <30 kg/m².

The non-obese controls for exome sequencing had scoliosis. Otherwise all other control subjects were healthy according to self-report. There was an overrepresentation of women in the studied cohorts. Obese women are more likely to search medical advice for their obesity. Scoliosis is more common among women. Subjects were of European ancestry and living in Sweden. The study was approved by the local Ethics Committees, and all subjects gave their informed consent to participation.

DNA preparation and pooling
Genomic DNA was prepared from PBMCs using the QiAmp DNA blood Maxi kit (cat no. 51194, Qiagen, Hilden, Germany). DNA purity and quality was confirmed by A260/280 ratio > 1.8 on Nanodrop (Thermo Fisher Scientific Inc., Waltham, MA, USA) and agarose gel electrophoresis. DNA concentration was measured by Qubit (Life Technologies, Stockholm, Sweden). Subsequently, we took 0.8 μg of each DNA sample and randomly divided them into 10 pools, each containing 10 samples of either obese cases or controls. The concentrations of pooled DNA samples were measured with Qubit, and the samples were run on agarose gel.

Exome sequencing
Exome sequencing was performed at the Science for Life Laboratory (SciLifeLab), Stockholm, Sweden. Each DNA library was prepared from 3 μg of the pooled genomic DNA. DNA was sheared to 300 bp using a Covaris S2 instrument and enriched by using the SureSelectXT Human All Exon 50 Mb kit and an Agilent NGS workstation according to the manufacturer’s instructions (SureSelectXT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing, version A, Agilent Technologies, Santa Clara, CA, USA).

The clustering was performed on a cBot cluster generation system using a HiSeq paired-end read cluster generation kit according to the manufacturer’s instructions (Illumina, San Diego, CA, USA). The samples were sequenced on an Illumina HiSeq 2000 as paired-end reads to 100 bp/reads (Illumina). All lanes were spiked with 1% phiX control library, except for lane 8, which had 2% phiX. The sequencing runs were performed according to the manufacturer’s instructions. Base conversion was done using Illumina’s OLB v1.9 (Illumina).

Sequencing read mapping, variant calling and functional annotation
Sequencing reads were aligned to the current human reference sequence (assembly hg19, NCBI build 37) (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/chromosomes/) by using Burrows-Wheeler Aligner (BWA version 0.6.1, http://bio-bwa.sourceforge.net/, Li and Durbin\textsuperscript{22}) with a read trimming parameter –q of 20. Sequence variants were called by using the multi-way pileup function of samtools-0.1.18 (http://samtools.sourceforge.net/), with a minimum mapping quality of 20 and minimum read depth of 5 × for filtering. PCR duplicates were removed using samtools prior to variant calling. For functional annotation of sequence variants, we employed annovar (http://www.openbioinformatics.org/annovar/\textsuperscript{23}, Wang et al\textsuperscript{23}) to integrate information from a variety of databases in public domains, such as gene reference (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/refGene.txt, 2013), dbSNPs (SNP135, http://hgdownload.soe.ucsc.edu/goldenPath/hg19/database/snp137.txt.gz) and the 1000 Genome project (http://www.1000genomes.org/).

Variant filtering and enriching for obesity
We used the following filtering criteria to select low-frequency and rare single-nucleotide variants (SNVs) from exome sequencing data. Those SNVs enriched in morbidly obese subjects were taken forward for validation (Figure 1). First, we filtered for variants with a depth of 5 × and mapping quality (MQ) ≥ 20. Then we looked for putatively functional variants, that is, variants from exonic regions, splicing sites or 5′ upstream regions. In the next step, we compared the occurrences of a variant between obese and control pools. Only variants called in ≥2 obese pools, but with no call or called once in control pools, were used in the further filtering step; this step was applied to avoid potential false-positive variants. The last filtering was based on allele frequency. We looked for low-frequency and rare SNVs, that is, variants that were not found in public databases or known SNPs with a MAF ≤ 5% in general populations (1000 Genomes project 2011 May release).

Variant validation by genotyping
Variant validation was focused on SNVs enriched by the above filtering steps that could fit into a manageable genotyping panel and was carried out in two steps. First, all selected variants were genotyped in 494 obese subjects and 496 controls (Table 1 and Figure 1) using the Illumina GoldenGate assay performed at the SNP/SEQ Technology Platform at Uppsala University, Sweden (http://www.molmed.medsci.uu.se/SEQ+Technology+PlatformGenotyping/, Fan et al\textsuperscript{14}). Secondly, genotyped variants showing nominal significant association with obesity were genotyped in an additional 1425 obese and 782 controls. Genotyping was performed using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry (SEQUENOM, Agena Bioscience, San Diego, CA, USA) at the Mutation Analysis core Facility (MAF) at Karolinska Institutet, Sweden.\textsuperscript{15} Multiplexed assays were designed using the MassARRAY Assay Design v4.0 Software (Agena Bioscience). Protocol for allele-specific base extension was performed according to Agena Bioscience’s recommendation. All genetic association results in validation cohorts 1 and 2 have been submitted to GWAS Central.

Statistical analysis of association
Subjects were characterized by BMI, and values are reported as mean ± SD for obese and controls separately. Genetic association analysis and odds ratio calculation were performed using PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/\textsuperscript{16}, Purcell et al\textsuperscript{16}). Hardy–Weinberg equilibrium (HWE) of the genotype frequencies among cases and controls were checked separately for each run of validations prior to association analysis. A HWE nominal P-value of
0.01 in controls was used as a cutoff for exclusion of variants from further analysis. BMI was analyzed using standard linear regression implemented in PLINK. Sex and age were used as covariates to evaluate their influence on the association of a variant with BMI.

RESULTS

Discovery of rare variants by exome sequencing
To identify potential functional low-frequency and rare variants associated with extreme obesity, we performed exome sequencing on pooled DNA from 100 subjects (10 pools) with severe morbid obesity and 100 control subjects (10 pools). Thus there was DNA from 10 subjects in each pool. The characteristics of subjects for the exome sequencing are shown in Table 1. The obese subjects were older than the controls, whereas the gender distribution was similar between cohorts, with an overrepresentation of women. In total, we obtained 4100 million reads from each pool (Supplementary Table S1). After trimming low-quality reads and removing PCR duplicates, 95% of reads were mapped to the current human genome reference (assembly hg19, NCBI build 37) except for one control pool that had a little lower mapping rate. The vast majority of reads were uniquely mapped on the same chromosome for both paired reads. Discordance between paired reads was rare. Reads were paired with higher rates (97%) in obese pools than in control pools (80–95%) (Supplementary Table S1).

More than 92% of target regions were covered by 5× reads, and >80% of target regions were covered by 30× reads (Supplementary Table S2).

Potentially interesting variants were identified through several filtering steps (Figure 1). All variants were called using a depth of 5× and MQ ≥ 20 as the first filtering criteria. In total, we got 114034–190130 variants from each of the 20 pools. About 95% of variants were known SNPs, that is, SNPs with rs numbers. The majority of variants were located in exons, introns or intergenic regions, whereas others were in nearby slicing sites or in regulatory regions (Table 2). We focused on putatively functional variants, that is, variants from exonic regions (excluding synonymous variants), slicing sites or 5’ upstream. We obtained 31520 such variants in obese pools and 26565 in non-obese pools. Those variants were found in at least one obese pool or control pool, respectively. In all, 20652 (65.5%) potentially functional variants were present in obese pools and 26565 in non-obese pools. Those variants were found in at least one obese pool or control pool, respectively. In all, 20652 (65.5%) potentially functional variants were present in both obese and control pools, and 10868 (34.5%) were singletons for obesity. Next we looked for variants appearing more often in obese pools. At this step, there were 5788 variants left, and they were detected in ≥2 obese pools but did not appear or appeared just once in control pools. Finally we filtered for low-frequency and rare SNVs, that is, SNVs that were not found in public databases or SNPs with a MAF ≤ 5% in

![Diagram](image-url)
general populations (1000 Genome project, http://www.1000genomes.org/) were taken forward. With these criteria, 1032 obesity-associated SNVs were obtained.

SNV validation and association analysis

From the 1032 SNV, we excluded insertion or deletion variants to avoid potential technical difficulties in genotyping. Then we filtered SNVs shared by at least two obese pools with the highest yield. After manually checking alignments of reads carrying the variants by using the IGV software (http://www.broadinstitute.org/igv/) to exclude potential artifacts, we selected 144 SNVs for genotyping (142 dbSNP and 2 unknown SNVs). Sequencing depths and numbers of shared obesity pools of the 144 SNVs are shown in Supplementary Table S3. Ten SNVs had difficulties in primer design for genotyping and were replaced with two unknown SNVs). Sequencing depths and numbers of shared obesity pools of the 144 SNVs are shown in Supplementary Table S3. Ten SNVs had difficulties in primer design for genotyping and were replaced with

SNV validations were divided into two steps. First, the 144 SNVs were genotyped in 494 morbidity obese subjects, and 496 non-obese controls by quantitative trait analysis in all genotyped subjects, rs62623713 showed strong association with BMI assuming an additive genetic model (beta 2.13, P-value 0.004) and rs35923425 (NM_024727:c.G1134C:p.L378F), showed nominally significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3).

Five out of the 98 SNVs showed significant allelic association with obesity with a nominal P-value 0.05 (Table 3, Supplementary Table S4). These five SNVs were subsequently subjected to the second run of genotyping for confirmation in a large case–control cohort consisting of 1425 morbidity obese subjects and 782 never-obese controls. Only one SNV, rs62623713 (NM_001040709:c.A296G:p.E99G), consistently showed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3).

SNV genotyping for confirmation in a large case–control cohort consisting of 1425 morbidity obese subjects and 782 never-obese controls. Only one SNV, rs62623713 (NM_001040709:c.A296G:p.E99G), consistently showed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3). However, in the final joint analysis of all genotyped subjects two SNVs, rs62623713 and rs35923425 (NM_001040709:c.A296G:p.E99G), consistently displayed a significant association with obesity (Table 3).
genetic variants. The association of rs62623713 with BMI was still performed a variety of tests to adjust for the analysis of multiple BMI disappeared after adjustment for age. Furthermore, we also associations with BMI after adjustment for age or gender added these variables as covariates in regression models and tested (Supplementary Tables S6 and S7).

To further evaluate any confounding effect of age or gender, we excluded 100 subjects used in exome sequencing for analyses among our obese subjects (MAF 8%). rs62623713 was not presented among our obese subjects (MAF 8%), rs62623713 was not covered by Illumina or Affymetrix genotyping arrays used in previous

**DISCUSSION**

Using exome sequencing followed by large-scale genotyping, we identified a low-frequency coding variant, rs62623713 (E99G) in exon 4 of the SYPL2 gene, consistently showing association with morbid obesity. rs62623713 has a MAF of 2.9% for the G allele in the general populations according to 1000 Genomes. The variant is overrepresented among our obese subjects (MAF 8%), rs62623713 was not covered by Illumina or Affymetrix genotyping arrays used in previous

| CHR | SNP     | A1 | Freq_case | Freq_control | A2 | P-value | OR | L95 | U95 | Gene     |
|-----|---------|----|-----------|--------------|----|---------|----|-----|-----|----------|
|     |         |    |           |              |    |         |    |     |     |          |
| First validation | 1 | rs62623713 | G | 0.100 | 0.077 | A | 0.044 | 1.38 | 1.01 | 1.88 | SYPL2   |
|     | 1 | rs41280330 | T | 0.017 | 0.0061 | C | 0.021 | 2.88 | 1.13 | 7.33 | GNAT2   |
|     | 3 | rs35923425 | G | 0.082 | 0.050 | C | 0.0042 | 1.70 | 1.18 | 2.45 | LRRC31  |
|     | 14 | rs34106261 | A | 0.054 | 0.033 | G | 0.0184 | 1.71 | 1.09 | 2.67 | RIPK3   |
|     | 19 | rs45546534 | G | 0.061 | 0.039 | A | 0.021 | 1.63 | 1.07 | 2.47 | TMPRSS9 |
| Second validation | 1 | rs62623713 | G | 0.077 | 0.054 | A | 0.0049 | 1.44 | 1.12 | 1.86 | SYPL2   |
|     | 1 | rs41280330 | T | 0.0075 | 0.0089 | C | 0.61 | 0.84 | 0.43 | 1.63 | GNAT2   |
|     | 3 | rs35923425 | G | 0.074 | 0.063 | C | 0.15 | 1.19 | 0.94 | 1.52 | LRRC31  |
|     | 14 | rs34106261 | A | 0.034 | 0.038 | G | 0.57 | 0.91 | 0.64 | 1.28 | RIPK3   |
|     | 19 | rs45546534 | G | 0.044 | 0.040 | A | 0.52 | 1.10 | 0.82 | 1.50 | TMPRSS9 |
| Final analysis | 1 | rs62623713 | G | 0.081 | 0.063 | A | 0.0063 | 1.32 | 1.08 | 1.61 | SYPL2   |
|     | 1 | rs41280330 | T | 0.0094 | 0.0079 | C | 0.51 | 1.20 | 0.69 | 2.08 | GNAT2   |
|     | 3 | rs35923425 | G | 0.076 | 0.058 | C | 0.0040 | 1.35 | 1.10 | 1.66 | LRRC31  |
|     | 14 | rs34106261 | A | 0.038 | 0.035 | G | 0.58 | 1.08 | 0.82 | 1.44 | RIPK3   |
|     | 19 | rs45546534 | G | 0.047 | 0.040 | A | 0.18 | 1.19 | 0.92 | 1.52 | TMPRSS9 |

**Table 3 Associations of five SNVs with obesity in the first and second runs of validation and in the final analysis**

Abbreviations: A1, minor allele (observed allele); A2, alternative allele (reference allele), see Supplementary Table S3 for more information about SNPs; CHR, chromosome; Freq, frequencies of minor allele A1; OR, odds ratio; L95 and U95, the upper and lower 95% confidence intervals.

The G allele is overrepresented in our obese subjects (MAF 8%), especially in the homozygous form (12 obese to 2 controls). The same tendency of associations was shown in females (beta 1.99, P-value 0.001), males (beta 2.42, P-value 0.016), obese subjects (beta 3.39, P-value 0.007) and non-obese subjects (beta 2.16, P-value 0.031) (Table 4). rs35923425 in LRRC31 was also associated with BMI (beta 1.5, P-value 0.60 × 10$^{-7}$). However, significant association was seen in females only.

The significance of associations of 5 SNVs with either obesity or BMI in the first and the final analysis were not influenced when we excluded 100 subjects used in exome sequencing for analyses (Supplementary Tables S6 and S7).

To further evaluate any confounding effect of age or gender, we added these variables as covariates in regression models and tested their individual effects separately. rs62623713 maintained strong associations with BMI after adjustment for age or gender (Supplementary Table S8). The association between rs35923425 and BMI disappeared after adjustment for age. Furthermore, we also performed a variety of tests to adjust for the analysis of multiple genetic variants. The association of rs62623713 with BMI was still significant with a Bonferroni adjusted P-value of 3.0 × 10$^{-4}$ and a genomic-control corrected P-value of 0.05. The association of rs35923425 became non-significant with genomic controls, whereas the Bonferroni corrected P-value was 0.029 (Supplementary Table S9).

To the best of our knowledge there are only three reports of obesity-associated variants detected by exome sequencing.17,18 Albrechtsen et al.17 analyzed a cohort with the metabolic syndrome and examined multiple metabolic phenotypes. The study by Huang et al.18 primarily analyzed type 2 diabetes. Gill et al.18 identified novel LEPR variants in severe childhood onset obesity using exome sequencing.

The hypothesis underlying our study is that more severe forms of common obesity might be due to low-frequency or rare variants with a larger impact on the phenotype. A minor portion of morbidly obese cases has been shown to be due to genetic variants with high penetrance. Variants in the melanocortin 4 receptor (MC4R) gene explain a few percentage of childhood obesity,19 and a copy number repeat on chromosome 1620 has a high penetrance in a few percentages of morbidly obese subjects with cognitive defects. We did not detect any variants in exonic or nearby upstream regions of the MC4R in our exome sequencing data, which may be due to the low number of investigated subjects. Our results, that is, the detection of one SNP in SYPL2 remaining significantly associated with obesity after adjustment for multiple testing, support the notion that low-frequency and rare variants contribute to morbid obesity. However, the present study lacks power to estimate to what extent rare genetic variants cause morbid obesity in the population.

The approach we report using pooled DNA for exome sequencing to get obesity-associated variants, which was followed by genotype validation, is a cost-effective method. Our study has its limitations. We did not have access to a reliable family history of our patients and therefore do not know whether their obesity displays patterns of monogenic inheritance. On the other hand, as obesity is common, pedigree information might be difficult to interpret due to heterogeneity, that is, different causes of obesity in different family members. We only genotyped validated variants detected in more than one sequencing pools to limit the number of false positives. Although by
Obese

Male

Final total

Female

Lean

Obese

Table 4 Quantitative trait analysis with BMI using regression in final validation

| CHR | SNP      | A1 | TEST No. | Beta  | L95     | U95     | P-value |
|-----|----------|----|----------|-------|---------|---------|---------|
|     |          |    |          |       |         |         |         |

|     |          |    |          |       |         |         |         |

Abbreviations: ADD, the additive effects of allele; A1, observed minor allele; CHR, chromosome; L95 and U95: the upper and lower 95% confidence intervals.

genotyping variants detected in one sequencing pool only we might have discovered additional rare obesity-associated variants, we did not apply this strategy because of lower sequencing overage in some regions and risk of false-positive variants. However, with the applied strategy we were able to detect true low-frequency SNVs (about 30% of SNVs with a MAF ≤ 0.05 and > 60% with a MAF ≤ 0.10) in obese pools. Estimated MAFs in exome sequencing data correlated with those based on individual sample genotyping (Supplementary Figure S1). In our study, failure in validation by genotyping seemed not due to low depth as major failed SNVs had an average depth ≥ 20 × (Supplementary Table S3). Imperfect primer designs in multiplex genotyping could be a major cause of genotyping failure.

SYPL2 (also called MG29) is expressed in the adipose tissue, brain, kidney, heart and cerebellum. Functionally, the SYPL2 protein has been proposed to participate in cellular calcium ion homeostasis. The protein is expected to be involved in transporter activity and to localize in various compartments, such as synaptic vesicle.22

Interestingly, mice lacking Sypl2/MG29 exhibit reduced body weight, abnormal skeletal muscle membranes and irregular skeletal muscle contractility.23 How the gene might influence adipose tissue formation and its potential role in obesity development is unclear and needs further investigation, which is beyond the scope of this study.

In previous reports, SYPL2 was associated with major depressive disorder in European population.24 An association between obesity and depression has repeatedly been established. In a meta-analysis, obesity was found to increase the risk of depression. In addition, depression was found to be predictive of obesity.25 Another study found that the association of obesity with depression was mainly confined to persons with severe obesity.26

SYPL2 belongs to the synaptophysin family. Synaptophysin regulates activity-dependent synapse formation in cultured hippocampal neurons,27 and it is required for kinetically efficient endocytosis of synaptic vesicles in cultured hippocampal neurons.28 Food intake is subject to a complex regulation by the hypothalamus and other brain centers, including the brain stem and the hippocampus. Thus one could hypothesize that SYPL2 is involved in the central regulation of food intake possibly affecting the central reward systems and the hedonic effects of food.

In conclusion, we identified a low-frequency obesity-associated coding variant in the SYPL2 gene using exome sequencing with pooled DNA from 100 morbidly obese and 100 non-obese subjects, which was followed by genotype validation in 3197 case-control subjects. Our results provide evidence of the existence of a coding variant associated with obesity although further functional studies of this genetic variant remain to be performed.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

The project was supported by grants from Stockholm Community (ALF), the Swedish Research Council, Novo Nordic Foundation, the Diabetes Strategic Research Program at Karolinska Institutet and the Erling-Persson Family Foundation. The authors are grateful to the mutation analysis facility at the Karolinska Institute and the SNP&SEQ Technology Platform at the Uppsala University for performing genotyping and for excellent technical assistance. We would like to acknowledge support from the Science for Life Laboratory, the National Genomics Infrastructure and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. We thank Anna Grauers for sclerosis sample collection and characterization. We are indebted to Elisabeth Dungner for DNA preparation.

1 Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH: The disease burden associated with overweight and obesity. JAMA 1999; 282: 1523–1529.
2 Bernt S, Gustafsson S, Magi R et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet 2013; 45: 501–512.
3 Frayling TM, Timpson NJ, Weedon MN et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007; 316: 889–894.
4 Speliotes EK, Willer CJ, Berndt SI et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010; 42: 937–948.
5 Walley AJ, Asher JE, Froude P: The genetic contribution to non-syndromic human obesity. Nat Rev Genet 2009; 10: 431–442.
6 Yang J, Benyamin B, McEvoy BP et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010; 42: 565–569.
7 Bamshad MJ, Ng SB, Bigham AW et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011; 10: 745–755.
8 Ng SB, Buckingham KJ, Lee C et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010; 42: 30–35.
9 Gill R, Him Cheung Y, Shen Y et al. Whole-exome sequencing identifies novel LEPR mutations in individuals with severe early onset obesity. Obesity (Silver Spring) 2013; 21: 576–584.
10 Jiao H, Arner P, Dickson SL et al. Genetic association and gene expression analysis identify FGFR1 as a new susceptibility gene for human obesity. J Clin Endocrinol Metab 2011; 96: E962–E966.

1 Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH: The disease burden associated with overweight and obesity. JAMA 1999; 282: 1523–1529.
2 Bernt S, Gustafsson S, Magi R et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. Nat Genet 2013; 45: 501–512.
3 Frayling TM, Timpson NJ, Weedon MN et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. Science 2007; 316: 889–894.
4 Speliotes EK, Willer CJ, Berndt SI et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet 2010; 42: 937–948.
5 Walley AJ, Asher JE, Froude P: The genetic contribution to non-syndromic human obesity. Nat Rev Genet 2009; 10: 431–442.
6 Yang J, Benyamin B, McEvoy BP et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet 2010; 42: 565–569.
7 Bamshad MJ, Ng SB, Bigham AW et al. Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 2011; 10: 745–755.
8 Ng SB, Buckingham KJ, Lee C et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010; 42: 30–35.
9 Gill R, Him Cheung Y, Shen Y et al. Whole-exome sequencing identifies novel LEPR mutations in individuals with severe early onset obesity. Obesity (Silver Spring) 2013; 21: 576–584.
10 Jiao H, Arner P, Dickson SL et al. Genetic association and gene expression analysis identify FGFR1 as a new susceptibility gene for human obesity. J Clin Endocrinol Metab 2011; 96: E962–E966.
11 Jiao H, Amer P, Hoffstedt J et al: Genome wide association study identifies KCNMA1 contributing to human obesity. BMC Med Genomics 2011; 4: 51.
12 Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26: 589–595.
13 Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38: e164.
14 Fan JB, Oliphant A, Shen R et al: Highly parallel SNP genotyping. Cold Spring Harb Symp Quant Biol 2003; 68: 69–78.
15 Jurinke C, van den Boom D, Cantor CR, Koster H: Automated genotyping using the DNA MassArray technology. Methods Mol Biol 2002; 187: 179–192.
16 Pursell S, Neale B, Todd-Brown K et al: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81: 559–575.
17 Albrechtsen A, Granup N, Li Y et al: Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. Diabetologia 2013; 56: 298–310.
18 Huang K, Nair AK, Muller YL et al: Whole exome sequencing identifies variation in CYB5A and RNF10 associated with adiposity and type 2 diabetes. Obesity (Silver Spring) 2013; 22: 984–988.
19 Farooqi IS, Keski JM, Yeo GS, Lank EJ, Cheetham T, O’Rahilly S: Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. N Engl J Med 2003; 348: 1085–1095.
20 Yang TL, Guo Y, Li SM et al: Ethnic differentiation of copy number variation on chromosome 16p12.3 for association with obesity phenotypes in European and Chinese populations. Int J Obes (Lond) 2013; 37: 188–190.
21 Kurebayashi N, Takeshima H, Nishi M, Murayama T, Suzuki E, Ogawa Y: Changes in Ca2+ handling in adult MG29-deficient skeletal muscle. Biochem Biophys Res Commun 2003; 310: 1266–1272.
22 Fernandez-Chacon R, Sudhof TC: Genetics of synaptic vesicle function: toward the complete functional anatomy of an organelle. Annu Rev Physiol 1999; 61: 753–776.
23 Nishi M, Komazaki S, Kurebayashi N et al: Abnormal features in skeletal muscle from mice lacking mitsugumin29. J Cell Biol 1999; 147: 1473–1480.
24 Shi J, Potash JB, Knowles JA et al: Genome-wide association study of recurrent early-onset major depressive disorder. Mol Psychiatry 2011; 16: 193–201.
25 Luppino FS, de Wit LM, Bouvy PF et al: Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. Arch Gen Psychiatry 2010; 67: 220–229.
26 Oryike CU, Crum RM, Lee HB, Lyketsos CG, Eaton WW: Is obesity associated with major depression? Results from the Third National Health and Nutrition Examination Survey. Am J Epidemiol 2003; 158: 1139–1147.
27 Tarsa L, Goda Y: Synaptophysin regulates activity-dependent synapse formation in cultured hippocampal neurons. Proc Natl Acad Sci USA 2002; 99: 1012–1016.
28 Kwon SE, Chapman ER: Synaptophysin regulates the kinetics of synaptic vesicle endocytosis in central neurons. Neuron 2011; 70: 847–854.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)