

ARTICLE

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

Hong Jiao^{*1}, Peter Arner², Paul Gerdhem^{3,4}, Rona J Strawbridge⁵, Erik Näslund⁶, Anders Thorell^{6,7}, 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 ($n = 10/\text{pool}$), 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. Out 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), $P = 6.28 \times 10^{-5}$) 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 (E99G) 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* has 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. We reasoned that in this way we would enrich for obesity genes among cases and filtered away such genes in the controls. Here we report a low-frequency obesity-associated variant in the coding region of the synaptophysin-like 2 (*SYPL2*) gene.

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

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, Ersta 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 5858 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

	Obese cases				Controls			
	Nationality	Female/male	BMI (kg/m ²)	Age (years)	Nationality	Female/male	BMI (kg/m ²)	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.^{10,11} 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 24.5 ± 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 (SciLife-Lab), 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/read (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 Durbin¹²) 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/>, Wang *et al*¹³) 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/SNP+SEQ+Technology+Platform/Genotyping/>, Fan *et al*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.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/>, Purcell *et al*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

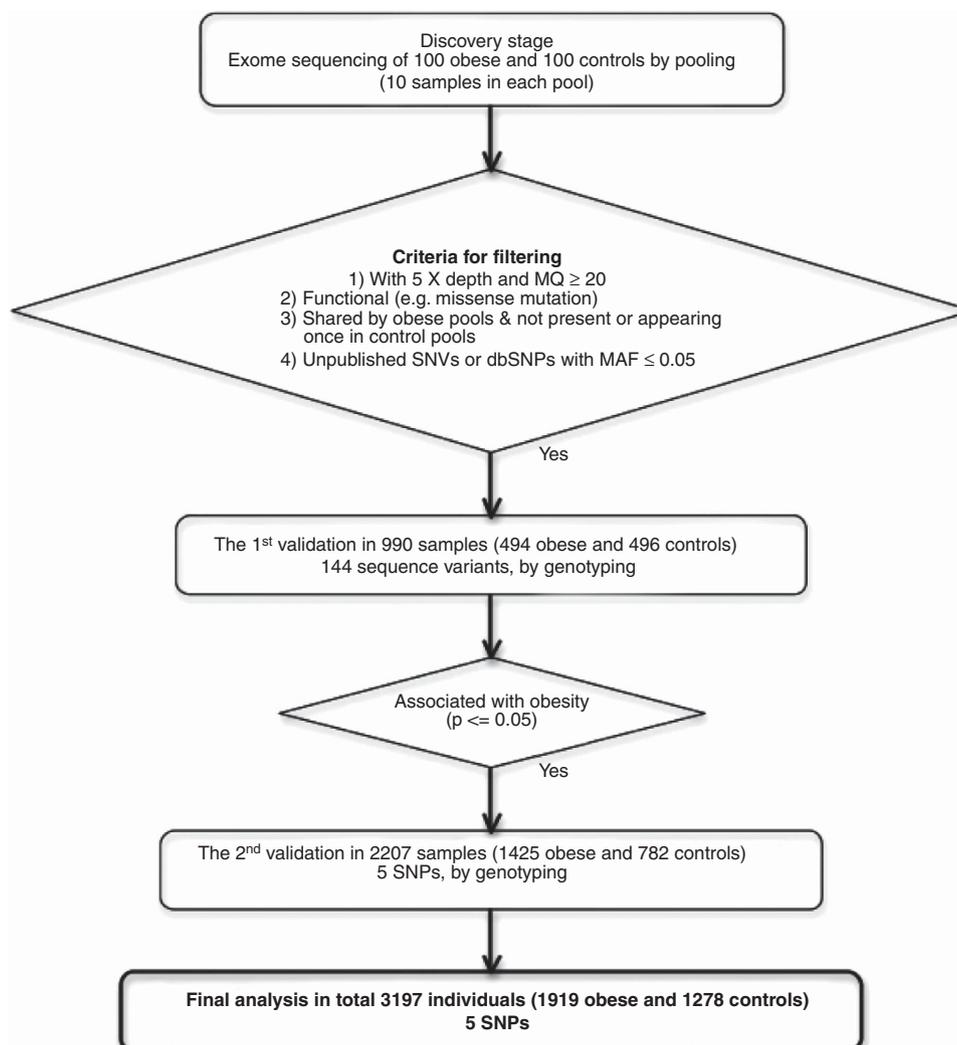


Figure 1 Overview of working processes. Unpublished SNVs: variants that are not included in SNP135.

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 >100 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 $5\times$ reads, and >80% of target regions were covered by $30\times$ reads (Supplementary Table S2).

Potentially interesting variants were identified through several filtering steps (Figure 1). All variants were called using a depth of $5\times$ and $MQ \geq 20$ as the first filtering criteria. In total, we got 114 034–190 130 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), splicing sites or 5' upstream. We obtained 31 520 such variants in obese pools and 26 565 in non-obese pools. Those variants were found in at least one obese pool or control pool, respectively. In all, 20 652 (65.5%) potentially functional variants were present in both obese and control pools, and 10 868 (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 $MAF \leq 5\%$ in

Table 2 Summary of variants with at least 5 × depth and MQ > 20 found in pools

Genome region	Pool_1	Pool_2	Pool_3	Pool_4	Pool_5	Pool_6	Pool_7	Pool_8	Pool_9	Pool_10
<i>Obese pools</i>										
Downstream	1041	1509	1031	906	1013	1150	1140	1864	1405	1073
Exonic	22 598	22 619	20 867	22 697	22 622	22 031	22 134	21 244	21 387	22 505
Exonic;splicing	339	309	303	341	322	312	311	298	306	324
Intergenic	25 008	18 871	22 830	22 817	25 879	29 673	28 890	30 407	34 951	26 365
Intronic	90 832	110 903	91 629	85 416	93 314	100 064	99 213	122 534	109 572	95 353
Splicing	103	127	113	99	110	100	105	124	107	104
Upstream	1570	3399	1622	1448	1658	1871	1818	4082	2222	1665
Upstream;downstream	101	219	121	79	94	107	103	228	132	90
UTR3	4077	6094	4074	3907	4235	4529	4493	6414	5262	4295
UTR5	1523	2737	1511	1394	1510	1643	1617	2931	1776	1522
UTR5;UTR3	4	7	3	4	3	3	5	4	3	3
Total	147 196	166 794	144 104	139 108	150 760	161 483	159 829	190 130	177 123	153 299
<i>Control pools</i>										
<i>Genome region^a</i>										
Downstream	1024	1047	901	1024	721	762	727	612	662	776
Exonic	21 755	20 916	21 605	21 259	20 885	21 415	21 831	21 913	20 981	21 608
Exonic;splicing	310	305	312	304	306	312	307	316	300	297
Intergenic	29 501	35 161	28 299	31 790	23 043	27 455	23 063	22 350	22 527	26 235
Intronic	92 012	88 906	83 922	90 250	70 290	74 229	75 347	63 912	69 481	74 572
Splicing	97	98	95	94	88	91	96	88	98	91
Upstream	1628	1546	1398	1654	1120	1170	1159	917	1077	1224
Upstream;downstream	94	89	86	91	58	69	64	47	53	65
UTR3	4018	3881	3746	3934	3124	3348	3440	2886	3126	3256
UTR5	1510	1405	1303	1461	1198	1235	1175	990	1147	1220
UTR5;UTR3	3	2	3	2	1	2	2	3	1	4
Total	151 952	153 356	141 670	151 863	120 834	130 088	127 211	114 034	119 453	129 348

^aOutput from annovar. http://www.openbioinformatics.org/annovar/annovar_gene.html.

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 for 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. Most SNVs were found in two (83) or three (31) pools. Ten SNVs had difficulties in primer design for genotyping and were replaced with singleton SNVs.

SNV validations were divided into two steps. First, the 144 SNVs were genotyped in 494 morbidly obese subjects, including the 100 exome-sequenced obese subjects, and 496 non-obese controls (Table 1). Genotyping of 20 SNVs failed. The average per SNV call rate for the remaining 124 SNVs was 97.95%. The concordance rate of MAFs estimated from exome sequencing data and the MAF based on genotype results was 0.83 (Supplementary Figure S1). Among the successfully genotyped SNVs, 21 were monomorphic. The remaining 103 variants were used for association analysis. HWE was checked prior to association analysis. After removal of 5 additional SNVs due

to an HWE P -value < 0.01, 98 SNVs were used in further association analysis.

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 morbidly obese subjects and 782 never-obese controls. Only one SNV, rs62623713 (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_024727:c.G1134C;p.L378F), showed nominally significant association with obesity in 1917 morbid obese and 1278 never-obese controls, with a P -value of 0.0063 (odds ratio (OR) = 1.32, MAF 0.081 in obese and 0.063 in controls) or a P -value of 0.004 (OR = 1.35, MAF 0.076 in obese, 0.058 in controls), respectively (Table 3). rs62623713 located at 110 019 439 base pair (bp) on chromosome 1 (hg19) is a missense variant in exon 4 of the SYPL2 gene. rs35923425 located at 169 569 432 bp on chromosome 3 is also a missense variant in exon 7 of the leucine-rich repeat containing 31 (LRRC31) gene.

By quantitative trait analysis in all genotyped subjects, rs62623713 showed strong association with BMI assuming an additive genetic model (beta 2.13, P -value of 6.28×10^{-5} , Table 4). The variant explained 0.5% BMI variation in genotyped subjects. Homozygous subjects for the rs62623713*G allele had higher BMI (mean = 45.56 kg/m²) when compared with subjects who were heterozygous or homozygous for the rs62623713*A allele (Supplementary Table S5).

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

CHR	SNP	A1	Freq _{case}	Freq _{control}	A2	P-value	OR	L95	U95	Gene
<i>First validation</i>										
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
<i>Second validation</i>										
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
<i>Final analysis</i>										
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

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 0.0007) 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^{-3}). 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).

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

GWA studies, and no other variants on those arrays was in strong linkage disequilibrium with rs62623713.

To the best of our knowledge there are only three reports of obesity-associated variants detected by exome sequencing.^{9,17,18} 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*⁹ 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

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

CHR	SNP	A1	TEST	No.	Beta	L95	U95	P-value
<i>Final total</i>								
1	rs62623713	G	ADD	3169	2.13	1.09	3.18	6.28E-05
1	rs41280330	T	ADD	3177	2.04	-0.82	4.90	0.1626
3	rs35923425	G	ADD	3174	1.50	0.43	2.58	0.006022
14	rs34106261	A	ADD	2782	1.09	-0.48	2.66	0.174
19	rs45546534	G	ADD	3171	0.90	-0.41	2.20	0.1799
<i>Female</i>								
1	rs62623713	G	ADD	2451	1.99	0.79	3.19	0.001182
1	rs41280330	T	ADD	2457	0.92	-2.36	4.19	0.583
3	rs35923425	G	ADD	2456	1.63	0.41	2.85	0.00897
14	rs34106261	A	ADD	2137	0.61	-1.17	2.39	0.5007
19	rs45546534	G	ADD	2451	0.46	-1.06	1.97	0.5532
<i>Male</i>								
1	rs62623713	G	ADD	718	2.52	0.48	4.57	0.01589
1	rs41280330	T	ADD	720	6.26	0.53	11.99	0.03253
3	rs35923425	G	ADD	718	1.00	-1.19	3.18	0.3723
14	rs34106261	A	ADD	645	2.76	-0.44	5.97	0.09143
19	rs45546534	G	ADD	720	2.02	-0.51	4.54	0.1176
<i>Lean</i>								
1	rs62623713	G	ADD	1269	0.48	0.05	0.92	0.0307
1	rs41280330	T	ADD	1270	0.75	-0.43	1.94	0.2125
3	rs35923425	G	ADD	1271	-0.07	-0.51	0.38	0.7708
14	rs34106261	A	ADD	1258	-0.01	-0.58	0.57	0.9769
19	rs45546534	G	ADD	1269	0.13	-0.40	0.66	0.6224
<i>Obese</i>								
1	rs62623713	G	ADD	1900	0.91	0.38	1.43	0.0007152
1	rs41280330	T	ADD	1907	1.37	-0.08	2.82	0.06345
3	rs35923425	G	ADD	1903	0.08	-0.46	0.62	0.7744
14	rs34106261	A	ADD	1524	1.19	0.35	2.04	0.005898
19	rs45546534	G	ADD	1902	0.04	-0.63	0.71	0.9119

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 coverage 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 $\geq 20\times$ (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.²¹ 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 scoliosis sample collection and characterization. We are indebted to Elisabeth Dungen for DNA preparation.

- 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.
- Berndt SI, 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.
- 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.
- 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.
- Walley AJ, Asher JE, Froguel P: The genetic contribution to non-syndromic human obesity. *Nat Rev Genet* 2009; **10**: 431–442.
- 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.
- Bamshad MJ, Ng SB, Bigham AW *et al*: Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet* 2011; **12**: 745–755.
- Ng SB, Buckingham KJ, Lee C *et al*: Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30–35.
- 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; **22**: 576–584.
- 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, Arner 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 Purcell 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, Grarup 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, Keogh 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 Onyike 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>)