Extreme obesity is associated with variation in genes related to the circadian rhythm of food intake and hypothalamic signaling

Edwin C. M. Mariman,1 Freek G. Bouwman,1 Erik E. J. G. Aller,1 Marleen A. van Baak,1 and Ping Wang1,2

1Department of Human Biology, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre, Maastricht, The Netherlands; and 2Laboratory of Biochemical Genetics, Department of Clinical Genetics, University Hospital Maastricht, Maastricht University Medical Centre, Maastricht, The Netherlands

Submitted 22 January 2015; accepted in final form 19 March 2015

Obesity nowadays is common in all Western and westernized countries, and its prevalence is gradually rising in other parts of the world. Because overweight and obesity are associated with increased risk for severe health complications like Type II diabetes and cardiovascular disorders, it leads to an alarming situation with respect to the current and future health of the global population. The etiology of obesity is complex with lifestyle, including eating behavior and physical activity, as a major contributor (23). In addition, successful maintenance of a healthy weight in an obesogenic environment clearly depends on the genetic background (20, 36).

Early onset is a hallmark for suspecting an underlying genetic cause for obesity. In several cases of (syndromic) childhood obesity, variants in candidate genes have been detected (29). Usually these variants have a low population frequency, have a high predicted or proven impact on protein function, and comply with a recessive mode of inheritance. Such variants have been found most frequently in the gene for MC4R, but also in other genes including the gene for the leptin receptor (LEPR) and single-minded family bHLH transcription factor 1 (SIM1) (22). Interestingly, it has been suggested that heterozygous variants with a predicted high impact on protein function may also confer risk for obesity. For instance, in the gene for MRAP2 that interacts with MC4R, three rare heterozygous variants with a predicted serious impact on protein function have been reported in early-onset obesity (4). It suggests that heterozygous variants can be involved in the genetic predisposition for early-onset obesity (18). Notably, all of these genes are expressed in the hypothalamus, a section of the brain that is involved in regulating eating behavior.

In the present study, we hypothesized that the genetic background of extreme obesity in young adults is constituted by a mixture of rare variants and common variants. The detailed exome sequencing data of 30 extreme obese subjects provided us the opportunity to explore the genetic background by a combination of methods used for analyzing rare variants and common variants. Because the hypothalamus seems to play a key role in the genetic predisposition of obesity, we focused on the variation of hypothalamus-related genes in relation to obesity. To analyze the variation of those hypothalamus-related genes, we searched first for unknown rare variants with a predicted damaging effect on the function of the protein. Second, we performed genetic association analysis of detected, already known variation.

MATERIALS AND METHODS

Subjects. Subjects were selected as described before (24). In short, 561 obese persons attended the private obesity clinic CO-EUR (http://www.co-eur.eu/), and from them we selected 30 (19 women, 11 men)
relatively young (average age 29.7 yr, range 19–40.4 yr) Caucasian subjects who were extremely obese (average BMI 51.1 kg/m², range 45.3–65.1 kg/m²). Information about weight during childhood was not collected, but eight subjects spontaneously mentioned that they had been overweight/obese at that time (present average age 31.2 yr, average BMI 54.2 kg/m²). Written informed consent for genetic studies was obtained from all participants, and permission was granted by the ethical committee of Maastricht University Medical Center.

**DNA isolation and sequencing.** DNA was isolated from peripheral blood leukocytes using the QIAamp DNA blood kit (Qiagen, Amsterdam, The Netherlands) and was outsourced for whole-exome sequencing in a CLIA-certified laboratory (EdgeBio, http://www.edgebio.com/). Sequencing was done on the Illumina HiSeq2000 after selection of genomic coding regions with the Nimblegen capture kit. Sequence results were compared with GRCh37 (b37) as the reference genome (http://www.ncbi.nlm.nih.gov/projects genome/assembly/grc/human). For every subject data were returned in excel-readable format (TSV) as a single nucleotide polymorphism (SNP) file and an indel file (insertion/deletion). Information was provided concerning genetic data (chromosome, cytoband, reference position, gene, zygosity, rs number, etc.), quality of the sequencing (quality score, depth), genetic variation in the 30 extreme obese subjects, we composed three data files with a read depth of >19 and a phred quality score >30, which is equivalent to a base call accuracy of >99.9%: one containing missense variants, one with nonsense variants, and one with indels. For the latter, only indels were included that related to a change of amino acid-codons, stop codons, splice sites, or to frameshifts. When allelic calls for “heterozygosity” [reference allele/referenced altered allele] were <0.25 or >0.75, or calls for “homozygosity” [altered allele/ altered + reference allele)] were <0.75, that information was regarded as incorrect and further ignored. The resulting three files were used to select all the variants in the listed hypothalamus-related genes. Selection of variants was done according to the scheme in Fig. 1. Rare variants are defined as those that have no rs number and were absent from the Genome of the Netherlands database (http://www.ncbi.nlm.nih.gov/projects/SNP/batchquery.html), and the Exome Variant Server (EVS; http://evs.gs.washington.edu/EVS/) for the European American population.

**Rare variant functional analysis.** For missense and nonsense variants, prediction of a damaging effect on the protein was based on a combination of a SIFT value of ≤0.05 and a PolyPhen2 classification of “P” (possibly damaging) or “D” (probably damaging). For indels, variants with either a “moderate” or “high” indication from SNP effect were taken into account. For variation already registered in public databases, allelic association analysis was performed.

**Association analysis of variants present in public databases.** For allelic association analysis, we applied a “case-control” approach in which the control group was represented by the general Dutch population. To take into account the small sample size, we used Fisher’s exact test instead of the χ²-test. The allele counts observed among the extremely obese subjects were compared with the expected allele

---

1 The online version of this article contains supplemental material.

---

**Fig. 1.** Overview of the approach used to select candidate variants for the genetic background of extreme obesity.
counts based on the allele frequency taken from the Genome of the Netherlands database. If not present there, the allele frequency was obtained from the dbSNP database and when also not present there, the allele frequency was obtained from the EVS for the European American population. Statistic significance was based on a false discovery rate (FDR) < 0.05 to correct for multiple testing (5).

RESULTS

Rare variants with a predicted impact on protein function.
The hypothalamus-related genes were screened for variants that did not have an rs number. For all missense variants, this resulted in the detection of 27 variants. They all occurred only once among the 30 subjects. Of those, eight were reported with a predicted damaging effect on protein function (SIFT ≤ 0.05 and P = D or P). One of these, NPC1 R714H, was recorded in the EVS database and had an allele frequency of 0.00012. The other indel is in the gene for LEPR (Table 1).

Search for allelic association. When our data files were screened for polymorphisms that were already known in population databases, we found 165 missense variants, 0 nonsense variants, and eight indels. Of those 173 polymorphisms, 146 (84%) were recorded in the Genome of the Netherlands database, one in the dbSNP data, and 26 on the EVS server. When we performed association analysis using population database allelic frequencies as a reference, we observed significant (FDR < 0.05) allelic association for six missense variants in different genes (Table 2). In the case of MBOAT4, there was

Table 1. Novel rare single indel and SNP variants with a predicted dramatic effect on the protein in hypothalamus-related genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Name</th>
<th>Known Functions (NCBI Gene Database)</th>
<th>Location</th>
<th>Variant Type</th>
<th>DNA Change</th>
<th>Protein Change</th>
<th>SIFT</th>
<th>PolyPhen2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAIAP3</td>
<td>BA11-associated protein 3</td>
<td>angiogenesis inhibitor, neurotransmitter secretion</td>
<td>16p13.3</td>
<td>SNP</td>
<td>AAG/ACG</td>
<td>K1181T</td>
<td>0.01</td>
<td>D</td>
</tr>
<tr>
<td>LEPR</td>
<td>leptin receptor</td>
<td>proline-rich coiled-coil 2A human major</td>
<td>1p31.3</td>
<td>indel</td>
<td>TAT/-</td>
<td>Y1078-</td>
<td>0.98</td>
<td>D</td>
</tr>
<tr>
<td>NBEA</td>
<td>Neurobeachin</td>
<td>human major histocompatibility complex class III region gene</td>
<td>6p21.33</td>
<td>SNP</td>
<td>CGG/TGG</td>
<td>R397W</td>
<td>0.00</td>
<td>P</td>
</tr>
<tr>
<td>PRRC2A</td>
<td>proline-rich coiled-coil 2A</td>
<td>human major histocompatibility complex class III region gene</td>
<td>19q13.2</td>
<td>SNP</td>
<td>AGC/AAC</td>
<td>S52N</td>
<td>0.01</td>
<td>D</td>
</tr>
<tr>
<td>RYR1</td>
<td>ryanodine receptor 1</td>
<td>calcium release channel</td>
<td>6q16.3</td>
<td>SNP</td>
<td>AAG/ATG</td>
<td>K4M</td>
<td>0.00</td>
<td>D</td>
</tr>
<tr>
<td>SIM1</td>
<td>single-minded family bHLH transcription factor 1</td>
<td>regulates trafficking and anorexigenic action</td>
<td>3q22.1</td>
<td>SNP</td>
<td>GCA/CGA</td>
<td>G408R</td>
<td>0.04</td>
<td>P</td>
</tr>
<tr>
<td>TRH</td>
<td>thyrotropin-releasing hormone</td>
<td>regulation and release of thyroid-stimulating hormone and prolactin</td>
<td>6p21.33</td>
<td>SNP</td>
<td>CGG/TGG</td>
<td>R231W</td>
<td>0.00</td>
<td>D</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DNA Change</th>
<th>Ref/Alt</th>
<th>Allele Freq. (GON)</th>
<th>Ref/Alt</th>
<th>Allele Freq. (this study)</th>
<th>Ref/Alt</th>
<th>P Value</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACT/GCT</td>
<td>T46A</td>
<td>0.034/0.966</td>
<td>2.2 x 10^-16</td>
<td>2.8 x 10^-14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAT/CGT</td>
<td>H215R</td>
<td>0.580/0.420</td>
<td>2.1 x 10^-4</td>
<td>5.4 x 10^-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAC/GCC</td>
<td>D149A</td>
<td>0.676/0.324</td>
<td>3.7 x 10^-6</td>
<td>3.1 x 10^-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CAG/GAG</td>
<td>Q339E</td>
<td>0.671/0.329</td>
<td>3.8 x 10^-9</td>
<td>2.4 x 10^-7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GCC/CCC</td>
<td>A962P</td>
<td>0.092/0.908</td>
<td>2.1 x 10^-4</td>
<td>5.4 x 10^-3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGT/CAT</td>
<td>R1740H</td>
<td>0.611/0.389</td>
<td>3.0 x 10^-9</td>
<td>3.0 x 10^-2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FDR, false discovery rate; RA, reference allele; GON, Genome of the Netherlands; Ref, reference allele; Alt, altered allele.
a significant disturbance of the Hardy-Weinberg equilibrium ($P < 10^{-6}$) with a considerable underrepresentation of heterozygotes in the present cohort. There was no indication of disequilibrium of this polymorphism in the Genome of the Netherlands general population. Also, this was not observed for other polymorphisms, making it very unlikely that the 30 subjects were selected from a population with a high degree of inbreeding.

**DISCUSSION**

**General features of this study.** Unraveling the genetic background of complex traits like obesity is usually done by GWAS with (very) large sample numbers. However, most of the genetic background cannot be explained in this way. This may be due to the involvement of relatively rare variants with a considerable impact on protein function. Therefore, other approaches are required that are able to extract information on both common and rare variants from affected subjects. Burden tests are nowadays often used, in which variants in a gene or a pathway are combined to ease the analysis of rare variants. However, this requires still hundreds of samples. On the other hand, whole genome/exome sequencing with rich genome data often is accompanied by a small sample size. Here we present an alternative approach for identifying candidate variants by making use of a homogenous extreme phenotype and a targeted biological process/pathway with a limited number of genes, in the context of rich sequence information obtained by whole exome sequencing. With this strategy we searched for novel candidate variants for extreme obesity in hypothalamus genes by combining screening methods for detecting novel rare variants and common variants. Indeed, both novel rare variants and (moderately) common polymorphisms (minor allele frequency $> 0.03$) were found as candidates for genetic predisposition to extreme obesity. Of course, the observed candidate variants need further verification in a larger sample with the same extreme phenotype.

It should be noted that for the association analysis we did not use a lean control group; rather, we compared variant frequencies in our study group with that of the general population. Primarily, we used the data of the Genome of the Netherlands for comparison. In that project 231 trios and 19 quartets with twins were collected from 11 of the 12 provinces of the Netherlands (8). This covered an area of 41,543 km$^2$ with more than 17 million inhabitants and included the province Limburg from where the present cohort was assembled. On the other hand, 27 (16%) of the 173 analyzed polymorphisms were compared with the dbSNP or EVS population allelic frequencies. None of those were found significantly associated with extreme obesity. In this regard, we cannot exclude the possibility that a difference in gene pool between those populations and our cohort limited the power for detecting obesity-contributing variation.

We did not investigate whether variants affected the risk for obesity in a positive or negative way. With regard to the rare variants, we selected for those with negative impact on the protein function, but in that way we may have missed rare variants with a positive effect on protein function. In this respect, the genes in which we detected the rare variants should have a protective function against obesity. For common variants, the observed altered allele frequencies of the associated variants (Table 2) were always smaller in our study group than in the general population. In the case of PRRC2A, the lower frequency of the altered allele 1740H with a predicted impaired function suggests that the risk for extreme obesity is related to a better performance of the 1740R form of the PRRC2A protein. For the other variants the PolyPhen2 prediction was B (benign), or no predictive information was available (NUCB2). For those genes we are not able to draw a conclusion on whether a better or worse protein performance is linked to the risk for obesity.

The rare variants detected in this study were all heterozygous. Interestingly, there is a growing awareness that also in the heterozygous state variants with a predicted high impact on protein function can influence a complex phenotype like obesity, and this awareness is supported by experimental evidence (4). Wheeler et al. (40) showed that a moderate frequency variant of LEPR was altered in expression and that this heterozygous variant was able to contribute to the genetic risk for severe early-onset obesity. Sequencing the SIM1 gene in 2,100 severe early-onset obese subjects and in 1,680 nonobese controls, Ramachandrappa et al. (30) detected 13 heterozygous missense variants. Nine of those variants, including one also present in the control group, had a decreased activity for transactivation of a reporter gene. Carriers of those variants showed increased food intake, indicating that heterozygous variants influenced the phenotype but possibly with decreased penetrance. Zegers et al. (44) did a similar analysis and found four heterozygous rare missense variants in SIM1 in 561 overweight/obese subjects, and two of those showed reduced transactivation activity in vitro.

**Comparison of the present results with GWAS loci and early-onset obesity genes.** So far GWAS, mainly in Caucasian cohorts, have revealed 56 loci linked to BMI and/or other obesity-related parameters (40). For 25 of those loci the associated SNP is located within the gene. Of those 25, seven (28%) were present in our list of hypothalamus-related genes: BDNF, DMN3, FTO, LEPR, MTCH2, NPC1, and NRXN3, underscoring the importance of the hypothalamus for the traits studied in the GWAS. Two of the seven genes also showed up in our present study: LEPR and NPC1. For LEPR the link with the risk for obesity has been well established (3, 39), but the 1078Y deletion has not been identified before. 1078Y is a putative phosphorylation site. In vitro studies with LEPR carrying a Y1078F mutation did not show a noticeable effect on pSHP-2 binding to the receptor and STAT3 mediated transcriptional activation. However, that does not preclude a role of 1078Y in other transduction reactions (7). NPC1 was first reported as the result of a GWAS of adult morbid obesity (26). The average BMI of the obese population was 47.3 kg/m$^2$, which is close to the present average BMI of 51.1 kg/m$^2$, and association was found in both studies with rs1805081 (H215R). According to the gene annotation portal BioGPS (http://biogps.org) NPC1 expression in the brain is highest in the hypothalamus. Mutations in the NPC1 gene cause the Niemann-Pick disease, an autosomal recessive lipid storage disorder characterized by progressive neurodegeneration. The protein is important for intracellular cholesterol trafficking, and recent genetic studies suggest a link between this gene and insulin homeostasis (2, 25, 32). However, metaanalysis by den Hoed et al. (11) indicates that NPC1 variants do not influence obesity-related traits in the general population.
In an Arabic cohort the H215R polymorphism appeared not to be associated with class I/II obesity (average BMI = 35.2 kg/m²) (2), leaving the possibility that the 215R allele is predisposing specifically to class III obesity (BMI ≥ 40 kg/m²) and that extreme obesity is a separate genetic entity.

Because the subjects of our cohort were relatively young, some of them might have had early-onset obesity. In the present study no variation of any kind was observed in the “classical” early-onset obesity genes MC4R and MRAP2, but two heterozygous rare variants were detected in the early-onset obesity gene for SIM1, which is linked to overfeeding (30), G408R and K4M. G408R was also one of the variants that were detected in the control group of the study by Zegers et al. (44). It suggests that the contribution of G408R to the risk for obesity is limited. On the other hand, the K4M variant has a strong deleterious score and is therefore more likely to affect the phenotype. Although no particular function has been ascribed to the NH2-terminal region of the SIM1 protein, it is remarkable that seven of 15 amino acids are lysine or arginine. This suggests a potential biological role for this basic segment and for the evolutionary conserved lysine at position 4. In addition, the presence of a second AUG codon close to the initiation codon may disturb the initiation of translation. A functional test is needed to provide convincing evidence in this case.

A GWAS among Chinese women has shown significant association of the PRRC2A gene locus on chromosome 6p with BMI, body weight, and the prevalence of obesity, a finding that could not be observed in children (34, 42). The expression of this gene, also referred to as BAT2, in the hypothalamus was lower in mice that had been on a high-calorie diet than in mice on a normal chow diet (43). It suggests that this gene in the hypothalamus may be under the control of the diet. It should be noted that the GWAS locus contains, besides PRRC2A, various other genes. The present study is the first indication that the genetic contribution of this locus is actually contributed by the PRRC2A gene.

Candidate genes and the circadian rhythm. PER1 is a central component of the biological clock. It is most highly expressed in the suprachiasmatic nucleus (SCN) of the hypothalamus and at lower levels in many other tissues. It is a key regulator of the activity of other genes in the body (45) and is suggested to play an important role in the coherence between the circadian clock and energy metabolism (19). Its expression pattern follows a 24 h cycle even without a light/dark signal. Therefore, variation in PER1 may cause shifting of the 24 h chronobiological cycle vs. the day/night cycling, and this may lead to disturbances of behavior and sleep, causing metabolic changes and making people prone to develop obesity (9, 19). Studies of nonfunctional variation in the PER1 gene showed association with extreme diurnal preferences and the timing of human behavioral rhythms (6, 17). Mice with a mutation of the S714 phosphorylation site in PER1 rapidly develop obesity on a high-fat diet (19). Our results are the first indication for a link between a functional variant in PER1 and extreme obesity in humans.

Also ryanodine receptors are located in the SCN and in the pineal gland. In fact, in humans the skeletal muscle type of ryanodine receptors, RYR1, has a relatively low expression in the hypothalamus but a high expression in the pineal gland. Rodent experiments performed with Ryr2 show that in these tissues ryanodine receptors regulate Ca²⁺ concentrations in response to external signals (1). Those signals may come from light through the retinohypothalamic tract. In that way, ryanodine receptor action helps to adjust the behavior and metabolism to the light and dark phases (12). Although a rare dramatic variant of RYR1 was detected in our cohort, no variants of RYR2 were observed that fulfilled our selection criteria.

Nucleobindin2 (NUCB2) is expressed in the paraventricular nucleus of the hypothalamus. In the rat it has been shown that NUCB2 and its processing product nesfatin-1 are involved in the circadian regulation of food intake (33). Gene expression increases during the early light phase and leads to suppression of food intake, whereas blunting the formation of nesfatin-1 leads to increased food intake. Altogether our findings seem to indicate the involvement of circadian regulators in the genetic predisposition for extreme obesity.

Candidate genes and hypothalamic signaling. Yet another process is put forward by our results, because five of the 12 genes are involved in signaling either as a receptor (RYR1, LEPR) or as a hormone or hormone modifier (MBOAT4, NPW, TRH). As mentioned above, RYR1 is a regulator of the calcium homeostasis in neurons and muscle cells. LEPR transfers the signal of the satiety hormone leptin to the POMC-positive neurons in the hypothalamus, which leads to an enhanced feeling of satiety. MBOAT4 is an enzyme that catalyzes the octanoylation of the hunger stimulating hormone ghrelin, which is secreted by gastrointestinal tissues. This posttranslational modification activates ghrelin and allows the hormone to bind to the type 1a growth hormone secretagogue receptor (16). Under certain conditions ghrelin is rapidly deacylated (15). Therefore, MBOAT4 in the hypothalamus may keep ghrelin in its active state, allowing it to efficiently report the hunger signal to the target neurons. The peptide hormone NPW is produced mainly in the paraventricular nucleus of the hypothalamus. It may have a role in transferring metabolic information from one set of neurons to another set via specific orphan G protein-coupled receptors located on corticotropin-releasing hormone-containing neurons (37). Injection into rats or mice induces reduced the food intake (27, 37). Thyrotropin-releasing hormone (TRH) has an anorexigenic activity in the hypothalamus and may activate similar signaling pathways as nesfatin-1 (NUCB2) and corticotropin-releasing hormone, which are involved in food intake regulation (13, 38). The heterozygous deleterious variant that we detected may, therefore, increase the risk of overfeeding. Altogether, signaling from and toward the hypothalamus also seems to be involved in the genetic predisposition of extreme obesity. Since this study focused on hypothalamic genes, it may not be surprising that the main functions of the hypothalamus are presented in the identified genes.

Other candidate genes. This study is the first report of a link between BAIAP3 and obesity. BAIAP3 is a p53-induced protein that binds to the brain-specific angiogenesis inhibitor 1. Genetic variation in BAIAP3 was found to be associated with anxiety and benzodiazepine abuse (41). Associations between anxiety disorders and obesity have been reported, but their interplay seems very complex. It is possible that the detection of BAIAP3 as a candidate gene may give direction to research aimed at better understanding this link. NBEA is a regulator of neuronal membrane protein trafficking and is important

Physiol Genomics • doi:10.1152/physiolgenomics.00006.2015 • www.physiolgenomics.org
for synapse development. Heterozygous knockout mice have enlarged fat depots, and increased food intake and in humans association was observed between intrinsic SNPs and BMI (28).

Here we have analyzed exome sequence data of 30 extremely obese subjects with the intention to identify genes with a functional link to the hypothalamus that are candidates for the genetic predisposition to extreme obesity. Twelve genes resulted from this analysis, some of which were already known to be related to obesity or adiposity parameters by human or animal experimental data. Yet most of the genes had not been reported before from GWAS or candidate gene analysis. This suggests that extreme obesity or class III obesity (BMI > 40 kg/m²) is a distinct genetic entity, which is in line with some previous studies (10, 21). Relevant processes for genetic predisposition seem to be regulation of the circadian rhythm of food intake and neuronal and metabolic signaling to and from the hypothalamus. Nevertheless, our findings do not exclude a role for other tissues/pathways, because our present analysis was focused on hypothalamus-related genes. Altogether, our findings underscore the complexity of the genetic background of obesity.

ACKNOWLEDGMENTS
We thank Jillian Loree for assisting in data analysis.

DISCLOSURES
No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

REFERENCES


