Genetic approach to unravel the molecular basis of severe early-onset obesity

Francesc Bou De Pieri

DOCTORAL THESIS UPF / 2020

THESIS SUPERVISORS

Prof. Luis A. Pérez Jurado

Dra. Clara Serra Juhé

Departament de Ciències Experimentals i de la Salut

Universitat Pompeu Fabra Barcelona

Al David,

al Joan,

a la Yaiza,

i a tots els que em feu sentir viu

ACKNOWLEDGEMENTS

Primer de tot, m'agradaria agrair al Luis i a la Clara l'oportunitat d'entrar a la Unitat de Genètica de la UPF i fer-me el seguiment durant aquests anys de doctorat. Al Luis, per l'orientació en l'àmbit científic i la confiança de deixar-me avançar el més autònomament possible. I a la Clara, per creure en mi des del primer moment. Per deixar-me formar part d'un projecte engrescador, ensenyar-me els misteris de la recerca i guiar-me en tot moment. Gràcies per la teva dedicació i l'amor per les coses ben fetes.

A Jesús y a Gaby, por haber compartido un gran proyecto. Por el titánico esfuerzo realizado cada día con los pacientes y los familiares, por las muestras y por las reuniones compartidas.

No voldria oblidar a tots els qui heu passat pel Lab422 i el Lab720 durant aquests anys. Per les estones compartides dins el laboratori i fora. Els partits de vòlei, les beer sessions, els open days, els dinars, els lab meetings, els congressos, les pràctiques a la facultat, les lluites compartides, els sopars de Nadal...

A la Raquel, per ser-hi sempre. Per les teves sàvies recomanacions, per sentir el laboratori com casa teva i per totes les hores dedicades a MEST.

A la Mar, per ser com ets. Per tenir sempre les idees clares, per la paciència amb mi i per la teva passió per l'assessorament.

A la Paula, per estar sempre disposada a tot. Per la teva dedicació, per compartir moments de tots colors i per les etiquetes dels primers.

Al Carlos, el Marcos i l'Armand, per ser els reis de les pantalles negres. Per iniciar-me en el món de la programació, per la paciència revisant codis i per la predisposició a ajudar sempre.

A l'Eulàlia, la Montse, l'Andrés i la Patrícia, per fer del despatx una extensió del laboratori. Per totes les gestions, les discussions constructives i les estones compartides.

A la Marivi, la Ivón i la Roser, per ser la veu de l'experiència, un referent i un puntal dins el laboratori.

Al Marc, per ajudar-me a entrar en el món de la clonació i els cultius cel·lulars.

A l'Aïda, la Judith, la Marta, la Tina, la Debora, la Maria i la Gaby, per haver compartit les meves primeres passes al laboratori. Per ensenyar-me a compartir els èxits i dividir els fracassos.

A tots els estudiants de màster que han passat. Al Rubén, per la teva espontaneïtat i dedicació. Al Marc, per totes les sondes de MLPA dissenyades. A la Marina, per aconseguir que anéssim al roco. A l'Arturo, pels riures infinits. Al Guillermo, per haver fet camí plegats. Al Ricard, per veure la recerca amb uns altres ulls. A la Mireia, per haver coincidit després de la carrera. I a tota la resta, per l'aprenentatge conjunt.

Als qGenomics, per haver compartit laboratori en els inicis, i especialment al Benja, per ajudar-me amb els pools.

També a la Mónica, l'Eva, la Sira i el Xavi, per formar part d'un gran equip de vòlei. Per les victòries, per les derrotes, però sobretot pels riures a la sorra.

A la cara d'una moneda.

I a totes les persones fora del PRBB amb qui he compartit un trosset de la meva vida i m'han acompanyat aquests anys.

Als amics del barri, als de tota la vida, per estar sempre propers i fer-me sentir viu. Per les partides clandestines, les orxates, les passejades per la Rambla, les múltiples escapades, els sopars que s'allarguen fins les tantes, les esquiades exprés i els viatges compartits per terres llunyanes.

A les Mambas, pels partits de dijous a la nit. Per ser la vàlvula d'escapament entre setmana, pels pastissos i per l'afició.

Als castellers de Barcelona, pels reptes aconseguits plegats, per les responsabilitats compartides, per les ganes de festa i per fer-me sentir part d'una gran família. A Nucli Dur, per acollir-me i fer-me descobrir persones precioses. No fan falta més paraules.

Al jovent de la UEC, per la seva passió per la muntanya, per caminar plegats i per les nits compartides sota les estrelles.

Als Canovins, per jugar i créixer mentre apreníem plegats el valor de les coses importants. I a tots els membres de les colònies familiars, per crear un ambient màgic on poder ser un mateix.

A les famílies de Pontons, per creure en un projecte de vida compartit.

Als genetistes Bang!, per haver compartit 4 anys de carrera, partides interminables, viatges a Mallorca i molt més.

Als monitors de l'esplai i als del MCECC, per creure en l'educació en el lleure fermament, per la infinitat d'hores dedicades amb passió i per tots els somriures regalats.

A la família, per ser incombustibles i ser-hi sempre. Al pare, per ser un referent de vida plena i ajudar-me a creure en mi mateix. A la mare, per ser propera i fer-me sentir estimat. A la Marta, per les argumentacions estúpides, pels balls i els riures. Al David, per fer-me costat sempre i per acompanyar-me en múltiples bogeries.

Moltes gràcies a tots.

ABSTRACT

Although obesity is clearly derived from an imbalance between energy intake and expenditure, and highly associated to lifestyle, it is considered a multifactorial disease with high heritability (50-75%), probably higher in early-onset cases. Different studies persistently describe a highly heterogeneous disease at the clinical and molecular levels. The genetic causes underlying the disease are still largely unknown and there is an important proportion of missing heritability. Different approaches have been used in the present thesis to unravel the molecular basis of severe early-onset obesity. Point mutations and copy number variants have been explored, and segregation studies have been performed for relevant alterations. Moreover, questionnaires to analyze the eating behaviors of mutation-carriers have been used. Results from the different approaches reveal a relevant role of genetic alterations in severe early-onset obesity, either as disease-causative mutations or susceptibility factors.

RESUM

Tot i que l'obesitat és clarament deguda a un desequilibri entre la ingesta d'energia i la despesa, i està molt associada a l'estil de vida, es considera una malaltia multifactorial amb una alta heretabilitat (50-75%), probablement superior en els casos d'inici precoç. Diferents estudis la descriuen de forma persistent com una malaltia altament heterogènia a nivell clínic i molecular. Les causes genètiques de la malaltia encara es desconeixen en gran mesura i hi ha una proporció important d'heretabilitat no explicada. En la tesi que es presenta s'han utilitzat diferents estratègies per esbrinar les bases moleculars de l'obesitat severa d'inici precoç. S'han explorat mutacions puntuals i variants en el número de còpies, i s'han realitzat estudis de segregació per a les alteracions rellevants. A més, s'han utilitzat qüestionaris per analitzar els comportaments alimentaris dels portadors. Els resultats dels diferents enfocaments revelen un paper rellevant de les alteracions genètiques en l'obesitat severa d'inici precoç, ja sigui com a mutacions causants de la malaltia o com a factors de susceptibilitat.

PROLOGUE

The development of new sequencing technologies in the recent years has been key to understand genetics and its relation to disease. Relying on the application of these platforms, a wide range of genotype-phenotype correlations have been identified and hundreds of new loci have been associated to complex and genetically heterogenic diseases, such as obesity.

This thesis expands the knowledge about the genetic basis of severe early-onset obesity, using various high-throughput technologies, including pooled DNA sequencing, SNP arrays and exome sequencing. In addition, it explores the contribution of maternally inherited variants to the phenotype and its implications in eating behavior.

The thesis is structured as follows:

The introduction gives a general overview of the obesity phenotype, focusing on its genetic causes. It also contains a brief description of the molecular techniques used. The last part provides a general picture about questionnaires assessing eating behaviors.

The main body of the thesis is divided in four chapters corresponding to the different studies, describing in detail the methods and results obtained.

The discussion aims to integrate and interpret the results obtained in previous chapters from a cohort perspective, as well as to place them into the framework of previous knowledge of the genetic basis of severe early-onset obesity. Following the discussions, the main findings of the thesis are summarized in the conclusions.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

1 OBESITY PHENOTYPE

1.1 DEFINITION AND CLASSIFICATION

According to the World Health Organization (WHO), obesity is defined as an abnormal or excessive fat accumulation which may impair health. Fat accumulation might be present in different parts of the body, including abdomen, hips, butt, thighs, calves, underarms, chest or others; and it can be measured using several methods [1,2]. The body mass index (BMI, calculated as weight/height^2) is the most widely used index to measure the relative weight and the nutritional status in adults. Other simple measurements used are the skinfold thickness, the waist circumference and the waist-to-hip ratio. Fat mass and distribution can also be estimated with predictive techniques such as bioelectrical impedance analysis, dual-energy X-ray absorptiometry, densitometry, hydrometry by isotope dilution and magnetic resonance imaging.

Although BMI does not measure body fat directly, it is the most accepted way to define obesity. On the one hand, BMI is an index with no direct relation with body composition, fat proportion or fat distribution, which are the main characteristics linked to the health risks of excess weight [3]. But, on the other hand, BMI is easily calculated and it has been shown to moderately correlate with more direct measures of body fat and appears to be strongly correlated with various adverse health outcomes [4,5]. According to BMI, adults with a BMI greater than or equal to 30 are classified as affected by obesity, whereas adults with a BMI in the range of 25-30 are considered overweight (Table 1). A BMI of 40 or higher (class 3 obesity) is sometimes categorized as extreme or severe obesity.

Category	BMI (kg/m^2)
Underweight	<18.5
Normal weight	18.5 to <25
Overweight	25 to <30
Obesity - Class 1	30 to <35
Obesity - Class 2	35 to <40
Obesity - Class 3	40 or higher

Table 1. Obesity classification according to BMI.

1.2 EPIDEMIOLOGY

Globally, in 2015, there were 1.9 billion overweight adults and 609 million adults with obesity, which represents around 39% of the world's population. During the last 35 years, the worldwide prevalence rates of overweight and obesity have almost doubled; from 25.4% to 38.5% in men (0.37%/y) and from 27.8% to 39.4% in women (0.33%/y) [6]. Some estimations point to more than half of the population being overweight or having obesity in 2030 if the current trends continue [7]. Although the absolute rates and trends in overweight and obesity vary substantially across regions and countries, generally, men have higher rates of overweight and women have higher rates of obesity, and its prevalence increases with age.

Actually, in Spain, there are around 71% of men and 50% of women affected by overweight or obesity. From 1987 to 2014, the prevalence of Spanish adults with overweight and obesity has increased 21.87% in men (0.81%/y) and 10.58% in women (0.39 %/y). In fact, the mean BMI is increasing by 0.10 $kg/m^2/y$ in men and 0.26 $kg/m^2/y$ in women [8].

Although obesity is seen as an epidemic restricted to industrialized societies, the estimations point to 115 million people suffering from obesity-related problems in developing countries. Indeed, patients with excess weight can be found worldwide and represent an important direct medical cost for the health system [9]. In Spain, these costs rise up to 1.95 billion €/y and represented 2% of the 2016 health budget [8].

1.3 CAUSES

Obesity is a multifactorial disease, mainly caused by a combination of environmental and genetic factors that lead to an imbalance between energy intake and expenditure. In first world countries, which have the majority of obesity cases, during the last decades, several factors favored a positive energy balance and weight gain. These factors include an increased availability of food supplies and consumption, especially of palatable food with high calories; decrease of jobs which need physical activity; substitution of leisure-time physical activities with more sedentary activities; inadequate sleep; and growing use of medicines that have weight gain as a side effect [10]. These and many other factors have set the basis for the current obesity epidemic. However, the exposure to these factors do not affect in the same way to all people, suggesting an individual variability explained by genetics. In fact,

different studies have shown that the heritability of obesity is high, ranging from 40% to 70% [11].

In general, genetic factors related to obesity are connected with the molecular components of the physiological system that regulates body weight, the leptin-melanocortin system (Figure 1), which control food intake and energy expenditure [12]. Leptin (LEP) is a hormone secreted by adipose cells and acts as an indicator of fat storage levels. An increase in body fat leads to increased levels of this hormone, which in turns reduces food intake. Decreased body fat is associated with decreased levels of leptin, which act to stimulate food intake. This compensatory mechanism allows weight to be maintained within a relatively narrow range, but leptin is only an indicator of nutritional status; the real regulation takes place in the arcuate nucleus of the hypothalamus. In this brain region there are two antagonistic neural populations responsible for integrating signals from peripheral hormones including insulin and leptin, as well as cognitive processes, the hedonic system and other stimuli. The first population expressing the leptin receptor (LEPR) is composed by anorexigenic neurons which have an appetite suppressing effect. These neurons express pro-opiomelanocortin (POMC), which is cleaved by proprotein convertases (PC1 and PC2) and generates the melanocyte-stimulating hormone alpha (alpha-MSH). Alpha-MSH is recognized by melanocortin receptors (MC4R and MC3R) expressed in neurons from the paraventricular nucleus of the hypothalamus and acts to reduce food intake. The other population expressing LEPR, together with the ghrelin receptor, is composed by orexigenic neurons which act as appetite stimulants. These neurons express the agouti-related protein (AgRP) and the neuropeptide Y (NPY), whose levels are increased by ghrelin and decreased by leptin. NPY acts directly through its receptor NPY1R and AgRP as a melanocortin receptor antagonist, both inhibiting the signals from anorexigenic neurons. Both neuronal populations project to neurons expressing melanocortin receptors located in the paraventricular nucleus of the hypothalamus. Transcription factors and neurotransmitters expressed in those neurons, such as single-minded homolog 1 (SIM1), brain-derived neurotrophic factor (BDNF) and tyrosine receptor kinase B (TrkB), are important modulators of the leptin-melanocortin system responsible for the physiological changes needed to maintain body weight regulation. Indeed, paraventricular neurons project to other satiety brain areas, such as the ventromedial nucleus and the parabrachial nucleus, responsible for the termination of feeding. Moreover, it may be involved in energy expenditure through BAT thermogenesis and physical activity control [13–15]. So, when fat stores are high, leptin levels are

increased leading to the activation of anorexigenic neurons which reduce food intake through the action of melanocortin receptors (Figure 1A). Contrary, when fat stores are low, leptin levels are decreased leading to the activation of orexigenic neurons which increase food intake (Figure 1B).

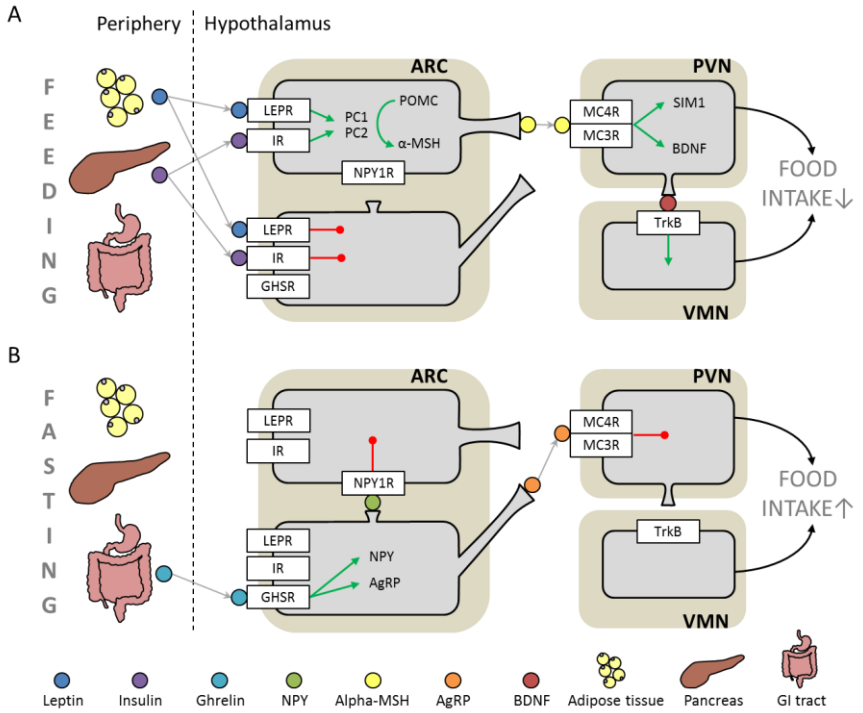

Figure 1. Schematic representation of the leptin-melanocortin system. Hypothalamic regions involved in the system are shown: arcuate nucleus (ARC), paraventricular nucleus (PVN) and ventromedial nucleus (VMN). Receptors present in these neurons are represented in white boxes: leptin receptor (LEPR), insulin receptor (IR), ghrelin receptor (GHSR), neuropeptide Y receptor type 1 (NPY1R), melanocortin receptors (MC4R, MC3R) and BDNF receptor (TrkB). Interactions between biological components from the system can be of activation (green arrows) or repression (red cropped arrows). **A:** Feeding state. When fat stores are high, leptin and insulin levels increase. Anorexigenic neurons induce the secretion of alpha-MSH, which activates the melanocortin receptors. Downstream signaling leads to food intake reduction. **B:** Fasting state. When fat stores are low, ghrelin levels are increased. Orexigenic neurons induce the secretion of neuropeptide Y and agouti-related protein, which repress the action from anorexigenic neurons. Lack of downstream signaling leads to an increase of food intake.

1.4 CONSEQUENCES

The main consequences of obesity are the significant morbidity and premature mortality associated with cardiovascular disease and type 2 diabetes, but also from cancer and chronic diseases, such as osteoarthritis, liver and kidney disease, sleep apnea and depression [16] (Figure 2). Indeed, compared with adults of normal weight, adults with severe obesity (BMI>40kg/m^2) have higher risk of several adverse outcomes including type II diabetes mellitus, high blood pressure, high cholesterol, asthma, arthritis, and generally poor health [17].

Cardiovascular disease in an obesity context can be explained by coronary heart disease, myocardial infarction, angina pectoris, congestive heart failure, hypertension and stroke. Several large prospective and observational studies have confirmed the increased risk of cardiovascular disease among patients with obesity and its marked adverse effects [18].

Body weight has been described as the major risk factor for type 2 diabetes and an important contributor to the insulin resistance that is often present in patients with obesity. In fact, recent estimates point to 89% of the US adult population with diagnosed diabetes being overweight or having obesity [19].

Increased body weight has been associated with increased death rates for several cancers, including colon, esophagus, liver, pancreas, kidney, among others. A large-scale prospective study has described increased death rates (52% higher for men and 62% higher for women) in patients with severe obesity (BMI>40) compared to normal weight adults. With the current observations, it is estimated that overweight and obesity could account for 14-20% of all deaths from cancer in adults [20].

A long-term positive energy balance leads to an excess adiposity and may be the cause of chronic diseases. Accumulation of lipids is accompanied by a volume increased in skeletal muscle, liver, and other organs. In obesity, subcutaneous adipose tissue store most of the excess lipids, but also visceral adipose tissue is enriched [21]. In specific organs and body locations the mass and the volume of adipocytes may lead to further adverse outcomes. In the liver, accumulations of lipids coming from plasma free fatty acids may be found in liposomes from hepatocytes which may lead to nonalcoholic fatty liver disease and cirrhosis [22]. In the kidney, renal compression can lead to an increase of blood pressure which may contribute to hypertension [23], and intraglomerular pressure may damage the kidneys and raise the risk of developing glomerulopathies, nephrolithiasis and malignancies [24]. In the lungs, breathing problems and poor airflow may appear due to asthma, obesity

hypoventilation syndrome and chronic obstructive pulmonary disease [25]. In the pharynx, airway blockage during sleep may lead to obstructive sleep apnea [26]. In the abdomen, an increase in the intraabdominal pressure may account for gastroesophageal reflux disease, Barrett's esophagus and esophageal adenocarcinoma [27]. In the pancreas, lipotoxicity may increase the incidence of acute pancreatitis [28]. In the gallbladder, high levels of cholesterol may form gallstones [29]. In the full body, an extra mechanical load on joints may lead to osteoarthritis and even gout [30]. In women, obesity has also been associated with menstrual dysfunction and infertility [31].

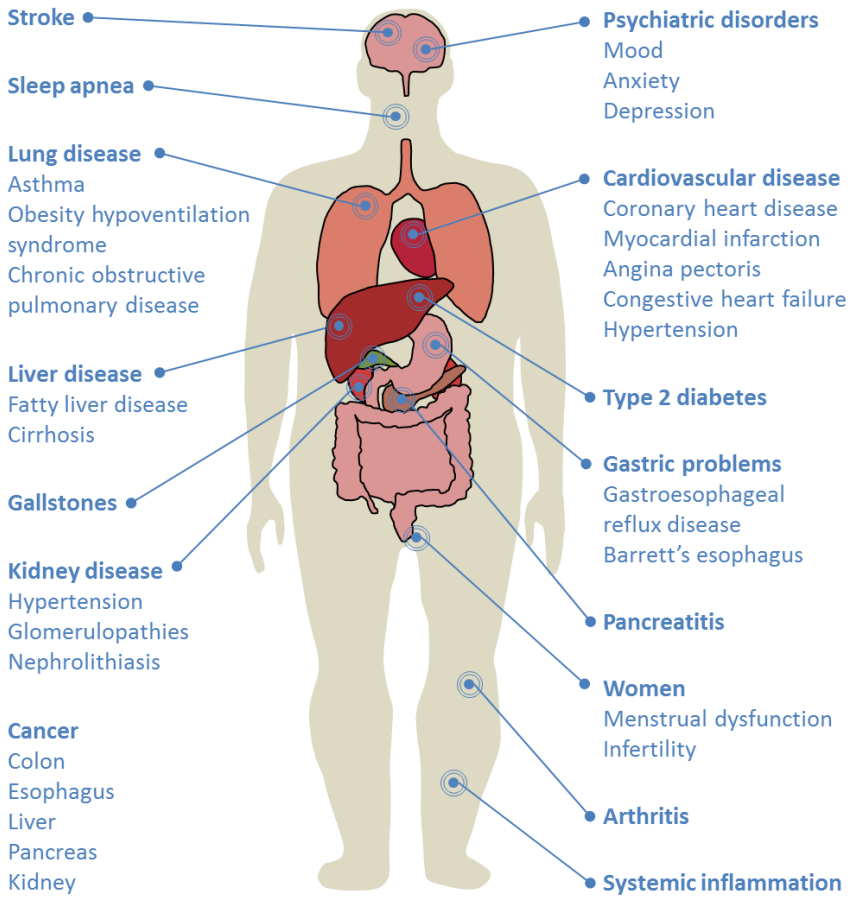

Figure 2. Major health problems caused by mechanical, metabolic, physiological and psychological effects of excess adiposity.

In addition, adipocytes synthesize proinflammatory adipokines, which lead to a low-grade systemic inflammatory state and contributes to the insulin resistance that is often present in patients with obesity [32]. All in all, hyperinsulinemia, together with hypertension and dyslipidemia represent the most important cardiovascular risk factors associated with the endocrine and metabolic changes found in obese patients [33].

Moreover, beyond the physiological problems described, obesity is associated with an increased prevalence of mood, anxiety, depression and other psychiatric disorders [34].

1.5 MANAGEMENT AND TREATMENT

According to guidelines, obesity needs to be addressed in a comprehensive way, from the diagnosis to the treatment [35,36]. Prevention of obesity has to be dealt in a multifaceted approach. From policy makers to individuals, going through states, local organizations, schools, childcare and healthcare professionals and families, all must work together to create an environment that support a healthy lifestyle. Weight management should be focused on a realistic weight loss to achieve a reduction in health risks and should include promotion of weight loss, maintenance of healthy weight and prevention of weight regain.

The first recommendation when treating obesity is to prescribe lifestyle modifications (dietary, physical activity, behavioral) to promote a decrease in BMI. Dietary changes have to be in line with healthy eating habits, such as decreased consumption of fast, processed and high-fat foods, avoidance of snaking between meals and decreased size of food portions. Regarding physical activity, sedentary behaviors must be limited and 20-60 minutes of moderate to vigorous physical activity should be carried out daily. Behavioral changes should be addressed in a family-based perspective encouraging healthy habits and improving self-esteem. Only when a formal program of intensive lifestyle modification fails to limit weight gain or to ameliorate comorbidities, pharmacotherapy or bariatric surgery may be used. Drug therapies for long-term weight management of patients with obesity and other co-morbidities include orlistat, lorcaserin, liraglutide, phentermine/topiramate, naltrexone/bupropion and other newer drugs [37,38]. Molecularly characterized obesity patients might be treated with more directed interventions, such as leptin therapy in leptin-deficient patients or MC4R

agonists in POMC-deficient patients [39,40]. Despite specific subpopulations, bariatric surgery has been proved the most effective treatment for morbid obesity in terms of long-term weight loss, improvements of co-morbidities and quality of life and decreases of overall mortality [41].

1.6 CHILDHOOD OBESITY

Excess body weight can also be manifested during pediatric age (up to 18 years old). As BMI during childhood is age- and sex-specific, obesity classification is done by percentiles (Table 2). Overweight is defined by a BMI at or above the 85th percentile and below the 95th percentile for children of the same age and sex, and obesity is defined by a BMI at or above the 95th percentile.

Category	BMI (percentile)
Underweight	<5th
Normal weight	5th to <85th
Overweight	85th to <95th
Obesity	95th or higher

Table 2. Childhood obesity classification according to BMI percentiles.

During the last 40 years a 10-fold worldwide increase has been observed in the number of school-age children and adolescents with obesity, rising from 11 million to 124 million [42]. Together with 216 million overweight children, childhood obesity represents a real global public health challenge that affects every country in the world. This increase of prevalence is a multifactorial phenomenon influenced by many factors, including food systems, commercial interests and social determinants [43].

Childhood obesity has been associated with comorbidities usually considered adult-specific diseases including type 2 diabetes, hypertension, dyslipidemia, fatty liver disease and obstructive sleep apnea [44]. In addition to health problems, childhood obesity may lead to social consequences. For example, children affected by overweight or obesity, compared with children with a healthy weight, may have lower self-esteem, higher likelihood of being bullied and poorer school attendance levels and school achievements; in addition to poorer employment prospects as an adult, and a lower-paid job [45]. Moreover, pediatric obesity may lead to obesity and related comorbidities into adulthood. In fact, 30% of adults with obesity were obese during pediatric age [46].

During pediatrics, special efforts must be done in a correct diagnosis using BMI, and genetic testing when needed [47]. Realistic lifestyle interventions should be prioritized as few data has been published regarding weight-loss medications and bariatric surgery outcomes in pediatric obesity [48].

Early-onset cases represent a relevant clinical subgroup of the obesity spectrum. The etiology of obesity is multifactorial, but early-onset severe cases are usually monogenic, while other forms of obesity are polygenic and are caused by the cumulative effect of multiple susceptibility genes involved in the regulation of energy intake and expenditure [49]. These premature cases may be less influenced by environmental factors and may be more likely explained by high penetrant genetic alterations. The study of this extreme subgroup, rather than adult population, is crucial to identify new genetic factors contributing into multifactorial and complex diseases, such as obesity [50].

2 GENETICS OF OBESITY

As mentioned before, some obesity cases may be explained partly, mainly or totally by genetic factors. Obesity can be manifested alone as an isolated form, however in some cases it is accompanied by other clinical features as part of a syndromic entity. Syndromic obesity has been widely studied and a great range of genes and genetic regions have been linked to the development of obesity together with other characteristics. Non-syndromic cases may be explained by rare highly penetrant variants in single genes (monogenic obesity) or may be the result of multiple common genetic susceptibility factors increasing the risk of developing obesity. Despite numerous genetic factors have been identified, we are still far from understanding the whole pathological mechanisms involved in obesity development.

2.1 SYNDROMIC OBESITY

Up to 79 obesity syndromes have been reported, but only 30 have been fully or partially genetically elucidated and 55 include obesity as a mandatory feature [51]. These syndromes can be caused either by chromosomal rearrangements, point mutations or epigenetic alterations, and should be correctly diagnosed as they are often associated with comorbidities that require additional evaluation and treatment. A description of the most frequent syndromic obesity forms is presented here.

2.1.1 Prader Willy syndrome

Prader-Willi syndrome (PWS) is a neurogenetic disorder considered the most common form of dysmorphic genetic obesity associated with intellectual disability. PWS is caused by paternal loss of function of imprinted genes in 15q11-q13, maternal uniparental disomy of the same region or mutations within the imprinting center [52,53].

2.1.2 Bardet-Biedl syndrome

Bardet-Biedl syndrome (BBS) is an autosomal recessive and genetically heterogeneous ciliopathy characterized by obesity, intellectual disability, renal anomalies, polydactyly, retinal degeneration and hypogenitalism [54]. BBS can be caused by mutations in up to 19 *BBS* genes mapped in different chromosomes [55]. A model of triallelic inheritance has been proposed for BBS,

where the presence of at least three pathogenic variants are required to manifest the phenotype [56].

2.1.3 Alström syndrome

Alström syndrome (ALMS) is an autosomal recessive disorder characterized by progressive cone-rod dystrophy leading to childhood obesity, blindness and sensorial hearing loss. ALMS is caused by mutations in the *ALMS1* gene on chromosome 2p13 [57].

2.1.4 Carpenter syndrome

Carpenter syndrome (CRPT) is an autosomal recessive disorder associated with the dysfunction of the primary cilium. Patients affected by CRPT may have craniosynostosis, polydactyly, soft-tissue syndactyly and obesity, due to mutations in the *RAB23* gene on chromosome 6p11.2 [58].

2.1.5 Cohen syndrome

Cohen syndrome (CS) is an autosomal recessive multisystem disorder characterized by facial dysmorphism, microcephaly, truncal obesity, intellectual disability, progressive retinopathy and intermittent congenital neutropenia. CS is caused by mutations in the *COH1* gene on chromosome 8q22 [59].

2.1.6 Beckwith-Wiedemann syndrome

Beckwith-Wiedemann syndrome (BWS) is an overgrowth syndrome with widely variable clinical phenotype that attenuates with age and involves a predisposition to tumor development [60]. BWS is caused by genetic and epigenetic alterations affecting the imprinted 11p15.5 locus. This region includes two genes, *H19* and *IGF2*, with parent-of-origin expression regulated by two differentially methylated imprinting control regions, IC1 and IC2. In normal conditions IC1 is methylated in the paternal allele and IC2 is methylated in the maternal allele. BWS can be caused by paternal UPD, hypermethylation in IC1, hypomethylation in IC2 or point mutation in active genes from the region, such as *CDKN1C*. All these alterations lead to overexpression of genes that stimulates growth and cell proliferation [61]. As patients with BWS may be misdiagnosed as isolated obesity, genetic testing should be considered because of the increased risk for malignancies [62].

2.1.7 WAGRO syndrome

WAGRO syndrome is a subcategory of WAGR syndrome. WAGR syndrome is a contiguous gene deletion syndrome on chromosome 11p13 characterized by Wilms' tumor, aniridia, genitourinary anomalies and intellectual disability. Patients with larger deletions encompassing the *BDNF* gene along with other critical genes also present obesity (WAGRO syndrome) [63].

2.1.8 Pseudohypoparathyroidism

Pseudohypoparathyroidism (PHP) is a heterogeneous group of disorders with resistance to parathyroid hormone as the common feature. Patients with PHP are characterized by short stature, obesity, round facies, subcutaneous ossifications, brachydactyly and other skeletal anomalies, which are referred as Albright hereditary osteodystrophy. PHP is caused by mutations in the *GNAS* genes on chromosome 20q13.32 (type Ia) or alterations in a differentially methylated region from the same locus (type Ib) [64].

2.2 MONOGENIC OBESITY

Around 5-10% of non-syndromic obesity cases can be explained by highly penetrant rare genetic variants affecting single genes. The leptin-melanocortin system is the main regulator of energy balance, so most of monogenic obesity forms are caused by mutations affecting genes involved in this pathway [65]. However, genes related to adipogenesis and other pathways may also cause Mendelian obesity. Moreover, new candidate genes are being identified every year and the percentage of explained obesity cases may still increase.

2.2.1 Leptin and leptin receptor

Mutations affecting *LEP* gene and its receptor, *LEPR*, were first reported in 1997 and have been associated with severe obesity and intense hyperphagia in humans and rodents [66,67]. Leptin deficiency and leptin resistance are considered rare autosomal recessive forms, as homozygous and compound heterozygous mutations have been described in few severe obesity patients.

2.2.2 Melanocortin receptor

MC4R mutations are the most common monogenic form of obesity accounting for up to 6% of non-syndromic cases [68]. MC4R deficiency

represents a serial disruption of the central energy-balance pathway. Patients with *MC4R* mutations are characterized by severe obesity, increased lean mass, increased linear growth, hyperphagia and severe hyperinsulinemia. Homozygous carriers of loss-of-function mutations are more severely affected compared to heterozygous carriers. Therefore, *MC4R* mutations are considered to be inherited in a codominant manner [68].

2.2.3 Melanocyte precursors

Mutations in *POMC* and *PCSK1* are associated with recessive forms of obesity due to decreased levels of functional alfa-MSH [69,70]. Patients with *POMC* mutations are characterized by obesity, hypocortisolism, hair and skin hypopigmentation, neonatal hypoglycemia, seizures, cholestasis and voracious appetite [69]. In addition to severe early-onset obesity, patients with *PCSK1* mutations display hyperphagia, hypoglycemia, hypogonadotropic hypogonadism, hypocortisolism, elevated plasma proinsulin and elevated POMC but low insulin concentrations [70].

2.2.4 Melanocortin signaling

BDNF, its receptor *NTRK2*, and *SIM1* are genes involved in the development and maintenance of the paraventricular nucleus of the hypothalamus, but also play an important role in the regulation of energy homeostasis [71]. Moreover, mutations in components of the signaling pathways around the melanocortin system have also been associated with obesity development, including *SH2B1*, *KSR2* and *ADCY3* [72–74].

Patients with mutations in *BDNF* and *NTRK2* are rare, but a great range of genetic alterations, including inversions, deletions and point mutations have been associated with obesity [75–77]. These patients are characterized by severe early-onset obesity, hyperphagia, developmental delay and cognitive impairment. Deletions [78], translocations [79] and loss-of-function mutations [80] affecting *SIM1* have also been reported as the cause of severe obesity and a Prader-Wili-like phenotype [81,82].

Moreover, mutations affecting the *SH2B1* gene, which codes for an adaptor protein involved in signaling pathways including the LEP and BNDF receptors, have been linked with autosomal dominant forms of early-onset obesity exhibiting hyperphagia, insulin resistance and short stature [83,84]. *KSR2* gene codes for a molecular scaffold that regulates some signaling cascades and activates AMPK, a key regulator of cell energy homeostasis. Several loss-of-

function *KSR2* variants have been reported in severe early-onset obesity patients [85]. *ADCY3* gene is involved in the formation of cAMP, an essential second messenger in intracellular signaling of key metabolic factors such as glucagon-like peptide 1, ghrelin and alfa-MSH. Few patients have been described with homozygous mutations affecting the *ADCY3* gene [86].

2.2.5 Candidate genes

Rare variants in candidate genes beyond the leptin-melanocortin system may be responsible for some obesity cases. Candidate genes can be identified by function (when are related direct or indirectly with the pathogenesis of the disease) or by chromosomal position (when are in genomic regions associated with the phenotype). A mix of both approaches can also be found, as in the case of *IRX3* gene, where BMI-associated *FTO* variants have been described to change IRX3 expression and shift adipocyte differentiation from energy-dissipating beige adipocytes to energy-storing white adipocytes [87,88]. Recent studies have described rare nonsynonymous variants in *IRX3* gene suggesting a possible link between IRX3 deficiency and human obesity [89].

Other candidate genes for obesity include *MC3R*, *NPY*, *PPARG*, *GRPR*, *GRIK1* and *GRM7*. The other melanocortin receptor expressed in the hypothalamus, the MC3R, has a complementary role to reduce body weight. Despite its contribution to obesity has been controversial for years and only few patients have been reported with pathogenic mutations, recent findings support its pathological contribution to obesity [90,91]. Gain of function mutations affecting *NPY* may be associated with obesity development [92,93]. A duplication affecting *NPY* gene has been described in a familial case with early-onset obesity and attention deficit hyperactivity disorder (as shown in Annex 1). PPARG is a regulator of adipocyte differentiation and has been implicated in the pathology of diseases such as obesity and diabetes. Mutations in the *PPARG* gene have been linked to accelerated differentiation into adipocytes and greater cell accumulation of triglycerides [94,95]. Receptors for gastrin and glutamate, among others, may also be relevant genes to explain some obesity cases [93] (as shown in Annex 1). Variants in those genes will require a functional validation to confirm its contribution to the phenotype.

2.3 SUSCEPTIBILITY FACTORS

Some familiar obesity cases may be explained by rare highly penetrant genetic variants. Nevertheless, as obesity is a continuous and complex phenotype and

only a few percentage of cases can be explained by monogenic causes, a multigenic or multifactorial inheritance may also be a good model to explain the etiology of some obesity cases. Genome-wide association studies (GWAS) conducted on obesity and related traits in humans have revealed common genetic variants with low penetrance involved in those phenotypes. However, the risk explained by known SNPs remains relatively modest compared with the estimated heritability of BMI. Other factors such as copy number variants (CNVs) and epigenetics could represent an additional source of the heritability that is missed by GWAS.

2.3.1 SNPs

Using large meta-analysis, tenths of common genetic variants have been associated to BMI [96]. The first obesity susceptibility locus identified by GWAS is a cluster of SNPs in intron 1 of fat mass and obesity associated (*FTO*) gene [97]. Other significantly associated SNPs for BMI are one 200kb downstream of *MC4R* and one near *TMEM18* [98]. Additional common genetic variants in genes possibly linked to energy balance have been related to BMI, such as *ADRB3*, *BDNF*, *GHSR*, *NEGR1*, *PCSK2*, *PPARG* and *SH2B1*. Moreover, an *AGRP* polymorphism has been associated with human fatness [99]. SNPs in these and other genes may lead to new candidate genes potentially involved in obesity development.

Cumulatively, the stronger effects of single common variants might explain about only 2% of the variance in BMI. The gap between the risk explained by known SNPs identified by GWAS and the estimated heritability from twin and familiar studies is defined as the missing heritability.

2.3.2 CNVs

Studies in complex traits have revealed that CNVs are important contributors to the genetic basis of complex disorders [100–102]. Obesity is not an exception and CNVs represent a genetic variation that deserves to be taken into account. Frequent CNVs acting as susceptibility factors and rare CNVs with high penetrance might be associated with extreme obesity phenotypes [103]. Susceptibility forms for obesity have been described all along the genome and include deletions in 1p31.1 upstream of *NEGR1* gene [104], deletions in 10q11.22 encompassing the *NPY4R* gene [105], duplications in 10q26.3 encompassing the *CYP2E1* gene [106], deletions in 11q11 encompassing different olfactory receptor genes [107] and deletions in 16p11.2 encompassing the *SH2B1* gene [108,109], among others. For example, patients with a 1p21.1

deletion encompassing the *AMY1* gene have been associated with higher BMIs compared to individuals without the deletion. This genomic alteration affecting the salivary amylase gene represented the first link between carbohydrate metabolism and BMI [110].

Despite the great range of susceptibility loci identified up to now, genetic studies revealed that common CNVs are not a major contributor to the severe obesity phenotype [107]. However, rare CNVs may account for part of the missing heritability of BMI. These less-frequent type of alterations may be highly penetrant and had been widely studied in syndromic forms of obesity. But, rare CNVs may also explain some familiar cases of non-syndromic obesity as a significantly increased burden of rare CNVs has been documented in patients with isolated severe early-onset obesity [103] (Annex 1).

2.3.3 Epigenetics

Epigenetics may not be the primary cause of obesity, but recent studies have pointed out its relevance in the current obesity pandemic. The rapid worldwide increase in obesity prevalence may be explained by epigenetic modifications, which are highly influenced by nutritional and environmental factors [111]. Multiple obesity-associated differentially methylated sites have been identified, but there are still no evidences linking global methylation and obesity [112]. Methylation alterations that change the expression of genes involved or candidates for obesity, such as *LEP*, *POMC*, *NPY*, *ADIPOQ*, *HIF3A*, *IGF-2*, *IRS-1*, *CLOCK*, have been linked to the development of obesity and metabolic disorders. In addition, histone modifications during adipocyte differentiation or hypothalamus development can play an important role in obesity development, and increased levels of specific miRNA have been associated with BMI in children [111,113].

3 EXPERIMENTAL TECHNIQUES

In order to study genetic alterations involved in a complex and multifactorial disorder, such as obesity, different approaches might be used, including sequencing and genotyping studies. In this section, different techniques relevant for this thesis are discussed.

3.1 TECHNIQUES TO STUDY POINT MUTATIONS

3.1.1 Sanger sequencing

Described in 1977, Sanger sequencing was the first DNA sequencing method [114] and nowadays is still widely used. It is based on the selective incorporation of chain-terminating dideoxynucleotides during an *in vitro* DNA replication. A conventional DNA replication process consists of a DNA polymerase that starting from a DNA primer incorporates deoxynucleotides complementary to a single-stranded DNA template. Sanger sequencing incorporates low concentrations of modified fluorescently-labeled dideoxynucleotides missing the 3'-OH group which do not allow further DNA elongation. DNA replication proceeds until a modified dideoxynucleotide is incorporated, which blocks DNA elongation. Several cycles of replication generate DNA fragments of different lengths depending on the position where the dideoxynucleotides has been incorporated. Fluorescent labels in the dideoxynucleotides are used for automatic detection after electrophoresis [115], which allows reading the sequence of the fragments (Figure 3).

Sanger sequencing is used to analyze a targeted genomic region and always needs previous information regarding the genomic location of the variant or the gene. Sanger sequencing has a limited throughput, so for large-scale genome analyses new methods are used, which are called next generation sequencing (NGS) [116]. However, Sanger sequencing stands out for long read length (up to 1kb approx.), excellent accuracy and high quality, which makes it a usual method for validation of NGS results.

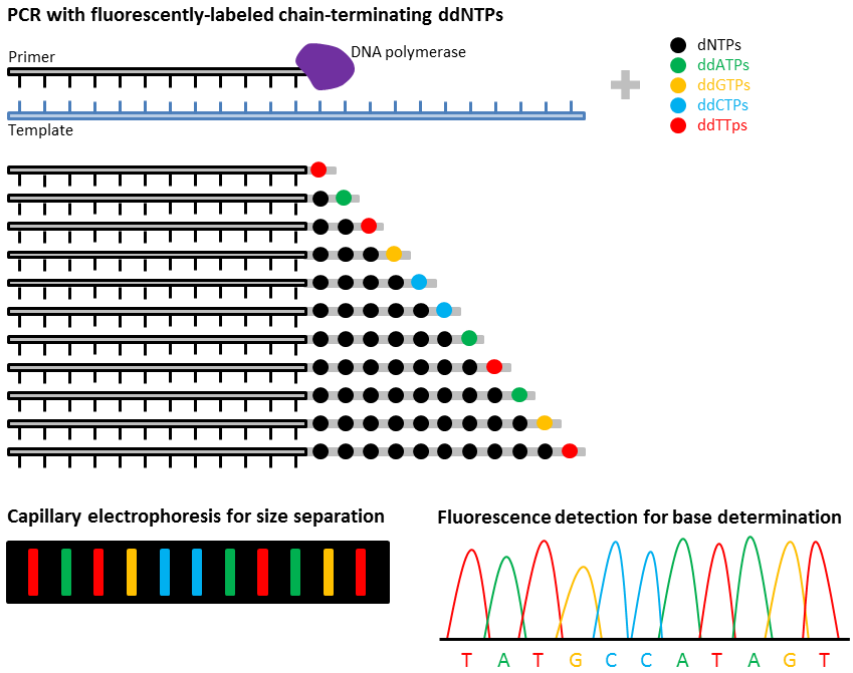

Figure 3. Schematic overview of Sanger sequencing. Amplification with chain-terminating dideoxynucleotides generates products of different size that can be separated by electrophoresis. Colors from each fluorescently-labeled dideoxynucleotide allow the reading of the sequence.

3.1.2 Next generation sequencing

NGS includes several technologies which are mainly characterized by a high-throughput parallel sequencing producing thousands or millions of sequences at once [117–120]. Despite accuracy is lower than Sanger sequencing and short reads produced by these technologies need to be mapped to a reference genome, the improvement in costs and times makes NGS the most widely used method to identify point mutations along the genome.

NGS methods consist on a random fragmentation of the DNA, the ligation of adaptors, the amplification of all fragments within the same PCR, an enrichment step if needed, the massive-parallel sequencing reactions, the mapping of reads to a reference genome and the calling of variants. NGS results can be very sensitive to false positives and false negatives [121,122]. Filters are used to reduce false positives and to identify the pathogenic variants among the many thousands variants detected. To prevent false negatives, low overall coverage, low accuracy and poor capture efficiency should be avoided.

The management and analysis of high amounts of genome-wide sequencing data is accomplished thanks to the rapid development of bioinformatics and the use of specific software.

As most of the variants detected in a whole genome analysis are difficult to interpret, other approaches using DNA enrichment are a valuable option [123]. Exome sequencing involves exome enrichment of DNA samples using hybrid or solution captures. The exome includes all protein coding genes and constitutes only approximately 1% of the human genome, but it harbors 85% of the mutations with large effects on disease-related traits. Indeed, several genes related to Mendelian diseases have been identified by using this method [124,125]. Moreover, DNA enrichment can be targeted to a specific panel of genes or genomic regions depending on the interest of researchers or clinicians.

3.1.3 Pooled DNA sequencing

In order to identify the genetic causes of multifactorial and highly heterogeneous disorders, a large number of patients is needed, which is translated into increased costs. In non-Mendelian disorders a minimum mutation burden in a specific gene is needed to prove its pathological relevance. Moreover, due to population stratification, controls from the same geographical origin are used to discriminate between variants associated with the phenotype and population variants. For these reasons, in sequencing projects of multifactorial or heterogeneous disorders, large numbers of samples must be analyzed. New approaches are being used in order to reduce costs without reducing the number of samples sequenced, such as pooled DNA sequencing [126].

Pooled DNA sequencing takes advantage from the deep sequencing capacity of NGS platforms to sequence multiple samples in the same sequencing reaction. In order to assign each variant detected to the correct sample, each sample must be present in two different pools (only sharing this sample) and variants are only real when detected in both pools. Common variants are detected in all, or nearly all, pools and the identification of the samples harboring the variants became impossible, so this pooling approach is only valid when looking for rare sequence variants.

The decrease in cost is proportional to the reduction of sequencing reactions, which depends on the number of samples included in each pool. For example, with 20 samples per pool, the cost is reduced one order of magnitude, which in projects with hundreds of samples is really relevant. However, additional

bioinformatics analyses are needed to tackle with the new complexity that represents pools.

3.2 TECHNIQUES TO STUDY GENETIC REARRANGEMENTS

3.2.1 MLPA

Multiplex Ligation-dependent Probe Amplification (MLPA) is mainly used to study copy number alterations, but it can also detect point mutations and methylation alterations. Among other CNV-detection methods, MLPA is a low cost, time-efficient and technically uncomplicated method and requires low DNA input for simultaneously detection of CNVs in several regions. Moreover, there is no need to know the exact breakpoint site to test the CNVs. However, its main limitations are the reduced number of regions that can be tested simultaneously and the necessity of including control samples as dosage references.

MLPA is based on the amplification of multiple MLPA probes that hybridize to target sequences using a single pair of PCR primers [127]. MLPA probes consist of two oligonucleotides complementary to adjacent genomic regions. Each oligonucleotide has a hybridization sequence and a sequence complementary to the universal PCR primers; moreover, one of the oligonucleotides has to be fluorescently labeled and can include a stuffer sequence. The MLPA reaction can be divided in several major steps (Figure 4). The first step of a MLPA reaction consist on the denaturalization of the DNA and the hybridization of the probes during an overnight incubation. The second step is the ligation of the adjacent probes. Ligation is only possible when the two probe oligonucleotides are both hybridized to their adjacent targets, creating a ligated probe containing both PCR primers binding sites. The third step consists on the amplification with the universal primers of all the ligated probes. The exponential amplification generates PCR products proportional to the number of target sequences in the sample, which are separated using capillary electrophoresis during the fifth step. Finally, the analysis of the signal allows identifying regions with reduced intensity (deletions) or increased intensity (duplications) in samples compared to a reference.

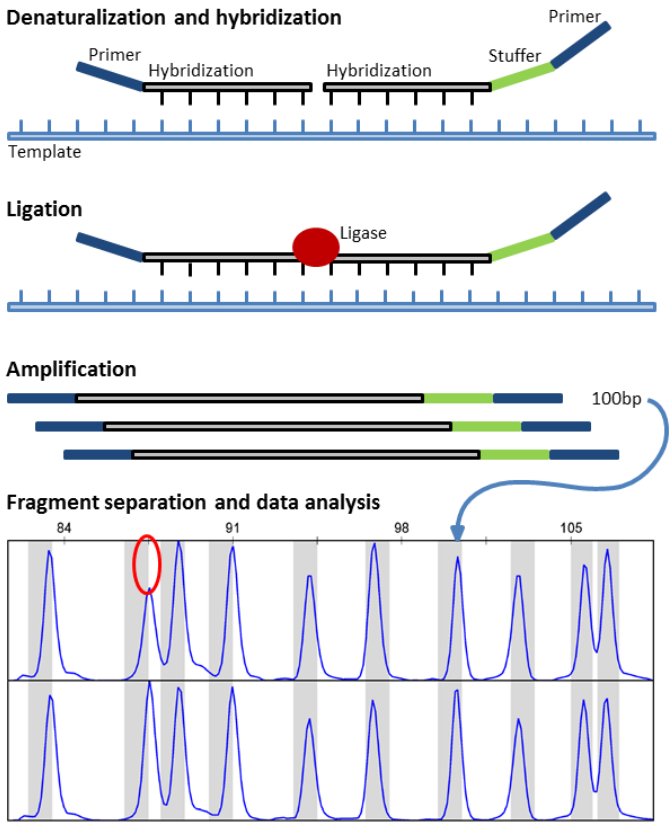

Figure 4. Schematic overview of MLPA. Amplification of ligated probes for each region of interest using one pair of primers generates products of specific lengths. The comparison of intensities between peaks and samples allows the identification of CNVs.

3.2.2 SNP arrays

SNP array platforms are used to genotype lots of SNPs along the genome, which might also be applied to study CNVs. The different platforms available differ by the number and the genomic location of the SNPs studied, and may have different sensitivities and specificities [128]. For every SNP, the intensity of each allele is measured to quantify the total intensity and the allele proportion, which are used to detect copy number alterations (Figure 5).

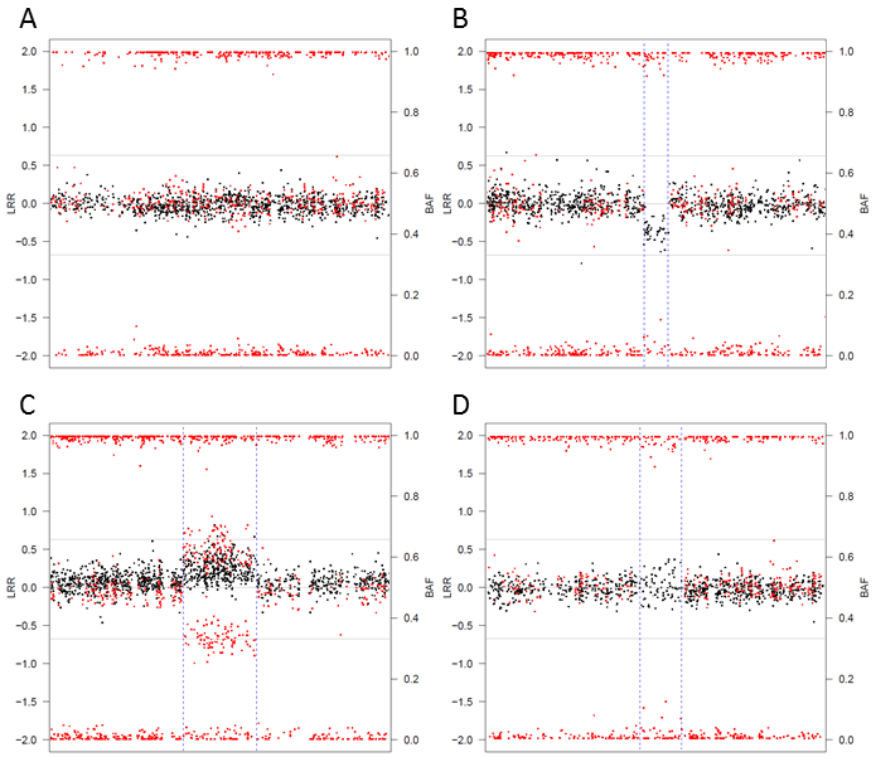

Figure 5. In SNP array plots, the LogR ratio and the BAF (B allele frequency) are shown for every studied SNP. The LogR ratio is a normalized value which measures the genetic dosage and is plotted in black; and the BAF is calculated as the B allele signal divided by the total signal from A and B alleles and is plotted in red. **A:** SNPs in diploid regions are expected to have a LogR ratio around zero and can be homozygous for the A allele (BAF=0), heterozygous (BAF=0.5) or homozygous for the B allele (BAF=1). **B:** SNPs in deletions have a decreased LogR ratio (negative value) and only hemizygous positions for each allele are detected (BAF=0 or BAF=1) as no heterozygotes exist. **C:** SNPs in duplications have an increased LogR ratio (positive value) and the BAF is altered because the total number of alleles is increased. **D:** Normal LogR ratio values with no heterozygous positions are indicative of homozygous regions.

SNP array data can also be used to perform association studies (comparing allele frequencies between groups). And despite, the main limitation of SNP arrays is the inability to detect balanced rearrangements, new algorithms are being developed to study inversions by measuring the linkage disequilibrium between SNPs.

4 BEHAVIORAL QUESTIONNAIRES

Specific questionnaires can be used to assess the possible implications of genetic variants into the eating and feeding attitudes of carriers. A glance of different questionnaires is presented here.

4.1 PRIMARY DIET QUESTIONNAIRE

As obesity is a multifactorial disease, not only genetic factors have to be taken into account when defining its etiology. Body weight is regulated mainly by adjusting energy intake and energy expenditure, so food consumption is half of the picture. Healthy dietary habits have been associated with reduced obesity incidence, whereas high fat diets are known as weight gain inducers [129]. Unhealthy habits, such as overeating and sedentary lifestyle, are the main targets of lifestyle interventions because are important non-genetic contributors to body weight deregulation. Moreover, we cannot discard that unhealthy habits could be the consequence of a behavior dysfunction due to genetic alterations affecting hypothalamic body weight regulators, which may establish a direct link between genetics and eating habits.

To evaluate eating behaviors many questionnaires have been developed. Primary diet questionnaires aim to quantify recent food consumption in order to classify diets as healthy or not. These questionnaires can be used to understand the dietary habits of diagnosed patients or to follow the evolution after lifestyle modifications. All in all, identifying dietary dysfunctions allow the characterization of other obesity contributors beyond genetic.

4.2 FAMILY ADAPTABILITY AND COHESION EVALUATION SCALE

Several studies pointed out the relevance of the familiar context in the individual's development and health. Each family is a different world, but can be generally defined using the three central dimensions of the Circumplex model: cohesion, adaptability and communication [130]. Family cohesion is the degree of affective union between family members and can define families as "disengaged" indicating considerable independent family relations, "separated" indicating less distancing among family members, "connected" indicating closer family relations and ties or "enmeshed" indicating relationships of dependence among family members. Family adaptability is

understood as the capacity of a family to be flexible and able to alter its structure based on its needs, and according to different levels of adaptability, families can be classified as "rigid" representing a centralized authority with no possible changes in the system, "structured" representing an open family with sharing of roles, "flexible" representing a more malleable family regarding rules and the distribution of activities or "chaotic" representing a family with no definition regarding the roles and activities of the members. Family communication is seen as a facilitating dimension, which helps the movement on the other two dimensions. The major hypothesis of the Circumplex model is that moderate levels of cohesion and adaptability are related to a good family functioning, whereas the two extremes of the dimensions are dysfunctional (Figure 6).

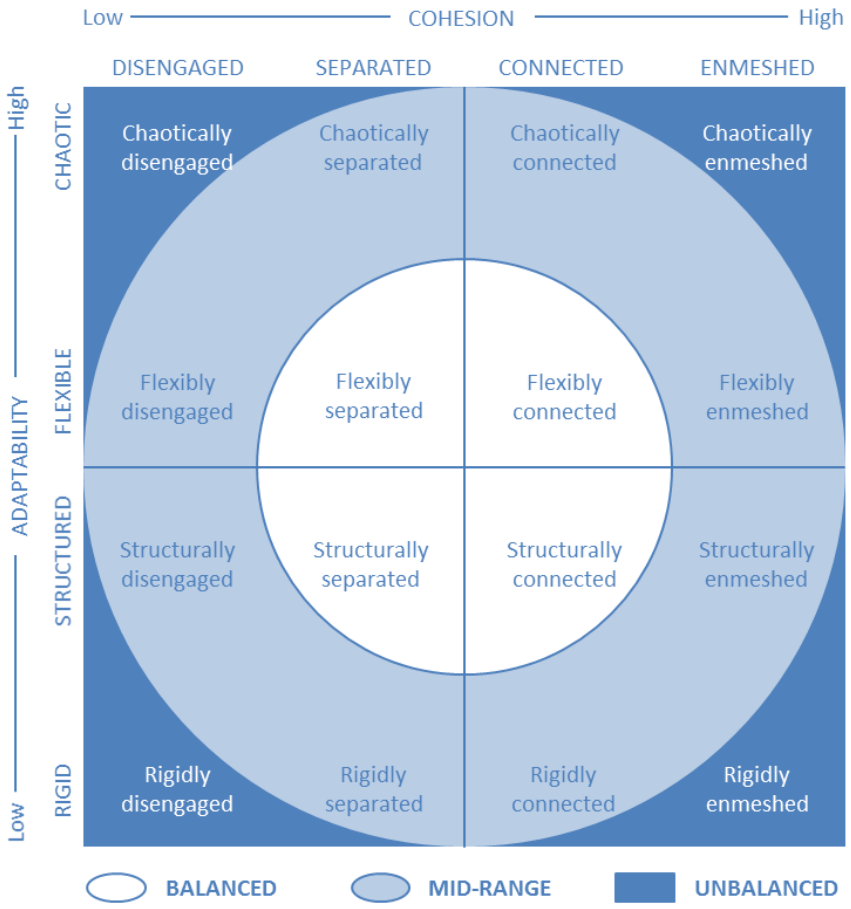

Figure 6. Circumplex model of family functioning (adapted from Olson et al. 1989).

The Family Adaptability and Cohesion Evaluation Scale (FACES) is used to evaluate family functioning using these two dimensions [131]. The FACES contains items about cohesion and adaptability which can be scored numerically where greater punctuations of each item indicate a greater characteristic of the dimension assessed. Most versions of FACES use a linear model to measure family functioning, which associate family balance with high cohesion and adaptability levels. Moreover, some versions of the instrument were designed to resemble the Circumplex model, adopting the idea that balanced family systems were considered more functional compared to unbalanced systems. In an obesity context, lifestyle interventions may be more effective in balanced functioning families, whereas in extreme families it may be more difficult to acquire and maintain healthy habits [132].

4.3 CHILD-FEEDING QUESTIONNAIRE

In addition to eating habits and familiar context, feeding practices from parents to sons may also be an important contributor to childhood obesity, especially in early ages. Indeed, it is widely accepted that parents influence the development of their children's eating habits and may control their diet, which can be determinant in the development of obesity [133,134]. The model of obesity proneness describes how parental control of children's eating disrupts a child's natural ability to self-regulate energy intake [135]. Pressuring children to eat and restricting the intake of palatable foods are paternal attitudes that can lead to children's inability to regulate their own food intake, which may dysregulate children's eating habits and increase predisposition to childhood obesity [136].

The Child-Feeding Questionnaire (CFQ) [137] assesses paternal feeding attitudes and practices focusing on the predisposition of children to obesity. The CFQ evaluates several paternal feeding factors that correlate with children's weight, such as perceived responsibility for feeding, perceived parent weight, perceived child weight, concern about child weight, restriction, pressure to eat and monitoring of food intake. Inappropriate parental behaviors can be recognized by the CFQ and may be targeted to prevent risky eating behaviors in their children. Feeding attitudes are part of the human behavior, and behavioral genetics supports the role of a lot of small-effect genes in its development [138]. Shared genetic variants between family members may predispose to similar dietary attitudes, such as eating disinhibition or self-control. Overall, genetic factors may account for a proportion of the individual behavior.

HYPOTESIS

1. Obesity is a heterogeneous and multifactorial disease with a clear genetic component.

2. Extreme obesity phenotypes may be caused by low-frequency high-penetrant genetic variants.

3. Alterations in genes from the leptin-melanocortin pathway, which is responsible for the hypothalamic control of body weight, may be responsible for some familiar obesity cases, but not all.

4. The identification of genetic factors causing obesity would make possible a more targeted therapeutic intervention in the clinics.

OBJECTIVES

The main objective of this thesis was to unravel the molecular mechanisms underlying severe early-onset obesity. Here, we used DNA samples from extreme obesity patients to explore known and new genetic alteration causing this phenotype.

To achieve this global goal, we proposed the following specific objectives:

1. Identify point mutations in patients with severe early-onset obesity affecting candidate genes for obesity with a pooled DNA sequencing approach.

2. Identify copy number variants in patients with severe early-onset obesity with SNP array platforms. Discover novel candidate genes affected by these alterations and analyze its contribution to the obesity phenotype.

3. Explore new candidate genes for obesity using exome sequencing of patients with a clear familiar inheritance pattern of obesity.

4. Investigate any paternal transmission bias of inherited variants in patients with severe early-onset obesity.

CHAPTER 1

Heterozygous rare genetic variants in non-syndromic early-onset obesity

Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Francesc Bou de Pieri, Raquel Flores, Julie A. Chowen, Luis A. Pérez-Jurado, Jesús Argente

Int J Obes (Lond). 2020 Apr: 44, 830-841

Several genes with strong effects have been identified as causative of obesity. Indeed, monogenic forms represent a relevant proportion of non-syndromic obesity cases. Understanding the genetic contributions to obesity has allowed developing novel therapeutic interventions for specific genetic forms.

A series of 463 severe early-onset obesity patients and 480 controls were analyzed using a pooled DNA sequencing approach. Rare single-nucleotide genetic variants were screened in 15 obesity candidate genes to define its contribution to the extreme phenotype. 293 additional patients from the "Viva la Familia" study were used as a replication dataset.

Serra-Juhé C, Martos-Moreno GÁ, Bou de Pieri F, Flores R, Chowen JA, Pérez-Jurado LA, Argente J. Heterozygous rare genetic variants in non-syndromic early-onset obesity. Int J Obes (Lond). 2020 Apr: 44, 830-841

International Journal of Obesity (2020) 44:830–841
https://doi.org/10.1038/s41366-019-0357-5

ARTICLE

Pediatrics

Heterozygous rare genetic variants in non-syndromic early-onset obesity

Clara Serra-Juhé[1,2] · Gabriel Á. Martos-Moreno[3,4] · Francesc Bou de Pieri[1,2] · Raquel Flores[1,2] · Julie A. Chowen[3,4] · Luis A. Pérez-Jurado[1,2,5] · Jesús Argente[3,4,6]

Received: 3 November 2018 / Revised: 7 January 2019 / Accepted: 7 February 2019 / Published online: 29 March 2019

Abstract

Background Obesity is a very heterogeneous disorder at both the clinical and molecular levels and with high heritability. Several monogenic forms and genes with strong effects have been identified for non-syndromic severe obesity. Novel therapeutic interventions are in development for some genetic forms, emphasizing the importance of determining genetic contributions.

Objective We aimed to define the contribution of rare single-nucleotide genetic variants (RSVs) in candidate genes to non-syndromic severe early-onset obesity (EOO; body mass index (BMI) >+3 standard deviation score, <3 years).

Methods Using a pooled DNA-sequencing approach, we screened for RSVs in 15 obesity candidate genes in a series of 463 EOO patients and 480 controls. We also analysed exome data from 293 EOO patients from the "Viva la Familia" (VLF) study as a replication dataset.

Results Likely or known pathogenic RSVs were identified in 23 patients (5.0%), with 7 of the 15 genes (*BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*) harbouring RSVs only in cases (3.67%) and none in controls. All were heterozygous changes, either de novo (one in *BDNF*) or inherited from obese parents (seven maternal, three paternal), and no individual carried more than one variant. Results were replicated in the VLF study, where 4.10% of probands carried RSVs in the overrepresented genes. RSVs in five genes were either absent (*LEP*) or more common in controls than in cases (*ADRB3*, *LEPR*, *PCSK1* and *PCSK2*) in both obese datasets.

Conclusions Heterozygous RSVs in several candidate genes of the melanocortin pathway are found in ~5.0% patients with EOO. These results support the clinical utility of genetic testing to identify patients who might benefit from targeted therapeutic intervention.

Introduction

Obesity is the most prevalent chronic childhood disease in the occidental world and a risk factor for later obesity- and metabolic-related disorders. It is a heterogeneous multifactorial disease with high heritability (50–75%) that is undoubtedly higher in severe early-onset cases [1–3]. The genetic causes underlying obesity remain largely unknown with an important proportion of information missing regarding its heritability.

A small percentage of obesity occurs as part of a syndromic entity [4–10], but in the majority, obesity is not accompanied by other specific phenotypes. Highly penetrant rare genetic variants, mainly autosomal recessive, affecting at least eight genes [*LEP* (MIM 164160), *LEPR* (MIM 601007), *MC4R* (MIM 155541), *PCSK1* (MIM 162150), *POMC* (MIM 176830), *MC3R* (MIM 155540),

These authors contributed equally: Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Francesc Bou de Pieri

These authors jointly supervised this work: Luis A Pérez-Jurado, Jesús Argente

Supplementary information The online version of this article (https://doi.org/10.1038/s41366-019-0357-5) contains supplementary material, which is available to authorized users.

✉ Jesús Argente
jesus.argente@uam.es

✉ Luis A. Pérez-Jurado
luis.perez@upf.edu

Extended author information available on the last page of the article

SIM1 (MIM 603128) or *NTRK2* (MIM600456)] are reportedly found in 2–5% of non-syndromic obese patients [11–17]. However, multigenic or multifactorial inheritance is the best model to explain the aetiology of most cases. Common genetic variants involved in these forms of obesity have been identified by genome-wide and candidate-gene association studies. Genetic variants in genes such as *ADRB3* (MIM 109691), *BDNF* (MIM 113505), *FTO* (MIM 610966), *GHSR*, *NEGR1* (MIM 613173), *PCSK2* (MIM 162151), *PPARG* (MIM 601487) and *TMEM18* (MIM 613220) [18–22], among others, have been related to body mass index (BMI), but the strongest effect of a single common variant might explain about only 2% of the variance in BMI. Similar to other complex traits [23, 24], copy number variants (CNVs) also contribute to the genetic basis of obesity [25–28]. Identification of a potential genetic cause of severe obesity is important not only for individualized follow-up and genetic counselling but also because novel therapeutic approaches for genetically defined obesities are becoming available [29].

Next-generation sequencing (NGS) has facilitated identification of the molecular basis of Mendelian diseases, even with a reduced number of samples [30, 31]. However, large numbers of patients are needed to identify a significant mutation burden in a specific gene and prove its pathophysiological relevance in multifactorial or highly heterogeneous genetic disorders. Given the known population stratification, appropriate controls from the same geographic origin are also needed to discard population variants that might not be disease causative. Strategies using pooled DNA-sequencing have proven effective in minimizing the costs of tackling multifactorial or heterogeneous disorders using large numbers of samples [32].

In highly heterogeneous diseases such as obesity, severe cases with early onset are more likely explained by highly penetrant rare genetic variants. We have used a pooled strategy to sequence a panel of selected candidate genes in 463 patients with non-syndromic severe early-onset obesity (EOO) and 480 controls searching for rare sequence variants (RSVs) with high penetrance.

Subjects and methods

Subjects and samples

Severe early-onset obese patients were selected by using the criteria of a BMI *Z*-score above three according to age and sex for the Spanish population [33] at their first examination and that visited the clinic due to obesity that was reported to have an onset before 3 years of age.

Patients underwent a detailed clinical evaluation, including oriented clinical history, physical examination and complementary tests to rule out syndromic, endocrine or secondary forms of obesity. These evaluations included an interview with a dietitian to determine eating habits such as the frequency, speed and amount of food ingested. A molecular karyotype by microarray and a methylation-sensitive multiple-ligation probe amplification to discard specific entities that might clinically overlap with isolated obesity were also performed [5].

A sample of 463 subjects with severe EOO was selected. Blood was collected from patients and both parents when possible to study the segregation of the genetic variants with the phenotype in the family. DNA was isolated from total blood using the Gentra Puregene Blood kit (Qiagen) according to the manufacturer's instructions.

A total of 480 adult subjects with normal weight (BMI <P75) from the same geographic origin, provided by Banco Nacional de ADN Carlos III (Universidad de Salamanca, Spain), were used as controls.

Candidate genes

To define the list of 15 candidate genes, a systematic literature review was done to identify: (1) genes with reported mutations in patients with obesity (*LEP*, *LEPR*, *MC3R*, *MC4R*, *PCSK1*, *NTRK2* and *SIM1)* [11–16, 34], and (2) genes with single-nucleotide polymorphisms associated to obesity in genome-wide association studies (*BDNF*, *FTO*, *NEGR1*, *GHSR*, *ADRB3*, *PPARG*, *PCSK2* and *TMEM18*A) [18–22].

Pooled DNA-sequencing

Targeted enrichment was used to capture coding regions from the 15 selected genes (SeqCap EZ Choice Enrichment Kits, Roche Sequencing). Instead of sequencing individual samples, a pooled DNA strategy was used. Each sample was included in two different pools and each pool had 20 different samples. A total of 48 pools from patients and 48 pools from controls were designed (Fig. 1). To identify which sample carried each variant, the results from each pool were overlapped to identify the two pools harbouring the same alterations. Using this pooling approach, one heterozygous variant in a sample is expected to appear in about 1/40 (2.5%) of the reads. To ensure that variants with a low percentage of reads were represented in the sequencing experiment, high coverage was mandatory. The massive sequencing was done with MiSeq (Illumina).

Variant calling

Reads were mapped to the human reference genome (UCSC hg19). We used the MUTEC software [35] to call RSVs represented in a low percentage of reads. Given the very

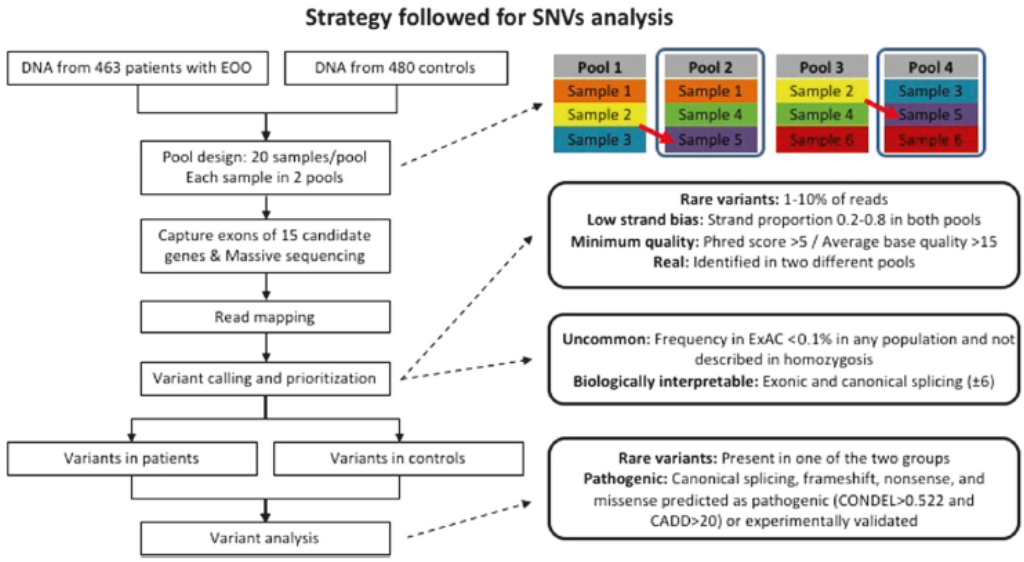

Fig. 1 Summary of the strategy followed for single-nucleotide variant (SNV) analysis and rare sequence variant (RSV) identification

high coverage, false-positive calls were expected in almost all positions mapped. Thus, several quality filters were used to discriminate real variants from false positives in an automatic manner (Fig. 1) and visual inspection of reads in the integrated genome viewer [36] allowed us to clearly discriminate in dubious cases. Six samples with seven previously detected variants were used for optimization of filtering parameters. Base quality index and strand bias were the best predictors to reduce false-positive detection rates. Since false positives tended to have much lower base quality values, only variants with individual Phred scores over 5 and an average base quality over 15 in the two different pools were selected. Variants with high strand bias were found to be false positives. Thus, variants with extreme strand bias (proportion of same strand reads <20% or >80%) were discarded as likely false positives.

Custom-made Perl scripts were used to automate sample identification crossing the pools where the same variant was detected. The proportion of variants in each pool was also considered in the pipeline as an estimation of allele frequency. Although obesity is a highly heterogeneous disorder, the same RSV might appear twice (present in two alleles from either the same or different samples), expected in 5% of the reads. Consequently, the first step was filtering the variants present in between 1 and 10% of the mapped reads. Variants with a lower percentage were probably artefacts and variants with a higher percentage (present in three or more samples in the same pool) were thought too common to be considered highly penetrant RSVs for this heterogeneous disease.

In positions where the coverage was under 1000×, it was particularly difficult to discriminate between real variants and false positives. Therefore, we individually assessed all variants in positions with coverage under 1000×.

Prioritization and analysis of variants

To establish the burden of RSVs identical approaches were used in each group (Fig. 1). Intronic variants were excluded due to the difficulty in interpreting their clinical significance and only exonic variants and those putatively affecting splicing (±6) were selected. We discarded variants with population frequencies in any public database (gnomAD, 1000 genomes, Kaviar, Spanish variant server) [37] higher than 1/1000 or described in a homozygous state in at least one individual, assuming that variants with a relatively high frequency in the general population are unlikely high penetrant variants related to the severe phenotype of EOO.

All nonsense or frameshift, canonical splicing and missense variants predicted as damaging (Condel >0.522 and CADD >20) [38, 39] were considered as highly likely to be pathogenic (from now on called likely pathogenic variants). All previously reported mutations with experimentally validated functional consequences were also included [40]. To define pathogenicity, we also considered the constraint metrics for variation tolerance of each gene. Variation-

intolerant genes were considered those with pLi (probability of loss-of-function intolerance) >0.50 and missense *Z*-score >0 [37].

Variant validation and co-segregation

To validate RSVs detected by pooled DNA-sequencing, PCR amplicons from individual DNA samples were obtained and sequenced by Sanger technology. The same technique was used to analyse parental samples when available to define inheritance pattern and possible co-segregation of the RSV with the phenotype in each family.

Replication cohort

Exomes from 293 unrelated patients with EOO from the Viva la Familia (VLF) study [41] were used to replicate the results found in our initial cohort. VLF comprises obese probands with BMI >95th percentile, between 4 and 19 years old, from Hispanic families in Houston. Sequence read archive (SRA) files were downloaded from publicly available servers (dbGAP: phs000616.v2.p2). Variant calling in the 15 candidate genes was done with GATK [42] and filtering was performed using the same criteria as above.

Ethics statement

The project was approved by the Medical Ethical Committee of Hospital Infantil Universitario Niño Jesús in Madrid, Spain and is in accordance with the "Ethical Principles for Medical Research Involving Human Subjects" adopted in the Declaration of Helsinki by the World Medical Association (64th WMA General Assembly, Fortaleza, Brazil, October 2013) and Spanish data protection act (Ley Orgánica 15/1999 de Protección de Datos). Written informed consent was obtained from all patients or their legal guardian after a complete description of the study.

Results

RSV detection sensitivity and specificity

The mean sequence coverage of patient pools was 2396× (SD: 574) and 3428× (SD: 660) in controls, with >98% of targeted sites showing coverage above 400×. The sensitivity and specificity of this approach were assessed. Six of the seven known variants in samples included in the study were blindly identified with the filters mentioned above, yielding a sensitivity of 86%. The only variant not identified had low-quality scores. As proof of concept for high specificity, 23 variants detected with the pooled DNA approach were selected and all (23/23, 100%) validated by Sanger sequencing. Although the pooled DNA approach may not detect a small percentage of real variants, it does not capture false positives. These results demonstrate that the strategy used is efficient to identify RSVs in large cohorts of samples.

Burden of RSVs in EOO

A total of 73 different RSVs were selectively identified in only one group (Table 1): 43 variants in 48 subjects of the patient group and 30 variants in 31 controls. This difference was statistically significant (odds ratio (OR) = 1.61; $p = 0.0342$; Table 1 and Supplemental Tables 1 and 2). We used very stringent criteria of pathogenicity to better define the putative clinically relevant variants: all loss-of-function (nonsense, frameshift and canonical splice-site variants) and missense variants predicted as pathogenic by two stringent algorithms (Condel and CADD) were considered as likely pathogenic. Likely or known pathogenic (previously validated) RSVs were found in 23 patients and 8 controls (OR = 2.98; $p = 0.0055$). Most RSVs were present in a single sample, but a few cases with two samples harbouring the same RSV were found and three patients shared a variant in *FTO*.

A subset of sequenced genes accumulated the differential mutational load found in patients compared to controls: *BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*. Analysing this group of genes separately, a total of 30 patients with RSVs were identified in patients versus four controls (OR = 7.78; $p < 0.0001$; Table 1) and four of these seven overrepresented genes (*BDNF*, *NEGR1*, *PPARG* and *SIM1*) are variation intolerant (Supplemental Table 3). Analysing the likely pathogenic variants in the subset of genes, RSVs were identified in 17 patients (3.67%) and no controls ($p < 0.0001$; Table 1). Two of the variants identified were present in two unrelated patients (Table 2). All 17 likely pathogenic variants identified in the overrepresented subset of genes heterozygous and no individual carried more than one likely pathogenic variant (Tables 1 and 2).

In the remaining subset of genes, 18 RSVs were identified in patients and 27 in controls (not statistically significant), including 6 likely pathogenic variants in patients and 8 in controls (not statistically significant; Supplemental Tables 1 and 2).

Heterozygous RSVs in patients with EOO

Among the genes with higher mutational load in patients, *MC4R* was the most frequent with seven RSVs in patients (1.51%). Of these, five have been reported as pathogenic by demonstration of partial (H76R) or complete (S127L,

Table 1 RSVs identified per gene in each one of the analysed groups: EOO-Sp, $n = 463$; VLF, $n = 293$; and controls, $n = 480$

	All RSVs			Likely pathogenic RSVs		
	EOO-Sp (%)	VLF (%)	Controls (%)	EOO-Sp (%)	VLF (%)	Controls (%)
ADRB3	1 (0.22)	1 (0.34)	4 (0.83)	–	–	1 (0.21)
BDNF	4 (0.86)	1 (0.34)	–	3 (0.65)	1 (0.34)	–
FTO	4 (0.86)	4 (1.37)	1 (0.21)	1 (0.22)	–	–
GHSR	2 (0.43)	6 (2.05)	3 (0.63)	–	1 (0.34)	–
LEP	–	–	–	–	–	–
LEPR	7 (1.51)	5 (1.71)	5 (1.04)	2 (0.43)	2 (0.68)	3 (0.63)
MC3R	3 (0.65)	3 (1.02)	–	2 (0.43)	1 (0.34)	–
MC4R	7 (1.51)	6 (2.05)	2 (0.42)	6 (1.30)	4 (1.37)	–
NEGR1	3 (0.65)	–	–	1 (0.22)	–	–
NTRK2	2 (0.43)	3 (1.02)	4 (0.83)	1 (0.22)	2 (0.68)	1 (0.21)
PCSK1	4 (0.86)	–	5 (1.04)	3 (0.65)	–	1 (0.21)
PCSK2	2 (0.43)	–	5 (1.04)	–	–	2 (0.42)
PPARG	3 (0.65)	–	–	2 (0.43)	–	–
SIM1	6 (1.30)	2 (0.68)	1 (0.21)	2 (0.43)	–	–
TMEM18	–	1 (0.34)	1 (0.21)	–	1 (0.34)	–
	48 (10.37)	**32 (10.92)**	**31 (6.46)**	**23 (4.97)**	**12 (4.10)**	**8 (1.67)**
Selected genes	30 (6.48)	16 (5.46)	4 (0.83)	17 (3.67)	6 (2.05)	–
Other genes	18 (3.89)	16 (5.46)	27 (5.63)	6 (1.30)	6 (2.05)	8 (1.67)

Bold represents the selected genes.

The number and percentage of individuals carrying rare and probably pathogenic genetic variants per gene are shown. The total burden of RSVs is significantly higher in EOO-Sp (48/463) than controls (31/480) (OR = 1.61; $p = 0.0342$), and in VLF compared to controls (32/293; OR = 1.69; $p = 0.0306$). The proportion of individuals carrying likely pathogenic RSVs was also significantly higher in EOO-Sp than in controls (23/463 vs. 8/480; OR = 2.98; $p = 0.0055$). Genes accumulating the RSV load in patients compared with controls are in bold: *BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*. The proportion of individuals carrying RSVs was significantly higher in EOO-Sp than in controls (30/463 vs. 4/480; OR = 7.78; $p < 0.0001$). This difference was also statistically significant in VLF (16/293; OR = 6.55; $p = 0.0002$). Considering only probably pathogenic variants, the proportion in patients (17/463 in EOO-Sp and 6/293 in VLF) was significantly higher than in controls ($p < 0.0001$ and $p = 0.0029$, respectively), where none were found (0/480).

OR odds ratio, *RSV* rare sequence variant, *OR* odds ratio, *EOO-Sp* early-onset obesity-Spain, *VLF* Viva la Familia

T150I, A259V and P272L) loss of function [13, 40]. The two RSVs not previously described include a nonsense change interrupting the protein in the 22nd codon (R22X) and a missense change (V52A) predicted as likely benign. In the five cases with parental samples available, the RSV was inherited from obese mothers. Two missense variants in *MC4R* were detected in the control group, both predicted as benign and not previously described in patients. Three patients were also found to have RSVs in a related gene, *MC3R*, with the likely pathogenic RSV (R302W) identified in two unrelated patients (Table 2).

SIM1 also showed a high mutational load in patients ($n = 6$) compared to controls ($n = 1$). Two of the six variants identified in patients are strongly predicted to be pathogenic. One affects a canonical splice site in intron 3 (c.352 + 1G>A) and the other is a missense variant on a highly conserved residue of the first Perl–Arnt–Sim of the protein (Q206K) found in two siblings concordant with the phenotype. These RSVs were inherited from an obese father (splicing variant) and obese mother (missense variant) (Table 2). The variant detected in a control subject and four of the RSVs in patients are not predicted to be pathogenic by our criteria.

In the *BDNF* gene, three RSVs were found in four patients and none in controls. Two of the variants are predicted to be pathogenic. One of these RSVs was not found in parental samples, indicating that it occurred de novo in the patient. Parenthood was proven using microsatellite markers. The other RSV (I231V) was found in two unrelated cases, paternally inherited in the case with parental samples available (Table 2).

We detected three RSVs (two likely pathogenic) in *PPARG* in patients, as well as one single likely pathogenic missense RSV in *FTO* and *NEGR1*, and none in controls.

Table 2 Main clinical features of the patients harbouring likely pathogenic RSVs in the overrepresented subset of genes

Gender	Gene	Type	gDNA	cDNA	Protein	Inheritance	Age at exam (years)	BMI (SDS)	Height (SDS)	Target height (SDS)	Bone age acceleration (months)	Newborn weight (SDS)	Cognition and behaviour	Hyperphagia	Metabolic disturbances	Other comorbidities	Familial obesity background
Male	*MC4R*	Missense*	Chr18:58038768	815C>T	P272L*	Maternal	6.84	+6.5	+4.3	+0.10	26	+3.2	NR	+	IR	Precocious pubarche	Mother (infancy) not adult
Male		Missense*	Chr18:58038807	776C>T	A259V*	Maternal	6.92	+7.0	+1.2	−1.52	−11	+2.0	NR	+	None	Macroorchydism	Both parents and families
Male		Missense*	Chr18:58039134	449C>T	T150I*	–	9.58	+5.8	+1.8	−0.90	42	−1.0	NR	+	IR + IGT + HU	Liver steatosis	Both parents and families
Male		Missense*	Chr18:58039203	380C>T	S127L*	–	14.25	+14.0	−1.3	−2.33	14	−0.4	NR	+++	IR + HT + HU	None	Both parents and families
Male		Missense*	Chr18:58039356	227A>G	H76R*	Maternal	14.00	+5.8	+1.1	−0.81	33	+0.1	NR	+++	IR + HU	Liver steatosis	Both parents and paternal family (grandfather and 3 aunts)
Male		Nonsense	Chr18:58039519	64A>T	R22X	Maternal	5.00	+18.8	+1.0	−2.06	51	+5.7	NR	+++	IR + HT + HU	Liver steatosis	Both parents and maternal family (morbid)
Female	*MC3R*	Missense	Chr20:54824803	904C>T	R302W	–	11.16	+4.4	+0.8	+0.57	10	−0.2	NR	+	IR + HU	None	Both parents and families
Male		Missense	Chr20:54824803	904C>T	R302W	–	12.00	+3.8	−0.9	−1.00	4	−0.4	NR	+	None	None	Both parents and families
Female	*SIM1*	Missense	Chr6:100896482	616C>A	Q206K	Maternal	8.75	+5.6	+1.8	+0.48	12	−1.1	NR	+	IR + IGT	None	Father and sister
Male		Splicing	Chr6:100898138	352+1G>A	–	Paternal	8.66	+4.3	+0.7	−0.01	12	+1.16	NR	+	IR + IGT	None	Father and paternal family (grandmother, 2 uncles and 1 aunt)
Male	*PPARG*	Missense	Chr3:12422920	326T>A	I109N	Maternal	9.33	+5.5	+4.2	+1.06	24	+5.1	NR	+	IR	None	Father
Male		Missense	Chr3:12434173	457C>T	R153W	–	11.00	+4.1	−0.4	−1.57	0	−2.0	NR	+	IR + IGT + HT	None	Father and sister
Male	*BDNF*	Missense	Chr11:27679421	691A>G	I231V	Paternal	6.00	+3.9	+0.3	−0.76	-6	−0.4	NR	+	None	None	Mother and maternal family (grandfather)
Female		Missense	Chr11:27679421	691A>G	I231V	–	11.75	+3.3	−0.1	−1.18	3	+0.1	NR	+++	IR + HU	Liver steatosis	Mother
Female		Missense	Chr11:27679691	421T>G	C141G	De novo	11.16	+6.1	+2.0	−0.41	17	+0.1	NR	+	IR + HU	Liver steatosis	Brother and paternal family (aunt, grandfather)
Male	*NEGR1*	Missense	Chr1:72400858	313A>G	I105V	Maternal	2.92	+8.8	+1.7	−0.11	3	+0.8	SP+BP	+++	HU	None	Both grandmothers
Female	FTO	Missense	Chr16:53859890	238C>T	R80W	Paternal	6.66	+3.8	+1.1	−0.57	14	−1.7	NR	+++	IR + HT + HU	Liver steatosis	Sister, both parents and maternal family (grandmother, 2nd degree aunt)

In columns "Type" and "Protein" the * indicates functional information available

SDS standard deviation score, *BMI* body mass index, *gDNA* genomic DNA, *cDNA* complementary DNA, *NR* nothing relevant, *SP* speech delay, *BP* behavioural disturbances, *IR* insulin resistance, *IGT* impaired glucose tolerance, *HU* hyperuricaemia, *HT* hypertriglyceridaemia

Clinical features of RSV carriers

The main phenotypic features of patients harbouring likely pathogenic RSVs in the selected genes are detailed in Table 2. The six patients harbouring pathogenic RSVs in *MC4R* were males with a BMI above + 5.5 standard deviation score (SDS), orexigenic impulsivity, hyperinsulinism, some overgrowth (height >+1 SDS over target height) and advanced bone age (14 months on average), a phenotype previously defined in patients with *MC4R* mutations [43, 44]. The patient with the A259V variant had macroorchydism (8 cc volume testes despite Tanner stage I), which is not a common feature of MC4R deficiency, given the lack of *MC4R* expression in the testis [45]. The patient with the novel nonsense mutation (R22X) had an extremely high BMI (+18.8 SDS), uncontrollable orexigenic impulsivity, overgrowth (+3 SDS), advanced bone age (51 months over chronological age) and hyperinsulinism. His mother, also a carrier of the R22X mutation, had a BMI of 51.3 kg/m^2. The patient with a predicted benign RSV at *MC4R* (V52A) was a girl with milder obesity (+4.2 BMI-SDS), no overgrowth (height at −1 SDS), average bone age, and hyperinsulinism.

The two patients with likely pathogenic RSVs in *MC3R* had milder obesity (BMI around +4 SDS) and no overgrowth. One had insulin resistance.

Patients harbouring *SIM1* RSVs had BMIs of 4.3 and 5.6 SDS, increased longitudinal growth (+0.7 and +1.3 SDS over standardized target height), mild hyperphagia, insulin resistance, and impaired glucose tolerance.

Patients with RSVs in the *BDNF* gene had BMIs ranging from +3.3 to +6.1 SDS. Patients with PPARG RSVs including one sibling (BMI +4.1 to +5.5 SDS) showed insulin resistance and three out of four had dyslipidaemia.

Replication cohort

The results of the VLF study are shown in Table 1 and Supplemental Table 4. The frequency of RSVs in VLF replicated the results for *BDNF*, *FTO*, *MC3R*, *MC4R* and *SIM1*. No RSVs with the established criteria were detected in *NEGR1* or *PPARG* in the VLF cohort, but likely pathogenic RSVs were identified in *GHSR* and *TMEM18*, absent in controls. The increased rate of RSVs between VLF subjects (32/293, 10.92%) and controls (31/480, 6.46%) in all genes was similar to that of the Spanish EOO dataset (OR = 1.69; $p = 0.0306$). The separate analysis of the overrepresented subset of genes (*BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*) yielded a significantly higher proportion of VLF patients with RSVs (16/293) than controls (4/480; OR = 6.55; $p = 0.0002$). This difference was also statistically significant when considering only likely pathogenic variants (6/293 VLF patients versus 0/480 controls; $p = 0.0029$; Supplemental Table 5).

Discussion

Analysis of an elevated number of individuals is necessary to identify genetic factors with strong effects in heterogeneous diseases with important environmental influence, such as obesity. Pooled DNA-sequencing reduces the number of sequencing reactions, thus reducing costs without decreasing the number of samples. This approach must be optimized to identify all relevant variants and discern between real variants and artefacts. Sequencing coverage above 400 reads in most positions would yield >10 reads of any single RSV diluted in a pool of 20 different samples. Since sequence artefacts are detected in most mapped positions, quality filters were required to discriminate between real variants and false positives. The percentage of reads with the variant, strand bias and Phred score had the greatest discriminatory ability. Full validation rate by Sanger sequencing (23 of 23 variants) proved the high specificity and suitability of the pooling strategy.

We failed to detect one of seven7 previously defined RSVs (sensitivity 86%). This incomplete detection rate might invalidate using pooling strategies for diagnostic purposes, but it should not affect the results of this study. The average coverage of the patient pools was lower than that of the control pools (2396× vs. 3428×). If this difference were to generate bias when detecting RSVs, it would favour RSV detection in controls. A more similar coverage between groups might even increase the difference in burden between patients and controls.

We focused on RSVs in 15 candidate genes that could have a strong effect in a small subgroup of patients according to information available at the onset of this study. This strategy might have missed some variants with higher frequencies that also contribute to obesity. RSVs were identified in 10.36% of patients with half of them (5.0%) most likely being pathogenic, a burden significantly higher in patients than in controls. This burden was attributable to a subset of seven genes (*BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*). The replication study corroborated these results indicating that highly penetrant RSVs in these genes contribute to part of the missing heritability of EOO.

When patients are grouped according to the gene mutated, most groups are too small to establish conclusive clinical molecular correlations, but some common features can be discerned. Among the overrepresented genes, *MC4R* was the most frequent (1.51% of patients), similar to reports revealing heterozygous and homozygous mutations in

MC4R accounting for 1–6% of severe obesity in humans [40, 43, 44]. Patients with RSVs in *MC4R* had hyperphagia with compulsive eating, very severe obesity (mean BMI + 8.9 SDS), overgrowth, accelerated skeletal maturation and metabolic disturbances including insulin resistance. Behavioural disturbances occurred mainly associated with food seeking. It is of note that all patients described here with RSVs at *MC4R* were male. The literature does not support the possibility that in obese subjects this gene is exclusively affected in males. However, there may be a global bias in the parental transmission of RSVs. To determine this possibility, a larger sample size is required, as well as additional investigation. It is also of note that there was a lack of significant differences in gene burden between the two relatively close ethnic groups compared, European Spaniards and Hispanics in the USA.

Like *MC4R*, *MC3R* has a critical role in regulating energy balance [46]. It was recently reported to regulate the upper and lower limits of an individual's homeostatic set-points and the response to external metabolic challenges [47]. MC3R-knockout mice were shown to have dysregulation of energy expenditure, feeding responses and neuroendocrine responses to metabolic challenges such as fasting, high fat diet intake, pregnancy and the loss of estrogens. For example, these mice lost more weight during fasting than wild-type mice, but gained more weight when put on a high fat diet [47]. The three patients (0.65%) with missense mutations in *MC3R* had a similar degree of obesity, which was milder than those harbouring RSVs in MC4R, and no overgrowth or advanced skeletal maturation. These patients did not exhibit measurable hyperphagia, similar to what was observed in MC3R-deficient mice [47]. However, as with mice, an inadequate diet could contribute to the excess weight gain. Only one of our patients had hyperinsulinism and hyperuricaemia; this patient also harboured a variant in the other allele that did not fulfil the defined criteria as its frequency is 1/945 in the European population. However, if this second variant is functionally relevant, inheritance would be compatible with a recessive or codominant model as reported for *MC4R*. In the codominant mode of inheritance, both monoallelic and biallelic mutations can cause the disorder and patients with biallelic mutations are more affected than their heterozygous relatives [48].

BDNF is a pro-survival factor in the brain that also participates in appetite regulation. Heterozygous BDNF-knockout mice are obese, hyperphagic and hyperactive [49] and disruption and deletion of *BDNF* have been described in patients with obesity. WAGR syndrome is a contiguous gene deletion syndrome on chromosome 11p13 characterized by Wilms' tumour, Aniridia, Genitourinary anomalies and Mental retardation, with obesity present in patients with larger deletions encompassing *BDNF* along with the other critical genes (WAGRO syndrome) [10]. Disruption of *BDNF* expression in a patient with a paracentric inversion and deletion of the entire *BDNF* locus in a mother and child were associated with hyperphagia, severe obesity and mildly impaired cognition with or without attention deficits and hyperactivity [50–52]. However, no point mutations in *BDNF* have been reported to date in obese patients. In our series, three patients (0.65%) harboured two different missense RSVs in this gene predicted as pathogenic, with none detected in controls. Two unrelated cases carried the same variant (I231V) and another carried a de novo variant (C141G9). Obese patients with *BDNF* variants presented mild to severe hyperphagia and insulin resistance with dyslipidaemia in two cases. No behavioural or learning issues detected.

Six RSVs in *SIM1* were found in our cohort and two were predicted as likely pathogenic. A third case was the affected brother of case 1782, concordant for the Q206K RSV. The phenotype of all three cases included a BMI-SDS of +4 to +5, no intellectual or behavioural abnormalities, and insulin resistance with impaired glucose tolerance. *Sim1* haploinsufficiency in mice induces hyperphagia, obesity and central nervous system developmental abnormalities [53]. Deletions [54, 55] and translocations [56] affecting *SIM1* cause severe obesity in association with intense orexigenic impulsivity and other features resembling Prader–Willi syndrome. Loss-of-function point mutations in *SIM1* with variable expressivity and incomplete penetrance also produce a "Prader–Willi-like" phenotype [57], but also non-syndromic obesity [58]. Thus the described RSVs could contribute to the postulated increased intra-family risk for non-syndromic obesity [58].

Significant differences in RSV load between patients and controls were detected in *NEGR1* (3 patients, 0 controls), *FTO* (4 patients, 1 control) and *PPARG* (3 patients, 0 controls), with five of the RSVs predicted as likely pathogenic. Although association studies have documented their role in obesity, no highly penetrant RSVs have been described to date in these genes. Our data indicate that some RSVs at *NEGR1*, *FTO* and *PPARG* may behave as highly penetrant alleles causative of EOO, although the number of patients is too small to define distinctive phenotypes. The patient with a likely pathogenic RSV at *NEGR1* had speech and behavioural problems and hyperuricaemia, while the three patients harbouring RSVs in *PPARG* (including one sibling) had insulin resistance and associated dyslipidaemia, in agreement with the role of *PPARG* variants in predisposition to metabolic syndrome [59].

Weaker evidence for definite implication was obtained for *GHSR*, *NTRK2* and *TMEM18*. Likely pathogenic RSVs at *NTRK2* were identified in one of our patients and two

from the VLF study, but also in one control. Single patients with RSVs at *GHSR* and *TMEM18*, absent in controls, were found in the VLF cohort. Whether these monoallelic RSVs contribute to the obesity phenotype awaits studies in larger series and functional studies. *De novo NTRK2* mutations, a highly intolerant gene to both loss-of-function (pLI = 1) and missense variants (miss-$z = 4.35$), have been previously associated with severe obesity and developmental delay [17]. Thus, some heterozygous partially functional variants could cause a phenotype of non-syndromic obesity without significant developmental delay.

RSVs in five genes were either absent (*LEP*) or more common in controls than in cases (*ADRB3*, *LEPR*, *PCSK1* and *PCSK2*) in both datasets. Biallelic mutations in *LEP*, *LEPR* and *PCSK1* with a recessive model of inheritance have been described in severely obese patients. In a study of 300 obese subjects, biallelic *LEP* mutations were identified with a frequency of 3% [48], but 3/4th of the patients with biallelic mutations were homozygous, with mutations in identical-by-descent regions in consanguineous families [48]. Thus, the prevalence of biallelic mutations in recessive genes in populations that are not inbred, such as the two cohorts studied here, might be quite low [60]. *POMC* was not included in this study based on the assumed adrenal insufficiency associated to pathogenic variants in this gene. However, *POMC* is now being analysed in this cohort in another study.

It is clear that functional analyses at the cellular, tissue and organism levels are required to define the physiopathogenic mechanisms of all variants reported here. However, we show strong epidemiological and statistical evidence by defining the genes harbouring heterozygous deleterious changes in our large cohort of EOO patients and then validating the results in a replication dataset. Moreover, there is experimental evidence for the abnormal function of some of the reported RSVs [13, 40, 61]. Many genes involved in appetite and metabolic control appear to be sensible to dosage [50, 62, 63], with rescue of haploinsufficiency of *Mc4r* and *Sim1* recently shown to revert the obese phenotype in mice [64]. Thus, although the molecular mechanisms underlying how these newly described RSVs affect body weight remain to be established, there is precedence for heterozygous affectation of these genes being functionally important. Moreover, the overall prevalence of monoallelic RSVs found in our cohort is similar to that of a recent publication [65] where a diagnosis rate of genetic obesity was reported in 3.9% of patients (48/1230) and 7.3% of the paediatric subgroup (12/164). It should also be emphasized that the interaction of these genetic variants with the environment is of utmost importance with morbid obesity being more likely to occur in these cases when dietary habits are inadequate.

In summary, we documented a higher burden of likely pathogenic heterozygous RSVs in several candidate genes in patients with severe EOO compared to controls. The yield for RSV detection is relatively high, around 5% of patients with EOO. Identification of a potential genetic cause of the phenotype in each case is very important for individualized follow-up and genetic counselling to the family. In addition, definition of the specific defect at the molecular level may be instrumental, as this group of individuals may be relatively refractory to weight loss through diet and exercise and novel therapeutic approaches for genetically defined obesities are becoming available. For instance, highly potent second-generation MC4R agonists lead to weight loss in individuals with MC4R deficiency [40]. Thus, several patients in this study could possibly benefit from this therapy. Our results reinforce the role of the leptin–melanocortin pathway in obesity and bring to light other genes that may carry highly penetrant obesogenic single allele variants, such as *FTO*, *NEGR1* and *PPARG*. The availability of an evidence-based algorithm for diagnostic analysis of early-onset obesity would be of great clinical interest. However, the obese phenotype of the children reported here, as well as in other studies, is nonspecific, and an algorithm that could accurately suggest which genes should be studied in each case is difficult even if phenotype–genotype relationships lead to a better definition of the different clinical conditions. Instead, given the currently available and affordable tools for diagnostic genomic analysis, a genotype-first approach using expanded panel of genes with effective copy number analyses should be indicated in these children with early-onset obesity.

Acknowledgements JA was funded by the Spanish Ministry of Health (FIS-PI13/02195 and PI16/00485, co-funded by FEDER), the Fundación de Endocrinología y Nutrición and the "Centro de Investigación Biomédica en Red" for obesity and nutrition (CIBEROBN) of the Instituto de Salud Carlos III, Spain. LAP-J was funded by the Spanish Ministry of Health (FIS-PI1302481, co-funded by FEDER), the Generalitat de Catalunya (2014SRG1468), the Institució Catalana de Recerca i Estudis Avançats (ICREA Academia programme), the Spanish Ministry of Economy and Competiveness "Programa de Excelencia María de Maeztu" (MDM-2014-0370) and the Centro de Investigación Biomédica en Red for rare diseases (CIBERER) of the Instituto de Salud Carlos III, Spain. JAC was funded by grants from the Spanish Ministry of Science and Innovation (BFU2017-82565-C21-R2). We would like to thank Francisca Díaz and Sandra Canelles for their excellent technical assistance.

Compliance with ethical standards

Conflict of interest LAPJ is a partner and scientific advisor of qGenomics Laboratory. The remaining authors declare that they have no conflict of interest.

References

1. Friedman JM. Obesity in the new millennium. Nature. 2000;40:632–4.
2. Walley AJ, Asher JE, Froguel P. The genetic contribution to non-syndromic human obesity. Nat Rev Genet. 2009;10:431–42.
3. Llewellyn CH, Trzaskowski M, Plomin R, Wardle J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. Int J Obes. 2013;37:1506–9.
4. Hatada I, Ohashi H, Fukushima Y, Kaneko Y, Inoue M, Komoto Y, et al. An imprinted gene p57KIP2 is mutated in Beckwith-Wiedemann syndrome. Nat Genet. 1996;14:171–3.
5. Martos-Moreno GÁ, Serra-Juhé C, Pérez-Jurado LA, Argente J. Underdiagnosed Beckwith–Wiedemann syndrome among early onset obese children. Arch Dis Child. 2014;99:965–7.
6. Nicholls RD, Knoll JHM, Butler MG, Karam S, Lalande M. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader–Willi syndrome. Lett Nat. 1989;342:281–5.
7. Buiting K, Gross S, Lich C, Gillessen-Kaesbach G, El-Maarri O, Horsthemke B. Epimutations in Prader–Willi and Angelman syndromes: a molecular study of 136 patients with an imprinting defect. Am J Hum Genet. 2003;72:571–7.
8. Sanz-Fernández M, Muñoz-Calvo MT, Pozo-Román J, Martos-Moreno GÁ, Argente J. Clinical and radiological findings in a case of pseudohypoparathyroidism type 1a: Albright hereditary osteodystrophy. An Pediatr. 2015;82:439–41.
9. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, et al. Triallelic inheritance in Bardet–Biedl Syndrome, a Mendelian recessive disorder. Science. 2001;293:2256–9.
10. Han JC, Liu QR, Jones M, Levinn RL, Menzie CM, Jefferson-George KS, et al. Brain-derived neurotrophic factor and obesity in the WAGR syndrome. N Engl J Med. 2008;359:918–27.
11. Ranadive S, Vaisse C. Lessons from extreme human obesity: monogenic disorders. Endocrinol Metab Clin North Am. 2008;37:733–51.
12. Farooqi IS, Matarese G, Lord GM, Keogh JM, Lawrence E, Agwu C, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. J Clin Invest. 2002;110:1093–103.
13. Granell S, Serra-Juhé C, Martos-Moreno GÁ, Díaz F, Pérez-Jurado LA, Baldini G, et al. A novel melanocortin-4 receptor mutation MC4R-P272L associated with severe obesity has increased propensity to be ubiquitinated in the ER in the face of correct folding. PLoS ONE. 2012;7:e50894.
14. Montague CT, Farooqi IS, Wareham NJ, Soos MA, Rau H, Wareham NJ, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. Nature. 1997;387:903–8.
15. Vaisse C, Clement K, Guy-Grand B, Froguel P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. Nat Genet. 1998;20:113–4.
16. Yeo GSH, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S. A frameshift mutation in MC4R associated with dominantly inherited human obesity. Nat Genet. 1998;20:111–2.
17. Yeo GS, Connie Hung CC, Rochford J, Keogh J, Gray J, Sivaramakrishnan S, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci. 2004;7:1187–9.
18. Willer CJ, Speliotes EK, Loos RJF, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet. 2009;41:25–34.
19. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, et al. A common genetic variant is associated with adult and childhood obesity. Science. 2006;312:279–83.
20. Turcot V, Lu Y, Highland HM, Schurmann C, Justice AE, Fine RS, et al. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. Nat Genet. 2018;50:26–41.
21. Meyre D, Delplanque J, Chèvre JC, Lecoeur C, Lobbens S, Gallina S, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet. 2009;41:157–9.
22. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet. 2009;41:18–24.
23. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. Nature. 2009;459:569–73.
24. Stefansson H, Rujescu D, Cichon S, Pietiläinen OP, Ingason A, Steinberg S, et al. Large recurrent microdeletions associated with schizophrenia. Nature. 2008;455:232–6.
25. Jarick I, Vogel CIG, Scherag S, Schäfer H, Hebebrand J, Hinney A, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. Hum Mol Genet. 2011;20:840–52.
26. Serra-Juhé C, Martos-Moreno GÁ, Bou de Pieri F, Flores R, González JR, Rodríguez-Santiago B, et al. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. PLoS Genet. 2017;13:1–19.
27. Wheeler E, Huang N, Bochukova E, Keogh JM, Lindsay S, Garg S, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nat Genet. 2013;45:513–7.
28. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature. 2010;463:666–70.
29. Grandone A, Di Sessa A, Umano GR, Toraldo R, Miraglia Del Giudice E. New treatment modalities for obesity. Best Pract Res Clin Endocrinol Metab. 2018;32:535–49.
30. Mardis ER. A decade's perspective on DNA sequencing technology. Nature. 2011;470:198–203.
31. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. J Genet Genom. 2011;38:95–109.
32. Zuzarte PC, Denroche RE, Fehringer G, Katzov-Eckert H, Hung RJ, McPherson JD. A two-dimensional pooling strategy for rare variant detection on next-generation sequencing platforms. PLoS ONE. 2014;9:e93455.
33. Hernández M, Castellet J, Narvaiza JL, Rincón JM, Ruiz I, Sánchez E, et al. Curvas y tablas de crecimiento. 2nd edn. Madrid: Instituto de Investigación sobre Crecimiento y Desarrollo, Fundación Faustino Orbegozo, Editorial Garsi; 1988.

34. Ristow M, Müller-Wieland D, Pfeiffer A, Krone W, Kahn R. Obesity associated with a mutation in a genetic regulator of adipocyte differentiation. N Engl J Med. 2015;339:953–9.
35. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31:213–9.
36. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genomics Viewer. Nat Biotechnol. 2011;29:24–26.
37. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.
38. González-Pérez A, López-Bigas N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. Am J Hum Genet. 2011;88:440–9.
39. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46:310–5.
40. Collet TH, Dubern B, Mokrosinski J, Connors H, Keogh JM, Mendes de Oliveira E, et al. Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency. Mol Metab. 2017;6:1321–9.
41. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. Am J Clin Nutr. 2006;84:646–54.
42. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2009;20:254–60.
43. Govaerts C, Srinivasan S, Shapiro A, Zhang S, Picard F, Clement K, et al. Obesity-associated mutations in the melanocortin 4 receptor provide novel insights into its function. Peptides. 2005;26:1909–19.
44. Martinelli CE, Keogh JM, Greenfield JR, Henning E, van der Klaauw AA, Blackwood A, et al. Obesity due to melanocortin 4 receptor (MC4R) deficiency is associated with increased linear growth and final height, fasting hyperinsulinemia, and incompletely suppressed growth hormone secretion. J Clin Endocrinol Metab. 2011;96:181–8.
45. Chhajlani V. Distribution of cDNA for melanocortin receptor subtypes in human tissues. Biochem Mol Biol Int. 1996;38:73–80.
46. Zegers D, Beckers S, Hendrickx R, Van Camp JK, Van Hoorenbeeck K, Desager KN, et al. Prevalence of rare MC3R variants in obese cases and lean controls. Endocrine. 2013;44:386–90.
47. Ghamari-Langroudi M, Cakir I, Lippert RN, Sweeney P, Litt MJ, Ellacott KLJ, et al. Regulation of energy rheostasis by the melanocortin-3 receptor. Sci Adv. 2018;4. https://doi.org/10.1126/sciadv.aat0866.
48. Farooqi IS, Wangensteen T, Collins S, Kimber W, Matarese G, Keogh JM, et al. Clinical and molecular genetic spectrum of congenital deficiency of the leptin receptor. N Engl J Med. 2007;356:237–47.
49. Fox EA, Biddinger JE, Jones KR, McAdams J, Worman A. Mechanism of hyperphagia contributing to obesity in brain-derived neurotrophic factor knockout mice. Neuroscience. 2013;229:176–99.
50. Gray J, Yeo GSH, Cox JJ, Morton J, Adlam AL, Keogh JM, et al. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. Diabetes. 2006;55:3366–71.
51. Jouan L, Gauthier J, Dion PA, Rouleau GA. Rare variants in complex traits: novel identification strategies and the role of de novo mutations. Hum Hered. 2013;74:215–25.
52. Harcourt BE, Bullen DVR, Kao KT, Tassoni D, Alexander EJ, Burgess T, et al. Maternal inheritance of BDNF deletion, with phenotype of obesity and developmental delay in mother and child. Am J Med Genet A. 2018;176:194–200.
53. Michaud JL, Boucher F, Melnyk A, Gauthier F, Goshu E, Lévy E, et al. Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. Hum Mol Genet. 2001;10:1465–73.
54. Villa A, Urioste M, Bofarull J, Martínez-Frías M. De novo interstitial deletionq16.2q21 on chromosome 6. Am J Med Genet. 1995;55:379–83.
55. Faivre L, Cormier-Daire V, Lapierre JM, Colleaux L, Jacquemont S, Geneviéve D, et al. Deletion of the SIM1 gene (6q16.2) in a patient with a Prader–Willi-like phenotype. J Med Genet. 2002;39:594–6.
56. Holder JL, Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. Hum Mol Genet. 2000;9:101–8.
57. Bonnefond A, Raimondo A, Stutzmann F, Ghoussaini M, Ramachandrappa S, Bersten DC, et al. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi – like features. J Clin Invest. 2013;123:3037–41.
58. Ramachandrappa S, Raimondo A, Cali AMG, Keogh JM, Henning E, Saeed S, et al. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. J Clin Invest. 2013;123:1–9.
59. Rocha RM, Barra GB, ÉCCC Rosa, Garcia ÉC, Amato AA, Azevedo MF. Prevalence of the rs1801282 single nucleotide polymorphism of the PPARG gene in patients with metabolic syndrome. Arch Endocrinol Metab. 2015;59:297–302.
60. Paolini B, Maltese PE, Del Ciondolo I, Tavian D, Missaglia S, Ciuoli C, et al. Prevalence of mutations in LEP, LEPR, and MC4R genes in individuals with severe obesity. Genet Mol Res. 2016;19:15. https://doi.org/10.4238/gmr.15038718.
61. Moore BS, Mirshahi T. Genetic variants help define the role of the MC4R C-terminus in signaling and cell surface stability. Sci Rep. 2018;8:10397.
62. Lee YS, Poh LK, Kek BL, Loke KY. The role of melanocortin 3 receptor gene in childhood obesity. Diabetes. 2007;56:2622–30.
63. Drabkin M, Birk OS, Birk R. Heterozygous versus homozygous phenotype caused by the same MC4R mutation: novel mutation affecting a large consanguineous kindred. BMC Med Genet. 2018;19:135.
64. Matharu N, Rattanasopha S, Tamura S, Maliskova L, Wang Y, Bernard A, et al. CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. Science. 2018. pii: eaau0629.
65. Kleinendorst L, Massink MPG, Cooiman MI, Savas M, van der Baan-Slootweg OH, Roelants RJ, et al. Genetic obesity: next-generation sequencing results of 1230 patients with obesity. J Med Genet. 2018;55:578–86.

Affiliations

Clara Serra-Juhé[1,2] · Gabriel Á. Martos-Moreno[3,4] · Francesc Bou de Pieri[1,2] · Raquel Flores[1,2] · Julie A. Chowen[3,4] · Luis A. Pérez-Jurado[1,2,5] · Jesús Argente[3,4,6]

[1] Genetics Unit, Universitat Pompeu Fabra, Hospital del Mar Research Institute (IMIM), C/Doctor Aiguader, 8, 08003 Barcelona, Spain

[2] Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, C/Sinesio Delgado, 4, 28029 Madrid, Spain

[3] Hospital Infantil Universitario Niño Jesús, Department of Endocrinology, Instituto de Investigación La Princesa, Universidad Autónoma de Madrid, Department of Pediatrics, Avenida Menéndez Pelayo, 65, 28009 Madrid, Spain

[4] CIBER de Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, C/Sinesio Delgado, 4, 28029 Madrid, Spain

[5] Women's and Children's Hospital, South Australia Medical and Health Research Institute (SAMHRI) and University of Adelaide, 72 King William Road, North Adelaide, SA 5006, Australia

[6] IMDEA Food Institute, CEIUAM + CSI, Crta. de Cantoblanco, 8, 28049 Madrid, Spain

CHAPTER 2

New insights in severe early-onset obesity from SNP array data

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

In preparation

Copy number variants have been associated with several complex and genetically heterogenic disorders and may contribute to the physiopathology of obesity. The study of the genetic content of these alterations, may contribute to new candidate genes for obesity.

To assess the burden of known and rare alterations, a case-control study was performed between 465 patients with severe early-onset obesity and 500 controls. As a replication group, 231 patients from the "Viva la Familia" study were used. The analysis of the genetic content of more than 250 rare copy number variants from obesity patients, allowed the identification of pathways and cellular processes involved in the development of obesity.

New insights in severe early-onset obesity from SNP array data

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

ABSTRACT

Genetic studies trying to elucidate the genetic factors involved in obesity development, persistently describe it as a multifactorial and genetically heterogeneous disease. Although rare monogenic forms and several genes and regions of susceptibility, including copy number variants (CNVs), have been described, there are still several mechanisms to be elucidated.

Here, we searched for CNVs in 465 Spanish children with non-syndromic early-onset obesity (EOO: body mass index >3 standard deviations above the mean at <3 years of age) using SNP array molecular karyotypes. Known susceptibility factors were analyzed and a case control study was performed with rare CNVs detected in 465 EOO cases and 500 non-obese controls. A group of 263 obesity patients from the *Viva la Familia* study was used to replicate our findings.

A higher burden of rare CNVs was detected in EOO cases. Genes from these CNVs showed a clear enrichment of obesity-candidate genes, reinforcing the contribution of CNVs in the pathophysiology of the disease. Highly penetrant alterations causing familiar forms of obesity were detected, including genes related to cyclogenesis, circadian rhythm, adipogenesis and inflammation, among others. Our data expands the current knowledge of the key players involved in the pathological process of obesity and will allow the development of new targeted therapies.

Our findings showed a higher burden of rare CNVs in patients with EOO compared to controls, revealing new genes and pathways that may be good candidates for extreme obesity phenotypes. The wide range of genes altered by the CNVs detected highlighted the high genetic heterogeneity among clinically selected obesity patients.

INTRODUCTION

Worldwide obesity is considered an epidemic as it has nearly tripled since 1975. In 2016, over 650 million adults had obesity and up to more than 1.9 billion adults were overweight [1]. Recent studies have shown that around 55% of children with obesity will still be obese in adolescence and 80% of adolescents with obesity will go on to be obese in adulthood. In fact, 30% of adults with obesity were obese in childhood or adolescence [2]. In the last 40 years, the number of school-age children and adolescents with obesity worldwide has risen from 11 million to 124 million, which is more than 10-fold, and childhood overweight has reached 216 million. Thus, in the 21st century, childhood obesity is considered a serious global public health challenge, affecting every country in the world [3]. In Spain, 17% of adults and 10% of children aged 2 to 17 years have obesity, and 37% of adults and 18% of children are overweight [4,5].

In adults, obesity is defined by a body mass index (BMI) of 30kg/m^2 or higher; and in children, by a BMI at or above the 95th percentile according to age and sex (two standard deviations over the mean). The main consequences of this serious medical condition are the significant morbidity and premature mortality associated with type II diabetes, hypertension, coronary artery disease and cancer [6]. Obesity is caused by an imbalance between energy intake and expenditure. Despite the key role of modern lifestyles in the current obesity epidemic, the role of genetics in the development of human obesity has been proven [7]. Although there is strong evidence for a high genetic component of obesity, the underlying genetic causes are largely unknown. The early-onset obesity (EOO) forms represent the most severe end of the obesity spectrum, which is more likely explained by genetic alterations. Studies focused on those extreme subgroups represent a valuable source of genetic information in complex disorders where genetic factors may play an important role.

Heterozygous mutations in the melanocortin-4 receptor gene are the most frequent genetic cause of isolated obesity, and may explain up to 6% of severe EOO cases [8]. During the last years, other genes and SNPs have been associated with BMI and obesity; however, there is an important proportion of missing heritability, mostly due to the multifactorial and heterogeneous nature of this disorder. The missing heritability is defined by the gap between the risk explained by known genetic causes (around 10%) [9] and the estimated heritability (50%-75% from twin and other family studies) [10]. New types of

analyses, such as structural variants, are helping to fill that gap [11]. Studies in other complex traits, such as autism [12], schizophrenia [13] and mental retardation [14], have pointed out the relevance of copy number variants (CNVs) in the genetic contribution of complex disorders. The diagnostic yield of chromosomal abnormalities in groups of patients with complex disorders using high resolution platforms is at least twice as high as that of conventional techniques, such as GTG-banded karyotyping [14]. Genetic causes in patients with extreme obesity phenotypes have been mainly associated with frequent CNVs acting as susceptibility factors and rare CNVs with high penetrance. These susceptibility factors are represented by deletions near the *NEGR1* gene in 1p31.1 [15], deletions in 10q11.22 including the *NPY4R* gene [16], deletions in 11q11 including different olfactory receptor genes [17], duplications in 10q26.3 [18] and deletions in 16p11.2 including the *SH2B1* gene [19,20], among others. Several studies have also analyzed the presence of rare CNVs in severe EOO patients compared to controls, supporting the hypothesis that CNVs affecting specific genes could be important contributors to isolated obesity [21].

Here, we used data from SNP array platforms to genotype rare and known CNVs in a large Spanish sample of patients with isolated severe EOO, in order to identify novel candidate genes contributing to the obesity phenotype.

MATERIALS AND METHODS

Subjects and samples

Patients were recruited in *Hospital Infantil Universitario Niño Jesús*, a reference center for pediatric endocrinology. Inclusion criteria for severe early-onset obesity (EOO) was body mass index (BMI) 3 standard deviations above the mean for age and gender at their first examination with a reported onset before 3 years of age. All studies were performed as part of a research project approved by the Medical Ethical Committee of the hospital, after receiving written informed consent from the families.

Detailed clinical examination and family history were recorded to discard syndromic forms of obesity in patients. Collection of blood samples from patients and parents (when available) was followed by DNA isolation with Gentra Puregene Blood kit (Qiagen) according to manufacturer's instructions. A custom-made panel of Methylation Specific Multiplex Ligation Dependent-Probe Amplification (MS-MLPA) was used to detect genomic and epigenetic

alterations associated with pseudohypoparathyroidism (MIM 103580), Prader-Willi (MIM 176270), Temple (MIM 616222) and Beckwith-Wiedemann (MIM 130650) syndromes, as reported in Martos-Moreno G. et al. (2014) [22]. Additionally, patients with *MC4R* mutations, the most frequent cause of monogenic obesity, were excluded from the study, as reported in Serra-Juhé et. al. (2020) [23].

A total of 465 unrelated EOO Spanish patients were included in this project (EOO-Sp group). CNV data from some of these patients (157 EOO patients) was previously reported in Serra-Juhé et. al. (2017) [24].

Molecular karyotyping

Patients were studied by Omni1-Quad (61 subjects), Omni Express SNP (93 subjects) and Omni Express-24 (311 subjects) platforms, Illumina. The 3 array platforms include between 1M and 700K SNPs along the genome and allow a high throughput genotyping. Copy number alterations were called using the PennCNV software [25]. To assess the burden of CNVs, a case-control study was performed between 465 EOO-Sp patients and 500 controls. To determine the CNVs in controls, 500 individuals from the Spanish Bladder Cancer/EPICURO study [26,27] were randomly selected. Filtering and prioritization criteria were the same between patients and controls. Rare CNVs and known genetic alterations of >100kb containing >10 SNPs in autosomes and chromosome X were analyzed in both groups.

Known genetic alterations

First, we explored the obesity susceptibility risk associated with common chromosomal alterations. We analyzed copy number alterations overlapping with known CNVs already associated with obesity or BMI, such as duplications and deletions at 10q26.3, 16p11.2 or 16p13.11, among others. Frequencies of these alterations were explored in both groups, patients and controls.

Rare CNVs

Rare CNV should include coding genes, overlap <50% with Segmental Duplications and be present in <5/9820 population controls, from now on rCNVs. Coding genes were annotated using BiomaRt [28] (protein coding, hg19). Chromosome, start and end coordinates of reported Segmental Duplications in UCSC Browser (Seg_Dups_hg19) were used to determine the overlapping percentage with the detected rCNVs. Population frequencies were

calculated using 1M Illumina SNP array data of 9820 samples from two datasets (8,329 individuals previously used as population controls for developmental anomalies [29] and 1,491 Spanish adult individuals from the Spanish Bladder Cancer/EPICURO study [27]).

Validation

Plots including the affected regions and 1Mb flanking downstream and upstream regions were generated. Log R Ratio (LRR) was used as a measure of genetic dosage and the B Allele Frequency (BAF) for genotyping. In duplications the LRR is increased and the BAF is altered, whereas in deletions the LRR is decreased and the heterozygous points of the BAF are lost. CNV calls with plots not reproducing the expected values were discarded as false positives.

To further validate the genetic alterations detected by SNP array platforms in patients, a Multiplex Ligation-Dependent Probe Amplification (MLPA) assay was designed. When available, the parental samples were also studied by MLPA to determine the inheritance of the alteration in families. Specific synthetic probes were designed to target the specific CNVs detected (sequence available upon request). All MLPA reactions were performed with 100ng of genomic DNA and analyzed on an *ABI PRISM 3100 Genetic analyzer* according to manufacturers' instructions. Each MLPA signal was normalized and compared to the corresponding peak height obtained in control samples.

CNV content

To explain the severe phenotype observed in patients and highlight new candidate genes for obesity, special attention was paid to the genetic content of all the rCNVs detected. We listed all coding genes affected by rCNVs excluding the ones present in multiple copy regions. Lots of genes were included in the rCNVs identified, so an intensive prioritization was performed. We used 11 categories to identify the most relevant genes from the rCNVs:

Described

A huge amount of genes has already been associated with obesity. We selected the top500 genes associated with Obesity in the DisGeNET database [30].

Recurrent

New obesity candidate genes may be recurrent in rCNVs from EOO patients. We selected genes present in at least 2 rCNVs from different patients.

Intolerant

Alterations affecting genes intolerant to loss-of-function mutations may be more likely involved in disease development. We selected genes with a pLI higher than 0.9.

Pathway

The analysis of pathways already associated with obesity may be a good way of identifying new candidate genes involved in this disorder. We selected genes sharing at least one pathway with one of 15 obesity-related genes (ORGs: *ADRB3*, *AGRP*, *BDNF*, *LEP*, *LEPR*, *MC3R*, *MC4R*, *NPY*, *NPY1R*, *NTRK2*, *PCSK1*, *PCSK2*, *POMC*, *PPARG*, *SIM1*) in the GeneMANIA prediction server [31].

Physical interactions

The presence of physical interactions between proteins is a clear indication for functional relationship. We selected genes with at least one physical interactions with the ORGs in the GeneMANIA prediction server [31].

Predicted

Predicted functional relationships between genes through protein interactions and orthology may be indicative of similar cellular roles. We selected genes with at least one predicted functional relationship with the ORGs in the GeneMANIA prediction server [31].

Co-expressions

Similar gene expression levels across conditions may set the bases for a functional relationship between different genes. We selected genes with at least three co-expressions with the ORGs in the GeneMANIA prediction server [31].

Shared protein domains

Similar functions between proteins may be explained by common protein domains. We selected genes with at least one shared protein domains with the ORGs in the GeneMANIA prediction server [31].

Co-localization

To assess the possible interaction between two proteins it is mandatory that they are expressed in the same tissue and are localized in the same cellular compartment. We selected genes with at least one co-localization with the ORGs in the GeneMANIA prediction server [31].

Genetic interactions

Genetic interaction may represents a functional association between genes and pathways and is present when the effects of perturbing one gen modifies a second gene. We selected genes with at least three genetic interactions with the ORGs in the GeneMANIA prediction server [31].

Enriched

An enrichment analysis with the ConsensusPathDB [32] was performed with all coding genes included in rCNVs. Significantly enriched pathways were considered if the p-value corrected for multiple testing using the false discovery rate (q-value) was lower than 0.05. We selected the genes included in these enriched pathways.

Replication group

A group of patients with EOO from *Viva la Familia* study (phs000616.v2.p2) was used to replicate the results found in our initial EOO Spanish group. This study comprises Hispanic families from Houston with probands between 4 and 19 years old affected by obesity defined by BMI>95th percentile [33]. LRR and BAF values were generated from SNP array raw data [34]. Low quality data in some patients generated unfaithful rCNV callings, so we excluded patients with more than 25 rCNV calls as low quality SNP arrays. From each family, one sibling with obesity was randomly selected, which lead to a total of 231 EOO patients (VLF group). Moreover, as VLF individuals were from a slightly different geographical context, during the filtering rCNVs present in more than 5 individuals were removed as population variants. Apart from that, the rCNV calling, filtering and prioritization was performed using the same criteria

as in the previous analysis. Data from both obesity groups (EOO-Sp and VLF) was also analyzed together (All Obe group).

RESULTS

CNV detection sensitivity

CNV validation was performed analyzing the plots of LRR and BAF. Additionally, to prove the effectiveness of this approach, a set of 129 CNVs were analyzed by MLPA. All CNVs selected were validated, yielding to a 100% sensitivity and supporting the idea that plots are sufficient to discriminate between real CNVs and technical artifacts.

Known genetic alterations

Among the alterations with higher frequencies in cases versus controls, we identified 7 patients (1754, 1700, 1572, 1727, 2117, 1541 and VF2058) with a 16p13.11 duplication including a common single copy region of 800kb between two Segmental Duplications and only one control with a much bigger duplication. We also identified one patient (1443) with a deletion in the 16p11.2 region. Other alterations found more frequently in patients were the 10q26.3 duplication (22 patients) and some distal alterations in 22q11.2 region (3 patients). None of the alterations reached significant values. The 15q13.3 duplication and one proximal deletion in the 15q11.2 region were found more frequently in controls compared to patients. Other genetic alterations were found in similar proportions between patients and controls.

Burden of rCNVs

A case-control study was performed between 465 EOO-Sp patients and 500 controls, and a higher burden of rCNVs was detected in patients with EOO compared to controls (Table 1). In the group of cases, 134 non-related EOO patients (28.8%) carried 164 rCNVs (Supplementary Table 1); and in the group of controls, 102 individuals (20.4%) carried 119 rare CNVs, being this difference statistically significant (OR=1.41; $p<0.0001$). From the alterations found in patients, 32 were deletions (6.9%) and 132 duplications (28.4%). In controls, 30 were deletions (6.0%) and 89 were duplications (17.8%). Whereas deletions were detected in similar rates, we observed a significant increase of gain-type rCNVs ($p=0.0047$). Regarding the co-occurrence of more than one

rCNV in the same subject, multiple hits were present in 5.8% of patients and 3% of controls. This difference was statistically significant (p=0.0396).

Data from the VLF replication group confirmed the contribution of rCNVs to the etiology of the disease. We detected 75 non-related patients (32.5%) carrying 93 rCNVs (Table 1). The results showed even an increased proportion of alterations in patients compared to controls (OR=1.59; p=0.0006). This increase was observed in both, deletions and duplications.

Group	rCNVs	Deletions	Duplications	Multiple hits	Samples
Cont	119	30 (6.0%)	89 (17.8%)	15 (3.0%)	102 (20.4%)
EOO-Sp	164	32 (6.9%)	132** (28.4%)	27 (5.8%)	134* (28.8%)
VLF	93	29 (12.6%)	64 (27.7%)	14 (6.1%)	75** (32.5%)
All Obe	257	61 (8.8%)	196 (28.2%)	41 (5.9%)	209** (30.0%)

Table 1. Comparison of the frequencies of rCNVs (autosomal and chromosome X) fulfilling the established criteria (>100 kb, containing coding genes, no overlapping with Segmental Duplications and present in <1/2000 population individuals) in the different groups: Controls (Cont, n=500), early-onset obesity Spanish patients (EOO-Sp, n=465), Viva la Familia patients (VLF, n=231), and both groups of patients together (All Obe, n=696). *p≤0.05; **p≤0.01

Segregation of rCNVs with obesity phenotype

Segregation analysis of 44 rCNVs in 34 trios revealed that 41 rCNVs in 34 cases were inherited and 3 were *de novo*. No significant differences were found between paternally and maternally inherited alterations (19 and 22, respectively). Detailed clinical data from these 34 patients is presented in Supplementary Table 2.

From the inherited rCNVs, 20 were transmitted from a progenitor affected by obesity and 23 from a non-obese progenitor. Segregation studies revealed that half of the alterations analyzed were compatible with the phenotype inheritance observed in the family, which may be relevant for the management of obesity in those families.

Genes in rCNVs

The 164 rCNVs identified in EOO-Sp patients include 515 coding genes, the 93 rCNVs from VLF patients include 286 genes, and the 119 rCNVs from controls include 360. If we listed the number of different single-copy coding

gens, we obtained 486 in EOO-Sp patients, 226 in VLF patients and 243 in controls. Although rCNVs from EOO-Sp patients compared to controls, on average, were bigger and had more genes, no statistically significant differences were found (Table 2). Analysis of rCNVs from the VLF group showed that alterations were smaller in size (p=0.0007), but with similar genetic content (Table 2) compared to the other groups.

Group	Size (kb)				Genes (genes/rCNV)			
	Mean	SD	N	p	Mean	SD	N	p
Cont	408.2	404.5	119		3.03	2.95	119	
EOO-Sp	440.1	766.0	146	0.6820	3.53	4.69	146	0.3108
VLF	251.5	190.0	93	0.0007	3.08	2.49	93	0.8959
All Obe	366.7	616.4	239	0.5061	3.35	3.98	239	0.429

Table 2. Analysis of the size and the genetic content of rCNVs from the different groups: Controls (Cont, n=500), early-onset obesity Spanish patients (EOO-Sp, n=465), Viva la Familia patients (VLF, n=231), and both groups of patients together (All Obe, n=696)..

From the 11 categories studied, 192 genes from rCNVs of EOO-Sp patients were present in at least one category, 102 genes from rCNVs of VLF patients and 80 genes from rCNVs of controls. The number of individuals carrying rCNVs affecting genes from those categories is listed in Table 3. We observed a significant increase of patients with rCNVs affecting prioritized genes compared to controls ($p<0.0001$). When analyzed individually, the category Described, which included genes already associated with obesity, had a clearly significant increase in the two cohorts studied. This fact represented a proof of concept for our study, as we were able to reproduce already known genetic factors. Other categories harboring the main differences were Recurrent and Intolerant, so special attention may be paid in genes present in alterations from multiple patients and genes intolerant to loss of function variation. No significant differences were found in Pathway, Physical interactions or Predicted.

Category	Cont	EOO-Sp	VLF	All Obe
Described	4 (0.8%)	19 (4.1%)**	12 (5.2%)**	31 (4.5%)**
Recurrent	14 (2.8%)	41 (8.8%)**	32 (13.9%)**	75 (10.8%)**
Intolerant	27 (5.4%)	58 (12.5%)**	32 (13.9%)**	90 (12.9%)**
Pathway	2 (0.4%)	7 (1.5%)	1 (0.4%)	8 (1.1%)
Physical interactions	2 (0.4%)	7 (1.5%)	1 (0.4%)	8 (1.1%)
Predicted	0 (0%)	3 (0.6%)	1 (0.4%)	4 (0.6%)
Co-expressions	21 (4.2%)	26 (5.6%)	20 (8.7%)*	46 (6.6%)
Shared protein domains	3 (0.6%)	16 (3.4%)**	6 (2.6%)*	22 (3.2%)**
Co-localization	10 (2%)	24 (5.2%)**	10 (4.3%)	34 (4.9%)*
Genetic interactions	14 (2.8%)	24 (5.2%)	18 (7.8%)**	42 (6%)**
Enriched	5 (1%)	2 (0.4%)	20 (8.7%)**	0 (0%)*
None (0)	436(87.2%)	346 (74.4%)**	155 (67.1%)**	501 (72%)**
Single (1)	55 (11%)	98 (21.1%)**	62 (26.8%)**	159 (22.8%)**
Some (2)	17 (3.4%)	37 (8%)**	30 (13%)**	65 (9.3%)**
Quite (3)	5 (1%)	14 (3%)*	3 (1.3%)	16 (2.3%)
A lot (4)	0 (0%)	3 (0.6%)	5 (2.2%)**	6 (0.9%)*
Any (>=1)	64 (12.8%)	119 (25.6%)**	76 (32.9%)**	192 (27.6%)**
At least two (>=2)	9 (1.8%)	53 (11.4%)**	37 (16%)**	89 (12.8%)**

Table 3. Number of individuals with rCNVs harboring prioritized genes from each category studied in the different groups: Controls (Cont, n=500), early-onset obesity Spanish patients (EOO-Sp, n=465), Viva la Familia patients (VLF, n=231), and both groups of patients together (All Obe, n=696). In the first part of the table are presented the 11 categories studied individually: Described, Recurrent, Intolerant, Pathway, Physical interactions, Predicted, Co-expressions, Shared protein domains, Co-localization, Genetic interactions and Enriched (see Materials and Methods). In the second part of the table are presented the number of categories grouped together: Patients with rCNVs affecting genes present in 0 categories (none), 1 category (single), 2 categories (some), 3 categories (quite), 4 categories (a lot), at least one category (any) and at least two categories (at least two). Notice that the "All Obe" column may not always correspond to the sum of both obesity groups (for example: one gene present in one rCNVs from EOO-Sp and one rCNV from VLF, will only be considered recurrent when analyzing both groups together). *p≤0.05; **p≤0.01

Functional group	Gene	Region	rCNV	Patients
Obesity	*NPY*	7p15.3	Comp. Gain	OBE03
	SH2B1	16p11.2[#]	Comp. Loss	1443
	KCNQ1	11p15.5	Part. Loss	2195
	PHF6	Xq26.2-q26.3	Part. Gain	1689
Cilia	*FOPNL*	16p13.11[#]	Comp. Gain	1572, 1727, 1754, 2117
	CEP19	3q29	Comp. Gain	1770
	TTC8	14q31.3	Comp. Gain	1734
Circadian rhythm	*ARNTL2*	12p11.23	Comp. Gain	1856
	PER1	17p13.1	Comp. Gain	2214
	RORB	9p21.13	Part. Loss	VF1827
Adipogenesis	*BMP5*	6p12.1	Part. Gain	1858
	HFE2	1q21.1	Comp. Loss	2080
	HFE2	1q21.1	Comp. Gain	OBE34
	GREM2	1q43	Part. Loss	2197
	SCUBE2	11p15.4	Comp. Gain	VF0628
	ECHS1, CYP2E1	10q26.3[#]	Comp. Gain	1758
	CYP2E1	10q26.3[#]	Comp. Loss	1689
	FZD10	12q24.33	Comp. Gain	1608
	FZD6	8q22.3	Comp. Gain	1934
	WNT2B	1p13.2	Comp. Gain	VF0662
Inflammation	*FPR1, FPR2, FPR3*	19p13.41	Comp. Gain	2058
	ITGA10	1q21.1	Comp. Loss	2080
	ITGA10	1q21.1	Comp. Gain	OBE34
	ITGA7	12q13.2	Comp. Loss	2154
	CCL23	17q12	Comp. Gain	1974
	NLRC4	2p22.3	Part. Gain	VF1667
Central nervous system	*SEMA3A*	7q21.11	Part. Gain	VF1248
	SEMA5A	5p15.31-p15.2	Part. Gain	VF2125
	SERPINI1	3q26.1	Comp. Loss	VF0897
	VGF	7q22.1	Part. Gain	VF1743
	DMXL2	15q21.2	Part. Gain	1985
	IRX4	5p15.33	Comp. Gain	VF1662

	IL1RAPL1	Xp21.3	Part. Gain*	1450
	CNTN3	3p12.3	Part. Gain	1966
	PCDH9	13q21.32	Part. Gain	VF0735
	EXOC3	5p15.33	Comp. Gain	1719
	NLGN1	3q26.31	Part. Loss	1753
	NLGN4X	Xp22.32-p22.31	Comp. Loss	VF1935 (M)
	ERBB4	2q34	Part. Gain	VF0745
	GSK3B	3q13.33	Part. Gain	VF1683
	KCND2	7q31.31	Part. Gain*	VF2036
	MECP2	Xq28	Comp. Loss	VF0659 (F)
PLC activity and Ca^{2+} homeostasis	*CALY, CYP2E1*	10q26.3#	Comp. Gain	1758
	ADORA2A	22q11.2#	Comp. Gain	1741
	GRPR	Xp22.2	Comp. Gain	OBE66
	GRM7	3p26.1	Part. Gain	1827, 2120
	GRM7	3p26.1	Part. Loss*	2059
	GRIK1	21q21.3	Part. Gain	1586
	PIGZ	3q29	Comp. Gain	1770
	ANXA10	4q32.3	Part. Loss	VF0579
	HTR1E	6q14.3	Comp. Gain	1974
	PLCB1, PLCB4	20p12.3-p12.2	Part. Gain	1534
	IMPA1	8q21.13	Comp. Gain	1555
	NTSR2	2p25.1	Comp. Gain	VF1090
	TANC1	2q24.1-q24.2	Part. Gain	1380, 1748, 2034
	CALN1	7q11.22	Part. Loss	VF0802

Table 4. Prioritized genes present in rCNVs from obesity patients by functional groups. Alterations with the breakpoint truncating the gene are reported as partial (Part.), whereas alterations including the whole gene are reported as complete (Comp.). The patients with rCNVs affecting candidate genes are also presented. Patients from VLF are coded as "VF". Sex of patients with rCNV on chrX is reported: male (M), female (F). #: rCNVs overlapping with known genetic alterations. *: rCNVs only affecting an intronic region.

From the huge amount of genes present in rCNVs from obesity patients we have prioritized some, which can be found in Table 4. We identified genes already associated with obesity, linked to ciliogenesis, circadian rhythm,

adipogenesis, inflammation, central nervous system, phospholipase C (PLC) activity and cellular calcium (Ca^{2+}) homeostasis.

Only 5 genes were included in rCNVs identified in both groups of patients (EOO-Sp and VLF). *C9orf66* and *DOCK8* genes were present in a 9p24.3 terminal duplication identified in patient 1798 and 4 patients from VLF. *FANCM* and *MIS18BP1* were present in different alterations in the 14q21.2 region, including a 8Mb deletion in patient 2117, and 2 identical duplications and 1 slightly-overlapping deletion in VLF patients. *TJP1* gene was present in 15q13.1 duplications in patient 2205 and 2 VLF patients. None of the proteins coded by those genes had ever been linked to obesity.

DISCUSSION

Obesity, as a complex and multifactorial disorder, represents a challenging target for genetic studies. Although there is strong evidence for a high genetic component in obesity, we are still far from understanding all the genetic causes underlying this disorder. Our study focuses on an extreme phenotype from the obesity spectrum. Severe early-onset cases are more likely explained by high penetrant rare genetic variants. The recruitment of an elevated number of this type of patients is not always easy, and the collaboration with reference centers is crucial. Focusing on the subgroup of EOO, we have analyzed the presence of CNVs using data from SNP array platforms in 465 Spanish patients. In addition, we have also analyzed a replication group of 231 individuals with the same phenotype (severe EOO), with a similar range of ages and population context (young Hispanic). The identification of rCNVs in VLF patients including genes related to syndromic forms, such as *SOS1*, *PTPN11* and *MECP2* (data not shown), reveal the difficulty to select only patients with real isolated obesity.

Up to 1% of obesity patients (EOO-Sp: 6/465, and VLF: 1/231) were found to carry 16p13.11 duplications affecting several genes. Some of the alterations included the *FOPNL* gene, a cilia-related gene. Moreover, the only control with a 16p13.11 duplication, carried a much bigger alteration. The frequency in patients affected by obesity observed in the Decipher dataset in 2020 (0.65%) [35] was also bigger than the reported frequencies in controls (Hannes (2009): 0.29% [36], and Eichler (2014): 0.20% [37]). Despite the increased frequency detected in patients, no further evidences were found to support a

clear pathogenic effect of this alteration, which is usually reported as a neurocognitive disorder susceptibility locus or as benign [36] .

We also identified one patient with a deletion in the 16p11.2 region including the *SH2B1* gene. This 200kb-alteration, despite sharing the same terminal breakpoint in the latter Segmental Duplication, was smaller than the typically reported 16p11.2 deletions, which have been associated with BMI and neuropsychiatric disorders [19,20,38].

In addition, one obesity patient was described with a 22q11.2 distal duplication of 1.2Mb including *SLC2A11* and *ADORA2A* genes, coding for a glucose transporter and a Ca^{2+} regulator, respectively. This and other smaller distal duplication in that region were found more frequently in patients than controls, indeed 22q11.2 duplications have been positively associated with BMI [39]. We also identified 10q26.3 duplications, which were detected in several patients and included genes like *CYP2E1* and *CALY*. CYP2E1 is a member of cytochrome P450 superfamily of enzymes which play an important role in fatty acid metabolism. Different studies suggest that CNVs including the *CYP2E1* gene may be related to hypertriglyceridemia and obesity [40].

In the case-control study performed between 465 patients and 500 controls a higher burden of rCNVs was detected in patients with EOO compared to controls. Our results are in accordance with other studies where an increased amount of rCNVs was detected in severe EOO patients compared to controls. The percentages of EOO patients carrying rCNVs are high (29% in our EOO-Sp group and 32% in the replication group) compared to 20% in controls. Other similar studies have also found big differences (19% in patients and 3% in controls) [21], supporting the hypothesis that rCNVs could be important contributors to isolated obesity.

Co-segregation studies of the variants with obesity were performed in cases where parental genotypes have been analyzed. Although no parental bias was observed in inherited alterations, half of the transmitting progenitors were affected by obesity, representing a compatible explanation for the extreme phenotype observed in cases. The availability of more parental samples and familiar histories could have allowed a deepen knowledge about the contribution of specific alterations to the obesity phenotype. However, the study of genes included in rCNVs detected in patients confirmed known genotype-phenotype correlations and revealed new candidate genes which may be linked to obesity development. The 56 genes from Table 4 are discussed below.

Obesity

A total of 31 patients from both obesity groups were harboring rCNVs including 22 genes already associated with obesity. Genetic alterations affecting reported obesity genes were identified in near 5% of EOO patients and in less than 1% of controls. This statistically significant enrichment represents a proof of concept for our study as we reproduced what it has already been described. From the alteration affecting these genes, a *NPY* duplication in a familial case with EOO and attention deficit hyperactivity disorder was a clear example of high penetrant genetic variants, as already reported in a previous study from our group [24]. Moreover, as mentioned before, we also identified one patient with a deletion in the 16p11.2 region, including the *SH2B1* gene. Another obesity patient had a small deletion in the *KCNQ1* locus, an imprinting control region which disruption has been associated with Beckwith-Wiedemann syndrome [41]. Further analysis showed that the alteration was not affecting the imprinting control region and no methylation anomalies were identified by MS-MLPA (data not shown). And finally, one patient was harboring a duplication including the *PHF6* gene, which has been associated with a syndromic obesity form [42]. Other genes already associated with obesity, together with new candidates, are mentioned in the following lines.

Cilia

Variation in cilia-related genes may be part of the multifactorial etiology of this complex condition [43]. Furthermore, ciliopathy syndromes such as Bardet-Biedl syndrome and Alström syndrome display obesity as a hallmark characteristic. Impaired cilia function has been connected to both syndromic and nonsyndromic obesity in both human patients and rodents [43]. Here we identified 6 obesity patients with rCNVs affecting 3 cilia related genes (*FOPNL*, *CEP19* and *TTC8*). *FOPNL* is required for the formation of the primary cilium, but no alterations have been linked to obesity so far; *CEP19* is involved in cilia assembly and a homozygous mutation had been described in one family as the cause of severe obesity and spermatogenic failure [44]; and *TTC8* is part of a protein complex required for ciliogenesis and homozygous alterations cause Bardet-Biedl syndrome 8 [45]. These rCNVs affecting those genes may be contributing to explain the phenotype observed in the patients and replicate the knowledge that cilia related genes may have a relevant role in obesity development.

Circadian rhythm

It is well known that several hormones that are related to obesity, such as cortisol, leptin and adiponectin, among others, display circadian rhythmicity. The circadian rhythm is regulated by the heterodimer of CLOCK and BMAL1 (*ARNTL1*), that binds to the promoter regions of *PER* and *CRY* genes and induce their expression, which in turn inhibit Clock's activity and close the circadian loop [46]. The alteration of this normal pattern is called chronodisruption, and it has been associated with temporal alterations in feeding behavior, increased weight gain and obesity [47,48]. Here we identified 3 obesity patients with rCNVs including gens involved in the circadian rhythm (*ARNTL2*, *PER1* and *RORB*). ARNT and PER proteins share one structural characteristic with SIM1: the PAS domain. The PAS domain, which acts as a molecular velcro, has been shown as the key structural motif involved in protein-protein interactions of the circadian clock. *SIM1* is a known gene involved in the leptin-melanocortin pathway, so this may indicate a possible direct link between hypothalamic energy balance regulation and circadian rhythm.

Adipogenesis

Up to 10 patients were identified with rCNVs affecting genes involved in adipogenesis, such as a bone morphogenetic proteins (BMPs) (*BMP5*, *HFE2*, *GREM2* and *SCUBE2*), fatty acid beta-oxidation (*ECHS1* and *CYP2E1*) or the WNT signaling (*FZD10*, *FZD6* and *WNT2B*). Genetic alterations affecting genes related to adipogenesis may reveal relevant candidate genes in the development of obesity.

Despite being first described as inducers of bone formation, BMPs are now known to be involved in morphogenetic activities and cell differentiation throughout the body, including the development of adipose tissue and adipogenic differentiation. *BMP5*, which encodes a protein mainly expressed in brown adipose tissue upon cold exposure and high-fat diet (HFD), may be linked to food intake [49]. The duplication identified in one patient may disrupt *BMP5* and alter adipogenesis. Other genes affected by rCNVs were *HFE2*, which encodes a coreceptor for BMPs that has been defined as a molecular link between obesity and iron homeostasis [50], *GREM2*, which encodes a BMP antagonist, and *SCUBE2*, which encodes an antagonist of BMP2. Despite the functional similarities, further investigation will be needed to clearly link the alterations detected with the phenotype.

Genes involved in steps from the fatty acid beta-oxidation may be associated with hypertriglyceridemia and obesity, such as *CYP2E1* [40]. However, *ECHS1* encodes an enzyme that catalyzes the second step in mitochondrial fatty acid beta-oxidation but is associated with a severe syndrome not related with obesity [51].

WNT signaling pathway regulates adult tissue maintenance and remodeling, also in adipose tissue, where is implicated in regulating adipogenesis [52]. Studies from animal models support the implication of WNT signaling to obesity [53,54] and one rare missense mutation affecting *WNT10B* gene has been reported to segregate in a family with early-onset obesity [55], but further studies will be required to establish additional causal mutation in humans, such as the duplications identified in 3 patients affecting genes from the WNT signaling pathway.

Inflammation

In the adipose tissue, infiltration of inflammatory macrophages contributes to inflammation and insulin resistance which correlates with hyperinsulinemia and insulin resistance in human [56]. Specific chemokine signatures have been associated with obesity [57], but no mutations have been directly linked with the phenotype in humans. Here we identified 6 obesity patients with rCNVs affecting genes related with the inflammatory response (*FPR1*, *FPR2*, *FPR3*, *ITGA10*, *ITGA7*, *CCL23* and *NLRC4*). Activation of formyl peptide receptors prevents the accumulation of macrophages in adipose tissue, improves glucose tolerance and restores systemic insulin sensitivity in obese mice [58,59]. Studies in mice showed that *FPR2* deletion reduced obesity and related comorbidities together with inflammation [59]. The duplication including 3 formyl peptide receptors genes identified in one obesity patient may indicate a relevant role of these receptors in obesity development. Alpha4 integrins contribute to the development of HFD-induced insulin resistance by mediating the trafficking of monocytes into adipose tissue; hence, blockade of alpha4 integrin signaling can prevent the development of obesity-induced insulin resistance [60]. No mutation have been linked to human obesity, but two patients were identified with duplications or deletions affecting integrin genes. Other inflammatory genes (*CCL23* and *NLRC4*) may also be linked with obesity development.

Central nervous system

Hypothalamic neurons are the main central energy balance regulators and may be affected by processes involving neurogenesis and neurotransmission. Here we identified 16 obesity patients with rCNVs affecting genes related to the neuronal development or the central nervous system. All these genes may contribute to expand the knowledge of obesity physiopathology and may be relevant in some families to explain their phenotype.

Hypothalamic Pomc neurons are important regulators of energy homeostasis. Semaphorins, neuroserpin and neuroendocrine peptides are involved in neuronal processes from these neurons. Semaphorins act as axonal growth cone guidance molecules and have been associated with adipose tissue inflammation and insulin resistance [61]. Recently, rare variants in genes from the SEMA3-mediated signaling pathway have been linked to the development of hypothalamic melanocortin circuits involved in energy homeostasis [62]. Moreover, neuroserpin has been shown to interact with the proprotein convertases involved in POMC processing [63]. In addition, the VGF-derived peptide TLQP-21 has been linked to energy homeostasis by increasing energy expenditure [64,65]. The duplication truncating semaphorin genes identified in two obesity patients, the deletion of the neuroserpin gene identified in one obesity patient and the duplication of *VGF* gene in another patient may support the idea that these proteins are important for developing Pomc neurons and regulating energy balance.

Another neuronal player involved in endocrine processes affected by a rCNV from one patient with obesity was the synaptic protein rabonectin-3α, encoded by the *DMXL2* gene. Rabonectin-3α is a key controller of neuronal and endocrine homeostatic processes and is involved in the regulation of the Notch signaling pathway where it may inhibit brown adipocytes in white adipose tissue and therefore decrease energy expenditure and promote obesity [66]. Two different studies have linked *DMXL2* variants with appetite or BMI: one homozygous alteration has been described in a family with profound hypoglycemia progressing to nonautoimmune insulin-dependent diabetes mellitus [67] and one mutation influenced weight and BMI in early childhood in a parent-of-origin-specific manner [68].

Additional genes related with neuronal processes were identified as possible new candidate genes for obesity: *IRX4* for the link between *IRX3* and BMI-associated FTO variants and its possible implication in human obesity [69–71]; *IL1RAPL1* for the possible relationship with intellectual disability and obesity

[72]; *CNTN3* for the unclear role of contactins in obesity risk [73,74]; *PCDH9K* for the association of protocadherins with childhood obesity [75]; *EXOC3* for the implication of some exocyst components in the reduction of free fatty acid uptake by adipocytes [76]; *NLGN1* and *NLGN4X* for the wide range of phenotypes associated with neuroligins deficiencies [77–80]; *ERBB4* for the predisposition to metabolic syndrome [81] and the presence of a near SNP that has been associated with BMI [82]; *GSK3* for the role in adipocyte inflammation [83]; and *KCND2* for the implication in regulating preadipocytes proliferation and differentiation [84].

Moreover, we have identified a deletion in one obesity patient from the replication cohort that may link *MECP2* gene with obesity-related features. MECP2 is required for maturation of neurons and despite *MECP2* mutations have been mainly associated with Rett syndrome, *MECP2* mutation carriers may also develop an Angelman-like phenotype with no features of Rett syndrome [85]. The identified alteration may support this idea, but no further clinical data was obtained from this patient.

PLC activity and Ca^{2+} homeostasis

We identified 17 obesity patients with rCNVs affecting several genes related to Phospholipase C (PLC) activity, cellular calcium (Ca^{2+}) homeostasis and cAMP regulation. PLC activity and Ca^{2+} concentration are tightly regulated as are involved in several signal transduction pathways [86]. Dysfunction of these systems have been associated with several pathological conditions, including obesity. The great amount of genes related with signal transduction pathways and its possible implications to obesity development may be further studied to identify causal alterations explaining the severe obesity phenotype of patients.

Here we reported two genes included in known altered regions, the *CALY* gene in the 10q26.3 region and the *ADORA2A* gene in the 22q11.2 region. Duplications in these regions have been associated with hypertriglyceridemia and increased BMI [39,40]. The study of the genetic content of known alterations acting as susceptibility factors may reveal new candidates for obesity involved in signalling pathways related with Ca^{2+} and cAMP.

Previous sequencing studies of EOO patients from our group revealed gastrin and glutamate receptors as relevant candidate gene for obesity by reporting pathogenic variants affecting *GRPR*, *GRM7* and *GRIK1* genes [24]. Here we identified up to 1% of patients from the EOO-Sp group with rCNVs affecting

these genes. Gastrin is a hormone implicated in the regulation of food ingestion and satiety, and its receptor is a known activator of PLC signaling pathway [87]. Glutamate is the main excitatory neurotransmitter, and its receptors are involved in the modulation of synaptic plasticity, ionotropic glutamate receptors form ion channel pores and metabotropic glutamate receptors mediate signal transduction cascades [88]. Additional studies regarding the functional implications of these alterations may link gastrin and glutamate receptors with obesity development.

Additional genes related with signal transduction pathways were identified as possible new candidate genes for obesity: *PIGZ* for the association of 3q29 duplications with BMI [39]; *ANXA10* for the implication of Annexins in obesity and related metabolic diseases [89]; *HTR1E* for the role of serotonin in feeding behavior [90]; *PLCB1* and *PLCB4* for the implication in thermogenesis and adipogenesis [91]; *IMPA1* for the presence of a near SNP that has been associated with obesity [92]; *NTSR2K* for the role of Neurotensin in feeding suppression [93]; *TANC1* for the role in PLC activation [94]; and *CALN1* for the putative implication in Ca^{2+} signal transduction pathway.

Conclusions

In summary, our findings showed a higher burden of rCNVs in patients with EOO compared to controls, revealing new genes and pathways that may be good candidates for extreme obesity phenotypes. It is of note that most of those rCNVs may be important contributors to the phenotype studied and worth to be explored deeply, but the lack of functional studies could not ensure its real contribution to obesity. However, the identification of rCNVs in 29%-32% of patients affected by obesity from both groups studied represents an undeniable fact to reaffirm the contribution of low-frequency high-penetrant alterations in complex disorders, such as obesity. Moreover, the finding that most of the alterations detected in patients where unique and that only 1% of the genes affected by these rCNVs were also found in rCNVs from the replication group, revealed the high genetic heterogeneity between obesity patients.

REFERENCES

1. World Health Organization. *Fact Sheet "Obesity and Overweight."*; 2018. http://www.who.int/mediacentre/factsheets/fs311/en/.

2. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: A systematic review and meta-analysis. *Obes Rev*. 2016;17(2):95-107. doi:10.1111/obr.12334
3. World Health Organization. *Taking Action on Childhood Obesity*.; 2018.
4. Instituto Nacional de Estadística. *Spanish National Health Survey*.; 2017.
5. Eurostat. *European Health Interview Survey - Spain*.; 2014.
6. Must A, Spadano J, Coakley EH, Field AE, Colditz G, Dietz WH. The disease burden associated with overweight and obesity. *J Am Med Assoc*. 1999;282(16):1523-1529.
7. Bell CG, Walley AJ, Froguel P. The genetics of human obesity. *Nat Rev Genet*. 2005;164(4):293-301. doi:10.1016/j.trsl.2014.05.010
8. Farooqi IS, Keogh JM, Yeo GSH, Lank EJ, Cheetham T. Clinical Spectrum of Obesity and Mutations in the Melanocortin 4 Receptor Gene. *N Engl J Med*. 2003;348(12):1085-1095.
9. Ranadive S, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am*. 2008;37(3):733-751.
10. Llewellyn CH, Trzaskowski M, Plomin R, Wardle J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. *Int J Obes*. 2013;37(11):1506-1509. doi:10.1038/ijo.2013.30
11. Waalen J. The genetics of human obesity. *Transl Res*. 2014;164(4):293-301. doi:10.1016/j.trsl.2014.05.010
12. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569-573. doi:10.1038/nature07953
13. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455(7210):232-236. doi:10.1038/nature07229
14. de Vries BBA, Pfundt R, Leisink M, et al. Diagnostic Genome Profiling in Mental Retardation. *Am J Hum Genet*. 2005;77(4):606-616. doi:10.1086/491719
15. Wheeler E, Huang N, Bochukova EG, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet*. 2013;45(5):513-517. doi:10.1038/ng.2607
16. Aerts E, Beckers S, Zegers D, et al. CNV analysis and mutation screening indicate an important role for the NPY4R gene in human obesity. *Obesity*. 2016;24(4):970-976. doi:10.1002/oby.21435

17. Jarick I, Vogel CIG, Scherag S, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet*. 2011;20(4):840-852. doi:10.1093/hmg/ddq518
18. Yang TL, Guo Y, Shen H, et al. Copy number variation on chromosome 10q26.3 for obesity identified by a genome-wide study. *J Clin Endocrinol Metab*. 2013;98(1):191-195. doi:10.1210/jc.2012-2751
19. Bochukova EG, Huang N, Keogh J, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. *Nature*. 2010;463(7281):666-670. doi:10.1038/nature08689
20. Walters RG, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010;463(7281):671-675. doi:10.1038/nature08727
21. Pettersson M, Viljakainen H, Loid P, et al. Copy number variants are enriched in individuals with early-onset obesity and highlight novel pathogenic pathways. *J Clin Endocrinol Metab*. 2017;102(8):3029-3039. doi:10.1210/jc.2017-00565
22. Martos-Moreno GÁ, Serra-Juhé C, Pérez-Jurado LA, Argente J. Underdiagnosed Beckwith-Wiedemann syndrome among early onset obese children. *Arch Dis Child*. 2014;99(10):965-967.
23. Serra-Juhé C, Martos-Moreno G, Bou de Pieri F, et al. Heterozygous rare genetic variants in non-syndromic early-onset obesity. *Int J Obes*. 2020;40:830-841. doi:10.1038/s41366-019-0357-5
24. Serra-Juhé C, Martos-Moreno G, Bou de Pieri F, et al. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet*. 2017;13(5):1-19. doi:10.1371/journal.pgen.1006657
25. Wang K, Li M, Hadley D, et al. PennCNV: An integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-1674. doi:10.1101/gr.6861907
26. Guey LT, García-Closas M, Murta-Nascimento C, et al. Genetic Susceptibility to Distinct Bladder Cancer Subphenotypes. *Eur Urol*. 2010;57(2):283-292. doi:10.1016/j.eururo.2009.08.001
27. Rodríguez-Santiago B, Malats N, Rothman N, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. *Am J Hum Genet*. 2010;87(1):129-138. doi:10.1016/j.ajhg.2010.06.002
28. Durinck S, Moreau Y, Kasprzyk A, et al. BioMart and Bioconductor:

A powerful link between biological databases and microarray data analysis. *Bioinformatics*. 2005;21(16):3439-3440. doi:10.1093/bioinformatics/bti525

29. Cooper GM, Coe BP, Girirajan S, et al. A Copy Number Variation Morbidity Map of Developmental Delay. *Nat Genet*. 2012;43(9):838-846. doi:10.1038/ng.909
30. Bauer-Mehren A, Rautschka M, Sanz F, Furlong LI. DisGeNET: A Cytoscape plugin to visualize, integrate, search and analyze gene-disease networks. *Bioinformatics*. 2010;26(22):2924-2926. doi:10.1093/bioinformatics/btq538
31. Mostafavi S, Ray D, Warde-Farley D, Grouios C, Morris Q. GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function. *Genome Biol*. 2008;9(54):1-15. http://genomebiology.com/2008/9/s1/S4.
32. Kamburov A, Wierling C, Lehrach H, Herwig R. ConsensusPathDB - A database for integrating human functional interaction networks. *Nucleic Acids Res*. 2009;37(SUPPL. 1):623-628. doi:10.1093/nar/gkn698
33. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: Genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr*. 2006;84(3):646-654. http://www.ajcn.org/cgi/reprint/84/3/646%5Cnhttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006549764.
34. Zhou W. Segmentation-Based Detection of Mosaic Chromosomal Abnormality in Bladder Cancer Cells Using Whole Genome SNP Arrays. 2014;(February).
35. Firth H V., Richards SM, Bevan AP, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524-533. doi:10.1016/j.ajhg.2009.03.010
36. Hannes FD, Sharp AJ, Mefford HC, et al. Recurrent reciprocal deletions and duplications of 16p13.11: The deletion is a risk factor for MR/MCA while the duplication may be a rare benign variant. *J Med Genet*. 2009;46(4):223-232. doi:10.1136/jmg.2007.055202
37. Coe BP, Witherspoon K, Rosenfeld JA, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet*. 2014;46(10):1063-1071.

doi:10.1038/ng.3092

38. Bachmann-Gagescu R, Mefford HC, Cowan C, et al. Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genet Med*. 2010;12(10):641-647. doi:10.1097/GIM.0b013e3181ef4286
39. Owen D, Bracher-Smith M, Kendall KM, et al. Effects of pathogenic CNVs on physical traits in participants of the UK Biobank 11. *BMC Genomics*. 2018;19(1):1-9. doi:10.1186/s12864-018-5292-7
40. Li JJ, Chen Y qin, Fan LL, Jin JY, Guo S, Xiang R. Microduplication of 10q26.3 in a Chinese hypertriglyceridemia patient. *Mol Cell Probes*. 2018;37(November 2017):28-31. doi:10.1016/j.mcp.2017.11.002
41. Valente FM, Sparago A, Freschi A, et al. Transcription alterations of KCNQ1 associated with imprinted methylation defects in the Beckwith–Wiedemann locus. *Genet Med*. 2019;21(8):1808-1820. doi:10.1038/s41436-018-0416-7
42. BÖRJESON M, FORSSMAN H, LEHMANN O. An X-linked, Recessively Inherited Syndrome Characterized by Grave Mental Deficiency, Epilepsy, and Endocrine Disorder. *Acta Med Scand*. 1962;171(1):13-22. doi:10.1111/j.0954-6820.1962.tb04162.x
43. Mariman ECM, Vink RG, Roumans NJT, et al. The cilium: a cellular antenna with an influence on obesity risk. *Br J Nutr*. 2016;116(4):576-592. doi:10.1017/s0007114516002282
44. Shalata A, Ramirez MC, Desnick RJ, et al. Morbid obesity resulting from inactivation of the ciliary protein CEP19 in humans and mice. *Am J Hum Genet*. 2013;93(6):1061-1071. doi:10.1016/j.ajhg.2013.10.025
45. Ansley SJ, Badano JL, Blacque OE, et al. Basal body dysfunction is a likely cause of pleiotropic Bardet-Biedl syndrome. *Nature*. 2003;425(October):628-633. doi:10.1038/nature01910.1.
46. Darlington TK, Wager-smith K, Ceriani MF, et al. Closing the Circadian Loop : CLOCK-Induced Transcription of Its Own Inhibitors per and tim. *Science (80-)*. 1998;280(June):1599-1603.
47. Lopez-Minguez J, Gómez-Abellán P, Garaulet M. Circadian rhythms, food timing and obesity. *Proc Nutr Soc*. 2016;75(4):501-511. doi:10.1017/s0029665116000628
48. Engin A. Circadian Rhythms in Diet-Induced Obesity. *Adv Exp Med Biol*. 2017;960:19-52. doi:10.1007/978-3-319-48382-5
49. Blázquez-Medela AM, Jumabay M, Boström KI. Beyond the bone: Bone morphogenetic protein signaling in adipose tissue. *Obes Rev*.

2019;20(5):648-658. doi:10.1111/obr.12822

50. Luciani N, Brasse-Lagnel C, Poli M, et al. Hemojuvelin: A new link between obesity and iron homeostasis. *Obesity*. 2011;19(8):1545-1551. doi:10.1038/oby.2011.12
51. Al Mutairi F, Shamseldin HE, Alfadhel M, Rodenburg RJ, Alkuraya FS. A lethal neonatal phenotype of mitochondrial short-chain enoyl-CoA hydratase-1 deficiency. *Clin Genet*. 2017;91(4):629-633. doi:10.1111/cge.12891
52. Christodoulides C, Lagathu C, Sethi JK, Vidal-Puig A. Adipogenesis and WNT signalling. *Trends Endocrinol Metab*. 2009;20(1):16-24. doi:10.1016/j.tem.2008.09.002
53. Fuster JJ, Zuriaga MA, Ngo DTM, et al. Noncanonical wnt signaling promotes obesity-induced adipose tissue inflammation and metabolic dysfunction independent of adipose tissue expansion. *Diabetes*. 2015;64(4):1235-1248. doi:10.2337/db14-1164
54. Wu Q, Tang SG, Yuan ZM. Gremlin 2 inhibits adipocyte differentiation through activation of Wnt/β-catenin signaling. *Mol Med Rep*. 2015;12(4):5891-5896. doi:10.3892/mmr.2015.4117
55. Christodoulides C, Scarda A, Granzotto M, et al. WNT10B mutations in human obesity. *Diabetologia*. 2006;49(4):678-684. doi:10.1007/s00125-006-0144-4
56. Weisberg SP, Leibel RL, Anthony W, et al. Obesity is associated with macrophage accumulation in adipose tissue. *J Clin Invest*. 2003;112(12):1796-1808. doi:10.1172/JCI200319246.Introduction
57. Ignacio RMC, Gibbs CR, Lee ES, Son DS. Differential chemokine signature between human preadipocytes and adipocytes. *Immune Netw*. 2016;16(3):189-194. doi:10.4110/in.2016.16.3.189
58. Wollam J, Riopel M, Xu YJ, et al. Microbiota-Produced N-Formyl Peptide fMLF Promotes Obesity-Induced Glucose Intolerance. *Diabetes*. 2019;68(7):1415-1426. doi:10.2337/db18-1307
59. Chen X, Zhuo S, Zhu T, et al. Fpr2 Deficiency Alleviates Diet-Induced Insulin Resistance Through Reducing Body Weight Gain and Inhibiting Inflammation Mediated by Macrophage Chemotaxis and M1 Polarization. *Diabetes*. 2019;68(6):1130-1142. doi:10.2337/db18-0469
60. Féral CC, Neels JG, Kummer C, Slepak M, Olefsky JM, Ginsberg MH. Blockade of α4 integrin signaling ameliorates the metabolic consequences of high-fat diet-induced obesity. *Diabetes*. 2008;57(7):1842-1851. doi:10.2337/db07-1751

61. Shimizu I, Yoshida Y, Moriya J, et al. Semaphorin3E-induced inflammation contributes to insulin resistance in dietary obesity. *Cell Metab*. 2013;18(4):491-504. doi:10.1016/j.cmet.2013.09.001
62. van der Klaauw AA, Croizier S, Mendes de Oliveira E, et al. Human Semaphorin 3 Variants Link Melanocortin Circuit Development and Energy Balance. *Cell*. 2019;176(4):729-742.e18. doi:10.1016/j.cell.2018.12.009
63. Wiesmüller FJA. *An Investigation of Interaction between Neuroserpin and Its Putative Targets PC1 / 3 , PC2 and Furin with a YFP-Protein Complementation Assay*.; 2016.
64. Hahm S, Mizuno TM, Wu TJ, et al. Targeted deletion of the Vgf gene indicates that the encoded secretory peptide precursor plays a novel role in the regulation of energy balance. *Neuron*. 1999;23(3):537-548. doi:10.1016/S0896-6273(00)80806-5
65. Bartolomucci A, La Corte G, Possenti R, et al. TLQP-21, a VGF-derived peptide, increases energy expenditure and prevents the early phase of diet-induced obesity. *Proc Natl Acad Sci U S A*. 2006;103(39):14584-14589. doi:10.1073/pnas.0606102103
66. Sethi N, Yan Y, Quek D, Schupbach T, Kang Y. Rabconnectin-3 is a functional regulator of mammalian notch signaling. *J Biol Chem*. 2010;285(45):34757-34764. doi:10.1074/jbc.M110.158634
67. Tata B, Huijbregts L, Jacquier S, et al. Haploinsufficiency of Dmxl2, Encoding a Synaptic Protein, Causes Infertility Associated with a Loss of GnRH Neurons in Mouse. *PLoS Biol*. 2014;12(9). doi:10.1371/journal.pbio.1001952
68. Li A, Robiou-du-Pont S, Anand SS, et al. Parental and child genetic contributions to obesity traits in early life based on 83 loci validated in adults: the FAMILY study. *Pediatr Obes*. 2018;13(3):133-140. doi:10.1111/ijpo.12205
69. Gómez-marín C, Aneas I, Credidio FL, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371-375. doi:10.1038/nature13138.Obesity-associated
70. Claussnitzer M, Dankel SN, Kim K-H, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*. 2015;373(10):895-907. doi:10.1056/NEJMoa1502214
71. Zou Y, Lu P, Shi J, et al. IRX3 Promotes the Browning of White Adipocytes and Its Rare Variants are Associated with Human Obesity Risk. *EBioMedicine*. 2017;24:64-75. doi:10.1016/j.ebiom.2017.09.010

72. Ramos-brossier M, Montani C, Lebrun N, et al. Novel IL1RAPL1 mutations associated with intellectual disability impair synaptogenesis. *Hum Mol Genet*. 2015;24(4):1106-1118. doi:10.1093/hmg/ddu523
73. Nikpay M, Šeda O, Tremblay J, et al. Genetic mapping of habitual substance use, obesity-related traits, responses to mental and physical stress, and heart rate and blood pressure measurements reveals shared genes that are overrepresented in the neural synapse. *Hypertens Res*. 2012;35(6):585-591. doi:10.1038/hr.2011.233
74. Buchner DA, Geisinger JM, Glazebrook PA, et al. The juxtaparanodal proteins CNTNAP2 and TAG1 regulate diet-induced obesity. *Mamm Genome*. 2012;23(7-8):431-442. doi:10.1007/s00335-012-9400-8
75. Moon S, Hwang MY, Jang HB, et al. Whole-exome sequencing study reveals common copy number variants in protocadherin genes associated with childhood obesity in Koreans. *Int J Obes*. 2017;41(4):660-663. doi:10.1038/ijo.2017.12
76. Inoue M, Akama T, Jiang Y, Chun TH. The exocyst complex regulates free fatty acid uptake by adipocytes. *PLoS One*. 2015;10(3):1-4. doi:10.1371/journal.pone.0120289
77. Reissner C, Klose M, Fairless R, Missler M. Mutational analysis of the neurexin/neuroligin complex reveals essential and regulatory components. *Proc Natl Acad Sci*. 2008;105(39):15124-15129. doi:10.1073/pnas.0801639105
78. Fitzgerald TW, Gerety SS, Jones WD, et al. Large-scale discovery of novel genetic causes of developmental disorders. *Nature*. 2015;519(7542):223-228. doi:10.1038/nature14135
79. Parente DJ, Garriga C, Baskin B, et al. Neuroligin 2 nonsense variant associated with anxiety, autism, intellectual disability, hyperphagia, and obesity. *Am J Med Genet Part A*. 2017;173(1):213-216. doi:10.1002/ajmg.a.37977
80. Jamain S, Quach H, Betancur C, et al. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat Genet*. 2003;34(1):27-29. doi:10.1038/ng1136
81. Zeng F, Wang Y, Kloepfer LA, Wang S, Harris RC. ErbB4 deletion predisposes to development of metabolic syndrome in mice. *Am J Physiol - Endocrinol Metab*. 2018;315(4):E583-E593. doi:10.1152/ajpendo.00166.2018
82. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1126/science.1249098.Sleep

83. Wang L, Wang Y, Zhang C, et al. Inhibiting glycogen synthase kinase 3 reverses obesity-induced white adipose tissue inflammation by regulating apoptosis inhibitor of macrophage/CD5L-mediated macrophage migration. *Arterioscler Thromb Vasc Biol*. 2018;38(9):2103-2116. doi:10.1161/ATVBAHA.118.311363
84. Hu H, He ML, Tao R, et al. Characterization of ion channels in human preadipocytes. *J Cell Physiol*. 2009;218(2):427-435. doi:10.1002/jcp.21617
85. Watson P, Black G, Ramsden S, et al. Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J Med Genet*. 2001;38(4):224-228. doi:10.1136/jmg.38.4.224
86. Arruda AP, Hotamisligil GS. Calcium homeostasis and organelle function in the pathogenesis of obesity and diabetes. *Cell Metab*. 2015;22(3):381-397. doi:10.1016/j.cmet.2015.06.010
87. Meek CL, Lewis HB, Reimann F, Gribble FM, Park AJ. The effect of bariatric surgery on gastrointestinal and pancreatic peptide hormones. *Peptides*. 2016;77:28-37. doi:10.1016/j.peptides.2015.08.013
88. Debanne D, Daoudal G, Sourdet V, Russier M. Brain plasticity and ion channels. *J Physiol Paris*. 2003;97(4-6):403-414. doi:10.1016/j.jphysparis.2004.01.004
89. Grewal T, Enrich C, Rentero C, Buechler C. Annexins in adipose tissue: Novel players in obesity. *Int J Mol Sci*. 2019;20(14). doi:10.3390/ijms20143449
90. van Galen KA, ter Horst KW, Booij J, la Fleur SE, Serlie MJ. The role of central dopamine and serotonin in human obesity: lessons learned from molecular neuroimaging studies. *Metabolism*. 2018;85:325-339. doi:10.1016/j.metabol.2017.09.007
91. Hirata M, Suzuki M, Ishii R, et al. Genetic defect in phospholipase Cδ1 protects mice from obesity by regulating thermogenesis and adipogenesis. *Diabetes*. 2011;60(7):1926-1937. doi:10.2337/db10-1500
92. Wu Y, Wang W, Jiang W, Yao J, Zhang D. An investigation of obesity susceptibility genes in Northern Han Chinese by targeted resequencing. *Med (United States)*. 2017;96(7):1-6. doi:10.1097/MD.0000000000006117
93. Schroeder LE, Leinninger GM. Role of Central Neurotensin in Regulating Feeding: Implications for the Development and Treatment of Body Weight Disorders. *Biochim Biophys Acta*. 2018;176(1):139-148. doi:10.1016/j.physbeh.2017.03.040

94. Suzuki T, Li W, Zhang JP, et al. A novel scaffold protein, TANC, possibly a rat homolog of Drosophila rolling pebbles (rols), forms a multiprotein complex with various postsynaptic density proteins. *Eur J Neurosci*. 2005;21(2):339-350. doi:10.1111/j.1460-9568.2005.03856.x

SUPPLEMENTARY MATERIAL

Supplementary Table 1
List of 164 rCNVs detected in 134 patients with severe early-onset obesity from the 465 studied.

Supplementary Table 2
Summary of 44 rCNVs detected in 34 patients with severe early-onset obesity with segregation studies available.

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
1719	F	Gain	1p36.13	242	chr1:17202355-17444769	0	ATP13A2, CROCC, MFAP2, PADI2, SDHB
OBE54	M	Gain	1p31.3	405	chr1:61704166-62109502	3	NFIA
1781	F	Loss	1p21.2-p21.1	252	chr1:102044634-102296279	3	OLFM3
2080	F	Loss	1q21.1	368	chr1:145394955-145762959	2	15 genes
OBE34	F	Gain	1q21.1	507	chr1:145394955-145901548	0	17 genes
1774	M	Gain	1q23.3	728	chr1:162807273-163535283	0	C1orf110, NUF2, RGS4, RGS5
1863	M	Gain	1q32.1	118	chr1:200662215-200780153	1	CAMSAP2
2197	M	Loss	1q43	134	chr1:240707951-240842076	0	GREM2
OBE67	M	Gain	1q44	458	chr1:247828797-248286356	0	17 genes
1312	F	Gain	2p21	315	chr2:45573241-45887931	2	PRKCE, SRBD1
1926	M	Gain	2p13.2	135	chr2:71589618-71724396	0	DYSF, ZNF638
1974	F	Loss	2p12	101	chr2:77029323-77129922	1	LRRTM4
OBE27	M	Gain	2q24.1	336	chr2:159176303-159512667	0	CCDC148, PKP4
1748	M	Gain	2q24.1-q24.2	304	chr2:159649340-159953466	1	DAPL1, TANC1
2034	F	Gain	2q24.1-q24.2	304	chr2:159649340-159953466	1	DAPL1, TANC1
1380	M	Gain	2q24.1-q24.2	131	chr2:159755842-159886836	1	TANC1
1343	F	Loss	3p26.1	166	chr3:4138813-4304897	1	SUMF1
OBE26	M	Gain	3p26.1	214	chr3:4272253-4486303	2	SETMAR, SUMF1
2120	M	Gain	3p26.1	1120	chr3:5822491-6942072	3	GRM7
2059	M	Loss	3p26.1	118	chr3:6950189-7068147	0	GRM7
1827	M	Gain	3p26.1	364	chr3:7660133-8024019	3	GRM7
1922	M	Gain	3p26.1-p25.3	114	chr3:8631179-8745493	.	SSUH2

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
1966	M	Gain	3p12.3	488	chr3:74348587-74836374	0	CNTN3
1734	F	Loss	3q12.3-q13.11	3590	chr3:101816344-105406145	0	ALCAM, CBLB, ZPLD1
1964	F	Loss	3q24	114	chr3:143241871-143355892	0	SLC9A9
1753	M	Loss	3q26.31	1134	chr3:172263372-173397582	0	ECT2, NCEH1, NLGN1, SPATA16
1770	F	Gain	3q29	1460	chr3:195887205-197346971	0	20 genes
1455	F	Gain	3q29	137	chr3:197564995-197701913	2	IQCG, LMLN, LRCH3, RPL35A
1650	F	Gain	3q29	243	chr3:197594967-197838262	2	IQCG, LMLN, LRCH3, RPL35A
2115	F	Gain	4p15.32	1093	chr4:15849801-16942990	0	CD38, FGFBP1, FGFBP2, LDB2, PROM1, TAPT1
1380	M	Gain	4p15.32	215	chr4:15970349-16185525	0	FGFBP2, PROM1, TAPT1
1417	F	Gain	4p15.2	773	chr4:24512590-25285913	0	CCDC149, DHX15, LGI2, PI4K2B, SEPSECS, SOD3
1850	F	Gain	4q12	361	chr4:53842714-54203701	0	SCFD2
OBE60	M	Gain	4q13.3	101	chr4:71055318-71156301	0	CSN3, FDCSP, ODAM
1673	M	Gain	4q21.21-q21.22	816	chr4:81667695-82484170	0	BMP3, C4orf22, PRKG2, RASGEF1B
1825	F	Gain	4q22.1	119	chr4:89828080-89947462	2	FAM13A
1718	M	Gain	4q31.3	859	chr4:152450126-153309538	0	FAM160A1, FBXW7, PET112
1781	F	Gain	4q35.2	202	chr4:187329245-187531022	2	FAT1, MTNR1A
1719	F	Gain	5p15.33	295	chr5:303686-598289	1	AHRR, C5orf55, EXOC3, PDCD6, SLC9A3
1876	M	Gain	5p15.33	142	chr5:851046-992859	0	BRD9, TRIP13, ZDHHC11
2157	M	Gain	5p14.3	185	chr5:20290034-20474634	1	CDH18
1717	F	Gain	5p13.2	111	chr5:37239240-37349895	1	C5orf42, NUP155
1722	F	Gain	5q12.1	277	chr5:61854650-62131163	3	IPO11, LRRC70

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
2083	F	Gain	5q14.1	237	chr5:80021168-80257798	1	MSH3, RASGRF2
OBE07	M	Loss	5q21.1	157	chr5:101620174-101776835	0	SLCO4C1, SLCO6A1
2177	M	Gain	5q21.1	113	chr5:102459019-102572418	0	PPIP5K2
1651	M	Gain	5q23.2	263	chr5:122198481-122461778	0	PPIC, PRDM6, SNX24
2217	F	Loss	5q23.2	129	chr5:126196217-126325510	0	MARCH3
1742	F	Gain	5q32	127	chr5:146370029-146496726	4	PPP2R2B
1753	M	Gain	5q35.3	397	chr5:179220638-179617799	0	C5orf45, LTC4S, MAML1, MGAT4B, RASGEF1C, RNF130, SQSTM1, TBC1D9B
2083	F	Gain	6p25.1	142	chr6:5256116-5398329	2	FARS2, LYRM4
1836	M	Gain	6p22.2	643	chr6:25862466-26505362	0	42 genes
1748	M	Gain	6p21.2	209	chr6:39846137-40055046	0	DAAM2, MOCS1
1858	M	Gain	6p12.1	981	chr6:54720443-55701367	0	BMP5, FAM83B, GFRAL, HCRTR2, HMGCLL1
1974	F	Gain	6q14.3	1410	chr6:86340237-87749815	0	HTR1E, SYNCRIP
1729	M	Gain	6q15	593	chr6:89200648-89793993	0	PNRC1, RNGTT
2039	M	Gain	6q26	944	chr6:162795058-163739362	2	PACRG, PARK2
1326	M	Gain	7p15.3	202	chr7:21858215-22059791	0	CDCA7L, DNAH11
OBE03	M	Gain	7p15.3	266	chr7:24129888-24395900	0	NPY
1654	F	Gain	7p15.2	228	chr7:26007582-26235905	1	HNRNPA2B1, NFE2L3
OBE72	F	Gain	7p14.1	105	chr7:40117098-40221714	1	CDK13, MPLKIP, SUGCT
2207	F	Gain	7q11.21	208	chr7:63758482-63966166	.	ZNF736
2073	M	Gain	7q22.1	137	chr7:101238449-101375291	1	MYL10
1650	F	Gain	7q31.1	185	chr7:111730967-111915984	3	DOCK4, ZNF277

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
1823	F	Gain	7q32.2	153	chr7:129927778-130080812	0	CEP41, CPA1, CPA2, CPA4, CPA5
1708	F	Gain	8p23.3	156	chr8:1870734-2027154	2	ARHGEF10, KBTBD11, MYOM2
1934	F	Gain	8q22.3	294	chr8:104278419-104572771	0	CTHRC1, DCAF13, FZD6, RIMS2, SLC25A32
1555	M	Gain	8q21.13	133	chr8:82483592-82617029	1	IMPA1, SLC10A5, ZFAND1
2216	M	Gain	8q24.3	487	chr8:142188039-142674928	1	DENND3, GPR20, PTP4A3, SLC45A4
1798	F	Gain	9p24.3	213	chr9:185632-398673	3	C9orf66, CBWD1, DOCK8
1797	F	Gain	9p24.3	448	chr9:396232-844001	2	DMRT1, DOCK8, KANK1
2169	M	Loss	9p22.2-p22.1	1064	chr9:17825216-18888920	2	ADAMTSL1
1443	M	Loss	9q32	109	chr9:115712068-115821485	4	ZFP37
1728	M	Gain	9q34.11	240	chr9:132552206-132792272	0	C9orf78, FNBP1, TOR1A, TOR1B, USP20
1791	F	Gain	9q34.3	115	chr9:138149166-138263731	.	C9orf62
1687	F	Gain	9q34.3	149	chr9:138149166-138298164	.	C9orf62
1966	M	Gain	9q34.3	157	chr9:138149166-138306060	.	C9orf62
OBE23	F	Gain	9q34.3	104	chr9:138180621-138284278	.	C9orf62
1814	F	Gain	10p14	1029	chr10:10743956-11773389	0	CELF2, USP6NL
1734	F	Loss	10p21.3	156	chr10:69418270-69574169	1	CTNNA3, DNAJC12
1683	F	Gain	10q22.1	105	chr10:72435657-72540534	0	ADAMTS14, TBATA
1555	M	Gain	10q24.2	115	chr10:100399366-100514409	0	HPSE2
1758	M	Gain	10q26.3	406	chr10:134987146-135393215	1	15 genes
1689	M	Loss	10q26.3	126	chr10:135252347-135378802	0	CYP2E1, SPRN, SYCE1
2195	F	Loss	11p15.5	124	chr11:2573121-2697595	3	KCNQ1

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
1855	F	Loss	11p15.4	103	chr11:4807054-4910224	1	MMP26, OR51F2, OR51H1P, OR51S1, OR51T1, OR52R1
1749	M	Gain	11p15.1	120	chr11:21560094-21680488	3	NELL1
2130	M	Loss	11p14.3	224	chr11:22077572-22301807	0	ANO5
OBE16	M	Gain	11p13	138	chr11:32977624-33116054	0	CSTF3, DEPDC7, QSER1, TCP11L1
1855	F	Gain	11p12	590	chr11:39772957-40362590	0	LRRC4C
1331	F	Gain	11p11.2	127	chr11:46596674-46723603	0	AMBRA1, ARHGAP1, ATG13, HARBI1, ZNF408
1790	M	Loss	11q11	171	chr11:55424207-55595081	3	OR4C6, OR5D13, OR5D14, OR5D18, OR5L1, OR5L2
1797	F	Gain	11q23.1	103	chr11:110429104-110532348	0	ARHGAP20
OBE25	F	Gain	12p12.3	179	chr12:19394160-19573650	0	AEBP2, PLEKHA5
1852	M	Gain	12p12.3	243	chr12:19405901-19649096	0	AEBP2, PLEKHA5
1750	F	Gain	12p12.3	148	chr12:19447611-19595630	0	AEBP2, PLEKHA5
1856	F	Gain	12p11.23	413	chr12:27355213-27767913	2	ARNTL2, PPFIBP1, SMCO2, STK38L
2085	F	Gain	12q13.11	586	chr12:47378610-47964979	1	AMIGO2, PCED1B
2154	F	Loss	12q13.2	208	chr12:56064079-56271677	0	BLOC1S1, CD63, DNAJC14, GDF11, ITGA7, METTL7B, MMP19, ORMDL2, RDH5, SARNP
OBE25	F	Gain	12q23.3	118	chr12:104473431-104591886	0	HCFC2, NFYB
1608	M	Gain	12q24.33	323	chr12:130391024-130713953	1	FZD10
1789	M	Gain	12q24.33	242	chr12:131596433-131838842	1	GPR133
1784	F	Gain	13q21.1	322	chr13:57585556-57907840	2	PRR20A, PRR20B, PRR20C, PRR20D, PRR20E
2117	M	Loss	14q21.1	198	chr14:38702361-38900700	0	CLEC14A

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
2117	M	Loss	14q21.1-q21.3	8119	chr14:39122277-47241493	0	17 genes
OBE27	M	Loss	14q24.1	213	chr14:68434826-68647631	0	RAD51B
OBE01	M	Gain	14q31.1	105	chr14:81204951-81309536	.	CEP128
1734	F	Gain	14q31.3	262	chr14:89171711-89433411	0	EML5, TTC8
1977	F	Gain	14q32.13	261	chr14:95887911-96149245	1	GLRX5, SYNE3
2205	M	Gain	15q13.1	129	chr15:29889806-30018627	2	TJP1
2065	F	Loss	15q13.3	142	chr15:31094479-32514341	0	CHRNA7, FAN1, KLF13, MTMR10, OTUD7A, TRPM1
1985	M	Gain	15q21.2	298	chr15:51498539-51796806	1	CYP19A1, DMXL2, GLDN
1427	M	Gain	15q25.2	418	chr15:84414592-84832932	1	ADAMTSL3
1793	F	Gain	15q26.3	200	chr15:99612980-99813270	2	LRRC28, SYNM, TTC23
1863	M	Loss	16p12.2	596	chr16:21839340-22435811	3	C16orf52, CDR2, EEF2K, NPIPB4, PDZD9, POLR3E, SDR42E2, UQCRC2, VWA3A
1934	F	Gain	16q24.1-q24.2	154	chr16:87046188-87200035	.	C16orf95
1754	F	Gain	16p13.11	1323	chr16:14968859-16291983	0	14 genes
1572	M	Gain	16p13.11	1111	chr16:15092778-16203345	0	11 genes
1727	F	Gain	16p13.11	1270	chr16:15092778-16363239	0	13 genes
2117	M	Gain	16p13.11	712	chr16:15493046-16205501	3	ABCC1, C16orf45, FOPNL, KIAA0430, MPV17L, MYH11, NDE1
1443	M	Loss	16p11.2	218	chr16:28825605-29043450	2	ATP2A1, ATXN2L, CD19, LAT, NFATC2IP, RABEP2, SH2B1, SPNS1, TUFM
2214	M	Gain	17p13.1	231	chr17:7929777-8161149	0	11 genes
1974	F	Gain	17q12	151	chr17:34323944-34475131	0	CCL14, CCL15, CCL15-CCL14, CCL18, CCL23, CCL3, CCL4

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
2106	M	Gain	17q22-q23.1	561	chr17:57148991-57709814	0	CLTC, DHX40, GDPD1, PRR11, SKA2, SMG8, TRIM37, YPEL2
1759	M	Gain	17q23.1-q23.2	248	chr17:58113570-58361461	0	CA4, HEATR6, USP32
OBE45	F	Gain	18p11.31	520	chr18:6033403-6553040	0	C18orf64, L3MBTL4
2046	M	Gain	18p11.31-p11.23	715	chr18:6881677-7597104	1	ARHGAP28, LAMA1, LRRC30, PTPRM
OBE61	F	Gain	18p11.21	183	chr18:12917703-13100451	0	CEP192, PTPN2, SEH1L
1900	F	Gain	18q21.31	700	chr18:55215573-55915325	0	ATP8B1, FECH, NARS, NEDD4L
1427	M	Gain	19p13.3	416	chr19:2234013-2650034	0	13 genes
2058	M	Gain	19p13.41	131	chr19:52227916-52359169	1	FPR1, FPR2, FPR3, ZNF577
2176	M	Loss	19q13.41	171	chr19:53243026-53413825	2	ZNF28, ZNF320, ZNF468, ZNF600
1337	F	Gain	19q13.42	119	chr19:54029007-54148045	1	DPRX, ZNF331
1534	M	Gain	20p12.3-p12.2	648	chr20:8724104-9371825	0	PLCB1, PLCB4
1790	M	Gain	20p12.1	277	chr20:13199299-13476090	0	ISM1, TASP1
2189	M	Gain	20p12.1	314	chr20:13582476-13896499	0	ESF1, NDUFAF5, SEL1L2, TASP1
OBE01	M	Gain	20p12.1	204	chr20:13867165-14070869	0	MACROD2, SEL1L2
1586	F	Gain	21q21.3	163	chr21:31280603-31443375	1	GRIK1
1876	M	Gain	22q11.21	136	chr22:18504801-18640300	0	MICAL3, PEX26, TUBA8, USP18
OBE42	F	Loss	22q11.22	259	chr22:22314463-22573637	2	TOP3B
1672	M	Loss	22q11.22	255	chr22:22314463-22569035	2	TOP3B
OBE53	F	Loss	22q11.2	660	chr22:22998050-23657613	0	BCR, GNAZ, IGLL5, RAB36, RTDR1
1741	M	Gain	22q11.2	1267	chr22:23729698-24996630	2	28 genes
1454	M	Gain	22q11.2	305	chr22:24333865-24639069	2	CABIN1, GGT5, GSTT1, SUSD2
1464	F	Gain	22q13.32-q13.33	604	chr22:49396413-50000058	0	C22orf34

Sample	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Freq. Controls	Coding Genes
OBE21	M	Loss	22q13.33	122	chr22:50342728-50464496	0	IL17REL, PIM3, TTLL8
1946	F	Gain	22q13.33	114	chr22:50468933-50582626	0	MLC1, MOV10L1, TTLL8
1672	M	Gain	Xp22.31	1678	chrX:6456940-8135053	3	HDHD1, PNPLA4, STS, VCX
2040	M	Gain	Xp22.31	1678	chrX:6456940-8135053	3	HDHD1, PNPLA4, STS, VCX
1417	F	Gain	Xp22.31	1092	chrX:6507158-8158977	3	HDHD1, STS
OBE42	F	Gain	Xp22.31	978	chrX:6457128-8135239	3	PNPLA4, STS, VCX
1520	F	Gain	Xp22.31	618	chrX:7517325-8135053	4	PNPLA4, VCX
OBE66	F	Gain	Xp22.2	309	chrX:15952591-16261676	1	GRPR, MAGEB17
1450	F	Gain	Xp21.3	199	chrX:28608136-28807442	0	IL1RAPL1
2145	F	Gain	Xq12	134	chrX:65813960-65947838	.	EDA2R
663	F	Gain	Xq22.1	136	chrX:100805631-100941810	0	ARMCX1, ARMCX2, ARMCX3, ARMCX6
1632	M	Gain	Xq26.1	450	chrX:129680222-130130267	0	ENOX2
OBE69	F	Gain	Xq26.2	621	chrX:130639755-131261195	0	FRMD7, OR13H1
1689	M	Gain	Xq26.2-q26.3	782	chrX:133561242-134343016	0	11 genes
1940	M	Gain	Xq26.3	309	chrX:134307006-134786450	.	DDX26B, ZNF449, ZNF75D
1826	M	Loss	Xq27.2	431	chrX:140348507-140779598	1	SPANXA1, SPANXA2
1492	M	Gain	Xq27.2	431	chrX:140348507-140779598	0	SPANXA1, SPANXA2
1837	F	Gain	Xq27.2	138	chrX:140694098-140832281	.	SPANXD
1512	F	Gain	Xq27.2	138	chrX:140694098-140832281	.	SPANXD

Supplementary Table 1. The 164 rCNVs detected in 134 patients with severe early-onset obesity from the 465 studied. Control frequency (Freq. Controls) refers to the frequency of the same type of rearrangement, gain or loss, in a control dataset of 9,820 subjects. Coding genes included in each alteration are listed (if there are more than 10, only the number is shown). F: female; M: male.

Case	Gender	Gain/Loss	Region	Inheritance	Age at exam (years)	BMI (SDS)	Height (SDS)	Target height (SDS)	Bone age increase (months)	Newborn weight (SDS)	ID or BD	HP	Metabolic disturbances	Progenitor phenotype
1719	F	Gain	1p36.13	Mat	3.83	+6.38	+1.37	-0.66	+7	-0.27	NR	Yes	HU	N
OBE54	M	Gain	1p31.3	Mat	7.83	+3.79	+1.36	-0.12	-12	+0.89	NR	Yes	None	N
1863	M	Gain	1q32.1	*De novo*	8.16	+4.9	0		+28	-0.6	Mild	Yes	IR	N
OBE27	M	Gain	2q24.1	Mat	8.25	+4.7	+0.92	-1.04	+30	+0.8	NR	Yes	IR	N
OBE26	M	Gain	3p26.1	Pat	5.42	+6.2	-0.24		0	-0.2	NR	Yes	None	N
1827	M	Gain	3p26.1	Mat	0.58	+3.2	-0.22	-0.18	+2	+0.74	NR	Yes	HDL	N
1734	F	Loss	3q12.3-q13.11	*De novo*	14.08	+4.35	+1.24	-0.43	+11	+1.26	NR	Yes	Low HDL	.
1753	M	Loss	3q26.31	*De novo*	10.66	+4.7	+0.58	-0.26	+28	-1.02	Moderate	Yes	IR + HU	.
1455	F	Gain	3q29	Mat	15.08	+6.5	+0.36	-0.39	NA	+0.57	NR	Yes	IR + IGT + low HDL+ high LDL	OB
1850	F	Gain	4q12	Pat	13.33	+0.43	+3.3	+0.57	0	+1.11	Moderate	Yes	IR + IGT + low HDL + HU	OB
OBE60	M	Gain	4q13.3	Mat	11.00	+2.33	+0.91	-0.89	0	+1.21	NR	No	None	N
1719	F	Gain	5p15.33	Mat	3.83	+6.38	+1.37	-0.66	+7	-0.27	NR	Yes	HU	N
OBE07	M	Loss	5q21.1	Pat	8.66	+2.85	+0.93	+2.44	+15	+0.93	NR	Yes	None	OB
1753	M	Gain	5q35.3	Pat	10.66	+4.7	+0.58	-0.26	+28	-1.02	Moderate	Yes	IR + HU	N
1729	M	Gain	6q15	Mat	14.33	+4.4	+0.93	-1.34	+36	+1.13	NR	No	IR + low HDL + HU	N
1326	M	Gain	7p15.3	Pat	12.25	+4.15	+1.39	-0.43	+3	+0.9	Mild	Yes	None	N
OBE03	M	Gain	7p15.3	Mat	3.00	+4.9	-0.97	-1.95	+6	-1.53	Mild	Yes	IR + low HDL	OB
OBE72	F	Gain	7p14.1	Pat	17.25	+8.76	-0.93	-1.23	NA	-0.22	NR	Yes	HT	OB

Case	Gender	Gain/Loss	Region	Inheritance	Age at exam (years)	BMI (SDS)	Height (SDS)	Target height (SDS)	Bone age increase (months)	Newborn weight (SDS)	ID or BD	HP	Metabolic disturbances	Progenitor phenotype
1797	F	Gain	9p24.3	Mat#	11.16	+6.1	+2.0	-0.41	+17	+0.1	NR	Yes	IR + HU	N
1734	F	Loss	10p21.3	Mat	14.08	+4.35	+1.24	-0.43	+11	+1.26	NR	Yes	Low HDL	OB
1749	M	Gain	11p15.1	Pat#	8.66	+4.3	+0.7	-0.01	+12	+1.16	NR	Yes	IR + IGT	OB
OBE16	M	Gain	11p13	Pat	10.02	+2.33	+1.5	+0.48	+6	+2.06	NR	Yes	IFG	OB
1790	M	Loss	11q11	Mat	6.58	+8.68	+2.3	-0.56	+23	+0.8	Severe	Yes	None	N
1797	F	Gain	11q23.1	Mat	11.16	+6.1	+2.0	-0.41	+17	+0.1	NR	Yes	IR + HU	N
OBE25	F	Gain	12q23.3	Pat	9.00	+4.66	+1.1	NA	0	-0.62	NR	No	IR + HU	N
1789	M	Gain	12q24.33	Pat	8.25	+5.5	+2.45	+0.56	+12	-0.8	NR	Yes	IR	N
2117	M	Loss	14q21.1-q21.3 *	Mat	10.5	+5.35	-0.95	-0.67	+12	-0.38	Mild	Yes	None	OB
OBE27	M	Loss	14q24.1	Pat	8.25	+4.7	+0.92	-1.04	+30	+0.8	NR	Yes	IR	OB
OBE01	M	Gain	14q31.1	Mat	11.00	+3.59	+0.46	-0.7	+18	+1.51	NR	Yes	IR + low HDL + HU	N
1734	F	Gain	14q31.3	Mat	14.08	+4.35	+1.24	-0.43	+11	+1.26	NR	Yes	Low HDL	OB
1427	M	Gain	15q25.2	Pat	6.16	+6.58	+1.78	-0.65	-7	+0.42	NR	Yes	High LDL + HU	OB
1863	M	Loss	16p12.2	Pat	8.16	+4.9	0		+28	-0.6	Mild	Yes	IR	OB
2214	M	Gain	17p13.1	Pat	6.00	+3.9	+0.3	-0.76	-6	-0.4	NR	Yes	None	N
OBE45	F	Gain	18p11.31	Mat	9.16	+4.11	+0.19	-1.08	-6	-1.02	NR	Yes	IR + HU	OB
OBE61	F	Gain	18p11.21	Mat	4.83	+8.46	+1.38	-0.5	+14	-0.45	NR	Yes	HT + low HDL + high LDL + HU	N
1900	F	Gain	18q21.31	Pat	9.16	+3.74	+2.42	-0.19	+34	-0.48	NR	No	IR + IGT	N
1427	M	Gain	19p13.3	Pat	6.16	+6.58	+1.78	-0.65	-7	+0.42	NR	Yes	High LDL + HU	OB

Case	Gender	Gain/Loss	Region	Inheritance	Age at exam (years)	BMI (SDS)	Height (SDS)	Target height (SDS)	Bone age increase (months)	Newborn weight (SDS)	ID or BD	HP	Metabolic disturbances	Progenitor phenotype
1790	M	Gain	20p12.1	Pat	6.58	+8.68	+2.3	-0.56	+23	+0.8	Severe	Yes	None	OB
OBE01	M	Gain	20p12.1	Pat	11	+3.59	+0.46	-0.7	+18	+1.51	NR	Yes	IR + low HDL + HU	OB
1586	F	Gain	21q21.3	Pat	14.00	+5.16	-0.07	-0.66	+6	+0.45	NR	No	IGT + IR	N
OBE42	F	Loss	22q11.22	Pat	8.93	+3.07	+1.7	+1.9	+12	+0.72	NR	Yes	Low HDL	OB
OBE21	M	Gain	22q13.33	Pat	6.41	+3.28	+1.72	-1.03	+13	+0.42	Mild	Yes	None	N
OBE66	F	Gain	Xp22.2	Mat	1.50	+3.21	-0.65	-0.03	-3	-0.48	NR	Yes	None	N
OBE69	F	Gain	Xq26.2	Pat	9.16	+6.56	+0.44	+1.39	+15	-2.41	Moderate	Yes	IR	OB

Supplementary Table 2. Summary of 44 rCNVs detected in 34 patients with severe early-onset obesity with segregation studies available. Genotypic and phenotypic data from the patients is shown. Progenitor phenotype refers only to the progenitor carrying the alteration (familial obesity background not shown). F: female; M: male; Mat: maternal; Pat: paternal; *:presence of two alterations from the same genomic region ; #: sibling with obesity carrying the same alteration; NA: not available; ID: intellectual disability; BD: behavioral disturbances; NR: nothing relevant; HP: hyperphagia; IR: insulin resistance; IGT: impaired glucose tolerance; HT: hypertriglyceridemia; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; HU: hyperuricemia; N: normal; OB: obesity.

CHAPTER 3

Exome sequencing in severe early-onset obesity

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

In preparation

When trying to elucidate the genetic causes of complex disorders, exome sequencing may be a valuable option in severe cases from selected families.

This work was designed to explore novel genetic alterations behind the typical monogenic obesity forms. Two families with a clear pattern of inheritance among severe obesity members and no genetic factors explaining it, were selected for exome sequencing.

Exome sequencing in severe early-onset obesity

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

ABSTRACT

Exome sequencing has been very efficient when unraveling the genetic causes of mendelian diseases, but it can also be used in specific subgroups of complex cases that may be monogenic. Severe early-onset obesity cases may represent an extreme phenotype mainly explained by highly penetrant genetic alteration, so exome sequencing may be effective to explore new genetic factors related with the phenotype.

The exome of 3 patients from 2 families with cases of severe early-onset obesity was sequenced, and candidate genes and variants were analyzed.

Four selected polymorphisms acting as susceptibility factors and four rare variants acting as strong candidates were identified. With the current data it was very difficult to establish a causal relationship between the variants identified and the phenotype in the families.

Our findings revealed the complexity to analyze this type of families even they were conscientiously selected. Genetic approaches including functional studies and additional families will be crucial to define the pathological effects of the identified variants and prove its real contribution into the obesity phenotype. However, complex disorders may still be challenging to characterize at a genetic level.

INTRODUCTION

Exome sequencing has been widely used to unravel the genetic causes of several mendelian diseases [1,2]. The ability to select phenotypically defined patients has been crucial to identify the genes responsible for these monogenic forms. However, complex diseases are characterized by a great diversity at the phenotypic and genotypic levels. Most of these complex cases are difficult to tackle using exome sequencing, but specific subgroups may be monogenic,

and thus affordable with this technique. Patients with extreme phenotypes might be the ones with a higher probability to carry a highly penetrant genetic alteration. Consequently, extreme cases are a good target for exome sequencing.

In the case of obesity, which is considered a multifactorial and genetically heterogeneous diseases, not all cases have been genetically characterized. Direct sequencing of genes related with the hypothalamic pathway of energy balance was effective to identify the first causal mutations in severe obesity patients [3]. This approach has been widely used in monogenic obesity forms, but it may only explain up to 5-10% of cases [4]. The current knowledge of the genetics of obesity includes genes mainly from the leptin-melanocortin system and related with adipogenesis [5]. However, there may be a fraction of monogenic forms that will be recurrently missed by sequencing panels of known obesity genes. To fully understand the molecular mechanisms involved in monogenic obesity forms not described yet, exome sequencing of extreme cases may be a valuable tool.

Here, we used exome sequencing in two families with cases of severe early-onset obesity to explore new genetic factors related with the phenotype.

MATERIALS AND METHODS

Subjects

Two families were selected for exome sequencing. Inclusion criteria were families with a clear inheritance pattern of obesity and patients affected by severe early-onset obesity with no candidate genetic alterations explaining the phenotype. Patients were explored to exclude other phenotypic alterations and were studied by aCGH. Moreover, known causes of obesity were discarded, such as syndromic forms or *MC4R* alterations. A specific written informed consent was elaborated for both families.

Family A

A.III3 is a female patient diagnosed with early-onset obesity. She was hospitalized at the age of 3 for respiratory distress and followed-up since then for severe obesity. Her sister A.III2 had a similar phenotype and together with her brother A.III1 were severely obese. Her mother was also affected by severe obesity and several cases were reported in the maternal family. No

antecedents were reported in the paternal family, except for the father with a milder phenotype (Figure 1A).

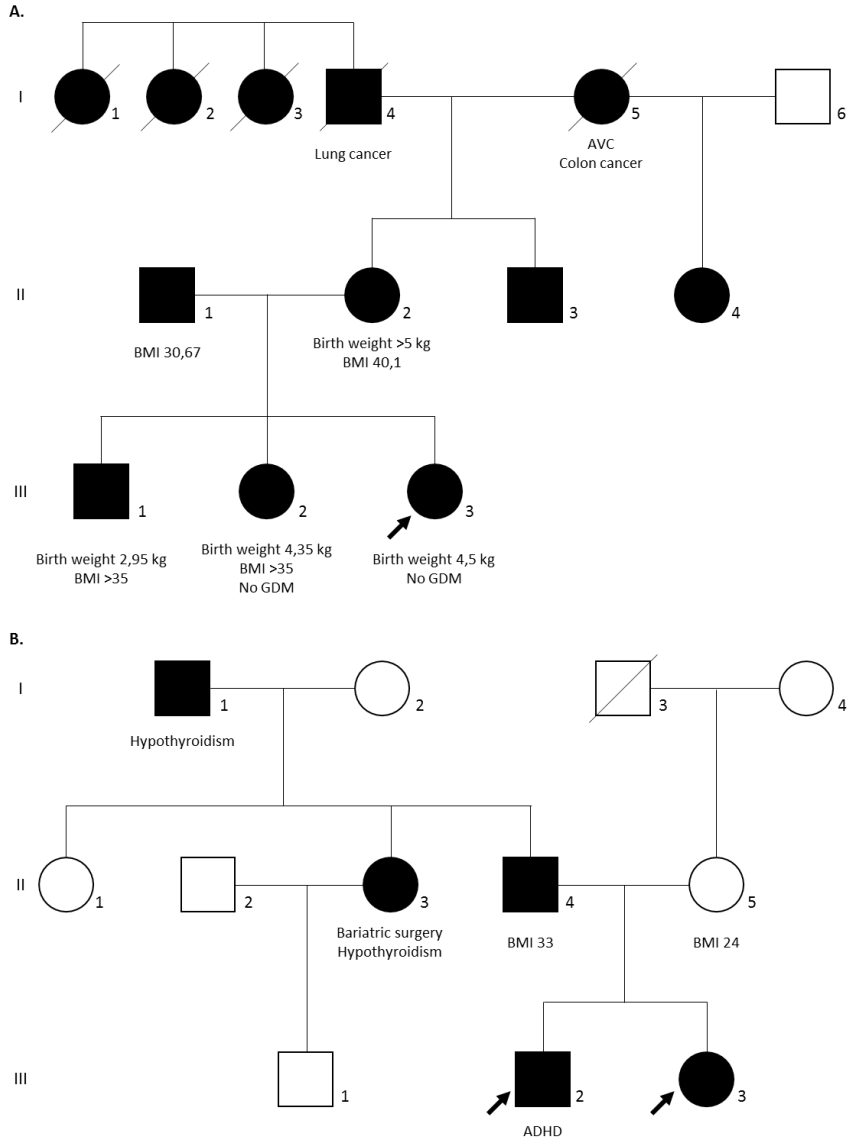

Figure 1. Pedigrees from both studied families. **A**: Family A. **B**: Family B. Black figures represent subjects with obesity (BMI > 30). Arrows indicate the patients whose exome has been sequenced. AVC: Aortic valve calcification. GDM: Gestational diabetes mellitus. ADHD: Attention deficit hyperactivity disorder.

Family B

B.III2 and B.III3 are siblings with severe obesity reported before 3 years of age. Both obese patients presented hyperphagia and insulin resistance with dyslipidaemia. Obesity was reported in several relatives from the paternal branch, including an aunt with bariatric surgery (Figure 1B).

Exome sequencing

Exome sequencing was performed in three obese patients (one from family A and two from family B). DNA was isolated with Gentra Puregene Blood kit (Qiagen) after collecting blood samples from patients and parents from both families, and additionally in family A, two siblings and a half-aunt from the maternal side. 3ug of DNA were shipped to qGenomics were exome sequencing was performed. The capture was done with the kit "SureSelect Human All Exon V5, 51 Mb" (Agilent) and the sequencing reactions were performed with HiSeq (Illumina). Reads were mapped to the reference genome with BWA-MEM software [6], BAM files were generated with Picard toolkit and the variant calling was performed with GATK software [7].

Variants filtering and prioritization

Variants with a coverage >20 and a genotype quality >20 we considered real variants and the rest were excluded as false positives. Only variants from exonic or splicing regions were analyzed.

Polymorphisms affecting candidate genes for obesity may be acting as susceptibility factors contributing to obesity development. We explored common variants affecting 15 obesity genes (*LEP*, *LEPR*, *MC4R*, *POMC*, *PCSK1*, *MC3R*, *BDNF*, *NTRK2*, *PPARG*, *SIM1*, *ADRB3*, *PCSK2*, *NPY*, *NPY1R* and *AGRP*). Polymorphisms associated to BMI were selected.

As obesity is a multifactorial and heterogeneous disorder, causal genetic variants were expected to have a low frequency and a high penetrance. Moreover, the extreme phenotype from patients may not be explained by common variants found in the general population, so these frequent variants were excluded. Rare variants with frequencies lower than 1/500 in different datasets available were selected.

Once the population variants were excluded, we select variants compatible with the recessive model and variants compatible with the dominant model. Despite the clear inheritance pattern affecting only one parental family,

homozygous and compound heterozygous variants (recessive model) were explored in all patients as may explain part of the phenotype observed. However, the best model fitting the pedigrees for both families was the dominant model were one heterozygous high-penetrant variant may explain the phenotype among carriers. Genetic variants are expected to be heterogenic, but these alterations should be very rare as the extreme obesity phenotype is not common in the general population. For this reason, we selected variants with frequencies lower than 1/100000 in different datasets available.

We used the IGV browser to manually inspect variants and discard false positives when there were no clear reads to call the alteration or discard compound heterozygotes when the two variants were present in the same read.

One first manual curation was performed to select candidate genes. The type of variant identified and the physiological function from each gene were used for the prioritization of the results. Stopgain, stoploss, frameshift indels and canonical splicing variants were considered more likely to be disease causative, whereas synonymous variants were not considered. Reported function from each gene was used to assess possible links with obesity. A set of candidate alterations was selected and studied for segregation in each family.

In family A, due to the pedigree structure, variants inherited from the mother were considered as strong candidates. In family B, paternally inherited variants were prioritized. *De novo* variants were also explored.

Finally, a last manual curation was performed to define the strongest candidates to be associated with the phenotype. Variants associated with other diseases not related with obesity were discarded and the predicted pathogenicity of selected alterations was used to identify probably pathogenic variants (CADD score over 20).

An overview of the steps of the prioritization can be found in Figure 2.

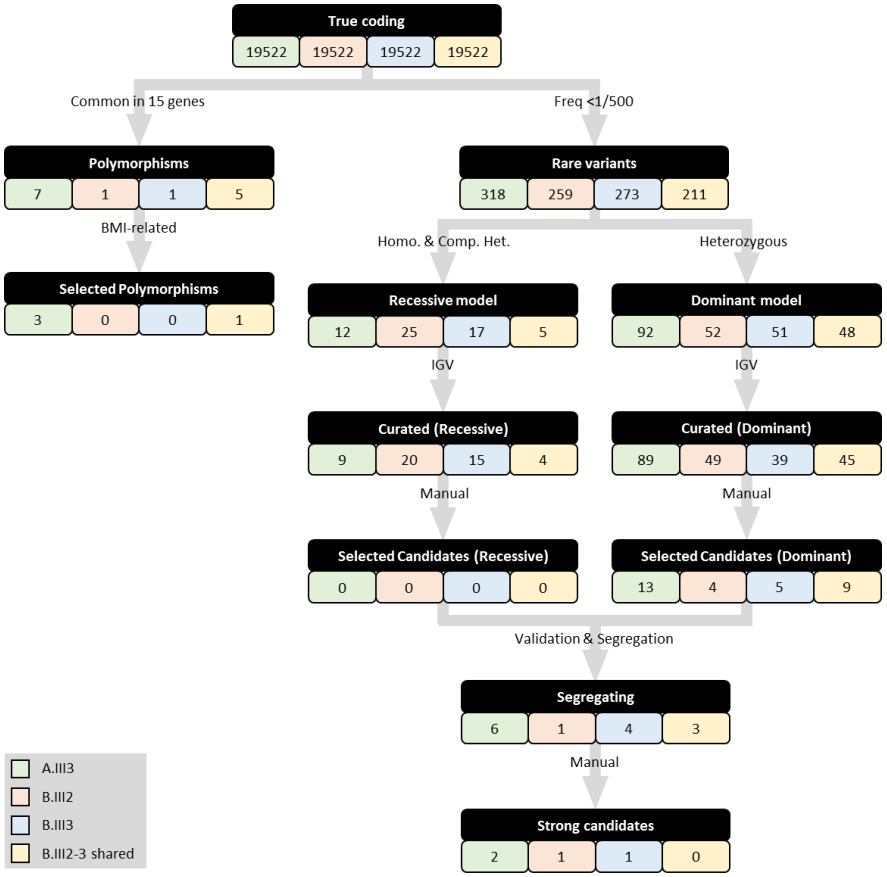

Figure 2. Pipeline followed for the prioritization of the variants identified by exome sequencing. The number of genetic variants identified in each step of the prioritization are shown for the individual studied: patient III3 from family A (A.III3, green), patient III2 from family B (B.III2, orange), patient III3 from family B (B.III3, blue) and shared between both patients from family B (B.III2-3 shared, yellow). Variants from B.III2-3 shared are not included in B.III2 and B.III3. Consult the Materials and Methods section to understand the criteria used in each step. Numbers from the recessive model correspond to genes, as multiple variants are included in each one.Before filtering for frequency, we studied common variants in a set of 15 obesity genes, and 3 polymorphisms linked with BMI or related traits were detected in family A and 1 in family B.

Validation and segregation studies

Validation of the selected variants was done with Sanger sequencing of PCR amplicons. Familiar samples were used to establish segregation patterns. In family A, we sequenced 6 individuals (patient A.III3, sister A.III2, brother A.III1, father A.II1, mother A.II2 and half-aunt A.II4) and in family B we sequenced 4 individuals (patient B.III2, patient B.III3, father B.II4 and mother B.II5).

RESULTS

Exonic variants

We identified 19522 true coding variants in patient A.III3, 20545 variants in patient B.III3 and 21604 variants in patient B.III2. The siblings B.III3 and B.III2 shared 16194 variants, so the remaining 4351 and 5410 variants were exclusive from each sibling. The number of variants from each step of the prioritization can be found in Figure 2.

Population frequency from reported databases allowed us identifying variants not common in the general population. Several genes harboring variants compatible with the recessive model were found, but special attention was paid to variants compatible with the dominant model.

IGV inspection and manual curation allowed us to discard some variants and elaborate a list of candidate alterations to be studied for segregation in the families.

After performing the validation and segregation studies, the last manual curation lead to 2 variants selected as strong candidates in family A and 3 variants in family B.

The identified polymorphisms and the strong candidate variants selected can be found in Table 1.

Fam	Carriers	Variant	Ref>Alt	Gene	Change	CADD	Frequency	Model
A	II2, III3	chr9:135374920	G>C	*CFAP77*	D189H	24,3	7,96E-06	Dom.
A	II2, III2, III3	chr5:176026756	A>G	*GPRIN1*	F27S	22,7	0	Dom.
A	II2, III3	chr6:100868721	G>A	*SIM1*	A371V	19,9	0,1737	Add.
A	II2, III3	chr6:100868779	G>T	*SIM1*	P352T	23,9	0,1736	Add.
A	II1, II2*, II3, III1*, III2, III3	chr1:66058513	A>G	*LEPR*	Q223R	13,2	0,5089	Add.
B	II4, III2	chr17:38856473	C>T	*KRT24*	1017+1G>A	25,6	3,98E-05	Dom.
B	III3	chr11:27679691	A>C	*BDNF*	C141G	24,0	0	Dom.
B	II4, II5, III2, III3	chr1:66075952	G>C	*LEPR*	K656N	6,7	0,1584	Add.

Table 1. Selected variants identified by exome sequencing. Carriers of each family are reported in accordance with the pedigrees in Figure 1. Carriers in bold correspond to the patients whose exome was sequenced. * represent individuals with homozygous variants. CADD was used as a predicted pathogenicity score, where 20 or higher scores were considered as highly pathogenic variants. Frequency from gnomAD was also reported. The genetic models in which the variants fitted the best were the dominant (Dom.) and the additive (Add.) models.

Family A

Family A was characterized by severe obesity phenotypes in patient A.III3 and her mother A.II2, so maternally-inherited variants in the patient were selected. Patient and mother shared a nonsynonymous variant in gene *CFAP77* (D153H) and two polymorphisms in linkage disequilibrium in gene *SIM1* (rs3734354 and rs3734355). Moreover, her sister and her brother were also presenting a similar phenotype, so shared variants among siblings inherited from the mother were also explored. Both sisters (A.III2 and A.III3) had a maternally-inherited variant in *GPRIN1* gene (F27S). No candidate genetic variants shared among the 3 siblings were identified. The two rare variants (*CFAP77* and *GPRIN1*) were predicted to be pathogenic and were reported in any or few individuals from the general population. Additionally, all family members studied (patient, sister, brother, father, mother and half-aunt) shared at least one G allele of a polymorphism in the *LEPR* gene (rs1137101).

Family B

Family B had two affected siblings and some cases in the paternal side, so paternally-inherited variants were the most likely explaining the phenotype in the family. Variants shared among both siblings inherited from the father were explored, but no candidates were identified. Both siblings were carriers of heterozygous variants associated with recessive diseases which may not be linked with the obesity phenotype. Instead, we identified different variants in each sibling. Patient B.III2 had a canonical splicing variant affecting the gene *KRT24* (1017+1G>A) inherited from the father and patient B.III3 was carrier of a *de novo* mutation affecting the *BDNF* (C141G) gene. Additionally, all family members studied (patients, father and mother) shared a polymorphism in gene *LEPRK* (rs1805094).

DISCUSSION

The use of exome sequencing is challenging in complex disorders, such as obesity, but can be used to genetically characterize patients with extreme phenotypes. These cases are more likely to be monogenic and exome sequencing can reveal novel obesity related genes. Familiar cases of severe obesity may indicate the presence of a shared variant explaining the phenotype. Here we used exome sequencing to sequence patients affected by severe early-onset obesity and with clear antecedents in the family. In the two families

studied, we have analyzed rare genetic variants shared between obese individuals and polymorphisms conferring susceptibility for increased body fat.

The recessive model should include inherited variants from both progenitors, so it may indicate that this model was not the best explanation for the phenotype, as families presented a clear inheritance pattern affecting only one parental branch. However, the dominant model was more likely to explain pathological effects of heterozygous variants found in the families.

Family A had cases of severe obesity in the maternal branch, so maternally-inherited variants were selected. The most severely affected individuals (the patient A.III3 and her mother A.II2) were carriers of a nonsynonymous heterozygous variant in gene *CFAP77*. *CFAP77* encodes a protein associated with cilia and flagella. Impaired cilia function has been associated with syndromic and nonsyndromic obesity [8], so alterations in *CFAP77* gene may be linked with the development of obesity. Despite the mutation is predicted as pathogenic, the expression of the gene has been reported mainly in testis and lungs and few bibliography was found to relate the variant with the phenotype in the family. Although this variant alone could be contributing to explain the severe obesity described in the patient and her mother, there may be other alterations contributing to the phenotype of the family. Indeed, her sister, who had a similar phenotype, was carrier a rare nonsynonymous heterozygous variant affecting *GPRIN1* gene. This alteration shared between the three female individuals (patient, sister and mother) was predicted as pathogenic and has not been described in the general population. *GPRIN1* encodes an inducer of neurite outgrowth, which is required for maturation of neurons, and its expression is activated by *MECP2* [9]. Despite *MECP2* mutations have been mainly associated with Rett syndrome, *MECP2* carriers may develop an Angelman-like phenotype with no features of Rett syndrome [10]. Obesity in *MECP2* defficient patients may be explained by neuronal alterations due to the lack of activation of *GPRIN1*. This finding may establish a causal link between *MECP2*, *GPRIN1* and obesity, but further cases will be needed to prove this relationship. Despite the variants found in females, no shared variants among the 3 siblings were identified as candiadtes.

In addition to the two rare variants inherited from the mother, there were multiple carriers of genetic polymorphisms that could have a role in the phenotype from the family. The patient A.III3 and her mother A.II2 were carriers of two polymorphisms in the *SIM1* gene (rs3734354 and rs3734355)

transmitted together. Both polymorphisms have been associated with Prader-Willi-like features [11] and increased BMI [12,13] and could explain part of the most severe phenotype observed in those individuals. Moreover, a polymorphism in the *LEPR* gene (rs1137101) was detected in the 6 samples studies from family A. All individuals were carrying at least one G allele, but the mother A.II2 and the older brother A.III1 were found to be homozygotes for the risk allele. This polymorphism has been associated with increased fat and bone mass [14,15]. The presence of this allele in the family could be explaining part of the severe phenotype of the brother, in which no variants compatible with the dominant model were identified.

Family B was characterized by two patients affected by extreme obesity and additional cases of obesity exclusive from the paternal branch. Despite both siblings were affected by severe early-onset obesity, no candidate paternally-inherited variants were identified in both siblings, so the phenotype may be explained by non-shared genetic alterations or other factors. On the one side, the boy was carrier of a rare heterozygous splicing variant in the *KRT24* gene inherited from the father. The mutation affected the splice donor site of intron 4 and may generate an aberrant mature transcript. Any relationship has been reported between keratin and obesity, so the effects of this variant are unknown. On the other side, the girl was carrier of a *de novo* variant affecting the *BDNF* gene. *BDNF* encodes a nerve growth factor that promotes neuronal survival in the brain and has been reported to participate in appetite regulation downstream of the melanocortin receptors. Alterations in *BDNF* have been described in patients with obesity and the WAGRO syndrome [16–18]. The alteration found only in the patient was predicted as pathogenic and has not been reported before. Despite this variant is not explaining the additional familiar cases of obesity, it may undoubtedly be contributing to the severe phenotype observed in the patient.

As in the other family, family B members were carriers of polymorphisms that could be explaining some cases of intermediate phenotypes. The 4 members studied were carriers of a polymorphism in the *LEPR* gene (rs1805094) that has been associated with impaired energy metabolism [19,20]. Probably this alteration was not the main responsible for the severe phenotype of both siblings, but may act as a susceptibility factor to promote an increased fat mas in the family.

The identified rare variants in the two families were affecting genes involved in neuronal development (*GPRIN1*, *BDNF*), ciliogenesis (*CFAP77*) and

intermediate filaments (*KRT24*). The only gene previously associated with obesity was the *BDNF* gene, but the other 3 genes have not been directly linked with obesity. The presence of shared heterozygous variants in the families may be indicative of a relevant role of these genes in the regulation of body weight, but with the current data no conclusive results could been established. Indeed, it was very difficult to establish a causal relationship in the families between the variants identified and the phenotype.

The families were selected with a clear mendelian pattern to maximize the chances of finding highly penetrant mutations, but the reality was far from expected. We have shown the complexity to analyze this type of families. First, by the pedigree itself. The study of a frequent phenotype such as obesity, makes impossible to establish the expected carriers of the pathogenic variant. Obesity observed in the families may be caused by genetic factors, but other confounding factors may also be taking part. And second, by the possible presence of phenocopies. In family B one sibling with severe obesity had a *de novo* mutation but the paternal aunt had needed a bariatric surgery. The shared extreme phenotype was not explained by this genetic factor, so other factors, such as environment, may be contributing in the family. In conclusion, the study of these cases proved the difficulty to use exome sequencing in a multifactorial disease.

It is of note, the complexity to define the exact contribution of the alterations detected with data from single families. Further functional studies will be crucial to understand the pathological effects of the variants identified and additional families with alterations affecting the same genes would be needed to prove its real contribution into the obesity phenotype. However, it will be still challenging to genetically characterize patients affected by complex disorders. Despite big efforts are being made to define the genetic etiology of obesity, we are still far for completely understanding its molecular basis and studies as the one presented here are a valuable approach to prove the complexity to deal with these complex cases.

REFERENCES

1. Hoischen A, van Bon B, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet.* 2010;42(6):483-485.
2. Gilissen C, Hoischen A, Brunner HG, Veltman J a. Unlocking

Mendelian disease using exome sequencing. *Genome Biol.* 2011;12(9):228. doi:10.1186/gb-2011-12-9-228

3. Friedman JM. Obesity in the new millennium. *Nature*. 2000;404:632-634.
4. Ranadive SA, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am*. 2008;37(3):733-751. doi:10.1016/j.ecl.2008.07.003
5. da Fonseca ACP, Mastronardi C, Johar A, Arcos-Burgos M, Paz-Filho G. Genetics of non-syndromic childhood obesity and the use of high-throughput DNA sequencing technologies. *J Diabetes Complications*. 2017;31(10):1549-1561. doi:10.1016/j.jdiacomp.2017.04.026
6. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-1760. doi:10.1093/bioinformatics/btp324
7. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2009;20:254-260. doi:10.1101/gr.107524.110.20
8. Mariman ECM, Vink RG, Roumans NJT, et al. The cilium: a cellular antenna with an influence on obesity risk. *Br J Nutr*. 2016;116(4):576-592. doi:10.1017/s0007114516002282
9. Chahrour M, Yun Jung S, Shaw C, et al. MeCP2, a Key Contributor to Neurological Disease, Activates and Represses Represses Transcription. *Science (80-)*. 2008;320:1224-1229.
10. Watson P, Black G, Ramsden S, et al. Angelman syndrome phenotype associated with mutations in MECP2, a gene encoding a methyl CpG binding protein. *J Med Genet*. 2001;38(4):224-228. doi:10.1136/jmg.38.4.224
11. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-424. doi:10.1038/gim.2015.30
12. Swarbrick MM, Evans DS, Valle MI, et al. Replication and extension of association between common genetic variants in SIM1 and human adiposity. *Obesity*. 2011;19(12):2394-2403. doi:10.1038/oby.2011.79
13. Windholz J, Kovacs P, Schlicke M, et al. Copy number variations in "classical" obesity candidate genes are not frequently associated with severe early-onset obesity in children. *J Pediatr Endocrinol Metab*.

2017;30(5):507-515. doi:https://doi.org/10.1515/jpem-2016-0435

14. Quinton ND, Lee AJ, Ross RJM, Eastell R, Blakemore AIF. A single nucleotide polymorphism (SNP) in the leptin receptor is associated with BMI, fat mass and leptin levels in postmenopausal Caucasian women. *Hum Genet*. 2001;108:233-236. doi:10.1007/s004390100468
15. Richert L, Chevalley T, Manen D, Bonjour JP, Rizzoli R, Ferrari S. Bone mass in prepubertal boys is associated with a Gln223Arg amino acid substitution in the leptin receptor. *J Clin Endocrinol Metab*. 2007;92(11):4380-4386. doi:10.1210/jc.2007-0932
16. Han JC, Liu Q-R, Jones M, et al. Brain-derived neurotrophic factor and obesity in the WAGR syndrome. *N Engl J Med*. 2008;359(9):918-927. doi:10.1056/NEJMoa0801119
17. Gray J, Yeo GSH, Cox JJ, et al. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*. 2006;55(12):3366-3371. doi:10.2337/db06-0550
18. Jouan L, Gauthier J, Dion PA, Rouleau GA. Rare variants in complex traits: Novel identification strategies and the role of de novo mutations. *Hum Hered*. 2013;74(3-4):215-225. doi:10.1159/000346478
19. Wauters M, Mertens I, Rankinen T, Chagnon M, Bouchard C, Van Gaal L. Leptin Receptor Gene Polymorphisms Are Associated with Insulin in Obese Women with Impaired Glucose Tolerance *. *J Clin Endocrinol Metab*. 2001;86(7):3227-3232.
20. Wauters M, Considine R V., Chagnon M, et al. Leptin levels, leptin receptor gene polymorphisms, and energy metabolism in women. *Obes Res*. 2002;10(5):394-400. doi:10.1038/oby.2002.54

CHAPTER 4

Maternal transmission bias of inherited genetic variants in severe early-onset obesity

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

In preparation

Dominant obesity forms can be explained by heterozygous high-penetrant mutations in several genes, mainly involved in the leptin-melanocortin system and the hypothalamic development. Incomplete penetrance and variable expressivity may explain differences among carriers but no special attention has been paid in gender bias until now. A recent observation from our previous works, has revealed that maternally inherited variants in those genes were more frequent than expected.

A meta-analysis with more than 200 inherited rare sequence variants was used to define any potential gender bias in phenotype penetrance, expressivity or inheritance pattern. Additional clinical and behavioral data were recorded to assess the contribution of mutations present in transmitting mothers in the development of obesity in her carrier children.

Maternal transmission bias of inherited genetic variants in severe early-onset obesity

Francesc Bou de Pieri, Clara Serra-Juhé, Gabriel Á. Martos-Moreno, Jesús Argente, Luis A. Pérez-Jurado

ABSTRACT

Despite obesity is highly influenced by diet and lifestyle, severe early-onset obesity (EOO) has a strong genetic component. In addition to several autosomal recessive forms, heterozygous rare genetic variants (RSVs) with dominant effect mostly affecting several genes of the leptin-melanocortin pathway and hypothalamic function have been reported in EOO. Considering a main maternal role during prenatal and early-life events, we aimed to define whether gender and/or parental inheritance could be biased in these genetic forms of EOO and the possible reasons.

A meta-analysis including a total of 210 pathogenic inherited RSVs in European non-syndromic patients with EOO was performed to analyze the proportion of maternally versus paternally inherited variants. Parental phenotype, proband's birth weight (BW) and child-feeding practices were analyzed in a subset to explore possible prenatal and postnatal effects.

A significant maternal overtransmission of RSVs in candidate genes was observed (62.4% vs 37.6%, p=0.0002). This proportion was even increased when analyzing RSVs inherited from a parent affected by severe obesity. Patients with maternally inherited RSVs had an increased BW. Transmitting mothers had lower perceived responsibility for child feeding and exercised less pressure to eat, which may be indicative of a parental role less committed to children eating habits.

Our data revealed a maternal transmission bias of inherited RSVs among patients with EOO. The presence of a shared variant in the mother and the patient may induce the patient to develop obesity in an earlier manner. This phenotype may arise before birth or due to deregulation of eating habits caused by parental carelessness.

INTRODUCTION

Obesity is defined as an abnormal or excessive fat accumulation which may impair health, and is considered one of the most serious public health problems of the 21st century. In adults obesity is defined by a BMI higher than 30kg/m^2, but in children and adolescents BMI has to be adjusted for age and gender, so childhood obesity is defined by a BMI at or above the 95th percentile. Early-onset cases are increasing worldwide, with an actual prevalence of more than 38 million children under 5 years of age who are overweight or obese [1]. Obesity in children has an increased morbidity and mortality associated to type 2 diabetes mellitus, dyslipidemia, hypertension, obstructive sleep apnea and fatty liver dysfunction, which represents an important extra cost for the healthcare system [2,3] with long-term health and social consequences [4].

The development of obesity is considered multifactorial, with influence of many factors. Most cases may be caused by a combination of environmental determinants in individuals with variable genetic susceptibility leading to imbalance between energy intake and expenditure over a prolonged period. However, genetic factors contributing to obesity may account for a high proportion of BMI variance. Obesity heritability is estimated to be high (40-70%)[5], but only a very small fraction of genes associated with obesity has been identified, mostly multiple gene variants of minor effect. Around 5-10% of non-syndromic severe early-onset obesity (EOO) cases are caused by gene variants with high penetrance in specific genes [6]. These forms are mainly caused by mutations in genes involved in the leptin-melanocortin pathway and the hypothalamic development [7]. Indeed, pathogenic mutations which disrupt protein functions in these circuits have been associated with the development of severe EOO in humans and represents a robust fact to demonstrate its critical role in energy balance [6,8].

Dominant obesity forms can be explained by heterozygous mutations in several genes. *MC4R* encodes for the receptor of α-MSH and mutations in this gene represent the most frequent cause of monogenic obesity (up to 5%) with more than 150 deleterious variants described [9–11]. *BDNF*, *NTRK2* and *SIM1* encode for important proteins responsible for the development of the hypothalamus, and haploinsufficiency of these genes have been associated with obesity, as well as cognitive impairment in humans [12–15]. Other genes harboring mutations in obesity patients are *MRAP2*, which encodes for an accessory protein that regulates melanocortin receptors [16]; *SH2B1*, which

encodes a cytoplasmatic adaptor protein that modulates leptin-melanocortin signaling [17,18]; and others.

Incomplete penetrance and variable expressivity are common in autosomal dominant forms of obesity. It has been reported that carriers of likely pathogenic rare sequence variants (RSVs) in *MC4R* are more likely than non-carriers to have a mother with obesity (75.0% vs 31.7%; $p<0.001$) [20], without testing the genotype of the mother. This potential gender bias in phenotype penetrance, expressivity or inheritance pattern could be observed at a pre or a postnatal level. Prenatal effects may be seen during pregnancy and postnatal effects due to eating and feeding habits of the family. Maternal attitudes towards food may induce infants to manifest obesity in an early manner. The idea that genetic factors have an influence to human behaviors is supported by several behavioral genetics studies [21]. Eating and feeding practices, as part of the human conduct, may also be influenced by genetic factors. Shared genetic alterations between family members may explain similar attitudes regarding food.

Here, we have tested whether there is a gender bias in the penetrance, expressivity or inheritance pattern of likely pathogenic RSVs associated to dominant forms of likely monogenic EOO, and whether the bias might be affecting either at a pre or a postnatal level.

MATERIALS AND METHODS

Meta-analysis

Previous studies from our group reported several inherited RSVs associated with severe EOO (BMI above 3 SDS documented before 3 years of age) in Spanish patients, 20 affecting genes from the leptin-melanocortin pathway [11,22,23]. In order to study the pattern of inheritance of heterozygous RSVs in other patients of European ancestry we reviewed articles reporting genetic heterozygous variants causing monogenic forms of non-syndromic obesity affecting the same genes, with segregation studies performed. Selected genes included: *MC4R*, *POMC*, *NPY*, *BDNF*, *SIM1*, *PPARG*, *NEGR1, FTO, BDNF*, *NTRK2, MRAP2* and *SH2B1*. We selected articles sequencing single (mostly *MC4R*) or multiple candidate genes for obesity in panels, as well as those based on exome data [24–26]. Variants affecting the listed candidate genes for obesity were selected if they were in heterozygous state (we expected

dominant forms with high penetrance) and had the inheritance pattern reported. Selected variants were classified as pathogenic or likely pathogenic if there were functional studies supporting its pathogenicity or evidences from computational predictors (CADD>20).

Finally, the meta-analysis was performed with 210 transmissions in 166 families from 37 articles with a total of 106 different RSVs. From each article, the genes studied, the number of variants included in the meta-analysis, the phenotype of the patients and the population origin can be found in Supplemental Table 1. Most index cases fulfilled criteria for severe EOO (n=172), while the phenotype was described as obesity or severe obesity with onset details in 38 cases. Half of the variants included in the meta-analysis where in *MC4R* (122 out of 210, 58%), the most frequently mutated gene in monogenic obesity.

From each proband and family, critical information was extracted mainly from tables and figures, but also from the text of the articles. Gender of the proband; inheritance pattern, type and pathogenic status of the variant; and phenotype of the transmitting progenitor were collected when possible and were used for the discussion of the results. According to the BMI reported, transmitting progenitors were classified as severely obese (BMI≥40), obese (BMI≥30), overweight (BMI≥25) or lean (BMI<25). The proportion of maternally inherited RSVs was analyzed in all the data collected.

Subjects

Severe EOO patients were recruited in *Hospital Infantil Universitario Niño Jesús*, after receiving written informed consent from the family. A total of 20 trios were selected from our previously studied cohorts by having segregation studies performed from RSVs in candidate genes for obesity. Children were split in two groups: 13 EOO children with RSVs inherited from their mothers (Mat) and 7 EOO children with RSVs not inherited from their mothers (Not-Mat, including paternally inherited and *de novo* variants).

Moreover, we recruited 23 mothers to analyze the contribution of the RSVs in the progenitors. Mothers were split in 3 groups: 8 transmitting mothers with obese children harboring maternally inherited variants (TM), 9 mothers with obese children with no reported genetic alterations or not inherited from the mother (Cont-Obe), and 6 mothers with non-obese children (Cont-Lean).

Birth weight

Clinical data from severe EOO patients was collected in their reference center for pediatric endocrinology. BW was obtained from clinical reports and BW SDS for gestational age was calculated from standard growth charts. Patients with +2 SDS in BW for gestational age were considered obese at birth. We had access to BW records in 13 patients from the Mat group and 6 patients from the No-Mat group.

Child feeding questionnaire

A complete Spanish questionnaire was created to evaluate eating and feeding behaviors or habits of mothers with obese children. The questionnaire was given in person or online after telephone contact and answers were collected by clinicians from the endocrinology service at *Hospital Infantil Universitario Niño Jesús*. Three groups of mothers answered the questionnaire (TM, Cont-Obe and Cont-Lean). The TM group was compared with both control groups and differences related to maternally inherited variants (TM vs. Cont-Obe) and childhood obesity phenotype (TM vs. Cont-Lean) were analyzed.

The questionnaire, consisting in 4 parts, was elaborated with the support of the "Cardiovascular risk and nutrition" group from the IMIM, Barcelona. The first part contained background information regarding the mother and her child; the second part asked for the primary diet of the mother; the third part contained the Family Adaptability and Cohesion Evaluation Scale (FACES); and the last part included the Child Feeding Questionnaire (CFQ) (Annex 1: CuestionarioESP, in Spanish).

Background questionnaire

The background questionnaire was used to collect information about the family: age, sex, birth place, education, occupation, anthropometric measures, obesity antecedents and cohabitation.

Primary diet questionnaire

The primary diet questionnaire aimed to define if eating habits of mothers were appropriate. We used a version from the IMIM-Hospital del Mar, which assesses dietary habits through 18 different items. A score between 0 and 36 is obtained, being 0-12 an inadequate diet (not healthy), 13-24 an adequate diet in some aspects, and 25-36 an adequate diet (healthy).

FACES

The FACES is a brief scale of analysis about family functioning [27]. We used a validated Spanish reduced version of FACES II that contained 10 statements about cohesion and 10 statements about adaptability which could be scored from 1 to 5 [28,29]. This FACES version used the linear model to explain family functioning, being high levels of cohesion and adaptability those more associated to well-functioning families.

CFQ

The CFQ assesses parent attitudes and practices in relation to infant feeding, focusing on the predisposition of children to obesity. We used a version of the CFQ [30] validated in Spanish population [31]. It contained 30 questions from 7 subscales which could be scored from 1 to 5. The subscales were perceived responsibility for feeding, perceived parent weight, perceived child weight, concern about child weight, restriction, pressure to eat and monitoring of food intake. The average values of each subscale were evaluated.

Statistical analyses

The proportion of maternally inherited variants was analyzed with a binomial test (one tailed) assuming that autosomal alleles are inherited from each parent with identical frequency, 50%.

Birth weight data were analyzed with a one-sample t-test by comparing the observed values with the population mean (0.0 SDS). The proportion of individuals with BW above 2 SDS was compared with the 2.5% proportion expected in a normal distribution with a binomial test (one tailed).

An unpaired t test was used to compare the means from the different groups analyzed in the child feeding questionnaire. Statistically significant values were considered when p was lower than 0.05.

RESULTS

Gender differences in EOO

Several studies previously published by our group had explored the contribution of rare sequence variants (RSVs) to the obesity phenotype in Spain [11,22,23]. Heterozygous variants affecting genes already reported as obesity-

causative and other candidate genes by function or animal models (*MC4R*, *NPY*, *BDNF*, *SIM1*, *PPARG*, *NEGR1* and *FTO*) were identified in severe EOO patients.

A meta-analysis with 37 articles revealed 210 probands in 166 families with EOO caused by heterozygous RSVs in any of the 12 candidate genes (106 different RSVs). Female patients were overrepresented in our meta-analysis (90 males and 120 females, p=0.0226), indicating a possible gender bias in the penetrance of the phenotype. However, if we excluded patients from studies with less stringent phenotype inclusion criteria (not severe EOO) this bias disappeared (85 males and 105 females, p=0.0840). Moreover, in our 20 Spanish cases the gender proportion was inversed (14 males and 6 females, p=0.0577).

Over-transmission of maternally inherited variants

Segregation studies revealed that 14 out of 20 RSVs in our Spanish group where inherited from the mother (70%, p=0.0577). This represents a huge difference from the 50% expected, considering that no imprinting was described in any of the genes.

The meta-analysis of 210 transmissions in 166 families detected that this bias respect the 50% expected is maintained in other studies. We observed that 62.4% of inherited heterozygous variants were of maternal origin, being a statistically significant difference (p=0.0002) (Table 1 and Supplementary Table 2).

When analyzing genes individually, only *MC4R* reached significant levels (p=0.0013) (Table 1). The other genes did not reached significant levels due to the small number of transmissions analyzed. However, the rest of genes together (all except *MC4R*) also showed a significant deviation (p=0.0347) (Table 1).

Despite female patients were overrepresented in our meta-analysis, the bias of maternally inherited variants was observed in both, males and females (Table 1).

At the light of the results, this bias was a general observation among both sexes and a great range of candidate genes involved in the leptin-melanocortin pathway, especially *MC4R*.

Genes	Maternal	Paternal	Total	% maternal	p
All	131	79	210	62.4%	0.0002
All - M	57	33	90	63.3%	0.0074
All - F	74	46	120	61.7%	0.0067
MC4R	78	44	122	63.9%	0.0013
MC4R - M	34	18	52	65.4%	0.0182
MC4R - F	44	26	70	62.9%	0.0207
Other	53	35	88	60.2%	0.0347
Other - M	23	15	38	60.5%	0.1279
Other - F	30	20	50	60.0%	0.1013
Pathogenic	103	61	164	62.8%	0.0006
Pathogenic - M	45	24	69	65.2%	0.0077
Pathogenic - F	58	37	95	61.1%	0.0198

Table 1. Proportion of maternally inherited rare sequence variants (RSVs) in candidate genes for obesity from patients with obesity. The contribution of the gender of the patient was also analyzed (M: male, F: female). All include the 17 candidate genes studied; other include all genes except MC4R; pathogenic include variants with functional studies or variants predicted to be deleterious. The p value calculated with one-tail binomial test (expected 0.5) is shown.

Phenotype of the transmitting progenitor

On average, from every 6 transmitting progenitors of the meta-analysis, 1 was lean, 1 overweight and 4 obese (16.8%, 16.8% and 66.3%, respectively). Considering only transmitting mothers, the proportion of obesity among them was similar to all progenitors (67.7%).

Moreover, the stratification of the progenitors according to the severity of the phenotype lead to interesting observations. If we selected only variants transmitted by obese progenitors the proportion of maternally inherited variants was similar to the overall reported (63.0%, p=0.0014), but if we selected variants transmitted by non-obese progenitors (lean and overweight) these differences were reduced (60.0%, p=0.0598) (Figure 1). Interestingly, if only variants transmitted by severely obese progenitors were selected, the proportion of maternally inherited variants was much higher (84.0%, p=0.0005) (Figure 1). Our data revealed a clear relationship between the severity of the phenotype among progenitors and the maternal bias of inherited RSVs.

Considering the phenotype of the proband, although all had been included in obesity studies, the maternally bias was not exclusive from the most severe cases and was observed along the whole obesity spectrum (data not shown).

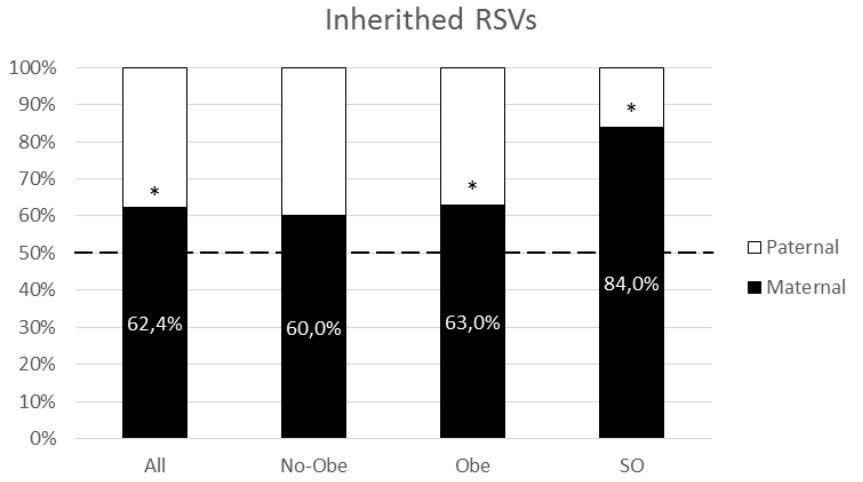

Figure 1. Origin of inherited RSVs by progenitor phenotype. Transmitting progenitors were grouped in 4 categories: progenitors with any phenotype (All, n=210), progenitors with a non-obese phenotype (No-Obe, n=70), progenitors with an obese phenotype (Obe, n=138), progenitors with a severe obesity phenotype (SO, n=25). The percentage of maternally inherited RSVs is shown in each category. Statistically significant deviations from the 50% expected by the binomial test are marked with an * ($p<0.05$).

Birth weight

The 13 patients from the Mat group had an average BW of +1.35 SDS, which represented a significantly higher observation compared to the population BW mean ($p=0.0348$), whereas the 6 No-Mat patients had an average BW of -0.27 SDS not significantly differing from the mean ($p=0.2526$). Interestingly, the top BW SDS values were exclusive from the Mat group, with 6 out of 13 patients classified as obese at birth (>2 BW SDS) and no patients in the No-Mat group. An observation beyond 2 SDS is expected in 2.5% of cases under normal distribution, but in the Mat group it was observed in 46.2% of patients ($p<0.0001$). These data showed that patients with RSVs inherited from the mother had an increased BW and that patients defined as obese at birth were over-represented in the maternally inherited group, suggesting a possible new role of the studied mutations before birth.

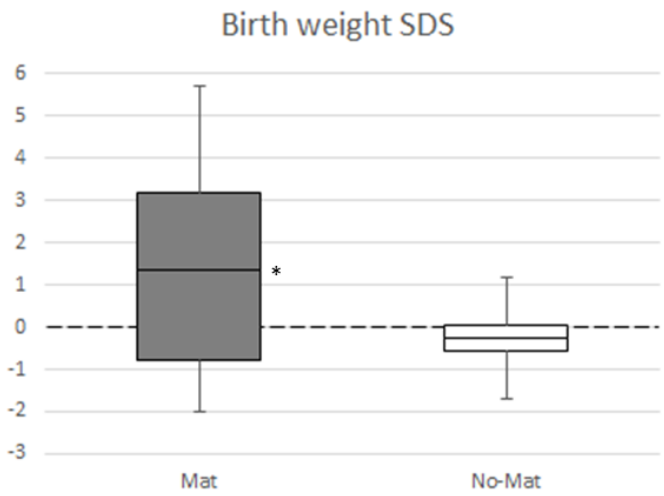

Figure 2. Birth weight SDS box plots from obese children. Patients were grouped according to the paternal origin of the RSVs: EOO children with RSVs inherited from their mothers (Mat, n=13) and EOO children with RSVs not inherited from their mothers (Not-Mat, n=6). The mean, the first and the third quartiles, the minimum and the maximum of the data are ploted. * represents statistically significant differences ($p<0.05$) with a one sample t test comparing the observed values with the population mean (0.0 SDS).

Child feeding behavior

The age of the responding mothers was between 35 and 55 years old, with children between 8 and 23 years old. TM mothers and their children were older than controls as they were first recruited some time ago ($p<0.05$). Children from TM and Cont-Obe groups had a mean BMI over 30 (with some cases of severe obesity), whereas all children from the Cont-Lean group were lean (mean BMI under 20). All but one TM mothers were overweight or obese (with half of them reporting childhood obesity), as well as the Cont-Obe mothers (with one third reporting childhood obesity), while none of the Cont-Lean mothers had reported obesity.

Diet data showed that mothers from all three groups were on average in the upper limit of "adequate diet in some aspects" (TM=22.6, Cont-Obe=23.4, and Cont-Lean=21.8) almost reaching the "adequate diet" threshold (Figure 3A). Similar diets between groups is the base for further comparisons with more precision and less confusing factors.

Familiar habits and functioning evaluated by FACES may be poorer in the TM group compared to control groups, although no significant differences were observed (TM=0.77, Cont-Obe=0.87, and Cont-Lean=0.88, $p_{(TM/Cont-Obe)}$=0.1037, $p_{(TM/Cont-Lean)}$=0.1557). Families from the TM group were characterized by lower cohesion scores (with two families reporting scores lower than 25 in a 10-to-50 scale) and lower adaptability scores compared to other families (Figure 3B), suggesting a weaker familiar equilibrium.

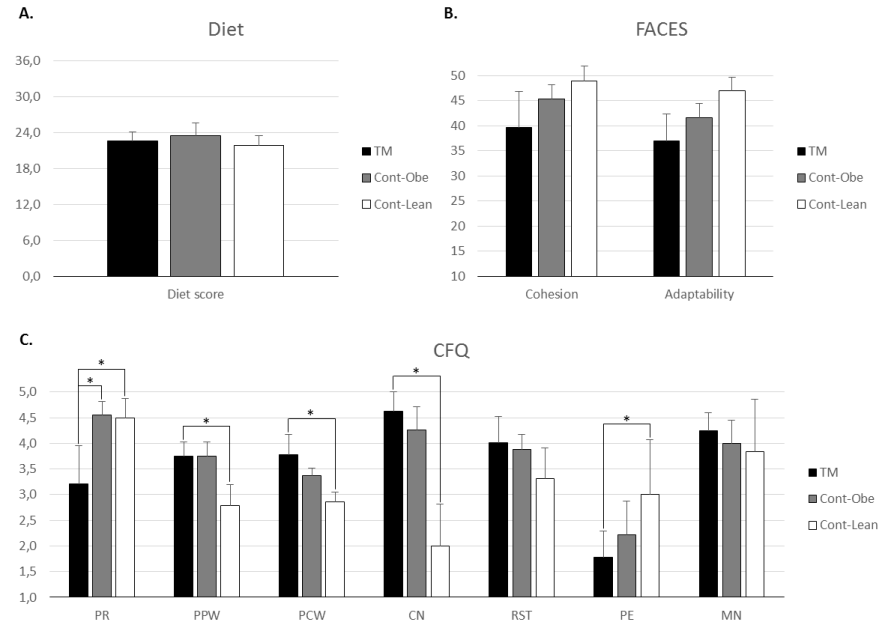

Figure 3. Data from the questionnaires answered by mothers from the 3 groups studied: transmitting mothers of obese children harboring maternally inherited variants (TM, n=8), mothers of obese children with no reported genetic alterations or not inherited from the mother (Cont-Obe, n=9), mothers of non-obese children (Cont-Lean, n=6). Error bars represent the 95% Confidence Interval and * represent statistically significant differences ($p<0.05$) with an unpaired t test between the case group and one control group. A: Diet scores from the primary diet questionnaire. Diet can be classified as inadequate (0-12), adequate in some aspects (13-24) and adequate (25-36). B: Family functioning from the FACES scale. High levels of cohesion and adaptability are associated with well-functioning families. C: Parent attitudes and practices in relation to infant feeding from the CFQ. Perceived responsibility for feeding (PR), perceived parent weight (PPW), perceived child weight (PCW), concern about child weight (CN), restriction (RST), pressure to eat (PE) and monitoring of food intake (MN) are evaluated.

Answers from the CFQ were used to compare the groups of mothers depending on child feeding believes, attitudes or practices (Figure 3C, Table 2). Perceived responsibility for feeding was lower in TM mothers compared to both controls ($p_{(TM/Cont\text{-}Obe)}$=0.0031, $p_{(TM/Cont\text{-}Lean)}$=0.0171), indicating a more relaxed role regarding child feeding practices. Perceived parent weight was higher in mothers of obese children (TM and Cont-Obe) compared to Cont-Lean mothers ($p_{(TM/Cont\text{-}Lean)}$=0.0013), which went in line with the phenotype of the mothers. Perceived child weight was especially high among TM mothers, but also in Cont-Obe ($p_{(TM/Cont\text{-}Lean)}$=0.0027), as expected by the severe phenotype observed among obese patients. Concern about child's weight was much higher in TM and Cont-Obe mothers compared to Cont-Lean mothers ($p_{(TM/Cont\text{-}Lean)}$<0.0001), indicating that mothers with children affected by obesity are more aware about weight issues. Restriction was similar among groups with higher scores among mothers with obese children (TM and Cont-Obe), which seemed more restrictive in some aspects. Pressure to eat was following an inverse effect with low values in TM mothers ($p_{(TM/Cont\text{-}Lean)}$=0.0477), suggesting that an increased weight in children may decrease eating exigency carried out by parents. Monitoring of food intake was similar between all groups.

Subscale	TM	Cont-Obe	Cont-Lean	Diff. TM/ Cont-Obe	Diff. TM/ Cont-Lean
PR	3,2 ± 1,1	4,6 ± 0,4	4,5 ± 0,5	-1,35*	-1,29*
PPW	3,8 ± 0,4	3,8 ± 0,4	2,8 ± 0,5	0,00	0,96*
PCW	3,8 ± 0,6	3,4 ± 0,2	2,9 ± 0,2	0,40	0,92*
CN	4,6 ± 0,5	4,3 ± 0,7	2,0 ± 1,0	0,37	2,63*
RST	4,0 ± 0,7	3,9 ± 0,4	3,3 ± 0,7	0,13	0,70
PE	1,8 ± 0,7	2,2 ± 1,0	3,0 ± 1,3	-0,44	-1,22*
MN	4,3 ± 0,5	4,0 ± 0,7	3,8 ± 1,3	0,25	0,42

Table 2. Results from the Child Feeding Questionnaire (CFQ) performed to mothers in relation to child feeding believes, attitudes or practice. Answers from the 3 groups studied: transmitting mothers of obese children harboring maternally inherited variants (TM, n=8), mothers of obese children with no reported genetic alterations or not inherited from the mother (Cont-Obe, n=9), mothers of non-obese children (Cont-Lean, n=6). Mean ± SD. The differences between means are also shown (Diff. TM/Cont-Obe and Diff. TM/Cont-Lean). The asterix symbol (*) represents statistically significant differences between groups ($p<0.05$) with an unpaired t test. The 7 subscales evaluated are PR (perceived responsibility for feeding), PPW (perceived parent weight), PCW (perceived child weight), CN (concern about child's weight), RST (restriction), PE (pressure to eat) and MN (monitoring of food intake).

DISCUSSION

Obesity is known as a multifactorial and genetically heterogeneous disorder, representing a challenging target for genetic studies. A wide range of genes and variants has been the focus of projects aiming to identify risk factors that could explain inter-individual differences in obesity development. Mutations affecting genes from the leptin-melanocortin pathway, especially *MC4R* (which is the most frequent cause of monogenic obesity), represent a deeply studied source of genetic causes for severe obesity [32]. Although several gene variants have been identified to cause recessive forms of obesity, heterozygosity for deleterious mutations in multiple genes of this pathway has also been clearly associated with severe early-onset obesity. This causative dominant genetic variants are frequently inherited from parents with milder or no phenotypes, showing incomplete penetrance and variable expressivity.

During several years, our group has put special efforts to genetically characterize a big Spanish cohort of patients affected by severe EOO. Among different genetic alterations, we explored RSVs, mainly represented by highly penetrant heterozygous variants acting in a dominant manner. After performing segregation studies in some patients, we observed an unexpected deviation in favor of maternally inherited RSVs (70%). With only our data, this difference was not statistically significant, but was surprising because no one has reported something similar.

In order to check if the increased proportion of maternally inherited RSVs in obesity patients was a real trend or only an artifact from our population, we performed a meta-analysis with reported variants in candidate genes for obesity. In total, 210 inherited monoallelic RSVs (likely dominant forms) were selected from 37 different European studies. *De novo* and biallelic variants were excluded. An increased proportion of maternally inherited variants (62.4%) was observed, representing a clear bias. Despite more female patients were included in the meta-analysis, the maternal bias was independent of the gender of the child. However, a clear trend was observed when stratifying by the phenotype of the transmitting progenitor. The proportion of maternally inherited variants increased with the severity of the phenotype, going from the 60% in non-obese progenitors, 63% in obese progenitors, to 84% in progenitors affected by severe obesity. This relationship highlighted the interesting role of maternally inherited variants in an obesity context.

Maternal over-transmission has also been reported in imprinted loci. Alterations in the maternal allele from the *GNAS* locus have been associated

with the development of syndromic obesity forms, such as pseudohypoparathyroidism [33]. To prevent any artifact in the meta-analysis, we have excluded RSVs affecting the *GNAS* gene, as it is a known imprinted gene.

The studied RSVs in patients may be explaining part of the extreme obesity phenotype observed among them. However, the overrepresentation of these variants in mothers may indicate that the presence of specific RSVs in the mother may increase the risk of the child to develop obesity in an earlier or more severe manner. So, the presence of these variants in both, child and mother, may be acting together to explain obesity in patients. The additive component of the maternally inherited variants may be due to effects from mothers to children at a prenatal or a postnatal level. Additional clinical and behavioral information was collected in order to elucidate possible mechanisms to explain the deviation observed.

At a prenatal level, we used the BW for gestational age to assess if the increased fat accumulation in patients was present even before birth. Children with maternally inherited variants showed significantly higher BW than the population average for gestational age. Moreover, children defined as obese at birth were clearly enriched in the group that had variants inherited from the mother. These observations pointed towards a possible role of shared RSVs between patients and mothers in inducing an earlier onset of obesity in children. Despite no causal information could be extracted from this data, it opened a new area of obesity research focusing on prenatal effects during pregnancy influenced by the presence of specific variants in the mother. Maternal overnutrition during pregnancy and/or epigenetic effects related to maternal obesity would also lead to prenatal overweight in siblings not inheriting the genetic variant. Maternal obesity during gestation is associated with increased fetal growth, although the effect on birth weight and adiposity seems significant only in girls [34].

Other possible links between childhood obesity and genetic variants in mothers may be the ones related with familiar attitudes and practices regarding food intake during the first years of live (postnatal). Feeding behaviors of mothers have been analyzed using a specific questionnaire that assesses specifically the predisposition of children to obesity, but we also collected data from diet and familiar functioning. Maternal diet was similar among groups, so differences in feeding behaviors may be independent of maternal eating habits, indicating that mothers may have a distinct role when eating

(themselves) and feeding (their children). Binge eating reported in *MC4R* carriers was not assessed in our study [20]. Moreover, poorer family functioning has been described as a risk factor for unhealthy parent feeding practices in children with overweight or obesity [35]; however, despite some divergences, no significant differences were found between groups. Thus, the observed differences in maternal feeding practices do not seem to be influenced by dietary habits or familiar functioning.

From the child feeding questionnaire, the first differences were observed due to the intrinsic characteristics of the groups analyzed. The perceived child weight was in agreement with the children's BMI, being high among mothers with obese children (TM and Cont-Obe); the concern about child's weight followed the same pattern, even with increased differences between groups; and the perceived parent weight revealed the differences expected as mothers with obese children (TM and Cont-Obe) were mostly affected by obesity and control mothers were not. Interestingly, despite no statistically significant differences were identified between TM and Cont-Obe mothers in these three subscales, scores from the TM group were always at the opposite extreme from the Cont-Lean group, with Cont-Obe mothers in the middle, which represented a trend. In addition, we observed a significantly lower perceived responsibility for feeding among TM mothers compared to both controls, as well as a reduced pressure to eat compared to Cont-Lean. These observations may be indicative of a parental role less committed to children eating habits. Parental involvement during the first years is essential for children to develop properly their eating control mechanisms [36], so carelessness in these aspects may be explaining dysregulation in eating habits in obesity patients with maternally inherited variants. Our data pointed to a negative association between child's BMI and pressure to eat and a positive association between child's BMI and concern about child weight, as previously reported [31]. No clear associations were found with factors related with feeding control, such as restriction or monitoring, but a clear trend was observed with TM mothers having the greatest values in these subscales.

The main limitation of the questionnaire performed in this study was the small sample size and reduced number of answers mostly due to difficulties to re-contact families that had been included in the study but were no longer in follow up. Nevertheless, it represented a valuable approach to quantify the genetic contribution into maternal eating behaviors. Larger datasets are required to validate or discharge some tendencies observed in our study.

In summary, we have described for the first time a significant increased proportion of maternally inherited variants in candidate genes for EOO. We propose that synergistic effects of mutations present in both, mothers and children, can contribute to EOO development in children, acting both pre- and postnatally. Although no causal evidences were identified, some insights were revealed, which may open a challenging new area of obesity for future studies linking maternal genotypes with children phenotypes.

ACKNOWLEDGEMENTS

We would like to thank Dra. Milà and Dr. Schröder for the support in the elaboration of the questionnaire.

REFERENCES

1. World Health Organization. *Fact Sheet "Obesity and Overweight."*; 2018. http://www.who.int/mediacentre/factsheets/fs311/en/.
2. Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *J Am Med Assoc*. 2003;289(1):76-79. doi:10.1001/jama.289.1.76
3. Hernáez Á, Zomeño MD, Dégano IR, et al. Excess Weight in Spain: Current Situation, Projections for 2030, and Estimated Direct Extra Cost for the Spanish Health System. *Rev Española Cardiol (English Ed*. 2019;72(11):916-924. doi:10.1016/j.rec.2018.10.010
4. Lobstein T, Baur L, Uauy R, Obesity I. Obesity in children and young people : a crisis in public health. *Obes Rev*. 2004;5:4-85.
5. Bray MS, Loos RJF, McCaffery JM, et al. NIH working group report - Using genomic information to guide weight management: From universal to precision treatment. *Obesity*. 2016;24(1):14-22. doi:10.1002/oby.21381
6. Ranadive SA, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am*. 2008;37(3):733-751. doi:10.1016/j.ecl.2008.07.003
7. da Fonseca ACP, Mastronardi C, Johar A, Arcos-Burgos M, Paz-Filho G. Genetics of non-syndromic childhood obesity and the use of high-throughput DNA sequencing technologies. *J Diabetes Complications*. 2017;31(10):1549-1561. doi:10.1016/j.jdiacomp.2017.04.026

8. Farooqi IS, O'Rahilly S. Monogenic obesity in humans. *Anu Rev Med*. 2005;56:443-458. doi:10.1146/annurev.med.56.062904.144924
9. Vaisse C, Clement K, Guy-Grand B, Froguel P. A frameshift mutation in human MC4R is associated with a dominant form of obesity. *Nat Genet*. 1998;20(october):113-114.
10. Yeo GSH, Farooqi IS, Aminian S, Halsall DJ, Stanhope RG, O'Rahilly S. A frameshift mutation in MC4R associated with dominantly inherited human obesity. *Nat Genet*. 1998;20(october):111-112.
11. Granell S, Serra-Juhé C, Martos-Moreno GÁ, et al. A novel melanocortin-4 receptor mutation MC4R-P272L associated with severe obesity has increased propensity to be ubiquitinated in the ER in the face of correct folding. *PLoS One*. 2012;7(12):e50894. doi:10.1371/journal.pone.0050894
12. Gray J, Yeo GSH, Cox JJ, et al. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*. 2006;55(12):3366-3371. doi:10.2337/db06-0550
13. Harcourt BE, Bullen DVR, Kao KT, et al. Maternal inheritance of BDNF deletion, with phenotype of obesity and developmental delay in mother and child. *Am J Med Genet Part A*. 2018;176(1):194-200. doi:10.1002/ajmg.a.38539
14. Yeo GSH, Connie Hung C-C, Rochford J, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci*. 2004;7(11):1187-1189. doi:10.1038/nn1336
15. Villa A, Urioste M, Bofarull J, Martínez-Frías M. De novo interstitial deletion q16.2q21 on chromosome 6. *Am J Med Genet*. 1995;55(3):379-383.
16. Baron M, Maillet J, Huyvaert M, et al. Loss-of-function mutations in MRAP2 are pathogenic in hyperphagic obesity with hyperglycemia and hypertension. *Nat Med*. 2019;25(11):1733-1738. doi:10.1038/s41591-019-0622-0
17. Walters RG, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010;463(7281):671-675. doi:10.1038/nature08727
18. Flores A, Argetsinger LS, Stadler LKJ, et al. Crucial role of the SH2B1 PH domain for the control of energy balance. *Diabetes*. 2019;68(11):2049-2062. doi:10.2337/db19-0608
19. Pearce LR, Atanassova N, Banton MC, et al. KSR2 mutations are

associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell.* 2013;155(4):765. doi:10.1016/j.cell.2013.09.058

20. Branson R, Potoczna N, Kral JG, Lentes K-U, Hoehe MR, Horber FF. Binge eating as a phenotype of melanocortin 4 receptor gene mutations. *N Engl J Med.* 2003;348(12):1096-1103. doi:10.1056/NEJM200308073490615
21. Fuller JL, Thompson W. *Behavior Genetics.*; 1960. doi:10.1002/9781444343120.ch10
22. Serra-Juhé C, Martos-Moreno G, Bou de Pieri F, et al. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet.* 2017;13(5):1-19. doi:10.1371/journal.pgen.1006657
23. Serra-Juhé C, Martos-Moreno G, Bou de Pieri F, et al. Heterozygous rare genetic variants in non-syndromic early-onset obesity. *Int J Obes.* 2020;40:830-841. doi:10.1038/s41366-019-0357-5
24. Kleinendorst L, Massink MPG, Cooiman MI, et al. Genetic obesity: next-generation sequencing results of 1230 patients with obesity. *J Med Genet.* 2018;55:578-586. doi:10.1136/jmedgenet-2018-105315
25. Hendricks AE, Bochukova EG, Marenne G, et al. Rare Variant Analysis of Human and Rodent Obesity Genes in Individuals with Severe Childhood Obesity. *Sci Rep.* 2017;7(1):1-14. doi:10.1038/s41598-017-03054-8
26. Philippe J, Stijnen P, Meyre D, et al. A nonsense loss-of-function mutation in PCSK1 contributes to dominantly inherited human obesity. *Int J Obes.* 2015;39:295-302. doi:10.1038/ijo.2014.96
27. Olson DH, Sprenkle DH, Russell CS. Circumplex model of marital and family systems: I. Cohesion and adaptability dimensions, family types, and clinical applications. *Fam Process.* 1979;18:3-28.
28. Martínez-Pampliega A, Iraurgi I, Galíndez E, Sanz M. Family Adaptability and Cohesion Evaluation Scale (FACES): Desarrollo de una versión de 20 ítems en español. *Int J Clin Heal Psychol.* 2006;6(2):317-338.
29. Martínez-Pampliega A, Iraurgi I, Sanz M. Structural validity of the FACES-20Esp: Family adaptability and cohesion evaluation scale, Spanish version of 20 items. *Rev Iberoam Diagnostico y Eval Psicol.* 2011;1(29):147-165.
30. Birch LL, Fisher JO, Grimm-Thomas K, Markey CN, Sawyer R, Johnson SL. Confirmatory factor analysis of the Child Feeding Questionnaire: A measure of parental attitudes, beliefs and practices

about child feeding and obesity proneness. *Appetite*. 2001;36(3):201-210. doi:10.1006/appe.2001.0398

31. Canals-Sans J, Blanco-Gómez A, Luque V, et al. Validation of the Child Feeding Questionnaire in Spanish Parents of Schoolchildren. *J Nutr Educ Behav*. 2016;48(6):383-391.e1. doi:10.1016/j.jneb.2016.03.017
32. Ranadive S, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am*. 2008;37(3):733-751.
33. Weinstein LS, Xie T, Qasem A, Wang J, Chen M. The role of GNAS and other imprinted genes in the development of obesity. *Int J Obes*. 2010;34(1):6-17. doi:10.1038/ijo.2009.222
34. Mitanchez D, Jacqueminet S, Nizard J, et al. Effect of maternal obesity on birthweight and neonatal fat mass: A prospective clinical trial. *PLoS One*. 2017;12(7):1-15. doi:10.1371/journal.pone.0181307
35. Warnick JL, Stromberg SE, Krietsch KM, Janicke DM. Family functioning mediates the relationship between child behavior problems and parent feeding practices in youth with overweight or obesity. *Transl Behav Med*. 2019;9(3):431-439. doi:10.1093/tbm/ibz050
36. Ek A, Sorjonen K, Eli K, et al. Associations between parental concerns about preschoolers' weight and eating and parental feeding practices: Results from analyses of the child eating behavior questionnaire, the child feeding questionnaire, and the lifestyle behavior checklist. *PLoS One*. 2016;11(1):1-20. doi:10.1371/journal.pone.0147257

SUPPLEMENTARY MATERIAL

Supplementary Table 1
Information from the 37 articles selected for the meta-analysis

Supplementary Table 2
Summary of the inherited variants included in the meta-analysis

Annex 1
CuestionarioESP

Reference	Genes	T-F-V	Phenotype	Country
Granell, et al. 2015	*MC4R*	1-1-1	SEOO	Spain
Serra-Juhé, et al. 2017	14	1-1-1	SEOO	Spain
Serra-Juhé, et al. 2020	15	18-18-15	SEOO	Spain
Kleinendorst, et al. 2018	52	18-17-12	O+1criteria	The Netherlands, Dutch Caribbean, UK and Finland
Hendricks, et al. 2017	32	21-14-12	SEOO*	UK
Philippe, et al. 2015	34	3-2-2	O, SO	France
Yeo, et al. 1998	*MC4R*	1-1-1	SEOO	UK
Vaisse, et al. 1998	*MC4R*	2-1-1	SO	France
Hinney, et al. 1999	*MC4R*	3-3-2	SEOO	Germany
Farooqi, et al. 2000	*MC4R*	2-2-2	SEOO	UK
Vaisse, et al. 2000	*MC4R*	5-4-4	SO	France
Dubern, et al. 2001	*MC4R*	4-4-4	SEOO	France
Miraglia Del Giudice, et al. 2002	*MC4R*	4-1-1	EOO	Italy
Lubrano-Berthelier, et al. 2003	*MC4R*	5-3-3	SEOO	France
Marti, et al. 2003	*MC4R*	1-1-1	O	Spain
Farooqi, et al. 2003	*MC4R*	20-11-9	SEOO	UK
Valli-Jaakola, et al. 2004	*MC4R*	3-3-3	SEOO, SO	Finland
Ochoa, et al. 2007	*MC4R*	4-4-4	EOO	Spain
Hainerová, et al. 2007	*MC4R*	8-4-3	SEOO	Czechia
Stutzmann, et al. 2008	*MC4R*	27-19-12	EOO, SO	France
Santoro, et al. 2009	*MC4R*	5-4-2	SEOO	Italy
Albuquerque, et al. 2014	*MC4R, LEPR*	6-6-6	SEOO	Spain, Portugal
Stanikova, et al. 2015	*MC4R*	2-2-2	SEOO	Slovakia
Vollbach, et al. 2016	*MC4R*	1-1-1	O	Germany
Gimeno-Ferrer, et al. 2019	*MC4R*	6-4-4	SO	Spain
Doche, et al. 2012	*SH2B*	3-3-3	SEOO	UK
Miraglia Del Giudice, et al. 2001	*POMC*	3-3-3	EOO	Italy
Challis, et al. 2002	*POMC*	4-1-1	SEOO	UK
Lee, et al. 2006	*POMC*	5-5-1	SEOO	UK
Biebermann, et al. 2006	*POMC*	3-2-1	EOO	Germany
Dubern, et al. 2008	*POMC*	1-1-1	EOO	France
Creemers, et al. 2008	*POMC*	2-2-2	SEOO	UK
Gray, et al. 2007	*NTRK2*	2-2-2	SO+Developmental delay	UK
Hung, et al. 2007	*SIM1*	1-1-1	SEOO	UK

Reference	Genes	T-F-V	Phenotype	Country
Bonnefond, et al. 2013	*SIM1*	2-2-1	SEOO, SO, PWL	France
Ramachandrappa, et al. 2013	*SIM1*	11-11-6	SEOO	UK
Schonnop, et al. 2016	*MRAP2*	2-2-2	SEOO	Germany

Supplementary Table 1. Information from the 37 articles selected for the meta-analysis. It included articles from our group (n=3), articles sequencing several candidate genes for obesity (n=3), articles sequencing MC4R gene (n=19) and articles sequencing other genes (n=12). When more than 2 genes were sequenced in the same study, the total number of genes and the number of candidate genes for this work are reported. The number of transmissions evaluated (T), families (F) and different variants (V) from each study is shown in column T-F-V. A total of 210 transmissions from 166 families and 106 different variants in obesity candidate genes were included in the meta-analysis. Inclusion criteria were referred by segregation studies (inheritance pattern reported), phenotype (obesity) and origin (European). Obesity phenotypes include, from less to more extreme: O=obesity, EOO=early-onset obesity, SO=severe obesity, SEOO=severe early-onset obesity. If more than 1 group of patients was studied, phenotypes of each group were shown separated by comas. PWL=Prader-Willi-Like syndrome. *Patients with MC4R mutations were excluded.

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	N74I	Missense	Pathogenic	Mat	O	Case #2	M	Granell, et al. 2015
NPY	V86D	Missense	VUS	Pat	O	Ob_158 (table 3)	F	Serra-Juhé, et al. 2017
MC4R	P272L	Missense	Pathogenic	Mat	EOO	OB48 (table 2)	M	Serra-Juhé, et al. 2020
MC4R	A259V	Missense	Pathogenic	Mat	O	2206 (table 2)	M	Serra-Juhé, et al. 2020
MC4R	H76R	Missense	Pathogenic	Mat	O	1698 (table 2)	M	Serra-Juhé, et al. 2020
MC4R	R22X	Nonsense	Pathogenic	Mat	O	1901 (table 2)	M	Serra-Juhé, et al. 2020
SIM1	Q206K	Missense	Pathogenic	Mat	L	1782 (table 2)	F	Serra-Juhé, et al. 2020
SIM1	c.352+1G>A	Splicing	Pathogenic	Pat	O	1749 (table 2)	M	Serra-Juhé, et al. 2020
PPARG	I109N	Missense	Pathogenic	Mat	L	1740 (table 2)	M	Serra-Juhé, et al. 2020
BDNF	I231V	Missense	VUS	Pat	L	2214 (table 2)	M	Serra-Juhé, et al. 2020
NEGR1	I105V	Missense	VUS	Mat	L	OB20 (table 2)	M	Serra-Juhé, et al. 2020
FTO	R80W	Missense	VUS	Pat	O	2153 (table 2)	F	Serra-Juhé, et al. 2020
NEGR1	A326T	Missense	VUS	Pat	L	OBE26 (text)	M	Serra-Juhé, et al. 2020
NEGR1	A326T	Missense	VUS	Mat	L	1860 (text)	M	Serra-Juhé, et al. 2020
BDNF	Q50R	Missense	VUS	Mat	L	OBE13 (text)	M	Serra-Juhé, et al. 2020
FTO	I492V	Missense	VUS	Pat	L	1863 (text)	M	Serra-Juhé, et al. 2020
FTO	I492V	Missense	VUS	Mat	L	1484 (text)	F	Serra-Juhé, et al. 2020
FTO	I492V	Missense	VUS	Mat	O	1698 (text)	M	Serra-Juhé, et al. 2020
SIM1	S348L	Missense	VUS	Mat	L	1931 (text)	F	Serra-Juhé, et al. 2020
SIM1	c.258+7G>T	Splicing	VUS	Mat	O	2069 (text)	F	Serra-Juhé, et al. 2020
MC4R	Y35X	Nonsense	Pathogenic	Mat	L	Patient 14 (table2)	F	Kleinendorst, et al. 2018

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	Y35X	Nonsense	Pathogenic	Mat	L	Patient 15 (table2)	F	Kleinendorst, et al. 2018
PCSK1	Y181H	Missense	VUS	Mat	O	Patient 31 (table2)	F	Kleinendorst, et al. 2018
POMC	P69fs	Frameshift	Pathogenic	Mat	O	Patient 32 (table2)	F	Kleinendorst, et al. 2018
POMC	R236G	Missense	Pathogenic	Mat	O	Patient 38 (table2)	F	Kleinendorst, et al. 2018
BDNF	S45R	Missense	VUS	Mat	OW	Patient 50 (tableS1)	F	Kleinendorst, et al. 2018
BDNF	T147X	Missense	VUS	Mat	L	Patient 51 (tableS1)	M	Kleinendorst, et al. 2018
CRHR2	R281H	Missense	VUS	Mat	OW	Patient 55 (tableS1)	M	Kleinendorst, et al. 2018
MC3R	C317F	Missense	VUS	Mat	OW	Patient 71 (tableS1)	F	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Pat	L	Patient 72 (tableS1)	M	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Mat	O	Patient 74 (tableS1)	M	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Mat	O	Patient 74m (tableS1)	M	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Mat	O	Patient 76 (tableS1)	F	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Pat	L	Patient 77 (tableS1)	F	Kleinendorst, et al. 2018
MCHR1	R317G	Missense	VUS	Mat	O	Patient 82 (tableS1)	F	Kleinendorst, et al. 2018
POMC	A56D	Missense	VUS	Pat	L	Patient 100 (tableS1)	F	Kleinendorst, et al. 2018
SIM1	R374Q	Missense	VUS	Pat	L	Patient 107 (tableS1)	F	Kleinendorst, et al. 2018
TBX3	Q231R	Missense	VUS	Mat	O	Patient 114 (tableS1)	F	Kleinendorst, et al. 2018
MC4R	I125K	Missense	Pathogenic	Mat	O	Case1 (figure2)	F	Hendricks, et al. 2017
MC4R	C271Y	Missense	Pathogenic	Mat	SO	Case2 (figure2)	F	Hendricks, et al. 2017
MC4R	D146G	Missense	Pathogenic	Mat	O	Case3a (figure2)	F	Hendricks, et al. 2017
MC4R	D146G	Missense	Pathogenic	Mat	NA	Case3b (figure2)	M	Hendricks, et al. 2017

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	E61K	Missense	Pathogenic	Mat	O	Case4a (figure2)	F	Hendricks, et al. 2017
MC4R	E61K	Missense	Pathogenic	Mat	O	Case4b (figure2)	M	Hendricks, et al. 2017
SH2B1	G59W	Missense	Pathogenic	Pat	OW	Case5 (figure2)	M	Hendricks, et al. 2017
SH2B1	G516S	Missense	Pathogenic	Mat	L	Case6 (figure2)	M	Hendricks, et al. 2017
SH2B1	R227C	Missense	Pathogenic	Pat	O	Case7 (figure2)	F	Hendricks, et al. 2017
SH2B1	R227C	Missense	Pathogenic	Pat	O	Case8 (figure2)	M	Hendricks, et al. 2017
SH2B1	A118V	Missense	Pathogenic	Pat	L	Case9 (figure2)	F	Hendricks, et al. 2017
SH2B1	A118V	Missense	Pathogenic	Mat	O	Case10a (figure2)	M	Hendricks, et al. 2017
SH2B1	A118V	Missense	Pathogenic	Mat	O	Case10b (figure2)	M	Hendricks, et al. 2017
SH2B1	A118V	Missense	Pathogenic	Mat	O	Case10c (figure2)	F	Hendricks, et al. 2017
BDNF	E183K	Missense	Pathogenic	Pat	O	Case11a (figure2)	M	Hendricks, et al. 2017
BDNF	E183K	Missense	Pathogenic	Pat	O	Case11b (figure2)	M	Hendricks, et al. 2017
BDNF	E183K	Missense	Pathogenic	Pat	O	Case11c (figure2)	F	Hendricks, et al. 2017
NTRK2	R696K	Missense	Pathogenic	Mat	SO	Case12 (figure2)	F	Hendricks, et al. 2017
NTRK2	P831L	Missense	Pathogenic	Mat	OW	Case14 (figure2)	F	Hendricks, et al. 2017
SIM1	L238R	Missense	Pathogenic	Mat	OW	Case15a (figure2)	M	Hendricks, et al. 2017
SIM1	L238R	Missense	Pathogenic	Mat	OW	Case15b (figure2)	F	Hendricks, et al. 2017
PCSK1	R80X	Nonsense	Pathogenic	Pat	O	Fam11 #III1-Patient (figure1)	F	Philippe, et al. 2015
PCSK1	R80X	Nonsense	Pathogenic	Mat	SO	Fam11 #II1-Father (figure1)	M	Philippe, et al. 2015
CADM2	R81H	Missense	Pathogenic	Mat	O	Fam9-Patient (figureS1)	M	Philippe, et al. 2015

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	L211fs (CTCTdel)	Frameshift	Pathogenic	Pat	SO	Patient II.1 (figure1)	M	Yeo, et al. 1998
MC4R	A244fs (GATTins)	Frameshift	Pathogenic	Mat	SO	FamDZ #III1-Proband (figure2)	F	Vaisse, et al. 1998
MC4R	A244fs (GATTins)	Frameshift	Pathogenic	Mat	SO	FamDZ #IV1-Niece (figure2)	F	Vaisse, et al. 1998
MC4R	L211fs (CTCTdel)	Frameshift	Pathogenic	Mat	O	Index A (text)	F	Hinney, et al. 1999
MC4R	Y35X	Nonsense	Pathogenic	Mat	SO	Index B (text)	F	Hinney, et al. 1999
MC4R	Y35X	Nonsense	Pathogenic	Mat	O	Index C (text)	F	Hinney, et al. 1999
MC4R	C279fs (GTins)	Frameshift	Pathogenic	Pat	SO	ID6 (figure2)	F	Farooqi, et al. 2000
MC4R	C271Y	Missense	Pathogenic	Mat	O	ID20 (figure2)	F	Farooqi, et al. 2000
MC4R	R165W	Missense	Pathogenic	Pat	EOO	Fam1143 #III.3-Proband (figure2)	F	Vaisse, et al. 2000
MC4R	R165W	Missense	Pathogenic	Mat	NA	Fam1143 #III.6-Cousin(figure2)	F	Vaisse, et al. 2000
MC4R	I170V	Missense	Pathogenic	Mat	OW	Fam1261 (figure2)	F	Vaisse, et al. 2000
MC4R	I301T	Missense	Pathogenic	Pat	O	Fam1283 (figure2)	F	Vaisse, et al. 2000
MC4R	W16fs (Gins47-48)	Frameshift	Pathogenic	Mat	L	Fam1417 (figure2)	F	Vaisse, et al. 2000
MC4R	S58C	Missense	Pathogenic	Pat	O	Fam1 (figure)	F	Dubern, et al. 2001
MC4R	V50M	Missense	Pathogenic	Pat	O	Fam2 (figure)	M	Dubern, et al. 2001
MC4R	I102S	Missense	Pathogenic	Mat	L	Fam3 (figure)	M	Dubern, et al. 2001

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	I170V	Missense	Pathogenic	Mat	O	Fam4 (figure)	M	Dubern, et al. 2001
MC4R	P48S	Missense	Pathogenic	Pat	O	Proband (figure2)	F	Miraglia Del Giudice, et al. 2002
MC4R	P48S	Missense	Pathogenic	Mat	O	Father (figure2)	M	Miraglia Del Giudice, et al. 2002
MC4R	P48S	Missense	Pathogenic	Pat	O	Cousin (figure2)	F	Miraglia Del Giudice, et al. 2002
MC4R	P48S	Missense	Pathogenic	Mat	O	Uncle (figure2)	M	Miraglia Del Giudice, et al. 2002
MC4R	S127L	Missense	Pathogenic	Mat	SO	Fam1 (figure2A)	M	Lubrano-Berthelier, et al. 2003
MC4R	S127L	Missense	Pathogenic	Pat	OW	Fam1m (figure2A)	F	Lubrano-Berthelier, et al. 2003
MC4R	A244E	Missense	Pathogenic	Pat	O	Fam2 (figure2B)	M	Lubrano-Berthelier, et al. 2003
MC4R	P299H	Missense	Pathogenic	Mat	O	Fam3a (figure2C)	M	Lubrano-Berthelier, et al. 2003
MC4R	P299H	Missense	Pathogenic	Pat	OW	Fam3b (figure2C)	F	Lubrano-Berthelier, et al. 2003
MC4R	W16X	Nonsense	Pathogenic	Mat	OW	Fam1 (figure1A)	F	Marti, et al. 2003
MC4R	112fs (Ains)	Frameshift	Pathogenic	Mat	EOO	Fam1 (figure2A)	M	Farooqi, et al. 2003
MC4R	112fs (Ains)	Frameshift	Pathogenic	Mat	EOO	Fam1-sister (figure2A)	F	Farooqi, et al. 2003
MC4R	L211fs (CTCTdel)	Frameshift	Pathogenic	Pat	EOO	Fam2 (figure2A)	M	Farooqi, et al. 2003
MC4R	C279fs (GTins)	Frameshift	Pathogenic	Pat	EOO	Fam3 (figure2A)	M	Farooqi, et al. 2003
MC4R	C279fs (GTins)	Frameshift	Pathogenic	Pat	EOO	Fam4 (figure2A)	F	Farooqi, et al. 2003
MC4R	I125K	Missense	Pathogenic	Mat	EOO	Fam5 (figure2A)	F	Farooqi, et al. 2003
MC4R	I125K	Missense	Pathogenic	Pat	EOO	Fam6 (figure2A)	M	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7 (figure2A)	F	Farooqi, et al. 2003

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-cousin1 (figure2A)	M	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-cousin2 (figure2A)	F	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-cousin3 (figure2A)	M	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-mother (figure2A)	F	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-aunt1 (figure2A)	F	Farooqi, et al. 2003
MC4R	C271Y	Missense	Pathogenic	Mat	EOO	Fam7-aunt2 (figure2A)	F	Farooqi, et al. 2003
MC4R	T11A	Missense	Pathogenic	Mat	EOO	Fam8 (figure2A)	M	Farooqi, et al. 2003
MC4R	T11A	Missense	Pathogenic	Mat	EOO	Fam8-sister (figure2A)	F	Farooqi, et al. 2003
MC4R	R165Q	Missense	Pathogenic	Pat	EOO	Fam9 (figure2A)	F	Farooqi, et al. 2003
MC4R	A175T	Missense	Pathogenic	Mat	EOO	Fam10 (figure2A)	M	Farooqi, et al. 2003
MC4R	I316S	Missense	Pathogenic	Pat	EOO	Fam11 (figure2A)	M	Farooqi, et al. 2003
MC4R	I316S	Missense	Pathogenic	Pat	EOO	Fam11-uncle (figure2A)	M	Farooqi, et al. 2003
MC4R	T112M	Missense	Pathogenic	Mat	SO	Fam2 (figure1)	F	Valli-Jaakola, et al. 2004
MC4R	S127L	Missense	Pathogenic	Pat	O	Fam4 (figure1)	M	Valli-Jaakola, et al. 2004
MC4R	I226T	Missense	Pathogenic	Mat	OW	Fam5 (figure1)	F	Valli-Jaakola, et al. 2004
MC4R	S30F	Missense	VUS	Pat	O	FamA (figure1)	M	Ochoa, et al. 2007
MC4R	T150I	Missense	Pathogenic	Pat	OW	FamB (figure1)	M	Ochoa, et al. 2007
MC4R	T162R	Missense	VUS	Pat	OW	FamC (figure1)	F	Ochoa, et al. 2007
MC4R	A244E	Missense	Pathogenic	Mat	L	FamD (figure1)	M	Ochoa, et al. 2007
MC4R	S19fs (delA)	Missense	VUS	Mat	L	Fam1a (figure1)	M	Hainerová, et al. 2007
MC4R	S19fs (delA)	Missense	VUS	Mat	O	Fam1b (figure1)	F	Hainerová, et al. 2007

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	S19fs (delA)	Missense	VUS	Mat	L	Fam2 (figure1)	F	Hainerová, et al. 2007
MC4R	S19fs (delA)	Missense	VUS	Mat	L	Fam2m (figure1)	F	Hainerová, et al. 2007
MC4R	C84R	Missense	Pathogenic	Pat	O	Fam3 (figure1)	F	Hainerová, et al. 2007
MC4R	C84R	Missense	Pathogenic	Pat	O	Fam3p (figure1)	M	Hainerová, et al. 2007
MC4R	R7C	Missense	Pathogenic	Pat	OW	Fam4 (figure1)	M	Hainerová, et al. 2007
MC4R	R7C	Missense	Pathogenic	Pat	O	Fam4p (figure1)	M	Hainerová, et al. 2007
MC4R	W16fs (47-48insG)	Frameshift	Pathogenic	Mat	L	Case1 (figureS1)	F	Stutzmann, et al. 2008
MC4R	I102T	Missense	Pathogenic	Mat	SO	Case2 (figureS1)	M	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Pat	O	Case3 (figureS1)	F	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Mat	O	Case3p (figureS1)	M	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Mat	O	Case4 (figureS1)	M	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Mat	L	Case5 (figureS1)	F	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Mat	O	Case5m (figureS1)	F	Stutzmann, et al. 2008
MC4R	T112M	Missense	Pathogenic	Mat	O	Case6 (figureS1)	M	Stutzmann, et al. 2008
MC4R	S127L	Missense	Pathogenic	Mat	SO	Case7 (figureS1)	M	Stutzmann, et al. 2008
MC4R	S127L	Missense	Pathogenic	Pat	OW	Case7m (figureS1)	F	Stutzmann, et al. 2008
MC4R	S127L	Missense	Pathogenic	Mat	SO	Case8 (figureS1)	F	Stutzmann, et al. 2008
MC4R	S127L	Missense	Pathogenic	Mat	O	Case8m (figureS1)	F	Stutzmann, et al. 2008
MC4R	T150I	Missense	Pathogenic	Pat	SO	Case9 (figureS1)	F	Stutzmann, et al. 2008
MC4R	T150I	Missense	Pathogenic	Mat	SO	Case10 (figureS1)	F	Stutzmann, et al. 2008

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	R165W	Missense	Pathogenic	Pat	O	Case11 (figureS1)	F	Stutzmann, et al. 2008
MC4R	I170V	Missense	Pathogenic	Mat	O	Case12 (figureS1)	F	Stutzmann, et al. 2008
MC4R	Del170V (CATdel170)	Frameshift	Pathogenic	Mat	OW	Case13 (figureS1)	F	Stutzmann, et al. 2008
MC4R	G252S	Missense	Pathogenic	Mat	SO	Case14 (figureS1)	F	Stutzmann, et al. 2008
MC4R	P260Q	Missense	Pathogenic	Mat	L	Case15 (figureS1)	F	Stutzmann, et al. 2008
MC4R	P260Q	Missense	Pathogenic	Pat	OW	Case16 (figureS1)	M	Stutzmann, et al. 2008
MC4R	P260Q	Missense	Pathogenic	Mat	OW	Case16p (figureS1)	M	Stutzmann, et al. 2008
MC4R	P299H	Missense	Pathogenic	Pat	O	Case17 (figureS1)	M	Stutzmann, et al. 2008
MC4R	P299H	Missense	Pathogenic	Mat	L	Case17p (figureS1)	M	Stutzmann, et al. 2008
MC4R	P299H	Missense	Pathogenic	Mat	O	Case18 (figureS1)	M	Stutzmann, et al. 2008
MC4R	P299H	Missense	Pathogenic	Pat	OW	Case18m (figureS1)	F	Stutzmann, et al. 2008
MC4R	R305S	Missense	Pathogenic	Mat	SO	Case19 (figureS1)	F	Stutzmann, et al. 2008
MC4R	R305S	Missense	Pathogenic	Mat	O	Case19m (figureS1)	F	Stutzmann, et al. 2008
MC4R	S127L	Missense	Pathogenic	Pat	O	Case 1 (figure1A)	F	Santoro, et al. 2009
MC4R	S127L	Missense	Pathogenic	Mat	OW	Case 1p (figure1A)	M	Santoro, et al. 2009
MC4R	S127L	Missense	Pathogenic	Pat	O	Case 2 (figure1B)	F	Santoro, et al. 2009
MC4R	S127L	Missense	Pathogenic	Mat	O	Case 3 (figure1C)	M	Santoro, et al. 2009
MC4R	Q307X	Missense	Pathogenic	Pat	O	Case 4 (figure2A)	F	Santoro, et al. 2009
MC4R	R7H	Missense	Pathogenic	Mat	SO	FamA (figure2)	M	Albuquerque, et al. 2014
MC4R	H76R	Missense	VUS	Pat	L	FamB (figure2)	F	Albuquerque, et al. 2014

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
MC4R	S127L	Missense	Pathogenic	Mat	SO	FamC (figure2)	F	Albuquerque, et al. 2014
MC4R	R147G	Missense	Pathogenic	Mat	SO	FamD (figure2)	M	Albuquerque, et al. 2014
MC4R	I251L	Missense	VUS	Pat	O	FamF (figure2)	F	Albuquerque, et al. 2014
MC4R	G323E	Missense	VUS	Pat	OW	FamG (figure2)	M	Albuquerque, et al. 2014
MC4R	S19fs (delA)	Frameshift	Pathogenic	Pat	O	MO56 (figure1)	F	Stanikova, et al. 2015
MC4R	S127L	Missense	Pathogenic	Mat	L	MO111 (figure1)	F	Stanikova, et al. 2015
MC4R	F152fs	Frameshift	Pathogenic	Pat	EOO	Case1 (text)	F	Vollbach, et al. 2016
MC4R	T101I	Missense	Pathogenic	Mat	O	Family1 (figure1A)	F	Gimeno-Ferrer, et al. 2019
MC4R	c.-24G>A	Promoter	Pathogenic	Pat	O	Family2 (figure1B)	M	Gimeno-Ferrer, et al. 2019
MC4R	c.-24G>A	Promoter	Pathogenic	Mat	O	Family2p (figure1B)	M	Gimeno-Ferrer, et al. 2019
MC4R	S30F	Missense	VUS	Mat	OW	Family3.1 (figure1C)	M	Gimeno-Ferrer, et al. 2019
MC4R	S30F	Missense	VUS	Pat	OW	Family3.2 (figure1C)	F	Gimeno-Ferrer, et al. 2019
MC4R	A259D	Missense	Pathogenic	Pat	OW	Family4 (figure1D)	F	Gimeno-Ferrer, et al. 2019
SH2B1	T175N	Missense	VUS	Mat	O	Case1 (table1)	F	Doche, et al. 2012
SH2B1	P322S	Missense	VUS	Mat	OW	Case2 (table1)	M	Doche, et al. 2012
SH2B1	F344fs	Frameshift	Pathogenic	Mat	SO	Case3 (table1)	F	Doche, et al. 2012
POMC	S7T	Missense	Pathogenic	Pat	OW	Case1 (figure1)	F	Miraglia Del Giudice, et al. 2001
POMC	S9L	Missense	Pathogenic	Pat	O	Case2 (figure2)	F	Miraglia Del Giudice, et al. 2001
POMC	R236G	Missense	Pathogenic	Pat	OW	Case3 (figure3)	F	Miraglia Del Giudice, et al. 2001
POMC	R236G	Missense	Pathogenic	Mat	O	Subject12-Patient (figure1B)	M	Challis, et al. 2002

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
POMC	R236G	Missense	Pathogenic	Mat	O	Subject3-Mother (figure1B)	F	Challis, et al. 2002
POMC	R236G	Missense	Pathogenic	Pat	O	Subject13-Cousin (figure1B)	M	Challis, et al. 2002
POMC	R236G	Missense	Pathogenic	Mat	O	Subject6-Uncle (figure1B)	M	Challis, et al. 2002
POMC	Y221C	Missense	Pathogenic	Pat	OW	Case1 (figure1B)	F	Lee, et al. 2006
POMC	Y221C	Missense	Pathogenic	Mat	O	Case2 (figure1B)	F	Lee, et al. 2006
POMC	Y221C	Missense	Pathogenic	Mat	O	Case3 (figure1B)	M	Lee, et al. 2006
POMC	Y221C	Missense	Pathogenic	Pat	OW	Case4 (figure1B)	F	Lee, et al. 2006
POMC	Y221C	Missense	Pathogenic	Mat	O	Case5 (figure1B)	F	Lee, et al. 2006
POMC	Y5C	Missense	Pathogenic	Pat	O	Case1 (figure1C)	F	Biebermann, et al. 2006
POMC	Y5C	Missense	Pathogenic	Mat	O	Case1p (figure1C)	M	Biebermann, et al. 2006
POMC	Y5C	Missense	Pathogenic	Mat	O	Case2 (figure1C)	F	Biebermann, et al. 2006
POMC	F144L	Missense	Pathogenic	Pat	O	Proband II1 (figure2)	F	Dubern, et al. 2008
POMC	C28F	Missense	VUS	Pat	OW	Case1 (figure1B)	M	Creemers, et al. 2008
POMC	L37F	Missense	VUS	Mat	OW	Case2 (figure1B)	F	Creemers, et al. 2008
NTRK2	I98V	Missense	VUS	Pat	OW	Case1 (text)	M	Gray, et al. 2007
NTRK2	T821A	Missense	VUS	Pat	L	Case2 (text)	M	Gray, et al. 2007
SIM1	I128T	Missense	VUS	Mat	L	Case1 (text)	F	Hung, et al. 2007
SIM1	T46R	Missense	Pathogenic	Mat	O	Participant1 (figure S2)	M	Bonnefond, et al. 2013
SIM1	T46R	Missense	Pathogenic	Mat	O	Participant2 (figure S2)	F	Bonnefond, et al. 2013

Gene	Variant	Type	Patho.	Origin	Progenitor phenotype	Case	Gender	Reference
SIM1	R171H	Missense	Pathogenic	Pat	SO	Case1 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	L238R	Missense	Pathogenic	Mat	SO	Case2 (figure2)	M	Ramachandrappa, et al. 2013
SIM1	P692L	Missense	Pathogenic	Pat	O	Case3 (figure2)	M	Ramachandrappa, et al. 2013
SIM1	T712I	Missense	Pathogenic	Pat	O	Case4 (figure2)	M	Ramachandrappa, et al. 2013
SIM1	R550H	Missense	Pathogenic	Mat	OW	Case5 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	R550H	Missense	Pathogenic	Pat	O	Case6 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	R550H	Missense	Pathogenic	Mat	SO	Case7 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	D707H	Missense	Pathogenic	Pat	O	Case8 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	D707H	Missense	Pathogenic	Mat	SO	Case9 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	D707H	Missense	Pathogenic	Mat	OW	Case10 (figure2)	F	Ramachandrappa, et al. 2013
SIM1	D707H	Missense	Pathogenic	Pat	O	Case11 (figure2)	F	Ramachandrappa, et al. 2013
MRAP2	A137T	Missense	VUS	Mat	L	Case1 (text)	M	Schonnop, et al. 2016
MRAP2	R125H	Missense	VUS	Mat	L	Case3 (text)	F	Schonnop, et al. 2016

Supplementary Table 2. Inherited rare sequence variants (RSVs) in candidate genes for obesity in patients with obesity. Each row corresponds to every transmission evaluated in the meta-analysis (n=210). Pathogenicity was assessed by functional studies or variants predicted to be deleterious Progenitor phenotype refers only to the progenitor carrying the alteration and includes, from less to more extreme: L=lean; OW=overweight, O=obesity, EOO=early-onset obesity, SO=severe obesity, NA=not available. The case identifier, the gender (M; male, F: female) and the reference are also included.

Hospital Infantil Universitario Niño Jesús
SaludMadrid
Comunidad de Madrid

Universitat Pompeu Fabra Barcelona

CUESTIONARIO

Este cuestionario ha sido diseñado en la Universidad Pompeu Fabra de Barcelona, en colaboración con el Hospital Infantil Universitario Niño Jesús de Madrid, en el marco de un proyecto de investigación sobre las causas genéticas de la obesidad de inicio precoz. Uno de los objetivos de este proyecto es definir la posible implicación de variantes genéticas y conducta alimentaria en el desarrollo de este trastorno.

Rellenando este formulario, autoriza el uso de los datos de forma anonimizada en dicho proyecto.
Muchas gracias por su colaboración.

Fecha:

Información sobre usted
Nombre:
Lugar de nacimiento:
Edad actual:
Edad al momento del nacimiento de su hijo/a:
Género: Hombre Mujer
Educación (marque el nivel más alto que haya completado):
__ Menos que la obligatoria
__ ESO, educación secundaria obligatoria
__ Bachillerato, BUP, COU
__ Formación profesional
__ Diplomatura universitaria
__ Licenciatura universitaria
__ Postgrado, máster, MIR, doctorado
Ocupación:
Estatura: metros (usar coma para los decimales)
Peso: kilogramos (usar coma para los decimales)
¿Tuvo obesidad o sobrepeso cuando era pequeño/a? SI NO

Información sobre su hijo/a
Nombre:
Fecha de nacimiento:
Edad actual:
Género: Hombre Mujer
Estatura: metros (usar coma para los decimales)
Peso: kilogramos (usar coma para los decimales)
¿Vive aun dentro la unidad familiar (o con padres)? SI NO (se marchó a los años)
Describa dicha unidad familiar: (p.ej.: padre, madre, hermano y abuela)

***En caso que su hijo/a no viva actualmente dentro la unidad familiar, responda a las preguntas del siguiente cuestionario pensando en cuando vivían juntos.**

CUESTIONARIO DE DIETA

Indique, por favor, la frecuencia con la que ha tomado cada alimento, en el último año, poniendo una X en el recuadro que mejor la defina:

1. Frecuencia diaria del consumo de los siguientes alimentos durante el último año:

Alimentos	Cantidad	Menos de 1 vez al día	1 vez al día	2 veces o más al día
Pan	1-2 rebanadas	☐	☐	☐
Verdura / Ensalada	1 plato/porción	☐	☐	☐
Fruta	1 pieza / porción	☐	☐	☐
Yogur o leche	1 tarrina / vaso	☐	☐	☐
Pasta o arroz	Un plato	☐	☐	☐
Aceite oliva / girasol	Una cucharada sopera	☐	☐	☐
Agua	3 vasos (250 ml/vaso)	☐	☐	☐

2. Frecuencia semanal del consumo de los siguientes alimentos durante el último año:

Alimentos	Cantidad	Menos de 4 veces a la semana	4-6 veces a la semana	7 veces o más a la semana
Carne	1 plato	☐	☐	☐
Embutidos	1-3 lonchas	☐	☐	☐
Queso	1 porción	☐	☐	☐
Bollería / Pastelería	1-2 piezas	☐	☐	☐
Mantequilla / Manteca	Una cucharada de café	☐	☐	☐
Bebida azucarada*	Un vaso o lata	☐	☐	☐
Comida rápida**	Una comida	☐	☐	☐

3. Frecuencia semanal del consumo de los siguientes alimentos durante el último año:

Alimentos	Cantidad	Menos de 2 veces a la semana	2-3 veces a la semana	4 veces o más a la semana
Pescado	1 plato	☐	☐	☐
Legumbres	1 plato / porción	☐	☐	☐
Frutos secos	1 puñado	☐	☐	☐

4. Frecuencia diaria del consumo de alcohol durante el último año:

Alimentos	Cantidad	Menos de 1 vez al día	1 vez al día	Más de 2 veces al día
Una bebida alcohólica	Un vaso / copa	☐	☐	☐

* Fanta, CocaCola, ... excepto gaseosa, tónica y bebida light
** McDonald, Burger King, ...

ESCALA DE ADAPTABILIDAD FAMILIAR Y EVALUACIÓN DE COHESIÓN (FACES)

A continuación se presenta una serie de enunciados relacionados con aspectos que se producen en las familias y entre los familiares. Indique, por favor, con qué frecuencia ocurren en su unidad familiar:

1.- Nunca o casi nunca 2.- Pocas veces 3.- A veces 4.- Con frecuencia 5.- Casi siempre

	1 2 3 4 5
1. Los miembros de la familia se sienten muy cercanos unos a otros.	1 2 3 4 5
2. Cuando hay que resolver problemas, se siguen las propuestas de los hijos.	1 2 3 4 5
3. En nuestra familia la disciplina (normas, obligaciones, consecuencias, castigos) es justa.	1 2 3 4 5
4. Los miembros de la familia asumen las decisiones que se toman de manera conjunta como familia.	1 2 3 4 5
5. Los miembros de la familia se piden ayuda mutuamente.	1 2 3 4 5
6. En cuanto a su disciplina, se tiene en cuenta la opinión de los hijos (normas, obligaciones).	1 2 3 4 5
7. Cuando surgen problemas, negociamos para encontrar una solución.	1 2 3 4 5
8. En nuestra familia hacemos cosas juntos.	1 2 3 4 5
9. Los miembros de la familia dicen lo que quieren libremente.	1 2 3 4 5
10. En nuestra familia nos reunimos todos juntos en la misma habitación (sala, cocina).	1 2 3 4 5
11. A los miembros de la familia les gusta pasar su tiempo libre juntos.	1 2 3 4 5
12. En nuestra familia, a todos nos resulta fácil expresar nuestra opinión.	1 2 3 4 5
13. Los miembros de la familia se apoyan unos a otros en los momentos difíciles.	1 2 3 4 5
14. En nuestra familia se intentan nuevas formas de resolver los problemas.	1 2 3 4 5
15. Los miembros de la familia comparten intereses y hobbies.	1 2 3 4 5
16. Todos tenemos voz y voto en las decisiones familiares importantes.	1 2 3 4 5
17. Los miembros de la familia se consultan unos a otros sus decisiones.	1 2 3 4 5
18. Los padres y los hijos hablan juntos sobre el castigo.	1 2 3 4 5
19. La unidad familiar es una preocupación principal.	1 2 3 4 5
20. Los miembros de la familia comentamos los problemas y nos sentimos muy bien con las soluciones encontradas.	1 2 3 4 5

CUESTIONARIO DE ALIMENTACIÓN INFANTIL (CFQ)

A continuación se presenta una serie de enunciados relacionados con creencias, actitudes y prácticas de los padres en relación a la alimentación infantil. Indique la respuesta que mejor describa cada declaración o pregunta (sobre usted).

Cuando su hijo/a está en casa:

1. ¿Con qué frecuencia es usted responsable de su alimentación?
   1 = nunca **2** = a veces **3** = la mitad de las veces **4** = la mayoría de las veces **5** = siempre
2. ¿Con qué frecuencia tiene usted la responsabilidad de decidir el tamaño de las porciones de su hijo/a?
   1 = nunca **2** = a veces **3** = la mitad de las veces **4** = la mayoría de las veces **5** = siempre
3. ¿Con qué frecuencia es usted responsable de decidir si su hijo/a está comiendo el tipo correcto de alimentos?
   1 = nunca **2** = a veces **3** = la mitad de las veces **4** = la mayoría de las veces **5** = siempre

Describa cuál cree que era su peso en las diferentes etapas de la vida:

4. Su infancia (5 a 10 años)
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
5. Su adolescencia
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
6. A los 20 años
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
7. Actualmente
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable

Describa cuál cree que era el peso de su hijo/a en las diferentes etapas de la vida:

8. Durante el primer año de vida (1 año)
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
9. Cuando era un/a niño/a pequeño/a (2-3 años)
   1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
10. Durante la educación infantil/preescolar (hasta los 5 años)
    1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
11. Durante la educación primaria (6-12 años)
    1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable
12. Durante la educación secundaria obligatoria (12-16 años)
    1 = peso notablemente bajo **2** = bajo peso **3** = normal **4** = sobrepeso **5** = sobrepeso notable

Describa su nivel de preocupación por las siguientes situaciones:

13. ¿Le preocupa que su hijo/a coma demasiado cuando usted no está presente?
    1 = nada **2** = un poco **3** = intermedio **4** = bastante **5** = mucho
14. ¿Le preocupa que su hijo/a deba realizar una dieta para mantener un peso deseable?
    1 = nada **2** = un poco **3** = intermedio **4** = bastante **5** = mucho
15. ¿Le preocupa que su hijo/a tenga sobrepeso?
    1 = nada **2** = un poco **3** = intermedio **4** = bastante **5** = mucho

Indique su grado de acuerdo respecto a las siguientes afirmaciones:

16. Debo asegurarme de que mi hijo/a no coma demasiados dulces (caramelos, helados o pasteles).
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
17. Debo asegurarme de que mi hijo/a no coma demasiados alimentos ricos en grasa.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
18. Debo asegurarme de que mi hijo/a no coma demasiado de sus comidas favoritas.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
19. Mantengo algunos alimentos fuera del alcance de mi hijo/a intencionalmente.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
20. Le ofrezco dulces (caramelos, helados, pasteles) como premio por un buen comportamiento.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
21. Le ofrezco a mi hijo/a sus alimentos favoritos a cambio de un buen comportamiento.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
22. Si no guiara o regulara la alimentación de mi hijo/a, comería demasiada comida basura.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
23. Si no guiara o regulara la alimentación de mi hijo/a, comería demasiado de sus comidas favoritas.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
24. Mi hijo/a siempre debe comer toda la comida que tiene en su plato.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
25. Debo tener especial cuidado para asegurarme de que mi hijo/a coma lo suficiente.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
26. Si mi hijo/a dice "no tengo hambre", intento que coma de todas formas.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo
27. Si no guiara o regulara la alimentación de mi hijo/a, comería mucho menos de lo que debería.
    1= en desacuerdo 2 = algo en desacuerdo 3 = neutral 4 = algo de acuerdo 5 = de acuerdo

Describa su grado de control sobre las siguientes situaciones:

28. ¿Me fijo en los dulces (caramelos, helados, pasteles) que come mi hijo/a?
    1 = nunca 2 = rara vez 3 = a veces 4 = casi siempre 5 = siempre
29. ¿Me fijo en las meriendas (patatas fritas, Doritos, bocaditos de queso) que come mi hijo/a?
    1 = nunca 2 = rara vez 3 = a veces 4 = casi siempre 5 = siempre
30. ¿Me fijo en los alimentos ricos en grasa que come mi hijo/a?
    1 = nunca 2 = rara vez 3 = a veces 4 = casi siempre 5 = siempre

Annex 1. CuestionarioESP. Spanish questionnaire used.

DISCUSSION

Extreme obesity cases to study genetic factors

Excessive fat accumulation represents the main feature observed in patients with obesity and is caused by a disequilibrium in energy homeostasis. High energy intake and low energy expenditure are mainly associated with high caloric diets and sedentariness, but lifestyle alone may not explain all obesity cases [10]. Several genetic factors have been described as causal for severe forms of obesity or as susceptibility factors for increased BMI. Most genetic alterations identified in patients with obesity affect genes from the leptin-melanocortin system, which is responsible for body weight regulation, but many other genes have been associated with obesity, such as genes involved in hypothalamic development, adipogenesis and cilia formation, among others [12]. All these variants define obesity as a very heterogeneous disease at the genetic level. As mentioned, some monogenic forms have also been described, but generally obesity is considered a multifactorial disease with contribution of both, genetic and environmental factors.

Elucidating the genetic factors contributing to a complex disease, such as obesity, may not always be straightforward. Familiar cases of severe obesity may be explained by shared genetic variants between family members, but usually obesity is manifested as a combination of multiple factors, which may be challenging when trying to study its effects independently. Obesity present in adult population is the best representation of this heterogeneity, where a combination of environmental and genetic factors play a significant role in its development. However, childhood obesity may be caused by factors inducing an earlier onset. Severe cases with an early-onset may be considered an extreme phenotype of the obesity spectrum mainly explained by genetic factors, as the environment has had no sufficient time to act. These extreme cases represents an excellent subpopulation for genetic studies.

Here, we studied a group of Spanish patients with severe early-onset obesity in order to unravel the molecular basis of obesity. Most genetic causes related with obesity development are very heterogeneous, and alterations in each individual gene may account for a little proportion of cases. Therefore, lots of cases are needed to prove the pathophysiological relevance of the variants detected. The collaboration with a reference center of pediatric endocrinology was of vital importance to recruit the patients and finally, more than 450 children were included in the genetic studies presented in this thesis. Moreover, to replicate the findings from our Spanish group we used genetic data from the *Viva la Familia* (VLF) study [139]. However, due to population

stratification, not only lots of patients are needed for genetic testing, but also control samples. Appropriate controls from the same geographic origin were used to discard population variants that might not be disease causative. All in all, a great number of samples from patients and control are required to deal with the genetic factors contributing to obesity development, a complex and highly heterogeneous disorder.

Pooled DNA sequencing to detect rare sequence variants

Genetic approaches trying to elucidate the molecular basis of complex disorders need lots of samples to be analyzed. In the study of rare sequence variants (RSVs) almost a thousand samples where sequenced, including 463 patients with obesity and 480 controls. Generating these huge amount of genetic data is not always affordable for all laboratories, so new approaches are being used to reduce costs by reducing the number of sequencing reactions without reducing the number of samples, such as pooled DNA sequencing [126]. In this technique the sequencing reaction is not performed in each sample individually, but to each pool, allowing the analysis of an elevated number of individuals in the same reaction. With 20 samples per pool we achieved a 10 times reduction in costs. Indeed, the savings are proportional to the number of samples in each pool, but excessive diluted samples may become undetectable. The optimal number of samples per pool is not known, but less than 5 samples per pool may not be cost-effective and more than 20 should be analyzed in each situation depending on the power of deep sequencing. In fact, in a pool with 20 samples, one heterozygous variant is detected in 1/40 reads (2.5%). To detect these low frequency variants in a reasonable way a huge coverage was mandatory and quality filters were needed to differentiate real variants from artifacts. Using the percentage of reads with the variant, the strand bias and the Phred score, we achieved a 100% specificity and 86% sensitivity. All in all, pooled DNA sequencing may have its limitations, such as the complexity to design and prepare the pools, the ability to detect only rare variants, the computational power needed for analysis, and the low sensitivity which may not recommend its use in the clinics; but it stands out by the important reduction of costs and the excellent specificity, which proves the high suitability of the pooling strategy to detect rare sequence variants in large groups of samples in a research context.

Penetrant genetic variants in patients with extreme obesity

Due to the high genetic heterogeneity and the low frequency of extreme obesity cases, it is very difficult to detect multiple patients with the same alteration. For that reason, extreme cases may be explained by low-frequency high-penetrant variants contributing to dysregulation of energy balance mechanisms. Here, we used different approaches to identify variants responsible for the phenotype in severe EOO patients. One targeted sequencing technology was used to identify point mutations in candidate genes, while other methods had a more genome-wide vision, such as SNP array platforms and exome sequencing.

Our first approach was to study the genetic contribution of a set of candidate genes for obesity. 15 genes were selected for sequencing, either already reported in obesity patients or with single nucleotide polymorphisms associated with obesity in genome-wide association studies. Focusing in those genes may increase the probability of identifying variants with a strong effect to the phenotype, but we may be missing alterations in other genes. Indeed, we identified RSVs in those genes in 10.4% of our obesity patients and 10.9% of patients from VLF, but only 6.5% of controls, which were statistically significant differences. In the group of Spanish patients, seven genes (*BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*) were responsible for the increased burden. Alterations in those genes were clearly over-represented in obesity patients, revealing its strong involvement with the phenotype. Moreover, 5.0% of obesity patients were carriers of likely pathogenic RSVs. Results from this study support the idea that highly penetrant RSVs in these genes are important contributors to part of the missing heritability of severe early-onset obesity.

Deficiencies in gens from the hypothalamic circuit that controls energy balance and regulates body weight, such as *LEP*, *LEPR*, *MC4R*, *POMC*, *PCSK1*, *PCSK2*, *BDNF*, *NTRK2* and *SIM1*, have been associated with severe forms of obesity [140]. Altogether, monogenic forms account for 5-10% of extreme obesity cases, but each gene individually may represent a small fraction of cases. Despite the global differences in the mutational burden between patients and controls, it was difficult to establish feasible genotype-phenotype correlations with specific genes due to the reduced number of carriers. When patients were grouped according to the gene mutated, differences didn't reach significant values, but some common clinical features were detected. *MC4R* gene, whose deficiency is the most frequent cause of

monogenic obesity [68], was the most mutated gene among our obesity patients (1.5%). Patients with *MC4R* mutations were characterized by extreme obesity, hyperphagia, overgrowth and insulin resistance. Moreover, three patients had mutations in *MC3R* gene, another melanocortin receptor with a critical role in regulating energy balance [141], but with milder obesity compared to *MC4R* mutation carriers. Patients with mutations affecting the *BDNF* gene presented mild to severe hyperphagia and insulin resistance. *SIM1* mutation carriers were moderately obese with insulin resistance and impaired glucose tolerance. Patients carrying mutation in *NEGR1*, *FTO* and *PPRAG* genes were also obese and may indicate a relevant role of these genes in obesity development, but more evidences will be needed to establish a causal relationship.

Secondly, we analyzed the presence of rare CNVs (rCNVs) along the genome to identify additional penetrant alterations beyond the known obesity genes studied, as in other complex disorders [100–102]. The differences between patients and controls were statistically significant, with 28.8% of the Spanish patients and 32.5% of VLF patients, compared to the 20.4% of controls harboring rCNVs. These results are in accordance with other studies where a higher burden of rCNVs has been detected in obesity patients [142]. The increase observed in our patients was attributable to gain-type rCNVs, whereas in VLF patients, the increase was observed in both, deletions and duplications. We identified rCNVs including genes involved in pathways like ciliogenesis, circadian rhythm, adipogenesis, inflammation, central nervous system, PLC activity and Ca^{2+} homeostasis. The clinical relevance of all the rCNVs detected needs to be further explored, but some of them may explain the phenotype observed in patients.

The genetic content of the rCNVs detected in patients has been used to replicate known genotype-phenotype correlations and to explore new candidate genes which may be linked to obesity development. A duplication of *NPY* gene or alterations affecting glutamate receptors (*GRM7* and *GRIK1*) represent some of the cases were new genetic factors may be explaining the phenotype of the family. Additional alteration affecting other genes may be contributing to obesity development, but the lack of functional studies represents a challenging step to establish new causal relationships. The genetic heterogeneity detected was remarkable, with more than 90% of the genes present in the rCNVs from Spanish patients found only in one alteration from one patient. Moreover, only 1% of the genes present in the rCNVs from Spanish patients were also included in rCNVs from VLF patients.

And finally, we used exome sequencing to study familiar cases not genetically elucidated and compatible with monogenic forms. Families with multiple cases of severe obesity following a clear inheritance pattern, may hide a shared variant that explains the phenotype. Some genetic alterations common among obesity family members have been detected in genes involved in neuronal development, ciliogenesis and intermediate filaments, but it was extremely difficult to establish clear pathogenic mechanisms to explain the role of these variants into the development of obesity. Indeed, it has been very challenging to prove the causality of the variants detected due to the complexity of the disease.

In summary, we reported rare variants in several candidate genes in patients with non-syndromic severe obesity that may be highly penetrant and could explain the phenotype present in some patients. The low genetic recurrence observed confirmed the high genetic heterogeneity between obesity patients. However, not all variants detected may be disease causative and further studies will be needed. The difficulty to assure the pathogenic role of the alterations detected was recurrent in our analysis, especially in rCNVs and variants from exome sequencing.

Susceptibility factors for obesity

Causal genetic alterations explaining the phenotype of patients are not always identified, indeed, only a small fraction of cases are molecularly elucidated. Point mutations have been widely studied in severe cases of obesity [143], and CNVs have been explored mainly in syndromic cases [144]. However, common genetic alterations may be acting as susceptibility factors and its additive effects may be explaining part of the phenotype from isolated obesity cases. In our study, we analyzed some of the genetic alterations known to be associated with BMI [108]. Despite not identifying statistically significant differences between patients and controls, these alterations may be contributing to the phenotype observed in the patients. Among the most frequent alterations in patients with obesity there where 16p13.11 duplications, 16p11.2 deletions, 10q26.3 duplications and some distal alterations at 22q11.2. Despite the extraordinary number of patients analyzed, it was insufficient to get enough statistical power for association studies. Further analysis with extended cohorts could explore the contribution of SNPs or inversions as additional sources of BMI variability. Some of the most

studied genetic predisposition factors include SNPs in *MC4R* and *FTO* genes [145] and inversions at 8p23.1 and 16p11.2 [146].

Maternally-inherited variants to explain childhood obesity

In adult population, the prevalence of obesity in women is higher than in men and this sex difference has remained remarkably constant over time. However, childhood obesity had a similar prevalence among both sexes, even with an inversed proportion respect to adult population [147]. Despite those observations, no genetic factors have been identified to justify these gender differences. Nowadays isolated obesity is considered a multifactorial disease affecting both sexes with no imprinting genes associated with. Only some syndromic forms have been related with imprinted genes, such as Beckwith-Wiedemann syndrome, Prader-Willi syndrome or pseudohypoparathyroidism. In this thesis, we analyzed the inheritance pattern of RSVs in genes related with obesity, such as *MC4R*, *NPY*, *BDNF*, *SIM1*, *PPARG*, *NEGR1* and *FTO*. These genes are responsible for energy homeostasis and have never been linked with imprinting mechanisms. Most of the variants were inherited from an overweight or obese progenitor, which went in line with other studies and represented a good evidence to link the variants detected with the development of obesity [12]. However, segregation studies of RSVs in these genes revealed an increased proportion of maternally inherited variants. A meta-analysis with more than 200 heterozygous variants in obesity-related genes has confirmed a statistically significant trend towards a maternal bias, with 2 out of 3 inherited variants being maternally inherited (62.4%). The analysis by gender or by subsets of candidate genes showed the same differences, so it really seemed a transversal observation including both sexes and different obesity-related genes, which has not been reported until now. However, stratification by the phenotype of the transmitting progenitor, showed even an increased proportion of maternally inherited variants among patients with severely obese progenitors (84%).

Mutations present in transmitting mothers may have a role in the development of obesity in her carrier children, so plausible explications have been explored. The increased fat accumulation present in the studied children had a clear early onset, but we wonder if it was already present before birth. For that reason, birth weight (BW) was checked in patients with maternally inherited variants and patients with non-maternally inherited variants. We observed a statistically significant increase of BW in patients with maternally inherited variants (+1.35

SDS) and an over-representation of children defined as obese at birth (46.2%). Patients with non-maternally inherited variants did not show these differences (-0.27 SDS and 0%), suggesting a specific role of the variants present in the mother to the phenotype before birth.

Mutations present in the transmitting mothers may predispose to specific behaviors contributing to the development of obesity in her children. Maternal attitudes towards food may induce infants to manifest obesity in an early manner, so we explored relevant conducts in transmitting mothers of children with obesity compared to control mothers (non-transmitting mothers of children with obesity and mothers of children without obesity). The main findings were regarding the perceived responsibility for feeding and the pressure to eat exercised on her children, but a general trend was also observed. Transmitting mothers had statistically significant lower scores in both scales, which were indicative of a more relaxed role regarding child feeding practices. These attitudes may be relevant in the development of obesity in their children as parental involvement during the first years is essential for children to develop properly their eating control mechanisms [148]. Parental careless may cause a dysregulation of eating habits in their children explaining the obesity phenotype. Despite the reduced sample size, significant levels were reached in these two subscales, and a clear trend was observed in the others subscales. Scores among transmitting mothers were always nearer to control mothers of children with obesity, and at the opposite side from control mothers of children without obesity. These observations may indicate that the presence of obesity in the children may affect the feeding attitudes from mothers, but a specific role of maternal variants should be considered due to the extreme values reported in transmitting mothers. When comparing transmitting mothers with both control groups, the cases were always in one extreme of the distribution, which reinforced the relevance of maternal variants in her feeding practices. Concern about child weight, restriction and monitoring of food intake followed this trend. Moreover, maternal diet and family functioning were also explored and no statistically significant differences were found between groups, although the same trend was observed in the dimensions evaluated in the family evaluation scale.

All in all, we reported a statistically significant overrepresentation of rare variants in candidate genes for obesity inherited from the mother in patients affected by severe early-onset obesity. Moreover, a synergistic effect of variants present in the mother and the child has been proposed to explain the early-onset obesity developed in children, including an increased birth weight

of the patient, an altered maternal perceived responsibility for child feeding and a low maternal implication in regulating food intake. Linking the maternal genotype with the child phenotype represents a relevant breakthrough that deserves to be further studied as may play an important role in the development of obesity.

Severe early-onset obesity: An integrative perspective

In this thesis a total of 489 patients with severe early-onset obesity have been genetically studied by a pooled DNA sequencing approach and/or by a SNP array platform. The vast majority of cases (90%) have been studied by both approaches and, as reported in the chapters of this thesis, RSVs and CNVs have been analyzed. Globally, we identified genetic alterations in 38.2% of the studied cases, with 1 every 4 patients harboring more than one alteration (Figure 1). On the one hand, the pooled DNA sequencing approach represented a targeted approximation in which alterations in 15 candidate genes have been explored. Patients harboring RSVs represented the 9.4% from all the cases. Despite being very restrictive in the criteria to detect RSVs, a significant proportion of cases have been identified to carry highly likely causative variants, which was indicative of the real contribution of the leptin-melanocortin pathway genes to the obesity phenotype. On the other hand, the SNP array approach represented a genome-wide approximation with a broader perspective focusing on rCNVs and known CNVs. Several rCNVs along the genome were detected in 26.8% of all cases. Pathogenic effects could not be attributed to all alterations, but the increased burden observed and the genes and pathways analyzed have revealed a valuable information in order to further explore this type of alterations in obesity patients. Moreover, susceptibility factors explained by known CNVs were identified in 7.6% of all cases. Some of these alterations have been associated with increased BMI and may be explaining part of the phenotype, other alterations may be acting as susceptibility factors but more research is needed to prove its real implications.

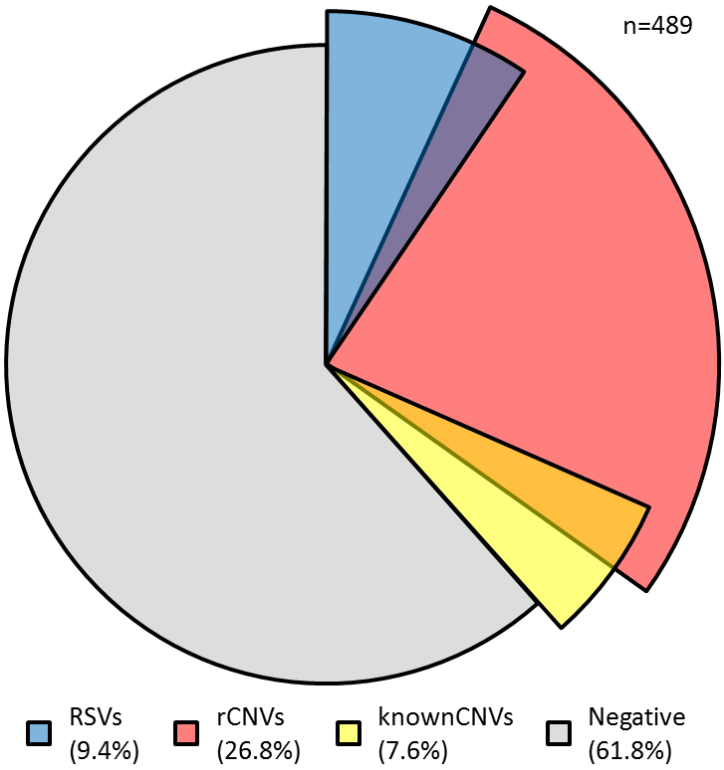

Figure 1. Percentage of obesity cases with genetic alterations from the 489 patients studied by at least one approach. The alterations identified were: rare sequence variants from the pooled DNA sequencing approach (RSVs, blue), rare copy number variants from SNP array (rCNVs, red) and known copy number variants acting as susceptibility factors from SNP array (knownCNVs, yellow). Overlapping between categories represent patients with multiple genetic alterations.

Despite the great number of variants detected, 61.8% of the patients studied had none of the searched genetic alterations. The big proportion of negative cases highlighted the complexity of obesity etiology and revealed that further research is needed to understand all its genetic factors. The great genetic heterogeneity and the interaction with environmental factors, makes obesity a challenging target for genetic studies. However, from the knowledge acquired during this thesis, the strategy that I would followed to tackle this disease at a genetic level is clear. First, the selection of extreme phenotypes is of great importance. Severe cases carefully characterized to discard other comorbidities are the basis for a great range of genetic studies. Reducing the variability at the phenotypic level is crucial to achieve conclusive results. Second, starting the genetic approach with a targeted sequencing of a small

group of genes it is a good approach to identify highly penetrant alterations with more chances to be actionable in the clinics. In this way, the most frequent altered genes can be screened in a cost-effective manner and some patients can quickly benefit from this information. Third, other approaches to expand the limits of the knowledge. It could include the study of more point mutations, CNVs, inversions or even epigenetics in order to identify new causes of obesity. These alterations must be studied in depth as its real contribution to the phenotype needs to be defined. And finally, functional studies. This thesis was focused in the discovery of genetic alterations, but did not explore the biological consequences of the variants identified. Getting involved in the functional analysis of some of the detected variants would have been fascinating. These type of analyses would be the final step in order to transfer to the clinics the knowledge acquired from research.

The molecular characterization of patients affected by obesity may represent a beneficial aspect in their therapeutic options. Usually, patients with obesity are treated with strict lifestyle interventions, including dietary control and promotion of physical activity [35]. However, in severe cases, additional measures are needed and pharmacotherapy or bariatric surgery may be used to limit weight gain or to ameliorate comorbidities. Most anti-obesity drugs are considered of limited efficacy and are usually prescribed as drug combinations and together with lifestyle interventions [149,150]. Moreover, patients with specific genetic defects may be treated with more directed interventions [39,40]. Indeed, the treatment of congenital leptin deficiency with leptin, represents a rare but powerful example of successful therapy emerging from a better understanding of the molecular mechanisms. To develop similar effective treatments, further research to understand these pathological processes is essential. However, this research is challenging and slow due to the great heterogeneity of the disorder. The identification of new genes related with obesity reported in this thesis, together with further investigation in novel therapies, may be an open door to more targeted therapeutic interventions for specific genetic defects.

Human obesity is a worldwide problem affecting an increasing proportion of the population. Excessive fat accumulation has been associated with a significant morbidity and a premature mortality mainly explained by cardiovascular disease and type 2 diabetes. The responsibility to deal with this condition needs to be addressed in a global perspective. Lifestyle attitudes have been proved as the basic modification to achieve weight loss, but some severe cases may need additional treatments. Despite high fat diets and

sedentary lifestyle are important contributors to the obesity development, the role of genetics should not be underestimated as it has been proved in several cases.

In summary, we have contributed to unravel the molecular basis of severe early-onset obesity using different genetic approaches. Genes, mainly from the leptin-melanocortin pathway, but also genes involved in other physiological processes, have been linked to obesity, which has been proved as multifactorial and genetically heterogeneous. Further studies including larger cohorts of phenotypically similar patients, together with functional analysis, will provide additional data to definitively establish the role of genetics in obesity development.

CONCLUSIONS

1. Patients affected by severe EOO represent an extreme subgroup from the obesity spectrum very valuable for genetic studies. It is of great importance the collaboration with pediatric endocrinology centers to obtain biological samples from children with isolated non-syndromic obesity and ensure a homogenous study group. The inclusion of big cohorts of subjects is key to analyze the contribution of genetic factors in genetically heterogeneous disorders, such as obesity.

2. We have proven that pooled DNA sequencing is as an efficient approach to identify RSVs in large groups of samples in a research context. A high coverage, a specific percentage of reads with the variant, a low strand bias and a minimum Phred score were identified as the crucial points to detect real variants with this strategy, which allowed a significant reduction of the sequencing costs.

3. We have identified RSVs in candidate genes for obesity in 10% of EOO patients with half of them (5%) carrying variants reported or predicted as pathogenic. In the VLF replication group we also detected an increased proportion of patients with RSVs affecting candidate genes for obesity compared to non-obese controls (statistically significant in both groups of patients).

4. Seven genes (*BDNF*, *FTO*, *MC3R*, *MC4R*, *NEGR1*, *PPARG* and *SIM1*) were responsible for the increased burden in the group of Spanish patients. Highly penetrant RSVs in these genes may be important contributors to part of the missing heritability of severe early-onset obesity, reinforcing the role of the leptin-melanocortin pathway in the development of the disease.

5. A statistically significant higher frequency of gain-type rCNVs was detected in Spanish patients with EOO compared to controls (28.8% vs 20.4%). Patients from VLF study also had an increased burden of rCNVs (32.5%) both, deletions and duplications.

6. Analyzing the genetic content of the rCNVs detected, we identified known and new candidate genes for obesity. We reported alterations affecting genes already associated with obesity development (*NPY* and *SH2B1*) and genes identified as promising candidates (*GRPR*, *GRM7*, *GRIK1*). Glutamate receptors may represent a functional link between neuronal regulation and obesity.

7. Genetic alterations affecting genes related with ciliogenesis, circadian rhythm, adipogenesis, inflammation, central nervous system, PLC activity and Ca^{2+} homeostasis deserve further investigation. The great diversity of genes and pathways identified highlighted the high genetic heterogeneity among EOO patients.

8. The duplication in 16p13.11 may be a susceptibility factor for increased BMI as it was detected more frequently in patients than controls, but further evidences will be needed. Additional alterations, including 16p11.2 deletions, 10q26.3 duplications and 22q11.2 alterations, may explain part of a multifactorial model with other genetic or environmental factors involved. More extended cohorts will allow a bigger statistical power for association studies.

9. The use of exome sequencing in selected families with cases of EOO has been challenging. From the two families studied, we identified rare candidate variants in genes involved in neuronal development (*GPRIN1*, *BDNF*), ciliogenesis (*CFAP77*) and intermediate filaments (*KRT24*). Establishing a causal relationship between the variants identified and the phenotype was very difficult, highlighting the complexity to analyze this type of families.

10. We reported a statistically significant enrichment of maternally inherited variants affecting obesity candidate genes in patients with EOO. From all the inherited RSVs studied, 62.4% were shared between the patient and the mother. The proportion of maternally inherited RSVs was even increased when analyzing variants inherited from a parent affected by severe obesity (84.0%). This divergence was observed independently of the gene analyzed or the gender of the patient, being a transversal observation not reported until now.

11. A synergistic effect of variants present in the mother and the child has been proposed. A prenatal effect has been explored to investigate if the onset of obesity was present even before birth. Patients with maternally inherited variants had a statistically significant increase of BW (+1.35 SDS) and were more frequently classified as obese at birth (46.2%). No differences were observed in patients with non-maternally inherited variants.

12. Postnatal effects may be explained by altered feeding practices. Transmitting mothers had a statistically significant lower perceived responsibility for child feeding and exercised a lower pressure to eat to her children, which may disrupt eating control mechanisms developed during the

first years of life. A clear trend was also observed among other feeding behaviors analyzed in transmitting mothers.

13. A total of 489 patients with severe EOO have been genetically studied. Genetic alterations were found in 38.2% of cases. In this thesis we have contributed to unravel the complexity of obesity etiology, but further research is still needed to understand all its molecular basis.

REFERENCES

1. Wells J, Fewtrell M. Measuring body composition. *Arch Dis Child*. 2006;91(7):612-617. doi:10.1136/adc.2005.085522
2. Peltz G, Aguirre MT, Sanderson M, Fadden MK. The role of fat mass index in determining obesity. *Am J Hum Biol*. 2010;22(5):639-647. doi:10.1002/ajhb.21056
3. Engin A. The Definition and Prevalence of Obesity and Metabolic Syndrome. *Adv Exp Med Biol*. 2017;960:1-17. doi:10.1007/978-3-319-48382-5_1
4. Freedman DS, Katzmarzyk PT, Dietz WH, Srinivasan SR, Berenson GS. Relation of body mass index and skinfold thicknesses to cardiovascular disease risk factors in children: the Bogalusa Heart Study. *Am J Clin Nutr*. 2009;90(1):210-216. doi:10.3945/ajcn.2009.27525
5. Willett K, Jiang R, Lenart E, Spiegelman D, Willett W. Comparison of bioelectrical impedance and BMI in predicting obesity-related medical conditions. *Obesity*. 2006;14(3):480-490. doi:10.1038/oby.2006.63
6. Chooi YC, Ding C, Magkos F. The epidemiology of obesity. *Metabolism*. 2019;92:6-10. doi:10.1016/j.metabol.2018.09.005
7. Kelly T, Yang W, Chen C-S, Reynolds K, He J. Global burden of obesity in 2005 and projections to 2030. *Int J Obes*. 2008;32(9):1431-1437. doi:10.1038/ijo.2008.102
8. Hernáez Á, Zomeño MD, Dégano IR, et al. Excess Weight in Spain: Current Situation, Projections for 2030, and Estimated Direct Extra Cost for the Spanish Health System. *Rev Española Cardiol (English Ed*. 2019;72(11):916-924. doi:10.1016/j.rec.2018.10.010
9. World Health Organization. *Taking Action on Childhood Obesity*.; 2018.
10. Heymsfield SB, Wadden TA. Mechanisms, Pathophysiology, and Management of Obesity. *N Engl J Med*. 2017;376(3):254-266. doi:10.1056/NEJMra1514009
11. Bray MS, Loos RJF, McCaffery JM, et al. NIH working group report-using genomic information to guide weight management: From universal to precision treatment. *Obesity*. 2016;24(1):14-22. doi:10.1002/oby.21381
12. Friedman JM. Obesity in the new millennium. *Nature*. 2000;404:632-634.
13. Sutton AK, Myers Jr MG, Olson DP. The Role of PVH Circuits in Leptin Action and Energy Balance. *Annu Rev Physiol*. 2016;78:207-221. doi:10.1146/annurev-physiol-021115-105347
14. Timper K, Brüning JC. Hypothalamic circuits regulating appetite and energy homeostasis: pathways to obesity. *Dis Model Mech*. 2017;10(6):679-389. doi:10.1242/dmm.026609
15. Baldini G, Phelan KD. The melanocortin pathway and control of appetite-progress and therapeutic implications. *J Endocrinol*. 2019;241(1):R1-R33. doi:10.1530/JOE-18-0596
16. Lawrence VJ, Kopelman PG. Medical consequences of obesity. *Clin*

Dermatol. 2004;22(4):296-302. doi:10.1016/j.clindermatol.2004.01.012
17. Mokdad AH, Ford ES, Bowman BA, et al. Prevalence of obesity, diabetes, and obesity-related health risk factors, 2001. *JAMA*. 2003;289(1):76-79. doi:10.1001/jama.289.1.76
18. Pi-Sunyer X. The Medical Risks of Obesity. *Postgrad Med*. 2009;121(6):21-33. doi:10.3810/pgm.2009.11.2074
19. Centers for Disease Control and Prevention. *National Diabetes Statistics Report 2020: Estimates of Diabetes and Its Burden in the United States*.; 2020.
20. Calle EE, Rodriguez C, Kimberly W-T, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. *N Engl J Med*. 2003;348(17):1625-1638. doi:10.1056/NEJMoa021423
21. Tchkonia T, Thomou T, Zhu Y, et al. Mechanisms and metabolic implications of regional differences among fat depots. *Cell Metab*. 2013;17(5):644-656. doi:10.1016/j.cmet.2013.03.008
22. McCullough AJ. The clinical features, diagnosis and natural history of nonalcoholic fatty liver disease. *Clin Liver Dis*. 2004;8(3):521-533. doi:10.1016/j.cld.2004.04.004
23. Hall JE, da Silva AA, do Carmo JM, et al. Obesity-induced hypertension: role of sympathetic nervous system, leptin, and melanocortins. *J Biol Chem*. 2010;285(23):17271-17276. doi:10.1074/jbc.R110.113175
24. Kovesdy P C, Furth L S, Zoccali C. Obesity and Kidney Disease: Hidden Consequences of the Epidemic. *Can J kidney Heal Dis*. 2017;4:1-10. doi:10.1177/2054358117698669
25. Zammit C, Liddicoat H, Moonsie I, Makker H. Obesity and respiratory diseases. *Int J Gen Med*. 2010;3:335-343. doi:10.2147/IJGM.S11926
26. Ashrafian H, Toma T, Rowland SP, et al. Bariatric Surgery or Non-Surgical Weight Loss for Obstructive Sleep Apnoea? A Systematic Review and Comparison of Meta-analyses. *Obes Surg*. 2015;25(7):1239-1250. doi:10.1007/s11695-014-1533-2
27. Hampel H, Abraham NS, El-Serag HB. Meta-analysis: obesity and the risk for gastroesophageal reflux disease and its complications. *Ann Intern Med*. 2005;143(3):199-211. doi:10.7326/0003-4819-143-3-200508020-00006
28. Khatua B, El-Kurdi B, Singh VP. Obesity and pancreatitis. *Curr Opin Gastroenterol*. 2017;33(5):374-382. doi:10.1097/MOG.0000000000000386
29. Erlinger S. Gallstones in obesity and weight loss. *Eur J Gastroenterol Hepatol*. 2000;12(12):1347-1352. doi:10.1097/00042737-200012120-00015
30. Goldring MB, Otero M. Inflammation in osteoarthritis. *Curr Opin Rheumatol*. 2011;23(5):471-478. doi:10.1097/BOR.0b013e328349c2b1
31. Özcan Dağ Z, Dilbaz B. Impact of obesity on infertility in women. *J Turkish Ger Gynecol Assoc*. 2015;16(2):111-117.

doi:10.5152/jtgga.2015.15232
32. Amin MN, Hussain MS, Sarwar MS, et al. How the association between obesity and inflammation may lead to insulin resistance and cancer. *Diabetes Metab Syndr*. 2019;13(2):1213-1224. doi:10.1016/j.dsx.2019.01.041
33. Must A, McKeown NM. The Disease Burden Associated with Overweight and Obesity. *Endotex*. 2010.
34. Berkowitz RI, Fabricatore AN. Obesity, psychiatric status, and psychiatric medications. *Physchiatric Clin North Am*. 2011;34(4):747-764. doi:10.1016/j.psc.2011.08.007
35. Yumuk V, Tsigos C, Fried M, et al. European Guidelines for Obesity Management in Adults. *Obes Facts*. 2015;8(6):402-424. doi:10.1159/000442721
36. Willumsen J, Bull F. Development of WHO Guidelines on Physical Activity, Sedentary Behavior, and Sleep for Children Less Than 5 Years of Age. *J Phys Act Health*. 2020;17(1):96-100. doi:10.1123/jpah.2019-0457
37. Saunders KH, Umashanker D, Igel LI, Kumar RB, Aronne LJ. Obesity Pharmacotherapy. *Med Clin North Am*. 2018;102(1):135-148. doi:10.1016/j.mcna.2017.08.010
38. Srivastava G, Apovian C. Future Pharmacotherapy for Obesity: New Anti-obesity Drugs on the Horizon. *Curr Obes Rep*. 2018;7(2):147-161. doi:10.1007/s13679-018-0300-4
39. Licinio J, Caglayan S, Ozata M, et al. Phenotypic effects of leptin replacement on morbid obesity, diabetes mellitus, hypogonadism, and behavior in leptin-deficient adults. *Proc Natl Acad Sci U S A*. 2004;101(13):4531-4536. doi:10.1073/pnas.0308767101
40. Collet T-H, Dubern B, Mokrosinski J, et al. Evaluation of a melanocortin-4 receptor (MC4R) agonist (Setmelanotide) in MC4R deficiency. *Mol Metab*. 2017;6(10):1321-1329. doi:10.1016/j.molmet.2017.06.015
41. Pories WJ. Bariatric surgery: risks and rewards. *J Clin Endocrinol Metab*. 2008;93(22 Supl 1):S89-S96. doi:10.1210/jc.2008-1641
42. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet*. 2017;390(10113):2627-2642. doi:10.1016/S0140-6736(17)32129-3
43. Spinelli A, Buoncristiano M, Kovacs VA, et al. Prevalence of Severe Obesity among Primary School Children in 21 European Countries. *Obes Facts*. 2019;12(2):244-258. doi:10.1159/000500436
44. Kumar S, Kelly AS. Review of Childhood Obesity: From Epidemiology, Etiology, and Comorbidities to Clinical Assessment and Treatment. *Mayo Clin Proc*. 2017;92(2):251-265. doi:10.1016/j.mayocp.2016.09.017

45. Lobstein T, Baur L. Obesity in children and young people: a crisis in public health. *Obes Rev*. 2004;5 Suppl 1:4-104. doi:10.1111/j.1467-789X.2004.00133.x
46. Simmonds M, Llewellyn A, Owen CG, Woolacott N. Predicting adult obesity from childhood obesity: a systematic review and meta-analysis. *Obes Rev*. 2016;17(2):95-107. doi:10.1111/obr.12334
47. Styne DM, Arslanian SA, Connor EL, et al. Pediatric Obesity-Assessment, Treatment, and Prevention: An Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2017;102(3):709-757. doi:10.1210/jc.2016-2573
48. Grandone A, Di Sessa A, Umano GR, Toraldo R, Miraglia Del Giudice E. New treatment modalities for obesity. *Best Pract Res Clin Endocrinol metamolism*. 2018;32(4):535-549. doi:10.1016/j.beem.2018.06.007
49. Proença da Fonseca AC, Mastronardi C, Johar A, Arcos-Burgos M, Paz-Filho G. Genetics of non-syndromic childhood obesity and the use of high-throughput DNA sequencing technologies. *J Diabetes Complications*. 2017;31(10):1549-1561. doi:10.1016/j.jdiacomp.2017.04.026
50. Kleinendorst L, van Haelst MM, van den Akker ELT. Genetics of Obesity. *Exp Suppl*. 2019;111:419-441. doi:10.1007/978-3-030-25905-1_19
51. Kaur Y, de Souza RJ, Gibson WT, Meyre D. A systematic review of genetic syndromes with obesity. *Obes Rev*. 2017;18(6):603-634. doi:10.1111/obr.12531
52. Horsthemke B, Wagstaff J. Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet*. 2008;146A(16):2041-2052. doi:10.1002/ajmg.a.32364
53. Nicholls RD, Knoll JH, Butler MG, Karam S, Lalande M. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. *Nature*. 1989;342(6247):281-285. doi:10.1038/342281a0
54. Beales PL, Elcioglu N, Woolf AS, Parker D, Flinter FA. New criteria for improved diagnosis of Bardet-Biedl syndrome: results of a population survey. *J Med Genet*. 1999;36(6):437-446.
55. Khan SA, Muhammad N, Khan MA, Kamal A, Rehman ZU, Khan S. Genetics of human Bardet-Biedl syndrome, an updates. *Clin Genet*. 2016;90(1):3-15. doi:10.1111/cge.12737
56. Katsanis N, Ansley SJ, Badano JL, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. *Science (80-)*. 2001;293(5538):2256-2259. doi:10.1126/science.1063525
57. Collin GB, Marshall JD, Ikeda A, et al. Mutations in ALMS1 cause obesity, type 2 diabetes and neurosensory degeneration in Alström syndrome. *Nat Genet*. 2002;31(1):74-78. doi:10.1038/ng867
58. Temtamy SA. Carpenter's syndrome: acrocephalopolysyndactyly. An autosomal recessive syndrome. *J Pediatr*. 1966;69(1):111-120.

doi:10.1016/s0022-3476(66)80368-2
59. Duplomb L, Duvet S, Picot D, et al. Cohen syndrome is associated with major glycosylation defects. *Hum Mol Genet*. 2014;23(9):2391-2399. doi:10.1093/hmg/ddt630
60. Weksberg R, Shuman C, Beckwith JB. Beckwith-Wiedemann syndrome. *Eur J Hum Genet*. 2010;18(1):8-14. doi:10.1038/ejhg.2009.106
61. Demars J, Gicquel C. Epigenetic and genetic disturbance of the imprinted 11p15 region in Beckwith-Wiedemann and Silver-Russell syndromes. *Clin Genet*. 2012;81(4):350-361. doi:10.1111/j.1399-0004.2011.01822.x
62. Martos-Moreno GÁ, Serra-Juhé C, Pérez-Jurado LA, Argente J. Underdiagnosed Beckwith-Wiedemann syndrome among early onset obese children. *Arch Dis Child*. 2014;99(10):965-967. doi:10.1136/archdischild-2014-307097
63. Han JC, Liu Q-R, Jones M, et al. Brain-derived neurotrophic factor and obesity in the WAGR syndrome. *N Engl J Med*. 2008;359(9):918-927. doi:10.1056/NEJMoa0801119
64. Mantovani G, Spada A. Mutations in the Gs alpha gene causing hormone resistance. *Best Pract Res Clin Endocrinol metamolism*. 2006;20(4):501-513. doi:10.1016/j.beem.2006.09.001
65. Paolacci S, Borrelli A, Stuppia L, et al. Mendelian obesity, molecular pathways and pharmacological therapies: a review. *Eur Rev Med Pharmacol Sci*. 2019;23(3):1357-1378. doi:10.26355/eurrev_201902_17031
66. Montague CT, Farooqi IS, P WJ, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature*. 1997;387(6636):903-908. doi:10.1038/43185
67. Clément K, Vaisse C, Lahlou N, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. *Nature*. 1998;392(6674):398-401. doi:10.1038/32911
68. Farooqi IS, Keogh JM, Yeo GSH, 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(12):1085-1095. doi:10.1056/NEJMoa1109400
69. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. *Nat Genet*. 1998;19(2):155-157. doi:10.1038/509
70. Jackson RS, Creemers JW, Ohagi S, et al. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat Genet*. 1997;16(3):303-306. doi:10.1038/ng0797-303
71. Nicholson JR, Peter J-C, Lecourt A-C, Barde Y-A, Hofbauer KG. Melanocortin-4 receptor activation stimulates hypothalamic brain-

derived neurotrophic factor release to regulate food intake, body temperature and cardiovascular function. *J Neuroendocrinol*. 2007;19(12):974-982. doi:10.1111/j.1365-2826.2007.01610.x

72. Ren D, Zhou Y, Morris D, Li M, Li Z, Rui L. Neuronal SH2B1 is essential for controlling energy and glucose homeostasis. *J Clin Invest*. 2007;117(2):397-406. doi:10.1172/JCI29417
73. Revelli J-P, Smith D, Allen J, et al. Profound obesity secondary to hyperphagia in mice lacking kinase suppressor of ras 2. *Obesity*. 2011;19(5):1010-1018. doi:10.1038/oby.2010.282
74. Hardie DG, Ross FA, Hawley SA. AMPK: a nutrient and energy sensor that maintains energy homeostasis. *Nat Rev Mol Cell Biol*. 2012;13(4):251-262. doi:10.1038/nrm3311
75. Gray J, Yeo GSH, Cox JJ, et al. Hyperphagia, severe obesity, impaired cognitive function, and hyperactivity associated with functional loss of one copy of the brain-derived neurotrophic factor (BDNF) gene. *Diabetes*. 2006;55(12):3366-3371. doi:10.2337/db06-0550
76. Harcourt BE, Bullen DVR, Kao K-T, et al. Maternal inheritance of BDNF deletion, with phenotype of obesity and developmental delay in mother and child. *Am J Med Genet Part A*. 2018;176(1):194-200. doi:10.1002/ajmg.a.38539
77. Yeo GSH, Hung C-CC, Rochford J, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. *Nat Neurosci*. 2004;7(11):1187-1189. doi:10.1038/nn1336
78. Villa A, Urioste M, Bofarull JM, Martínez-Frías ML. De novo interstitial deletion q16.2q21 on chromosome 6. *Am J Med Genet*. 1995;55(3):379-383. doi:10.1002/ajmg.1320550326
79. Holder Jr JL, Butte NF, Zinn AR. Profound obesity associated with a balanced translocation that disrupts the SIM1 gene. *Hum Mol Genet*. 2000;9(1):101-108. doi:10.1093/hmg/9.1.101
80. Ramachandrappa S, Raimondo A, Cali AMG, et al. Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest*. 2013;123(7):3042-3050. doi:10.1172/JCI68016
81. Faivre L, Cormier-Daire V, Lapierre JM, et al. Deletion of the SIM1 gene (6q16.2) in a patient with a Prader-Willi-like phenotype. *J Med Genet*. 2002;39(8):594-596. doi:10.1136/jmg.39.8.594
82. Bonnefond A, Raimondo A, Stutzmann F, et al. Loss-of-function mutations in SIM1 contribute to obesity and Prader-Willi-like features. *J Clin Invest*. 2013;123(7):3037-3041. doi:10.1172/JCI68035
83. Doche ME, Bochukova EG, Su H-W, et al. Human SH2B1 mutations are associated with maladaptive behaviors and obesity. *J Clin Invest*. 2012;122(12):4732-4736. doi:10.1172/JCI62696
84. Pearce LR, Joe R, Doche ME, et al. Functional characterization of obesity-associated variants involving the α and β isoforms of human SH2B1. *Endocrinology*. 2014;155(9):3219-3226. doi:10.1210/en.2014-

1264
85. Pearce LR, Atanassova N, Banton MC, et al. KSR2 mutations are associated with obesity, insulin resistance, and impaired cellular fuel oxidation. *Cell*. 2013;155(4):765-777. doi:10.1016/j.cell.2013.09.058
86. Saeed S, Bonnefond A, Tamanini F, et al. Loss-of-function mutations in ADCY3 cause monogenic severe obesity. *Nat Genet*. 2018;50(2):175-179. doi:10.1038/s41588-017-0023-6
87. Smemo S, Tena JJ, Kim K-H, et al. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*. 2014;507(7492):371-375. doi:10.1038/nature13138
88. Claussnitzer M, Dankel SN, Kim K-H, et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med*. 2015;373(10):895-907. doi:10.1056/NEJMoa1502214
89. Zou Y, Lu P, Shi J, et al. IRX3 Promotes the Browning of White Adipocytes and Its Rare Variants are Associated with Human Obesity Risk. *EBioMedicine*. 2017;24:64-75. doi:10.1016/j.ebiom.2017.09.010
90. Cummings DE, Schwartz MW. Melanocortins and body weight: a tale of two receptors. *Nat Genet*. 2000;26(1):8-9. doi:10.1038/79223
91. Ehtesham S, Qasim A, Meyre D. Loss-of-function mutations in the melanocortin-3 receptor gene confer risk for human obesity: A systematic review and meta-analysis. *Obes Rev*. 2019;20(8):1085-1092. doi:10.1111/obr.12864
92. Lesch K-P, Selch S, Renner TJ, et al. Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. *Mol Psychiatry*. 2011;16(5):491-503. doi:10.1038/mp.2010.29
93. Serra-Juhé C, Martos-Moreno GÁ, Bou de Pieri F, et al. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet*. 2017;13(5):e1006657. doi:10.1371/journal.pgen.1006657
94. Ristow M, Müller-Wieland D, Pfeiffer A, Krone W, Kahn CR. Obesity associated with a mutation in a genetic regulator of adipocyte differentiation. *N Engl J Med*. 1998;339(14):953-959. doi:10.1056/NEJM199810013391403
95. Wang XL, Oosterhof J, Duarte N. Peroxisome proliferator-activated receptor gamma C161>T polymorphism and coronary artery disease. *Cardiovasc Res*. 1999;44(3):588-594. doi:10.1016/s0008-6363(99)00256-4
96. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015;518(7538):197-206. doi:10.1126/science.1249098.Sleep
97. Loos RJF, Yeo GSH. The bigger picture of FTO: the first GWAS-identified obesity gene. *Nat Rev Endocrinol*. 2014;10(1):51-61. doi:10.1038/nrendo.2013.227
98. Waalen J. The genetics of human obesity. *Transl Res*. 2014;164(4):293-

301. doi:10.1016/j.trsl.2014.05.010
99. Argyropoulos G, Rankinen T, Neufeld DR, et al. A polymorphism in the human agouti-related protein is associated with late-onset obesity. *J Clin Endocrinol Metab*. 2002;87(9):4198-4202. doi:10.1210/jc.2002-011834
100. Glessner JT, Wang K, Cai G, et al. Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature*. 2009;459(7246):569-573. doi:10.1038/nature07953
101. Stefansson H, Rujescu D, Cichon S, et al. Large recurrent microdeletions associated with schizophrenia. *Nature*. 2008;455(September):232-237. doi:10.1038/nature07229
102. de Vries BBA, Pfundt R, Leisink M, et al. Diagnostic genome profiling in mental retardation. *Am J Hum Genet*. 2005;77(4):606-616. doi:10.1086/491719
103. Wheeler E, Huang N, Bochukova EG, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. *Nat Genet*. 2013;45(5):513-517. doi:10.1038/ng.2607
104. Willer CJ, Speliotes EK, Loos RJF, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. 2009;41(1):25-34. doi:10.1038/ng.287
105. Aerts E, Beckers S, Zegers D, et al. CNV analysis and mutation screening indicate an important role for the NPY4R gene in human obesity. *Obesity*. 2016;24(4):970-976. doi:10.1002/oby.21435
106. Yang T-L, Guo Y, Shen H, et al. Copy number variation on chromosome 10q26.3 for obesity identified by a genome-wide study. *J Clin Endocrinol Metab*. 2013;98(1):E191-E195. doi:10.1210/jc.2012-2751
107. Jarick I, Vogel CIG, Scherag S, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. *Hum Mol Genet*. 2011;20(4):840-852. doi:10.1093/hmg/ddq518
108. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C. Large, rare chromosomal deletions associated with severe early- onset obesity. *Nature*. 2010;463(7281):666-670. doi:10.1038/nature08689.Large
109. Walters RG, Jacquemont S, Valsesia A, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature*. 2010;463(7281):671-675. doi:10.1038/nature08727
110. Falchi M, El-Sayed Moustafa JS, Takousis P, et al. Low copy number of the salivary amylase gene predisposes to obesity. *Nat Genet*. 2014;46(5):492-497. doi:10.1038/ng.2939
111. Lopomo A, Burgio E, Migliore L. Epigenetics of Obesity. *Prog Mol Biol Transl Sci*. 2016;140:151-184. doi:10.1016/bs.pmbts.2016.02.002
112. van Dijk SJ, Molloy PL, Varinli H, Morrison JL, Muhlhausler BS. Epigenetics and human obesity. *Int J Obes*. 2015;39(1):85-97.

doi:10.1038/ijo.2014.34
113. Thaker V V. Genetic and Epigenetic causes of obesity. *Adolesc Med state art Rev*. 2017;28(2):379-405.
114. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-5467. doi:10.1073/pnas.74.12.5463
115. Prober JM, Trainor GL, Dam RJ, et al. A system for rapid DNA sequencing with fluorescent chain-terminating dideoxynucleotides. *Science (80-)*. 1987;238(4825):336-341. doi:10.1126/science.244397
116. Mardis ER. Next-generation sequencing platforms. *Annu Rev Anal Chem*. 2013;6:287-303. doi:10.1146/annurev-anchem-062012-092628
117. Lin B, Wang J, Cheng Y. Recent Patents and Advances in the Next-Generation Sequencing Technologies. *NIH Public Access*. 2008;2008(1):60-67.
118. Zhang J, Chiodini R, Badr A, Zhang G. The impact of next-generation sequencing on genomics. *J Genet genomics*. 2011;38(3):95-109. doi:10.1016/j.jgg.2011.02.003
119. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet*. 2011;52:413-435. doi:10.1007/s13353-011-0057-x
120. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*. 2012;2012(251364). doi:10.1155/2012/251364
121. Rieber N, Zapatka M, Lasitschka B, et al. Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies. *PLoS One*. 2013;8(6). doi:10.1371/journal.pone.0066621
122. Park M-H, Rhee H, Park JH, et al. Comprehensive analysis to improve the validation rate for single nucleotide variants detected by next-generation sequencing. *PLoS One*. 2014;9(1). doi:10.1371/journal.pone.0086664
123. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet*. 2011;48:580-589.
124. Hoischen A, van Bon B, Gilissen C, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet*. 2010;42(6):483-485.
125. Gilissen C, Hoischen A, Brunner HG, Veltman J a. Unlocking Mendelian disease using exome sequencing. *Genome Biol*. 2011;12(9):228. doi:10.1186/gb-2011-12-9-228
126. Zuzarte PC, Denroche RE, Fehringer G, Katzov-Eckert H, Hung RJ, McPherson JD. A two-dimensional pooling strategy for rare variant detection on next-generation sequencing platforms. *PLoS One*. 2014;9(4):e93455. doi:10.1371/journal.pone.0093455
127. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals G. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. *Nucleic Acids Res*.

2002;30(12):e57. doi:10.1093/nar/gnf056
128. Tucker T, Montpetit A, Chai D, et al. Comparison of genome-wide array genomic hybridization platforms for the detection of copy number variants in idiopathic mental retardation. *BMC Med Genomics*. 2011;4:25. doi:10.1186/1755-8794-4-25
129. National Institutes of Health. *The Practical Guide: Identification, Evaluation, and Treatment of Overweight and Obesity in Adults*.; 2000. https://www.nhlbi.nih.gov/files/docs/guidelines/prctgd_c.pdf.
130. Olson DH, McCubbin HI, Barnes HL, Larsen AS, Muxen MJ, Wilson MA. Families: What makes them work. *Sci Res*. 1983.
131. Olson DH, Sprenkle DH, Russell CS. Circumplex model of marital and family systems: I. Cohesion and adaptability dimensions, family types, and clinical applications. *Fam Process*. 1979;18:3-28.
132. Cohen DM, Lumley MA, Naar-King S, Partridge T, Cakan N. Child behavior problems and family functioning as predictors of adherence and glycemic control in economically disadvantaged children with type 1 diabetes: a prospective study. *J Pediatr Psychol*. 2004;29(3):171-184. doi:10.1093/jpepsy/jsh019
133. Johannsen DL, Johannsen NM, Specker BL. Influence of parents' eating behaviors and child feeding practices on children's weight status. *Obesity*. 2006;14(3):431-439. doi:10.1038/oby.2006.57
134. Kiefner-Burmeister AE, Hoffmann DA, Meers MR, Koball AM, Musher-Eizenman DR. Food consumption by young children: a function of parental feeding goals and practices. *Appetite*. 2014;74:6-11. doi:10.1016/j.appet.2013.11.011
135. Costanzo PR, Woody EZ. Domain-specific parenting styles and their impact on the child's development of particular deviance: The example of obesity proneness. *J Soc Clin Psychol*. 1985;3(4):425-445. doi:10.1521/jscp.1985.3.4.425
136. Fisher JO, Birch LL. Restricting access to palatable foods affects children's behavioral response, food selection, and intake. *Am J Clin Nutr*. 1999;69(6):1264-1272. doi:10.1093/ajcn/69.6.1264
137. Birch LL, Fisher JO, Grimm-Thomas K, Markey CN, Sawyer R, Johnson SL. Confirmatory factor analysis of the Child Feeding Questionnaire: A measure of parental attitudes, beliefs and practices about child feeding and obesity proneness. *Appetite*. 2001;36(3):201-210. doi:10.1006/appe.2001.0398
138. Fuller J, Thompson W. *Behavior Genetics*. Wiley, New York; 1960. https://www.gwern.net/docs/genetics/heritable/1960-fuller-behaviorgenetics.pdf.
139. Butte NF, Cai G, Cole SA, Comuzzie AG. Viva la Familia Study: Genetic and environmental contributions to childhood obesity and its comorbidities in the Hispanic population. *Am J Clin Nutr*. 2006;84(3):646-654. http://www.ajcn.org/cgi/reprint/84/3/646%5Cnhttp://ovidsp.ovid

.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed7&NEWS=N&AN=2006549764.

140. Farooqi IS. Monogenic human obesity. *Front Horm Res*. 2008;36:1-11. doi:10.1159/000115333
141. Zegers D, Beckers S, Hendrickx R, et al. Prevalence of rare MC3R variants in obese cases and lean controls. *Endocrine*. 2013;44:386-390. doi:10.1007/s12020-012-9862-1
142. Pettersson M, Viljakainen H, Loid P, et al. Copy number variants are enriched in individuals with early-onset obesity and highlight novel pathogenic pathways. *J Clin Endocrinol Metab*. 2017;102(8):3029-3039. doi:10.1210/jc.2017-00565
143. Ranadive SA, Vaisse C. Lessons from extreme human obesity: monogenic disorders. *Endocrinol Metab Clin North Am*. 2008;37(3):733-751. doi:10.1016/j.ecl.2008.07.003
144. Vuillaume M-L, Naudion S, Banneau G, et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. *Am J Med Genet*. 2014;164A(8):1965-1975. doi:10.1002/ajmg.a.36587
145. Hinney A, Hebebrand J. Polygenic obesity in humans. *Obes Facts*. 2008;1(1):35-42. doi:10.1159/000113935
146. González JR, Ruiz-Arenas C, Cáceres A, et al. Polymorphic Inversions Underlie the Shared Genetic Susceptibility of Obesity-Related Diseases. *Am J Hum Genet*. 2020;106(6):846-858. doi:10.1016/j.ajhg.2020.04.017
147. Afshin A, Forouzanfar MH, Reitsma MB, et al. Health effects of overweight and obesity in 195 countries over 25 years. *N Engl J Med*. 2017;377(1):13-27. doi:10.1056/NEJMoa1614362
148. Ek A, Sorjonen K, Eli K, et al. Associations between parental concerns about preschoolers' weight and eating and parental feeding practices: Results from analyses of the child eating behavior questionnaire, the child feeding questionnaire, and the lifestyle behavior checklist. *PLoS One*. 2016;11(1):1-20. doi:10.1371/journal.pone.0147257
149. Bhat SP, Sharma A. Current Drug Targets in Obesity Pharmacotherapy - A Review. *Curr Drug Targets*. 2017;18(8):983-993. doi:10.2174/1389450118666170227153940
150. Gouni-Berthold I, Berthold HK. Current Options for the Pharmacotherapy of Obesity. *Curr Pharm Des*. 2019;25(18):2019-2032. doi:10.2174/1381612825666190708192630

ANNEXES

Serra-Juhé C, Martos-Moreno GÁ, Bou de Pieri F, Flores R, González JR, Rodríguez-Santiago B, Argente J, Pérez-Jurado LA. Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. PLoS Genet. 2017 May; 13(5): e1006657

PLOS | GENETICS

RESEARCH ARTICLE

Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants

Clara Serra-Juhé[1,2,3☯], Gabriel Á. Martos-Moreno[4,5,6☯], Francesc Bou de Pieri[1,2,3], Raquel Flores[1,2,3], Juan R. González[2,7,8], Benjamín Rodríguez-Santiago[9], Jesús Argente[4,5,6,10*], Luis A. Pérez-Jurado[1,2,3*]

1 Genetics Unit, Universitat Pompeu Fabra, Barcelona, Spain, **2** Hospital del Mar Research Institute (IMIM), Barcelona, Spain, **3** Centro de Investigación Biomédica en Red de Enfermedades Raras (CIBERER), Instituto de Salud Carlos III, Barcelona, Spain, **4** Departments of Pediatrics & Pediatric Endocrinology, Hospital Infantil Universitario Niño Jesús, Universidad Autónoma de Madrid, Madrid, Spain, **5** Hospital de la Princesa Research Institute, Madrid, Spain, **6** Centro de Investigación Biomédica en Red de fisiopatología de la obesidad y nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain, **7** Center for Research in Environmental Epidemiology (CREAL), Barcelona, Spain, **8** Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP), Instituto de Salud Carlos III, Barcelona, Spain, **9** Quantitative Genomic Medicine Laboratories (qGenomics), Esplugues de Llobregat, Spain, **10** IMDEA Food Institute, CEI UAM & CSIC, Madrid, Spain

☯ These authors contributed equally to this work.
* luis.perez@upf.edu (LAPJ); jesus.argente@uam.es (JA)

OPEN ACCESS

Citation: Serra-Juhé C, Martos-Moreno GÁ, Bou de Pieri F, Flores R, González JR, Rodríguez-Santiago B, et al. (2017) Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. PLoS Genet 13(5): e1006657. https://doi.org/10.1371/journal.pgen.1006657

Editor: Xavier Estivill, Centre for Genomic Regulation (CRG), SPAIN

Received: August 1, 2016

Accepted: February 26, 2017

Published: May 10, 2017

Data Availability Statement: SNP array data are available in the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) with the code E-MTAB-5680.

Funding: LAPJ was funded by the Spanish Ministry of Health (FIS-PI1302481, co-funded by FEDER), the Generalitat de Catalunya (2014SRG1468), the Institució Catalana de Recerca i Estudis Avançats (ICREA Academia program), and the Spanish Ministry of Economy and Competiveness

Abstract

Obesity is a multifactorial disorder with high heritability (50–75%), which is probably higher in early-onset and severe cases. Although rare monogenic forms and several genes and regions of susceptibility, including copy number variants (CNVs), have been described, the genetic causes underlying the disease still remain largely unknown. We searched for rare CNVs (>100kb in size, altering genes and present in <1/2000 population controls) in 157 Spanish children with non-syndromic early-onset obesity (EOO: body mass index >3 standard deviations above the mean at <3 years of age) using SNP array molecular karyotypes. We then performed case control studies (480 EOO cases/480 non-obese controls) with the validated CNVs and rare sequence variants (RSVs) detected by targeted resequencing of selected CNV genes (n = 14), and also studied the inheritance patterns in available first-degree relatives. A higher burden of gain-type CNVs was detected in EOO cases versus controls (OR = 1.71, p-value = 0.0358). In addition to a gain of the *NPY* gene in a familial case with EOO and attention deficit hyperactivity disorder, likely pathogenic CNVs included gains of glutamate receptors (*GRIK1*, *GRM7*) and the X-linked gastrin-peptide receptor (*GRPR*), all inherited from obese parents. Putatively functional RSVs absent in controls were also identified in EOO cases at *NPY*, *GRIK1* and *GRPR*. A patient with a heterozygous deletion disrupting two contiguous and related genes, *SLCO4C1* and *SLCO6A1*, also had a missense RSV at *SLCO4C1* on the other allele, suggestive of a recessive model. The genes identified showed a clear enrichment of shared co-expression partners with known genes strongly related to obesity, reinforcing their role in the pathophysiology of the disease. Our data reveal a higher burden of rare CNVs and RSVs in several related genes in patients

"Programa de Excelencia María de Maeztu" (MDM-2014-0370). JA was funded by the Spanish Ministry of Health (FIS-PI13/02195 & PI16/00485, co-funded by FEDER) and the Fundación de Endocrinología y Nutrición. JRG was supported by the Spanish Ministry of Science and Innovation (MTM2011-26515) and Statistical Genetics Network - GENOMET (MTM2010-09526-E). The "Centro de Investigación Biomédica en Red" for rare diseases (CIBERER), obesity and nutrition (CIBEROBN), and epidemiology and public health (CIBERESP) are initiatives of the Instituto de Salud Carlos III, Spain. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: BRS and LAPJ are employee and scientific advisor, respectively, of qGenomics, a privately held company that provides genomics services to the scientific and medical community.

with EOO compared to controls, and implicate *NPY*, *GRPR*, two glutamate receptors and *SLCO4C1* in highly penetrant forms of familial obesity.

Author summary

Although there is strong evidence for a high genetic component of obesity, the underlying genetic causes are largely unknown, mostly due to the highly heterogeneous nature of the disorder. In this work, we have focused on the most severe end of the spectrum, severe obesity with early-onset in childhood, which is more likely due to genetic alterations. We screened for rare copy number variation (CNV) a sample of 157 Spanish children with early-onset obesity using molecular karyotypes and then studied the genes altered by CNVs in 480 cases and 480 non-obese controls. We identified a higher burden of gain-type CNVs in cases as well as several CNVs and sequence variants that were specific of the obese population. Interestingly, the genes identified shared co-expression partners with known obesity genes. Among those, the genes encoding the neuropeptide Y (*NPY*), two glutamate receptors (*GRIK1*, *GRM7*), the X-linked gastrin-peptide receptor (*GRPR*), and the organic anion transporter (*SLCO4C1)* are novel obesity candidate genes that may contribute to highly penetrant forms of familial obesity.

Introduction

Early-onset overweight (body mass index [BMI] $\geq$ 85th percentile for age and sex) and obesity (BMI $\geq$ 95th percentile for age and sex) currently affects 27.8% of children in Spain (Spanish National Health Survey, 2011–2012), being the most prevalent chronic disorder in childhood and adolescence. In the United States, 17.3% of children aged 2 to 19 years are obese, 5.9% meet criteria for class 2 obesity (BMI $\geq$ 120% of the 95th percentile or BMI $\geq$ 35), and 2.1% have class 3 obesity (BMI $\geq$ 140% of the 95th percentile or BMI $\geq$ 40) [1]. Early-onset obesity (EOO) entails several comorbidities and predisposes to obesity and related diseases during adulthood, being one of the most important health problems in developed countries.

Single gene alterations with Mendelian inheritance account for less than 5% of non-syndromic cases of severe EOO [2], including mutations in the *LEP* (MIM 164160) or *LEPR* (MIM 601007) genes [3–5], as well as in *MC4R* (MIM 155541) [6,7] which are the most common cause of monogenic obesity. Genetic, genomic and epigenetic alterations have also been identified in syndromic forms of obesity, such as Bardet-Biedl syndrome (MIM 209900) [8], Prader-Willi syndrome (MIM 176270) [9], Beckwith-Wiedemann syndrome (MIM 130650) [10] and other rare diseases. However, obesity is generally considered a multifactorial disorder with high heritability (50–75%), probably higher in early-onset cases [11]. To date multiple studies have tried to elucidate genetic factors contributing to the etiopathogenesis of obesity, and relevant SNPs in more than 100 loci have been identified by Genome Wide Association Studies (GWAS), including those near genes such as *FTO* (MIM 610966), *MC4R*, *NEGR1* (MIM 613173) or *TMEM18* (MIM 613220) [12–15]. Nevertheless, the fraction of BMI variance explained by these GWAS top hits is estimated to be only around 2% [16]. Even the infinitesimal model, that combines the effect of all common autosomal SNPs, only explains ~17% of the variance in BMI [17]. Gene-based meta-analysis of GWAS allowed the identification of regions with high allelic heterogeneity and new loci involved in obesity [18,19].

In addition, several common and rare copy number variants (CNV) contributing to the heritability of BMI and obesity have been reported, including deletions upstream of the *NEGR1* gene [13], proximal and distal deletions at 16p11.2 [20], gains at 10q26.6 containing the *CYP2E1* gene (MIM 124040) [21], and homozygous deletions at 11q11 encompassing olfactory receptor genes [22], among others. While several studies in large datasets led to the conclusion that common CNVs are not a major contributor [22], a significantly increased burden of rare CNVs was documented in cases of severe obesity with and without associated developmental delay [15,23]. Specifically, a significant enrichment for CNVs larger than 100 Kb and with a population frequency lower than 1% was identified in subjects with isolated severe EOO when compared to controls [15].

In the present study, we have analyzed the contribution to the phenotype of rare and common CNVs as well rare sequence variants (RSVs) in CNV-related genes in a large Spanish sample of patients with isolated severe EOO using case-control and family-based approaches, with the goal to identify novel genes involved in the pathophysiology of severe obesity.

Results

We used a sequential strategy to identify genes potentially related to EOO through the analysis of CNVs by molecular karyotyping and subsequent mutation screening using a DNA pooled approach in a subset of selected genes. The strategy, including the samples used for each step, is summarized in Fig 1.

CNV burden in EOO

A total of 42 autosomal CNVs fulfilling the established criteria (>100 kb, gene containing and present in <1/2000 population individuals) were identified in 36 cases (22.9%). We detected 7 deletions and 35 gains (100.1–3,590kb in length), with 5 samples harboring more than one rearrangement (Table 1). MLPA was used for validation (42/42) and determination of the

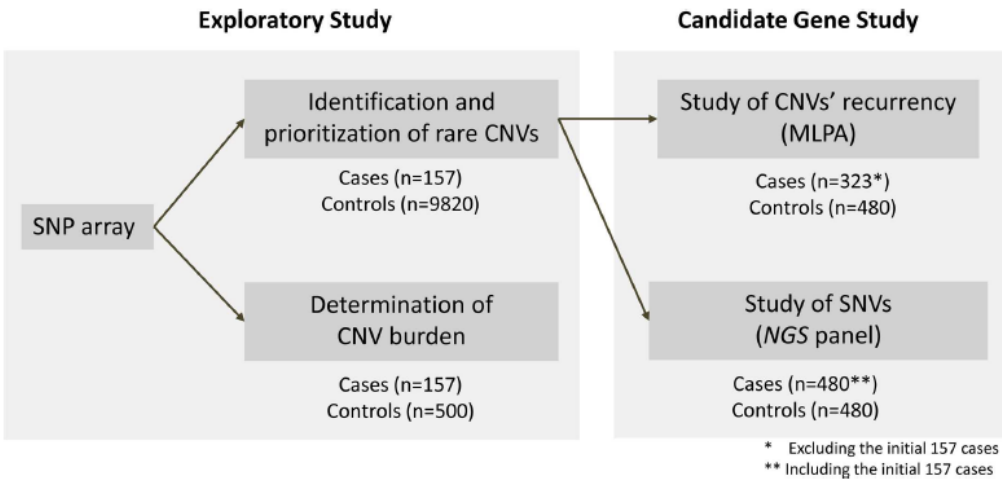

Fig 1. Workflow of the approach followed to study CNVs and RSVs in a subset of selected genes. The different control cohorts used for each of the analysis are shown, in addition to the number of obese patients and controls studied in each step.

https://doi.org/10.1371/journal.pgen.1006657.g001

Table 1. Summary of copy number variations detected in 157 samples of patients with severe obesity. Control frequency refers to the frequency of the same type of rearrangement, deletion or duplication, in the control cohort (9,820 subjects). Progenitor phenotype refers only to the progenitor carrying the alteration. Hg19 assembly. F: female; M: male; Mat: maternal; Pat: paternal; N: normal; OW: overweight; OB: obesity; NA: not available; +: description of a mouse model with a phenotype related to body mass index (S1 Table). The genes interrupted at any of the breakpoints of the CNV are shown in **boldface**.

Case	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Genes	Inheritance	Progenitor phenotype	Control frequency	Mouse model
Recurrent alterations in obese patients										
Ob_1	F	Gain	9q34.3	149	chr9:138149166–138298164	*C9orf62*	Mat	N	1	
Ob_2	F	Gain	9q34.3	139	chr9:138149166–138288052	*C9orf62*	Pat	OW	1	
Ob_3	F	Gain	9q34.3	139	chr9:138149166–138288052	*C9orf62*	Mat	OW	1	
Ob_4	M	Gain	7p22.1	106	chr7:5785086–5891221	***RNF216,ZNF815***	Mat	OB	4	
Ob_5	F	Gain	7p22.1	106	chr7:5785086–5891221	***RNF216,ZNF815***	Pat	OB	4	
Ob_6	F	Loss	11p15.4	104	chr11:4583029–4687238	*TRIM68* and 5 more genes	Pat	OB	2	
Obese specific alterations, not found in 10,320 controls, co-segregating with the phenotype in the family										
Ob_7	F	Gain	3q29	1169	chr3:196533320–197701913	*DLG1*,***PAK2,LMLN*** and 10 more genes	Mat	OB	0	+
Ob_8	F	Gain	4q12	361	chr4:53842714–54203701	***SCFD2***	Pat	OB	0	
Ob_9	F	Loss	6q23.2	166	chr6:133386327–133552737	*LINC00326*	Mat	OB	0	
Ob_10	M	Loss	5q21.1	157	chr5:101620174–101776835	***SLCO4C1,SLCO6A1***	Pat	OB	0	
Ob_11	M	Gain	7p15.3	202	chr7:21858215–22059791	*CDCA7L*,***DNAH11***	Pat	OB	0	
Ob_12	M	Gain	7p15.3	137	chr7:24258773–24395900	*NPY*	Mat	OB	0	+
Ob_13	F	Gain	10p14	1029	chr10:10743956–11773389	*CELF2* and 4 more genes	Mat	OB	0	
Ob_14	M	Gain	14q31.1	105	chr14:81204951–81309536	***CEP128***	Mat	OB	0	
Ob_15	F	Gain	14q31.3	274	chr14:89171711–89445350	***EML5***,*TTC8*	Mat	OB	0	+
Ob_16	M	Gain	19p13.3	428	chr19:2221792–2650034	*AMH*,***DOT1L,GNG7*** and 14 more genes	Pat	OB	0	
Ob_17	M	Gain	20p12.1	220	chr20:13255679–13476090	***ISM1***,***TASP1***	Pat	OB	0	+
Copy number variations detected in single cases										
Ob_18	F	Gain	1p36.13	223	chr1:17202355–17425829	*ATP13A2*,***PADI2*** and 3 more genes	Mat	N	0	
Ob_19	M	Gain	1p31.3	405	chr1:61704166–62109502	***NFIA***	Mat	OW	3	
Ob_20	M	Gain	2q24.1	336	chr2:159176303–159512667	***CCDC148,PKP4***	Mat	N	0	
Ob_21	M	Gain	3p26.1	214	chr3:4272253–4486303	*SETMAR*,***SUMF1***	Pat	N	2	
Ob_22	M	Gain	3p26.1	364	chr3:7660133–8024019	***GRM7***	Mat	OW	3	
Ob_15	F	Loss	3q12.3	3590	chr3:101816344–105406145	*ZPLD1, ALCAM,* ***CBLB***	*De novo*	-	0	

(*Continued*)

Table 1. (*Continued*)

Case	Gender	Gain/ Loss	Region	Length (kb)	Hg19 coordinates	Genes	Inheritance	Progenitor phenotype	Control frequency	Mouse model
Ob_23	M	Gain	4q13.3	101	chr4:71055318–71156301	*C4orf7,CSN3,ODAM*	Mat	N	0	
Ob_18	F	Gain	5p15.33	295	chr5:303686–598237	***AHRR,PDCD6,*** *and 6 more*	Mat	OW	3	
Ob_24	M	Gain	5q35.3	397	chr5:179220638–179617799	*C5orf45,**RASGEF1C*** *and 6 more*	Pat	N	0	
Ob_25	M	Gain	6q15	445	chr6:89349438–89793993	***PNRC1,RNGTT***	Mat	OW	0	
Ob_26	F	Gain	7p14.1	105	chr7:40117098–40221714	***C7orf10****,C7orf11,* ***CDK13***	Pat	OB	1	
Ob_27	F	Gain	9p24.3	448	chr9:396232–844001	***DMRT1,DOCK8,*** *KANK1*	Mat	N	2	
Ob_15	F	Loss	10q21.3	156	chr10:69418270–69574169	***CTNNA3,DNAJC12***	Mat	OB	1	
Ob_28	M	Gain	11p13	129	chr11:32986850–33116054	***CSTF3,QSER1*** and 3 more genes	Pat	N	0	
Ob_29	F	Gain	12q23.3	116	chr12:104476277–104591886	***HCFC2****,NFYB*	Pat	NA	0	
Ob_30	M	Gain	12q24.33	218	chr12:131620578–131838842	***GPR133****,LOC116437*	Pat	OB	1	
Ob_16	M	Gain	15q25.2	418	chr15:84414592–84832932	***ADAMTSL3****,* *EFTUD1P1*	Pat	OB	1	
Ob_31	M	Gain	17q23.1	248	chr17:58113570–58361461	*CA4,**USP32*** and 5 more genes	NA	NA	0	
Ob_32	F	Gain	18p11.31	245	chr18:6308208–6553040	***L3MBTL4****,* ***LOC100130480****,* *MIR4317*	Mat	OW	0	
Ob_5	F	Gain	18p11.21	262	chr18:12917703–13180058	*CEP192,SEH1L*	Mat	N	2	
Ob_14	M	Gain	20p12.1	204	chr20:13867165–14070869	***MACROD2,SEL1L2***	Pat	N	0	
Ob_33	F	Gain	21q21.3	163	chr21:31280603–31443375	***GRIK1***	Pat	OB	1	+
Ob_34	F	Loss	22q11.22	261	chr22:22312292–22573637	***TOP3B***	Pat	OB	2	
Ob_35	F	Loss	22q11.22	676	chr22:22981587–23657613	***BCR*** *and 7 more*	Mat	NA	0	
Ob_36	M	Gain	22q13.33	198	chr22:50342728–50584201	*IL17REL,MLC1,* ***MOV10L1****,PIM3*	Pat	NA	0	
Copy number variations detected in X chromosome										
Ob_37	F	Gain	Xp22.2	305	chrX:15952591–16257827	*GRPR*	Mat	N	*1*	+
Ob_6	F	Gain	Xq26.2	369	chrX:130639755–131008470	*LOC286467,OR13H1,* ***IGSF1***	Pat	OB	*0*	

https://doi.org/10.1371/journal.pgen.1006657.t001

inheritance pattern (41/42): all tested rearrangements were inherited except for the larger deletion. We also detected and validated 2 additional CNVs on the X-chromosome for a total of 44 CNVs.

Clinical data about the parental phenotype was available in all but two families with CNVs. The progenitor harboring the CNV was obese (defined as BMI >30) in 21 cases (53.8%), was overweight (BMI 25–30) in 7 cases (17.9%) and had a BMI in the normal range in 11 cases

Table 2. Comparisons of the frequency of rare copy number changes in autosomal chromosomes >100kb detected in the patient cohort and in the cohort of 500 controls. In brackets the proportion of samples with the CNV and the proportion of the specific type of rearrangement.

Group	Alterations	Deletions	Duplications	Double hit	Samples
Controls (500)	85	19 (3.8% / 22.4%)	66 (12.6% / 77.6%)	5 (1.0%)	79 (15.8%)
Obese patients (157)	42	7 (4.5% / 16.7%)	35* (19.7% / 83.3%)	5 (3.2%)	36 (22.9%)

* p-value<0.05

https://doi.org/10.1371/journal.pgen.1006657.t002

(28.2%). More than half of the rare CNVs (25 of 44), 21 gains and 4 deletions, were not found in any of the 9,820 adult population controls.

In order to analyze the global burden of rare CNVs in EOO, we compared the amount, type and length of autosomal CNVs in patients (157) with respect to 500 Spanish population controls (Table 2). Rare CNVs were found in 15.8% of controls with respect to the 22.9% frequency found in patients ($p = 0.053$). Rare CNVs were predominantly gains in both cohorts (83.3% in EOO patients and 77.6% in controls). When the frequency of deletions and gains was analyzed separately, no differences were observed in deletion-type CNVs (3.8% in controls and 4.5% in patients), while a statistically significant difference in the frequency of gains was detected ($p = 0.0358$). Thus, there is a higher burden of CNVs in EOO patients due to rare gain-type CNVs.

If we consider specific CNVs as those not described in the initial 9,820 subjects used to establish the frequency of each alteration in the population, 8.2% control individuals carried a CNV fulfilling this criteria while the frequency was 14.0% in EOO patients, with this difference being statistically significant ($p = 0.0422$).

Regarding the co-occurrence of more than one CNV in the same subject, two or three hits were present in 3.2% patients and 1% controls. This difference was not statistically significant ($p = 0.0644$) likely due to the small sample size. The inheritance pattern of these alterations was established in patients; in two cases each alteration was inherited from a different parent and in the remaining three both rearrangements were inherited from the same progenitor. Case Ob_15 presented a third *de novo* event additionally to the two rearrangements inherited from her obese mother.

Potentially pathogenic CNVs

CNVs were considered to have a higher probability to be pathogenic when they were exclusive of the EOO population, co-segregated with the phenotype in the family, disrupted known genes for the disorder and/or were found in more than one case.

Nine duplications and two deletions were absent in 9,820 population controls and co-segregated with the phenotype in the family (Table 1). One of them was a 137kb gain in 7p15.3 containing a single coding gene, *NPY* (MIM 162640), identified in a male case (Ob_12) presenting with EOO and attention deficit hyperactivity disorder (ADHD) (Fig 2A). The CNV was inherited from the also obese mother (Fig 2B). Additional cases of severe EOO and ADHD were identified in the maternal branch of this family by report (Fig 2C), but unfortunately no additional samples or clinical data could be obtained.

Some CNVs overlapped with previously described microdeletion/microduplication syndromes (Table 1). Rearrangements partially overlapping with the critical region of the 22q11.2 distal deletion syndrome were identified in two patients. Ob_35 carried a 676kb deletion encompassing several genes including *RSPH14* (MIM 605663) and *GNAZ* (MIM 139160), while a more proximal deletion including *TOP3B* (MIM 603582) was detected in Ob_34. A

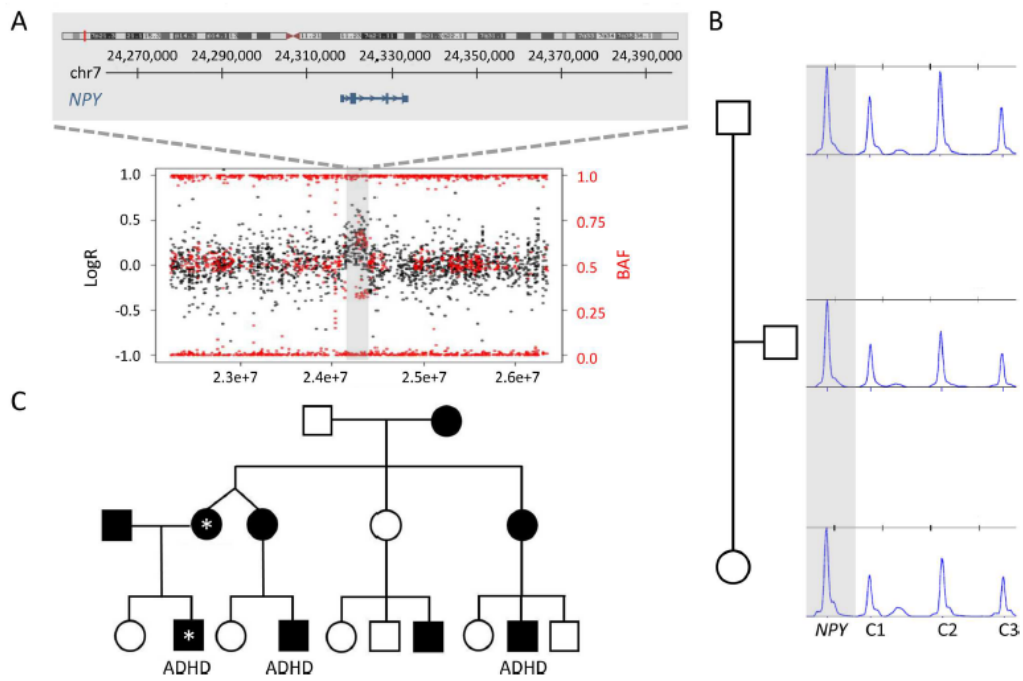

Fig 2. Detection, validation and inheritance of the duplication encompassing *NPY* in case Ob_12 and his family. A: Ideogram showing the location of the CNV and the specific genomic interval included in the duplication. The plot shows the results of SNP array with the Log R Ratio represented by black dots and the B Allele Frequency (BAF) represented by red dots. Hg19 assembly. B: MLPA of the trio showing the maternal inheritance of the rearrangement represented by a single probe (indicated). C: Pedigree of the family showing several cases with severe obesity, as well as ADHD (attention-deficit/hyperactivity disorder). The two individuals carrying the duplication are labeled by an *; samples from additional relatives were not available.

https://doi.org/10.1371/journal.pgen.1006657.g002

gain of 348kb at 1q21.1 overlapping with the region of Thrombocytopenia-Absent Radius syndrome (MIM 274000) was detected in case Ob_39; as its frequency in controls was 1/1,720 it was not included in the subset of selected CNVs.

Two CNVs fulfilling the established criteria were identified in more than one patient (Table 1). A gain of 139kb in 9q34.3 only including *C9orf62* was found in three cases (Ob_1, Ob_2, Ob_3). However, the parents also carrying the CNV had either overweight or normal weight. Another gain of 106kb in 7p22.1 encompassing *RNF216* (MIM 609948) and *ZNF815P* was identified in two cases (Ob_4, Ob_5), inherited from obese parents. We then completed the analysis of the CNVs identified in the entire sample of obese individuals (n = 480) and the Spanish adult non-obese controls (n = 480) by MLPA. All rare CNVs were patient-specific except for a second patient with a deletion at 11p15.4. None of the rare CNVs were identified among controls except for the 106kb gain at 7p22.1 that was found in 5 controls. The re-analysis of SNP array data unraveled the complexity of mapping this region due to small segmental duplications and was used to determine the real frequency of the rearrangement, which was above the established threshold of the study (1/2,000).

Association study with common CNVs

We also explored more common CNVs already described in association with obesity. The gain in 10q26.3 including *CYP2E1* was more common in patients than in controls (6.4% vs 3.6%) as previously described [21], but did not reach significance (OR: 2.01, CI95% 0.93–4.36, p-value = 0.075). The frequency of the homozygous deletion encompassing olfactory receptors in 11q11 was 5.1% in cases, which was slightly lower than the frequency in the control cohort (6.7%). Therefore, our data did not replicate the previous findings that indicate a preferable transmission of the 11q11 deletion to obese children [22]. In addition, in this study we did not detect alterations in the 16p11.2 region, including or next to the *SH2B1* (MIM 608937) gene.

Identification of RSVs in novel genes by targeted capture sequencing of pooled DNA

All genes included and/or disrupted by CNVs found in more than one patient and/or co-segregating in a familial case were selected for sequence analysis (n = 14): *SCFD2*, *NPY*, *ISM1* (MIM 615793), *TASP1* (MIM 608270), *GRM7*, *LOC401164*, *TRIML1*, *SLCO4C1*, *SLCO6A1*, *C11orf40*, *TRIM68* (MIM 613184), *GRIK1*, *TOP3B* and *GRPR*. In order to sequence the total number of patients (480) and controls (480) in a cost-efficient manner, pools of 20 DNA samples were sequenced with each DNA sample located in two pools.

We first validated the suitability and specificity of the pipeline to detect real variants among the pools. RSVs were considered when they had a frequency below 1/1.000 in the public database of the Exome Sequencing Consortium (ExAC) representing more than 60,000 exomes [24]. We selected 23 alterations predicted to be in a single sample and reanalyzed the same sample by Sanger sequencing. All 23 RSVs were validated in the specific samples.

We then compared the total burden of RSVs per gene between patients and controls. Significant differences were identified in a few loci, namely *NPY*, *GRIK1* (MIM 138245) and *GRPR* (MIM 305670) (Table 3). A single missense RSV in *NPY* (p.V86D) was identified in patient Ob_158, while no RSVs of this gene were found in controls. Although the residue is not evolutionarily conserved and is located outside the main functional domain, the change is likely to affect the shape and the affinity of the NPY protein and has not been described in ExAC. The study of parental samples revealed that the RSV was inherited from the obese father (BMI 34.3 kg/m^2). The low frequency of missense variants in this gene in the ExAC database (only 27 among 118.884 alleles) further reinforces its functional relevance.

A nonsense mutation (p.R897X) was identified in *GRIK1*, encoding the ionotropic glutamate receptor 1, in patient Ob_163. This nonsense variant has a frequency below 1/15000

Table 3. Point mutations detected by pooled DNA sequencing in the cohort of 480 patients. Control frequency refers to the allele frequency of the same variant in the subjects included in the ExAC database (60,706 unrelated individuals). Progenitor phenotype refers only to the progenitor carrying the alteration. Hg19 assembly. F: female; M: male; Mat: maternal; Pat: paternal; NA: not available.

Gene name	Variant	cDNA level	Protein level	Phylo P	Control frequency	Case	Gender	Inheritance	Progenitor phenotype
NPY	chr7:24329186	c.T257A	p.V86D	0.798641	0	Ob_158	F	Pat	Obesity
GRPR	chrX:16142093	c.G17C	p.C6S	0.992202	0.00095	Ob_159	F	NA	NA
GRPR	chrX:16142335	c.C259A	p.L87M	0.948124	0.00097	Ob_160	M	Mat	Obesity
GRPR	chrX:16170380	c.T767C	p.I256T	0.997201	0	Ob_161	F	NA	NA
GRPR	chrX:16168672	c.G658A	p.V220I	0.787201	0.00049	Ob_38	F	NA	NA
						Ob_162	M	Mat	Normal
GRIK1	chr21:30909580	c.C2689T	p.R897X	0.999731	0.00006	Ob_163	M	NA	NA
SLCO4C1	chr5:101606433	c.A697C	p.I233L	0.974709	0	Ob_10	M	Mat	Normal

https://doi.org/10.1371/journal.pgen.1006657.t003

alleles and, generally, nonsense and frameshift variants at the *GRIK1* gene are rare, representing less than 1/3000 alleles in the ExAC database. Additional missense mutations were identified in both glutamate receptors (*GRIK1* and *GRM7* (MIM 604101)), but with no significant differences between cases and controls.

Four different missense mutations were detected in *GRPR* in five patients with obesity, while no mutations in this gene were found in controls (5/662 alleles vs 0/726 alleles; p = 0.0245). One of the mutations has never been found in ExAC, while the remaining three had frequencies <1/1000 alleles and all are predicted to result in significant functional consequences.

Finally, a RSV in *SLCO4C1* (MIM 609013) was identified in a patient harboring a deletion encompassing the same gene previously identified by CMA. The RSV (p.I233L) has not been described previously and affects a highly conserved amino acid (phylo P = 0.975). The frequency of the deletion in the control cohort is 0/9,820. Parental studies confirmed that each progenitor had transmitted one of the alterations; the deletion was inherited from the obese father and the RSV from the non-obese mother. These findings are compatible with a recessive pattern of inheritance or a two-hit mechanism, with a major contribution of the CNV (inherited from an obese progenitor) and an additional and milder effect of the RSVs (inherited from a non-obese progenitor).

Co-expression enrichment analyses

We focused our subsequent analysis on four novel candidate genes considering our CNVs and RSVs findings: *GRIK1*, *GRM7*, *GRPR* and *SLCO4C1*. To explore their possible role in obesity, we looked for co-expressions with a stringent list of 15 genes previously related with obesity. We selected a total 10 genes with described highly penetrant mutations in severely obese patients, all of them coding for proteins of the leptin-melanocortin pathway: *LEP*[4], *LEPR*[5], *MC4R*[6], *POMC* (MIM 176830) [25], *PCSK1* (MIM 162150) [26], *MC3R* (MIM 155540) [27,28], *BDNF* (MIM 113505) [29], *NTRK2* (MIM600456) [30], *PPARG* (MIM 601487) [31] and *SIM1* (MIM 603128) [20,32]. We also included 5 additional genes with a relevant intermediary role in the same pathway: *ADRB3* (MIM 109691) [33], *PCSK2* (MIM 162151) [34,35], *NPY*[36], *NPY1R* (MIM 162641) [37], *AGRP* (MIM 602311) [36,38] (Fig 3A).

A total of 10 shared co-expressed partners were identified in the analysis between our 4 novel strongest candidate genes (*GRIK1*, *GRM7*, *GRPR* and *SLCO4C1*) and the set of 15 obesity-related genes described in the literature. The maximum number of shared co-expressed partners between our 4 genes and 500 genes sets randomly selected was 7, being the empirical p-value of this difference 0.002 (Fig 3B and 3C).

Discussion

Our results reveal a relevant contribution of rare CNVs to the etiology of severe EOO with a significantly higher burden of gain-type CNVs in patients compared to controls (p = 0.0358). Among relatively common CNVs we only detected a non-significant higher frequency of the gain in 10q26.3 containing *CYP2E1* [21]. Previous studies reported a higher frequency of deletion-type CNVs in patients with severe EOO with and without developmental delay [23]. The sample studied here was stringently selected based on clinical exam and targeted genetic testing in order to exclude subjects with syndromic obesity. Thus, all patients presented isolated EOO without comorbid phenotypes such as developmental delay. This difference in the range of phenotype severity could explain the difference in the type of rearrangements found enriched in these cohorts. Although only one of the CNVs had occurred *de novo*, the progenitor carrying the alteration also presented overweight or obesity in 71.8% of cases, reinforcing

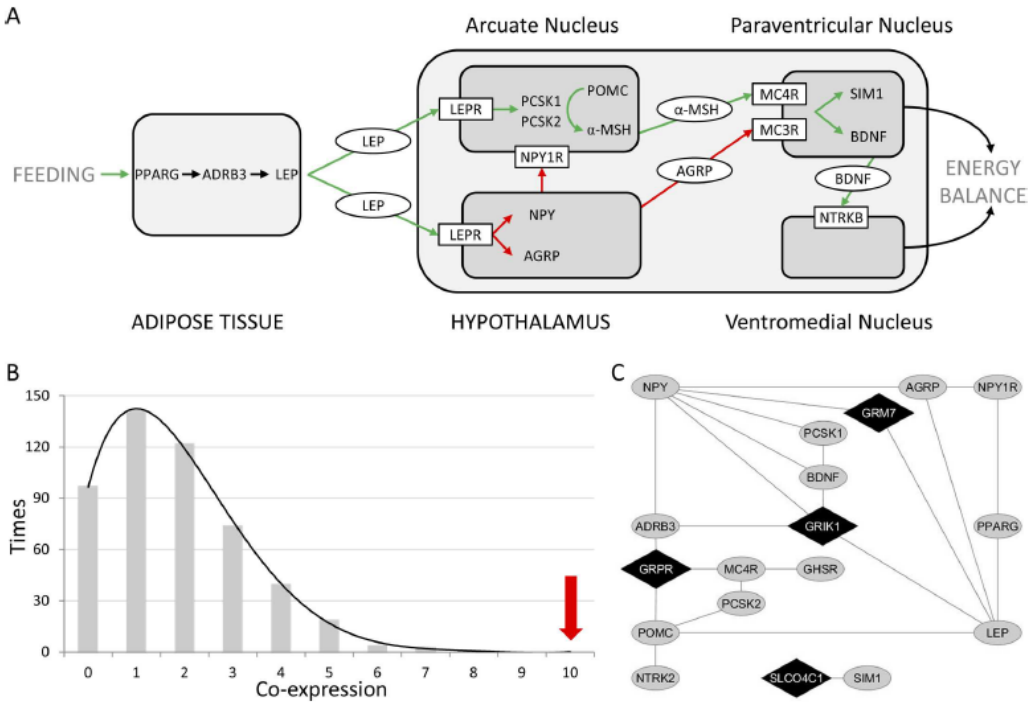

Fig 3. Enrichment analyses of shared co-expressed partners between 14 known obesity-related genes and the candidate genes from this study. A: Key players on the regulation of food intake and energy expenditure mostly act through the leptin-melanocortin pathway. B: Shared co-expressed partners between the four candidate genes (*GRIK1*, *GRM7*, *GRPR*, *SLCO4C1*) and 500 randomly generated gene sets. The number shared co-expressed partners of the candidate genes with the 14 obesity-related genes are indicated by a red arrow. C: Co-expression network including the selected obesity-related genes and *GRIK1*, *GRM7*, *GRPR*, *SLCO4C1*. This network was visualized using Cytoscape.

https://doi.org/10.1371/journal.pgen.1006657.g003

the potential role of many of these genetic alterations in the pathophysiology of the disorder. The only *de novo* alteration (a 3.6Mb deletion encompassing only 3 genes) was detected in a girl with two additional rearrangements inherited from her obese mother.

We have also screened for point mutations using DNA pools in a subset of selected genes located in obese-specific CNVs [39]. Interestingly, we found additional RSVs in patients in 4 of the selected genes (*NPY*, *GRPR*, SLCO4C1 and *GRIK1)* reinforcing their putative role in the pathophysiology of obesity. Although the pooled DNA strategy might have some limitations such as underdetection of relatively common variants, the complete validation rate (100%) demonstrates its high specificity.

A maternally inherited gain in 7p15.3 only encompassing the *NPY* gene was identified in a patient with EOO and ADHD. The mother also presented severe obesity, as did several relatives from the maternal branch including two male cousins with associated ADHD. Additionally, a missense RSV also inherited from an obese progenitor was identified in another patient, while no alterations were identified in controls (480). A larger gain of approximately 3Mb on chromosome 7p15.2–15.3 encompassing NPY and other genes was previously described in all

affected individuals of an extended pedigree presenting ADHD, increased BMI, and elevated NPY levels in blood [40]. Therefore, the gain encompassing only the *NPY* gene in patient Ob_12 and his obese mother, the point mutation in patient Ob_158 and her obese father while only 27 missense variants have been described among 118.884 alleles in ExAC are strong evidences supporting that gain of function mutations of *NPY* can cause severe obesity and ADHD.

NPY is a hypothalamic orexigenic peptide with neuromodulator functions in the control of energy balance and food intake. NPY is overproduced in the hypothalamus of leptin deficient ob/ob mice [41]; when depleted by genetic manipulation, ob/ob mice showed reduced food intake, increased energy expenditure and less obesity [42]. On the other hand, overexpression of NPY in noradrenergic neurons caused diet- and stress-induced gain in fat mass in a gene-dose-dependent fashion [43].

In humans, despite some conflictive reports, *NPY* gene variants have been significantly associated with weight changes from young adulthood to middle age and with risk of obesity [44]. NPY is widely expressed throughout the central nervous system (CNS) and a systematic review and meta-analyses of drug naïve case-control studies also suggested its implication in ADHD [45]. In addition, increased central availability of NPY by intracerebroventricular administration in male rats resulted in a shift of metabolism towards lipid storage and increased carbohydrate use, along with enhanced locomotor activity and body temperature [46].

Among other genes altered by the CNVs identified, we considered as probably pathogenic those exclusive of the EOO population that also presented exclusive RSVs co-segregating with the phenotype in the family. To further assess the possible implication of these strong candidate genes (*GRIK1*, *GRM7*, *GRPR* and *SLCO4C1*), we determined the co-expression patterns between them and 15 well-defined genes from the leptin-melanocortin pathway previously related to obesity. This analysis consistently identified a significant enrichment of co-expression shared partners among our genes and the subset of obesity related genes when compared to 500 randomly generated gene sets, reinforcing the possible role of those genes in the pathophysiology of EOO.

The alterations affecting glutamate receptors identified in two EOO patients were a partial gain of the gene encoding the ionotropic glutamate receptor *GRIK1* (Ob_33) and a gain partially encompassing the gene encoding the metabotropic glutamate receptor *GRM7* (Ob_22). L-glutamate is one of the main excitatory neurotransmitter in the CNS and activates both ionotropic and metabotropic glutamate receptors. A nonsense mutation was found in an additional patient in *GRIK1*. The metabotropic glutamate receptor 5 (mGluR5) plays a relevant role in energy balance and feeding. Adult mice lacking mGlu5 weighed significantly less than littermate controls and resisted diet-induced obesity [47]. Pharmacological approaches have described a reduction of food intake in response to antagonists of mGluR5 in a baboon model of binge-eating disorder [48] and in mGluR5+/+, but not mGluR5-/- mice [47]. On the contrary, dose-dependent stimulation of food intake has been described in rodents after injection of a mGluR5 agonist [49]. Moreover, the metabolic status and leptin can modify astrocyte-specific glutamate and glucose transporters, indicating that metabolic signals influence glutamatergic synaptic efficacy and glucose uptake [50]. Interestingly, *GRM7* is likely a loss of function intolerant gene given the difference between expected and observed frequency of loss of function variants in ExAC (25 expected, 1 observed). Partial gains, depending on the location, might act as loss of function alterations when disrupting the gene. Considering these data, glutamate receptors are promising candidates in the pathophysiology of obesity.

Several alterations affecting *GRPR* gene were identified, including a gain encompassing the whole gene and 4 point mutations (present in 5 subjects, two males and three females) while

none were found in controls. The male patients with hemizygous *GRPR* RSVs had inherited the variant from heterozygous mothers. Both patients had very early onset obesity in infancy presenting a quite severe phenotype at diagnosis (+5SD and +9SD respectively). One of the mothers (patient Ob_162) had a BMI within the normal range while the other presented adult-onset obesity. Thus, the phenotype of both males is more severe than the phenotype of their mothers, consistent with X-linked inheritance. *GRPR* encodes the receptor of gastrin-releasing peptide. Gastrin is a hormone secreted by the gastric antrum and duodenum in response to gastric distension and the presence of food in the stomach. This hormone increases the production of hydrochloric acid, pepsinogen, pancreatic secretions and bile to facilitate food digestion and also promotes satiety [51]. It is a hormone directly implicated in the regulation of food ingestion and satiety and, thus, a candidate to be associated with obesity (directly or by an alteration of a gene included in the pathway, such as *GRPR*).

A possible recessive pattern of inheritance or a double hit mechanism was identified in a patient who harbors a deletion partially encompassing *SLCO4C1* and *SLCO6A1* (MIM 613365) and a RSV in *SLCO4C1*, each alteration inherited from one of the progenitors. Considering that the CNV was inherited from an obese progenitor and the RSV from the non-obese mother, we postulate a major contribution of the CNV and an additional but likely milder effect of the RSVs. The SLCO4C1 belongs to the organic anion transporter family and is involved in the membrane transport of thyroid hormones, among others. Interestingly, none homozygote subjects for loss of function variants has been described in ExAC.

Other rearrangements were found in single EOO patients, including those in regions previously associated to disease, such as 22q11.2 or 1q21.1. However, the evidence to link these genomic regions to obesity susceptibility is still weak and further data will be needed.

Except for the patient with a biallelic alteration in *SLCO4C1* and the *de novo* deletion, all CNVs and RSVs identified are heterozygous in the patients and inherited from one of the parents. Parents carrying the allele also showed an obese phenotype as well in most cases. Thus, a dominant effect (either hypo or hypermorphic) for these rare genetic variants with additive effects is suggested, leading to a more severe phenotype in the younger generation. This effect has also been found in other studies [52] and can be due to the more "obesogenic" environment that has developed in industrialized societies during the last two decades. Our results, along with previous genetic, family-based and epidemiologic studies, further indicate that EOO etiology is complex and mostly multifactorial, with the presence of some alleles that can behave as highly penetrant susceptibility variants or monogenic forms of obesity.

In summary, our findings reveal a higher burden of rare CNVs in patients with EOO compared to controls, including novel CNVs likely associated with familial obesity. Dosage sensitive genes altered by these CNVs are candidates for contributing to the pathogenesis of EOO. Some of these genes also harbor patient-specific RSVs, reinforcing their putative role in the pathophysiology of obesity. *NPY*, *GRPR*, *SLCO4C1* and glutamate receptors emerge as novel candidate genes involved in monogenic familial obesity.

Materials and methods

Subjects

Criteria for severe EOO was a BMI more than three standard deviation measures above the mean for age and gender with onset earlier than 3 years of age. All cases underwent a detailed clinical examination as well as family history in search of syndromic forms of obesity, which were discarded. All studies were performed as part of a research project approved by the

Medical Ethical Committee of the Hospital Infantil Universitario Niño Jesús, after receiving written informed consent from the family.

Blood samples from patients were collected. Parental blood samples were also collected in cases in which an alteration was identified. DNA from patients and parents was isolated from total blood using the Gentra Puregene Blood kit (Qiagen) according to manufacturer's instructions. We excluded genomic and epigenetic alterations associated with pseudohypoparathyroidism (MIM 103580), Prader-Willi, Temple (MIM 616222) and Beckwith-Wiedemann syndromes with a custom-made panel (S2 Table) of Methylation Specific Multiplex Ligation Dependent-Probe Amplification (MS-MLPA) [53].

A total of 480 unrelated subjects with severe EOO were included in the study. As controls for CNV and RSV association analyses, we studied 480 adult individuals of Spanish origin with a current BMI lower than 25 and no known history of childhood obesity, obtained from the National DNA Bank from the University of Salamanca (Spain).

Molecular karyotyping

An initial sample of 157 probands was studied by using Omni1-Quad (64 subjects) or Omni Express SNP (93 subjects) platforms, Illumina. Copy number changes were identified using the PennCNV software with stringent filtering, as previously described [54]. CNVs encompassing known genes (RefSeq hg19), longer than 100 kb and with a frequency in control samples lower than 1/2,000 were selected. The frequency of each CNV in the control population was determined using 1M Illumina SNP array data of a total of 9,820 samples from two databases: 1) 8,329 individuals previously used as population controls for developmental anomalies [55] (81.2% of European descent, 2% African, and 16.5% other/mixed ancestry), and 2) 1,491 Spanish adult individuals from the Spanish Bladder Cancer/EPICURO study, which includes 1034 patients with urothelial cell carcinoma of the bladder and 457 hospital-based generally healthy controls with a mean age of 63.7 years [54]. To determine the frequency of CNV in the X chromosome, only the Spanish controls were considered, as data from the other cohort was not available. Given the size of the Spanish control sample (1,491), alterations in the X-chromosome absent in controls or only present in one subject were considered as rare. Briefly, a Hidden Markov Model (HMM) based on both allele frequencies and total intensity values was used to identify putative alterations, followed by manual inspection in conjunction with user guided merging of nearby (<1 Mbp between for arrays with <1 million probes and <200 kbp for arrays with >1 million probes) calls, which represent a single region broken up by the HMM, or gaps. All samples on arrays with densities <1M probes were filtered by a maximal genome-wide LogR ratio standard deviation of 0.25, while the high density 1.2 million probe WTCCC2 data was filtered using an increased standard deviation cut-off of 0.37. Mosaic alterations were excluded. For the two datasets where the Illumina array mapping corresponded to build35 (NHGRI), we utilized the autosomal calls generated previously [40] and mapped the coordinates to build36 using the UCSC LiftOver tool [56].

Estimation of rare CNV burden

In order to compare the global burden of rare CNVs in patients and controls, data from 500 individuals randomly selected from the Spanish Bladder Cancer/EPICURO study and not included as controls for the CNVs frequency determination [54] were used. For the comparison, only CNVs in autosomal chromosomes with a minimum length of 100 kb, altering genes, and a frequency in control samples lower than 1/2,000 were considered (S3 Table). Alterations totally overlapping with segmental duplications were excluded to minimize biases due to the different probe coverage among microarray platforms.

Multiplex Ligation-Dependent Probe Amplification (MLPA)

An MLPA assay was designed to validate genetic alterations detected by SNP platforms and to study inheritance in families (available upon request). A total of 100 ng of genomic DNA from each sample was subject to MLPA using specific synthetic probes (sequence available upon request) designed to target the specific CNV detected. All MLPA reactions were analyzed on an ABI PRISM 3100 Genetic analyzer according to manufacturers' instructions. Each MLPA signal was normalized and compared to the corresponding peak height obtained in control samples [57]. The MLPA assay was also used to analyze the frequency of the CNVs identified in the entire cohort (480 subjects) and in the control population (480 individuals).

Targeted capture sequencing of pooled DNA for RSV identification

To study RSVs in the genes included in the CNVs, an enrichment kit was designed to capture all the coding regions of the selected genes (n = 14). The targeted enrichment was done with SeqCap EZ Choice Enrichment Kits (Roche Sequencing) and the massive sequencing with MiSeq (Illumina).

In order to sequence a high number of patients and controls (960 in total) in a cost-efficient manner, a pooled DNA approach was used [39]. Each sample was included in two different pools, and each pool contained 20 samples, avoiding two samples sharing both pools. A priori, any heterozygous RSV should be present in approximately 1 every 40 reads (2.5%). Thus, to ensure the identification of all RSVs (expected to be found in just one or few individuals) a high coverage was required.

To discriminate real variants from false positives due to extremely high coverage, we optimized the analysis pipeline. Variant calling was done with MuTect [58] to detect variants in a low proportion of reads. We also considered the quality of reads (base quality >15 in each pool) and the absence of strand bias (between 0.2 and 0.8) to define potential real variants from false positives.

To analyze the results and compare patients and controls, we focused on RSVs. We first established the frequency of each variant in the general control population using EXAC as the reference database, composed of 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies. All alterations present in more than 1/1,000 alleles in EXAC in any of the populations included in the dataset were excluded. We specially focused on sequence changes with potential functional consequences, including loss of function variants (nonsense, frameshift and splice sites), missense variants predicted as pathogenic and changes in highly conserves residues. To search for recessive patterns of inheritance, we explored biallelic changes and RSVs that might act as second-hits in patients with previously identified CNVs.

Sanger sequencing

To validate the RSVs detected by NGS and to define the segregation in each family, we designed primers to amplify an amplicon encompassing the variant and sequenced the amplicon by Sanger technology (available under request).

Co-expression enrichment analyses

Using Genemania, we explored the co-expression between our candidate genes and the selected subset. To test if there was an enrichment of shared co-expressed partners, 500 sets of 15 genes with expression data available were randomly selected with Molbiotools (http://www.molbiotools.com/). For each set of genes the number of shared co-expressed partners was determined and

compared with the interactions between our candidates and the set of obesity-related genes. The empirical p-value was calculated based on the fraction of shared co-expressed partners.

Supporting information

S1 Table. Description of mouse models with a phenotype related to body mass index caused by alterations in genes included in CNVs detected in the patients' cohort.
(XLSX)

S2 Table. Probes included in the custom-made panel of Methylation Specific Multiplex Ligation Dependent-Probe Amplification (MS-MLPA) to exclude genomic and epigenetic alterations associated with pseudohypoparathyroidism, Prader-Willi, Temple and Beck-with-Wiedemann syndromes.
(XLSX)

S3 Table. CNVs in autosomal chromosomes with a minimum length of 100 kb, altering genes, and with a frequency in control samples lower than 1/2,000 detected in the control cohort.
(XLSX)

Acknowledgments

We would like to thank the patients and families for their participation in this study as well as to Francisca Díaz for her technical assistance.

Author Contributions

Conceptualization: CSJ GÁMM LAPJ JA.

Data curation: CSJ FBdP JRG BRS.

Formal analysis: CSJ FBdP JRG BRS.

Funding acquisition: LAPJ JA.

Investigation: CSJ FBdP RF GÁMM JA.

Methodology: CSJ LAPJ.

Project administration: CSJ LAPJ JA.

Resources: GÁMM JA.

Software: CSJ FBdP BRS.

Supervision: LAPJ JA.

Validation: CSJ FBdP RF.

Visualization: CSJ FBdP.

Writing – original draft: CSJ GÁMM LAPJ.

Writing – review & editing: JA CSJ GÁMM FBdP RF JRG BRS JA LAPJ.

References

1. Skinner AC, Skelton JA. Prevalence and Trends in Obesity and Severe Obesity Among Children in the United States, 1999–2012. JAMA Pediatr [Internet]. 2014 Apr 7 [cited 2014 May 23]; 168(6):561–6.

Available from: http://www.ncbi.nlm.nih.gov/pubmed/24710576 https://doi.org/10.1001/jamapediatrics.2014.21 PMID: 24710576

2. Ranadive SA, Vaisse C. Lessons from extreme human obesity: monogenic disorders. Endocrinol Metab Clin North Am [Internet]. 2008 Sep [cited 2013 Nov 29]; 37(3):733–51, x. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18775361 https://doi.org/10.1016/j.ecl.2008.07.003 PMID: 18775361
3. Farooqi IS, Matarese G, Lord GM, Keogh JM, Lawrence E, Agwu C, et al. Beneficial effects of leptin on obesity, T cell hyporesponsiveness, and neuroendocrine/metabolic dysfunction of human congenital leptin deficiency. J Clin Invest [Internet]. 2002 Oct [cited 2013 Nov 27]; 110(8):1093–103. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=150795&tool=pmcentrez&rendertype=abstract https://doi.org/10.1172/JCI15693 PMID: 12393845
4. Montague CT, Farooqi IS, Whitehead JP, Soos MA, Rau H, Wareham NJ, et al. Congenital leptin deficiency is associated with severe early-onset obesity in humans. Nature [Internet]. 1997 Jun 26 [cited 2013 Nov 20]; 387(6636):903–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9202122 https://doi.org/10.1038/43185 PMID: 9202122
5. Clément K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature [Internet]. 1998 Mar 26 [cited 2016 Nov 23]; 392(6674):398–401. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9537324 https://doi.org/10.1038/32911 PMID: 9537324
6. Farooqi IS, Keogh JM, Yeo GSH, Lank EJ, Cheetham T, O'Rahilly S. Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene. N Engl J Med [Internet]. 2003 Mar 20 [cited 2013 Nov 22]; 348(12):1085–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12646665 https://doi.org/10.1056/NEJMoa022050 PMID: 12646665
7. Granell S, Serra-Juhé C, Martos-Moreno GÁ, Díaz F, Pérez-Jurado L a, Baldini G, et al. A novel melanocortin-4 receptor mutation MC4R-P272L associated with severe obesity has increased propensity to be ubiquitinated in the ER in the face of correct folding. PLoS One [Internet]. 2012; 7(12):e50894. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3520997&tool=pmcentrez&rendertype=abstract https://doi.org/10.1371/journal.pone.0050894 PMID: 23251400
8. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, et al. Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science [Internet]. 2001 Sep 21 [cited 2013 Nov 19]; 293(5538):2256–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11567139 https://doi.org/10.1126/science.1063525 PMID: 11567139
9. Nicholls RD, Knoll JH, Butler MG, Karam S, Lalande M. Genetic imprinting suggested by maternal heterodisomy in nondeletion Prader-Willi syndrome. Nature [Internet]. 1989 Nov 16 [cited 2013 Nov 28]; 342(6247):281–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/2812027 https://doi.org/10.1038/342281a0 PMID: 2812027
10. Ramesar R, Babaya M, Viljoen D. Molecular investigation of familial Beckwith-Wiedemann syndrome: a model for paternal imprinting. Eur J Hum Genet [Internet]. 1993 Jan [cited 2013 Nov 28]; 1(2):109–13. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8055321 PMID: 8055321
11. Llewellyn CH, Trzaskowski M, Plomin R, Wardle J. Finding the missing heritability in pediatric obesity: the contribution of genome-wide complex trait analysis. Int J Obes (Lond) [Internet]. 2013 Mar 26 [cited 2013 Nov 11]; 37(11):1506–9. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3826033&tool=pmcentrez&rendertype=abstract
12. Meyre D, Delplanque J, Chèvre J-C, Lecoeur C, Lobbens S, Gallina S, et al. Genome-wide association study for early-onset and morbid adult obesity identifies three new risk loci in European populations. Nat Genet [Internet]. 2009 Feb [cited 2013 Nov 27]; 41(2):157–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19151714 https://doi.org/10.1038/ng.301 PMID: 19151714
13. Thorleifsson G, Walters GB, Gudbjartsson DF, Steinthorsdottir V, Sulem P, Helgadottir A, et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. Nat Genet [Internet]. 2009 Jan [cited 2013 Nov 22]; 41(1):18–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19079260 https://doi.org/10.1038/ng.274 PMID: 19079260
14. Willer CJ, Speliotes EK, Loos RJF, Li S, Lindgren CM, Heid IM, et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. Nat Genet [Internet]. 2009 Jan [cited 2013 Nov 10]; 41(1):25–34. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2695662&tool=pmcentrez&rendertype=abstract https://doi.org/10.1038/ng.287 PMID: 19079261
15. Wheeler E, Huang N, Bochukova EG, Keogh JM, Lindsay S, Garg S, et al. Genome-wide SNP and CNV analysis identifies common and low-frequency variants associated with severe early-onset obesity. Nat Genet [Internet]. 2013 May [cited 2013 Nov 6]; 45(5):513–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23563609 https://doi.org/10.1038/ng.2607 PMID: 23563609
16. Speliotes EK, Willer CJ, Berndt SI, Monda KL, Thorleifsson G, Jackson AU, et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. Nat Genet [Internet]. 2010 Nov [cited 2013 Nov 9]; 42(11):937–48. Available from: http://www.pubmedcentral.nih.gov/

articlerender.fcgi?artid=3014648&tool=pmcentrez&rendertype=abstract https://doi.org/10.1038/ng.686 PMID: 20935630

17. Yang J, Manolio TA, Pasquale LR, Boerwinkle E, Caporaso N, Cunningham JM, et al. Genome partitioning of genetic variation for complex traits using common SNPs. Nat Genet [Internet]. 2011 Jun [cited 2014 May 26]; 43(6):519–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21552263 https://doi.org/10.1038/ng.823 PMID: 21552263

18. Pei Y-F, Zhang L, Liu Y, Li J, Shen H, Liu Y-Z, et al. Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity. Hum Mol Genet [Internet]. 2014 Feb 1 [cited 2016 Jul 25]; 23(3):820–30. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24064335 https://doi.org/10.1093/hmg/ddt464 PMID: 24064335

19. Hägg S, Ganna A, Van Der Laan SW, Esko T, Pers TH, Locke AE, et al. Gene-based meta-analysis of genome-wide association studies implicates new loci involved in obesity. Hum Mol Genet [Internet]. 2015 Dec 1 [cited 2016 Jul 25]; 24(23):6849–60. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26376864 https://doi.org/10.1093/hmg/ddv379 PMID: 26376864

20. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, et al. A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. Nature [Internet]. 2010 Feb 4 [cited 2013 Nov 7]; 463(7281):671–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2880448&tool=pmcentrez&rendertype=abstract https://doi.org/10.1038/nature08727 PMID: 20130649

21. Yang T-L, Guo Y, Shen H, Li J, Glessner JT, Qiu C, et al. Copy number variation on chromosome 10q26.3 for obesity identified by a genome-wide study. J Clin Endocrinol Metab [Internet]. 2013 Jan [cited 2013 Nov 27]; 98(1):E191–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23175694 https://doi.org/10.1210/jc.2012-2751 PMID: 23175694

22. Jarick I, Vogel CIG, Scherag S, Schäfer H, Hebebrand J, Hinney A, et al. Novel common copy number variation for early onset extreme obesity on chromosome 11q11 identified by a genome-wide analysis. Hum Mol Genet [Internet]. 2011 Feb 15 [cited 2013 Nov 27]; 20(4):840–52. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3024044&tool=pmcentrez&rendertype=abstract https://doi.org/10.1093/hmg/ddq518 PMID: 21131291

23. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, et al. Large, rare chromosomal deletions associated with severe early-onset obesity. Nature [Internet]. 2010 Feb 4 [cited 2013 Nov 12]; 463(7281):666–70. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3108883&tool=pmcentrez&rendertype=abstract https://doi.org/10.1038/nature08689 PMID: 19966786

24. Exome Aggregation Consortium (ExAC). No Title [Internet]. [cited 2016 Jun 1]. Available from: http://exac.broadinstitute.org

25. Krude H, Biebermann H, Luck W, Horn R, Brabant G, Grüters A. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nat Genet [Internet]. 1998 Jun [cited 2016 Nov 23]; 19(2):155–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9620771 https://doi.org/10.1038/509 PMID: 9620771

26. Jackson RS, Creemers JWM, Ohagi S, Raffin-Sanson M-L, Sanders L, Montague CT, et al. Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. Nat Genet [Internet]. 1997 Jul [cited 2016 Nov 23]; 16(3):303–6. https://doi.org/10.1038/ng0797-303 PMID: 9207799

27. Lee Y-S, Poh LK-S, Loke K-Y. A Novel Melanocortin 3 Receptor Gene (*MC3R*) Mutation Associated with Severe Obesity. J Clin Endocrinol Metab [Internet]. 2002 Mar [cited 2016 Nov 23]; 87(3):1423–6. https://doi.org/10.1210/jcem.87.3.8461 PMID: 11889220

28. Mencarelli M, Dubern B, Alili R, Maestrini S, Benajiba L, Tagliaferri M, et al. Rare melanocortin-3 receptor mutations with in vitro functional consequences are associated with human obesity. Hum Mol Genet [Internet]. 2011 Jan 15 [cited 2016 Nov 23]; 20(2):392–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21047972 https://doi.org/10.1093/hmg/ddq472 PMID: 21047972

29. Gray J, Yeo GSH, Cox JJ, Morton J, Adlam A-LR, Keogh JM, et al. Hyperphagia, Severe Obesity, Impaired Cognitive Function, and Hyperactivity Associated With Functional Loss of One Copy of the Brain-Derived Neurotrophic Factor (BDNF) Gene. Diabetes [Internet]. 2006 Dec 1 [cited 2016 Nov 23]; 55(12):3366–71. https://doi.org/10.2337/db06-0550 PMID: 17130481

30. Yeo GSH, Connie Hung C-C, Rochford J, Keogh J, Gray J, Sivaramakrishnan S, et al. A de novo mutation affecting human TrkB associated with severe obesity and developmental delay. Nat Neurosci [Internet]. 2004 Nov [cited 2016 Nov 23]; 7(11):1187–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15494731 https://doi.org/10.1038/nn1336 PMID: 15494731

31. Ristow M, Müller-Wieland D, Pfeiffer A, Krone W, Kahn CR. Obesity associated with a mutation in a genetic regulator of adipocyte differentiation. N Engl J Med [Internet]. 1998 Oct 1 [cited 2013 Nov 27];

339(14):953–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9753710 https://doi.org/10.1056/NEJM199810013391403 PMID: 9753710

32. Michaud JL, Boucher F, Melnyk A, Gauthier F, Goshu E, Lévy E, et al. Sim1 haploinsufficiency causes hyperphagia, obesity and reduction of the paraventricular nucleus of the hypothalamus. Hum Mol Genet [Internet]. 2001 Jul 1 [cited 2016 Nov 23]; 10(14):1465–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/11448938 PMID: 11448938

33. Mantzoros CS, Qu D, Frederich RC, Susulic VS, Lowell BB, Maratos-Flier E, et al. Activation of beta(3) adrenergic receptors suppresses leptin expression and mediates a leptin-independent inhibition of food intake in mice. Diabetes [Internet]. 1996 Jul [cited 2016 Nov 23]; 45(7):909–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8666142 PMID: 8666142

34. Ohagi S, LaMendola J, LeBeau MM, Espinosa R, Takeda J, Smeekens SP, et al. Identification and analysis of the gene encoding human PC2, a prohormone convertase expressed in neuroendocrine tissues. Proc Natl Acad Sci U S A [Internet]. 1992 Jun 1 [cited 2016 Nov 24]; 89(11):4977–81. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1594602 PMID: 1594602

35. Furuta M, Carroll R, Martin S, Swift HH, Ravazzola M, Orci L, et al. Incomplete processing of proinsulin to insulin accompanied by elevation of Des-31,32 proinsulin intermediates in islets of mice lacking active PC2. J Biol Chem [Internet]. 1998 Feb 6 [cited 2016 Nov 24]; 273(6):3431–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9452465 PMID: 9452465

36. Baver SB, Hope K, Guyot S, Bjorbaek C, Kaczorowski C, O'Connell KMS. Leptin Modulates the Intrinsic Excitability of AgRP/NPY Neurons in the Arcuate Nucleus of the Hypothalamus. J Neurosci [Internet]. 2014 Apr 16 [cited 2016 Nov 23]; 34(16):5486–96. https://doi.org/10.1523/JNEUROSCI.4861-12.2014 PMID: 24741039

37. Herzog H, Hort YJ, Ball HJ, Hayes G, Shine J, Selbie LA. Cloned human neuropeptide Y receptor couples to two different second messenger systems. Proc Natl Acad Sci U S A [Internet]. 1992 Jul 1 [cited 2016 Nov 24]; 89(13):5794–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1321422 PMID: 1321422

38. Chai B-X, Pogozheva ID, Lai Y-M, Li J-Y, Neubig RR, Mosberg HI, et al. Receptor–Antagonist Interactions in the Complexes of Agouti and Agouti-Related Protein with Human Melanocortin 1 and 4 Receptors †, ‡. Biochemistry [Internet]. 2005 Mar [cited 2016 Nov 23]; 44(9):3418–31. https://doi.org/10.1021/bi0478704 PMID: 15736952

39. Wang SR, Carmichael H, Andrew SF, Miller TC, Moon JE, Derr MA, et al. Large-scale pooled next-generation sequencing of 1077 genes to identify genetic causes of short stature. J Clin Endocrinol Metab [Internet]. 2013 Aug [cited 2016 Jul 29]; 98(8):E1428–37. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23771920 https://doi.org/10.1210/jc.2013-1534 PMID: 23771920

40. Lesch K-P, Selch S, Renner TJ, Jacob C, Nguyen TT, Hahn T, et al. Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. Mol Psychiatry [Internet]. 2011 May [cited 2013 Nov 28]; 16(5):491–503. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20308990 https://doi.org/10.1038/mp.2010.29 PMID: 20308990

41. Stephens TW, Basinski M, Bristow PK, Bue-Valleskey JM, Burgett SG, Craft L, et al. The role of neuropeptide Y in the antiobesity action of the obese gene product. Nature [Internet]. 1995 Oct 12 [cited 2013 Nov 28]; 377(6549):530–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/7566151 https://doi.org/10.1038/377530a0 PMID: 7566151

42. Erickson JC, Hollopeter G, Palmiter RD. Attenuation of the obesity syndrome of ob/ob mice by the loss of neuropeptide Y. Science [Internet]. 1996 Dec 6 [cited 2013 Nov 28]; 274(5293):1704–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8939859 PMID: 8939859

43. Vähätalo LH1, Ruohonen ST1, Ailanen L2 SE. Neuropeptide Y in noradrenergic neurons induces obesity in transgenic mouse models. Neuropeptides. 2016; 55:31–7. https://doi.org/10.1016/j.npep.2015.11.088 PMID: 26681068

44. Karvonen MK, Pesonen U, Koulu M, Niskanen L, Laakso M, Rissanen A, et al. Association of a leucine (7)-to-proline(7) polymorphism in the signal peptide of neuropeptide Y with high serum cholesterol and LDL cholesterol levels. Nat Med [Internet]. 1998 Dec [cited 2013 Nov 28]; 4(12):1434–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9846584 https://doi.org/10.1038/4027 PMID: 9846584

45. Scassellati C1, Bonvicini C, Faraone SV GM. Biomarkers and attention-deficit/hyperactivity disorder: a systematic review and meta-analyses. J Am Acad Child Adolesc Psychiatry. 2012; 51(10):1003–19. https://doi.org/10.1016/j.jaac.2012.08.015 PMID: 23021477

46. Su Y, Foppen E, Fliers E, Kalsbeek A. Effects of intracerebroventricular administration of neuropeptide Y on metabolic gene expression and energy metabolism in male rats. Endocrinology [Internet]. 2016 Jun 6 [cited 2016 Jul 27];en20161083. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27267712

47. Bradbury MJ, Campbell U, Giracello D, Chapman D, King C, Tehrani L, et al. Metabotropic glutamate receptor mGlu5 is a mediator of appetite and energy balance in rats and mice. J Pharmacol Exp Ther [Internet]. 2005 Apr [cited 2013 Nov 28]; 313(1):395–402. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15590770 https://doi.org/10.1124/jpet.104.076406 PMID: 15590770

48. Bisaga A, Danysz W, Foltin RW. Antagonism of glutamatergic NMDA and mGluR5 receptors decreases consumption of food in baboon model of binge-eating disorder. Eur Neuropsychopharmacol [Internet]. 2008 Nov [cited 2013 Nov 28]; 18(11):794–802. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2591926&tool=pmcentrez&rendertype=abstract https://doi.org/10.1016/j.euroneuro.2008.05.004 PMID: 18573641

49. Ploj K, Albery-Larsdotter S, Arlbrandt S, Kjaer MB, Skantze PMC, Storlien LH. The metabotropic glutamate mGluR5 receptor agonist CHPG stimulates food intake. Neuroreport [Internet]. 2010 Jul 14 [cited 2013 Nov 28]; 21(10):704–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20505551 https://doi.org/10.1097/WNR.0b013e32833b4fe7 PMID: 20505551

50. Fuente-Martín E, García-Cáceres C, Granado M, de Ceballos ML, Sánchez-Garrido MÁ, Sarman B, et al. Leptin regulates glutamate and glucose transporters in hypothalamic astrocytes. J Clin Invest [Internet]. 2012 Nov 1 [cited 2013 Nov 28]; 122(11):3900–13. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3484452&tool=pmcentrez&rendertype=abstract https://doi.org/10.1172/JCI64102 PMID: 23064363

51. Meek CL1, Lewis HB2, Reimann F2, Gribble FM2 PA. The effect of bariatric surgery on gastrointestinal and pancreatic peptide hormones. Peptides. 2016; 77:28–37. https://doi.org/10.1016/j.peptides.2015.08.013 PMID: 26344355

52. Stutzmann F, Tan K, Vatin V, Dina C, Jouret B, Tichet J, et al. Prevalence of Melanocortin-4 Receptor Deficiency in Europeans and Their Age-Dependent Penetrance in Multigenerational Pedigrees. Diabetes [Internet]. 2008 Sep 1 [cited 2016 Nov 24]; 57(9):2511–8. https://doi.org/10.2337/db08-0153 PMID: 18559663

53. Martos-Moreno GÁ, Serra-Juhé C, Pérez-Jurado LA, Argente J. Underdiagnosed Beckwith-Wiedemann syndrome among early onset obese children. Arch Dis Child [Internet]. 2014 Oct [cited 2016 Jul 27]; 99(10):965–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25085994

54. Rodríguez-Santiago B, Malats N, Rothman N, Armengol L, Garcia-Closas M, Kogevinas M, et al. Mosaic uniparental disomies and aneuploidies as large structural variants of the human genome. Am J Hum Genet [Internet]. 2010 Jul 9 [cited 2013 Nov 6]; 87(1):129–38. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2896781&tool=pmcentrez&rendertype=abstract https://doi.org/10.1016/j.ajhg.2010.06.002 PMID: 20598279

55. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, et al. A copy number variation morbidity map of developmental delay. Nat Genet [Internet]. 2011 Sep [cited 2013 Nov 6]; 43(9):838–46. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3171215&tool=pmcentrez&rendertype=abstract https://doi.org/10.1038/ng.909 PMID: 21841781

56. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res [Internet]. 2004 Jan 1 [cited 2014 Jun 4]; 32(Database issue): D493–6. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=308837&tool=pmcentrez&rendertype=abstract https://doi.org/10.1093/nar/gkh103 PMID: 14681465

57. Rodríguez-Santiago B, Brunet A, Sobrino B, Serra-Juhé C, Flores R, Armengol L, et al. Association of common copy number variants at the glutathione S-transferase genes and rare novel genomic changes with schizophrenia. Mol Psychiatry [Internet]. 2010 Oct [cited 2013 Nov 21]; 15(10):1023–33. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19528963 https://doi.org/10.1038/mp.2009.53 PMID: 19528963

58. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol [Internet]. 2013 Mar [cited 2016 Jul 25]; 31(3):213–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23396013 https://doi.org/10.1038/nbt.2514 PMID: 23396013