



Genome-wide copy number variation analysis identifies novel candidate loci associated with pediatric obesity

Thanuja Selvanayagam^{1,2} · Susan Walker^{1,2} · Matthew J. Gazzellone^{1,2} · Barbara Kellam^{1,2} · Cheryl Cytrynbaum^{1,3,4} · Dimitri J. Stavropoulos⁵ · Ping Li⁶ · Catherine S. Birken^{7,8} · Jill Hamilton^{6,8} · Rosanna Weksberg^{1,3,4,8} · Stephen W. Scherer^{1,2,4,9}

Received: 21 August 2017 / Revised: 10 April 2018 / Accepted: 15 May 2018 / Published online: 5 July 2018
© European Society of Human Genetics 2018

Abstract

Obesity is a multifactorial condition that is highly heritable. There have been ~60 susceptibility loci identified, but they only account for a fraction of cases. As copy number variations (CNVs) have been implicated in the etiology of a multitude of human disorders including obesity, here, we investigated the contribution of rare (<1% population frequency) CNVs in pediatric cases of obesity. We genotyped 67 such individuals, including 22 with co-morbid developmental delay and prioritized rare CNVs at known obesity-associated loci, as well as, those impacting genes involved in energy homeostasis or related processes. We identified clinically relevant or potentially clinically relevant CNVs in 15% (10/67) of individuals. Of these, 4% (3/67) had 16p11.2 microdeletions encompassing the known obesity risk gene *SH2B1*. Notably, we identified two unrelated probands harboring different 6p22.2 microduplications encompassing *SCGN*, a potential novel candidate gene for obesity. Further, we identified other biologically relevant candidate genes for pediatric obesity including *ARID5B*, *GPR39*, *PTPRN2*, and *HNF4G*. We found previously reported candidate loci for obesity, and new ones, suggesting CNV analysis may assist in the diagnosis of pediatric obesity.

These authors contributed equally: Thanuja Selvanayagam, Susan Walker

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41431-018-0189-0>) contains supplementary material, which is available to authorized users.

✉ Rosanna Weksberg
rweksb@sickkids.ca

✉ Stephen W. Scherer
stephen.scherer@sickkids.ca

¹ Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada

² The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, ON, Canada

³ Division of Clinical & Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada

⁴ Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada

Introduction

Obesity is a complex condition that results from an imbalance between energy intake and expenditure [1]. Body mass index (BMI) percentiles and z-scores relative to age- and gender-specific reference values are frequently used to classify obesity in the pediatric population. According to the World Health Organization (WHO), obesity is defined as having a BMI >99.9%ile (+3 SD) for age and gender in children aged 0–5 years and a BMI >97%ile (+2 SD) in

⁵ Department of Pediatric Laboratory Medicine, Genome Diagnostics, The Hospital for Sick Children, Toronto, ON, Canada

⁶ Division of Endocrinology, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada

⁷ Division of Pediatric Medicine, Hospital for Sick Children, Toronto, ON, Canada

⁸ Department of Pediatrics, University of Toronto, Toronto, ON, Canada

⁹ McLaughlin Centre, University of Toronto, Toronto, ON, Canada

children/adolescents aged 5–19 years [2]. Obesity is associated with an increased risk of multiple chronic conditions, including coronary heart disease, hypertension, and type 2 diabetes, which result in a significant health and economic burden.

Twin, family, and adoption studies show that BMI is highly heritable and strongly influenced by genetic factors [3]. Genome-wide association studies (GWAS) have been successful in identifying multiple associated loci, many of which are involved in the control of body weight and appetite via the central nervous system [4]. Nevertheless, loci identified through GWAS account for less than 10% of the heritability [4]. In recent years, rare copy number variations (CNVs) have been implicated in the etiology of numerous conditions, including autism spectrum disorder [5], schizophrenia [6], type I diabetes [7], congenital heart disease [8], and pediatric growth disorders [9]. CNVs have also been reported in the pathogenesis of obesity highlighting the potential role of another type of genetic contribution [10], but to date, few studies have performed genome-wide analyses. The most commonly reported CNV to be associated with early-onset obesity is highly penetrant deletions at 16p11.2 encompassing *SH2B1* (which account for ~0.7% of individuals with obesity) [11].

Here, we use high-resolution microarrays to scan the genome to investigate the contribution of rare CNVs in a pediatric cohort comprised of 67 individuals with obesity of unknown cause including 22 probands with co-occurring developmental delay (DD).

Materials and methods

Study cohort

Our study included 67 children with obesity, aged 2–17.5 years, from two different cohorts. The first group consisted of 52 individuals who were recruited through the SickKids Team Obesity Management Program (STOMP) at the Hospital for Sick Children (SickKids) in Toronto, ON, Canada. STOMP patients were 12–17.5 years old with a BMI >99%ile or a BMI >97%ile with a minimum of one significant obesity-related co-morbidity or significant co-existing chronic illness, such as technology-dependent sleep-disordered breathing, obesity-associated type 2 diabetes, hypertension, pseudotumor cerebri, dyslipidemia, psychiatric illness, severe psychological impairment, or obesity-associated orthopedic complications. All families that were interested in participating were enrolled. These individuals underwent microarray analysis at The Centre for Applied Genomics (TCAG) at SickKids and were designated the “research microarray cohort”. One additional child (6-0234-03) who was 2 years old at the time of clinical

assessment was also included in this group due to an early presentation of obesity, but was not seen through the STOMP program. Whole blood samples were collected from probands during the enrollment for DNA extraction. Subsequently, saliva kits were mailed to families to obtain parental samples, but only a small number were returned. In 47/51 (92%) of probands from this cohort, there exists a positive family history of obesity (family history of obesity is unknown for one proband). This study was approved by The Research Ethics Board at SickKids (REB#1000007909). Informed consent was obtained from all families for participation in this study.

A second group of patients comprised of 15 individuals aged 2–6 years, presenting with obesity and DD, and were enrolled through the Division of Clinical and Metabolic Genetics at SickKids in Toronto, ON, Canada. We designated this group of individuals the “clinical genetics cohort”. Probands within this group all had a BMI >99.9%ile (+3 SD) except for one 3-year-old patient who had a BMI of 99.5%ile (+2.56 SD). All patients underwent assessment by a clinical geneticist; clinical features are presented in Table 1. Clinical microarray analysis was performed in the Genome Diagnostics Cytogenetics Laboratory at SickKids. In 10/12 (83%) of probands from this cohort, there is a positive family history of obesity (family history of obesity is unknown for three probands). BMI percentiles and *z*-scores of probands (according to age and gender) were calculated using WHO’s Anthroplus software for both cohorts [12].

The majority of probands did not undergo genetic testing for known pediatric obesity syndromes. For those who did

Table 1 Clinical feature summary of the study cohort

Study cohort characteristics	Research microarray cohort (<i>N</i> = 52)	Clinical genetics cohort (<i>N</i> = 15)
Average age in years at assessment (range)	14.87 ± 2.55 (12–17)	4.64 ± 1.11 (2–6)
<i>Sex</i>		
Male	19 (37%)	10 (67%)
Female	33 (63%)	5 (33%)
<i>Average BMI</i>		
Male	+4.45 SD	+5.43 SD
Female	+4.29 SD	+5.09 SD
Developmental delay ^a	7 (15%)	15 (100%)
Family history of obesity ^b	47 (92%)	10 (83%)
Clinically relevant CNVs	NA	3 (20%)
Potentially clinically relevant CNVs	7 (13%)	NA

Abbreviations: NA not available; no CNVs identified

^aDD diagnosis was available for 62 individuals

^bInformation regarding family history of obesity was available for 63 individuals

Table 2 Summary of clinically relevant and potentially clinically relevant CNVs found in our probands

Proband ID	Gen-der	Cytoband	CNV Type	Coordinates (hg19)	Size (kb)	Genes	Inheritance	Other rare CNVs (Size-Gene)	Relevant clinical details
P0310	F	16p11.2	Del	g.(28810350_28819028)_(29051191_29056973)del	232	<i>NPIPL1</i> , <i>AK125489</i> , <i>ATXN2L</i> , <i>TUFM</i> , <i>MIR4721</i> , <i>SH2B1</i> , <i>ATP2A1</i> , <i>LOC100289092</i> , <i>ATP2A1</i> , <i>RABEP2</i> , <i>CD19</i> , <i>NFATC2IP</i> , <i>MIR4517</i> , <i>LAT</i>	Unknown	NA	Obstructive sleep apnea, mild diffuse fatty liver infiltration, insulin resistance, hyperphagia
P0821 ^a	M	16p11.2	Del	g.(28825481_28843754)_(29044850_29076193)del	201	<i>NPIPL1</i> , <i>AK125489</i> , <i>ATXN2L</i> , <i>TUFM</i> , <i>MIR4721</i> , <i>SH2B1</i> , <i>ATP2A1</i> , <i>LOC100289092</i> , <i>ATP2A1</i> , <i>RABEP2</i> , <i>CD19</i>	Maternal	NA	Anxiety, elevated cholesterol and triglycerides, hyperphagia, sleep fragmentation
P0652 ^a	M	16p11.2	Del	g.(28825481_28843754)_(29044850_29076193)del	201	<i>NPIPL1</i> , <i>AK125489</i> , <i>ATXN2L</i> , <i>TUFM</i> , <i>MIR4721</i> , <i>SH2B1</i> , <i>ATP2A1</i> , <i>LOC100289092</i> , <i>ATP2A1</i> , <i>RABEP2</i> , <i>CD19</i>	Maternal	NA	
6-0197-03	F	22q11.21	Dup	g.(20727338_20733495)_(21464479_21608479)dup	730	<i>PI4KA</i> , <i>CRKL</i> , <i>THAP7</i> , <i>SLC7A4</i> , <i>SNAAP29</i> , <i>MED15</i> , <i>AIFM3</i> , <i>LZTR1</i> , <i>SERPIND1</i> , <i>P2RX6</i> , <i>SCARF2</i> , <i>ZNF74</i>	Unknown	(i) 8q12.1 Del (18kb- <i>XKRR4</i>) (ii) 1p31.1 Del (10kb- <i>LRRRC7</i>)	Mildly enlarged liver, increased PTH levels, insulin resistance
6-0282-03	F	10q21.2-21.3	Del	g.(62989120_62994557)_(67398310_67399447)del	4403	<i>ARID5B</i> , <i>C10orf107</i> , <i>EGR2</i> , <i>RTKN2</i> , <i>ZNF365</i> , <i>TMEM26</i> , <i>JMJD1C</i> , <i>ADO</i> , <i>REEP3</i> , <i>NRBF2</i>	De novo	NA	Renal tubular acidosis, growth hormone deficiency, delayed puberty, autoimmune hepatitis, elevated cholesterol, and LDL
6-0210-03	M	2q21.2	Dup	g.(132714357_132719550)_(13354301_13354793)dup	634	<i>GPR39</i>	Maternal	NA	Learning difficulties, elevated triglycerides, insulin resistance
6-0208-03	M	6p22.2	Dup	g.(25268508_25276941)_(25896712_25903308)dup	619	<i>SLC17A4</i> , <i>LRRRC16A</i> , <i>SLC17A1</i> , <i>SLC17A3</i> , <i>SCGN</i> , <i>HIST1H2BA</i> , <i>HIST1H2AA</i>	Unknown	NA	NASH, type 2 diabetes, obstructive sleep apnea, hypertension, elevated LDL, triglycerides, and cholesterol
6-0279-03	F	6p22.2	Dup	g.(25669833_25670946)_(25738956_25740502)dup	68	<i>SCGN</i> , <i>HIST1H2AA</i> , <i>HIST1H2BA</i>	Unknown	11p15.5 Dup (63kb- <i>SYT8</i> , <i>LSP1</i> , <i>TNNI2</i>)	Elevated triglycerides and cholesterol, insulin resistance, mild learning disability, episodes of binge eating, acanthosis nigricans
6-0191-03	M	8q21.11	Dup	g.(76470447_76474058)_(76489424_76500123)dup	15	<i>HNF4G</i>	Maternal	NA	NASH, elevated LDL, and cholesterol, insulin resistance
6-0248-03	F	7q36.3	Dup	g.(157299433_157306440)_(157372132_157373059)dup	65	<i>PTPRN2</i>	Unknown	14q23.1 Del (32kb- <i>PPM1A</i>)	Elevated LDL, triglycerides, and cholesterol, insulin resistance

Abbreviations: *M* male, *F* female, *Del* deletion, *Dup* duplication, *NA* not available; no CNVs identified, *PTH* parathyroid hormone, *LDL* low-density lipoprotein, and *NASH* nonalcoholic steatohepatitis

^aSiblings

undergo additional genetic testing ($N = 14$), the results were all negative (details in Supplementary Table 1).

DNA microarray analysis

DNA was extracted from whole blood samples for 52 probands from the research microarray cohort and genotyped on three different microarray platforms; 16 samples with the Affymetrix Genome-wide Human SNP Array 6.0 (Affymetrix, Santa Clara, CA, USA), 31 samples with the Illumina HumanOmni2.5 (Illumina, San Diego, CA, USA), and five samples with the Affymetrix CytoScan HD (Affymetrix, Santa Clara, CA, USA), according to the manufacturer's guidelines. Microarray data are deposited in Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE101418>). The DNA samples from the research microarray cohort were collected over an extended period of time, during which time microarray technology evolved and, as in other studies, samples were genotyped using the relevant platform at the time (three different platforms, see Supplementary Table 2 for more details). Multiple algorithms were used for CNV calling (for details regarding algorithms, see also Supplementary Table 2). We defined CNVs as high quality (predicted by at least two algorithms, spanning at least five consecutive probes, greater than 10 kb in size and overlap <75% of segmental duplication) and rare (present in <0.1% of population controls) according to previously described criteria [13–16]. To identify rare CNVs, we first compared proband CNV calls with platform-matched control data sets and then subsequently with remaining control data sets. Our control dataset comprises of 10,851 unique subjects, majority of them with European ancestry, genotyped on multiple platforms including the Affymetrix Genome-wide Human SNP Array 6.0, Illumina HumanOmni2.5, and Affymetrix CytoScan HD. Further details regarding these samples are described elsewhere [15, 16]. DNA was also extracted from the saliva of eight parents of the 52 probands (both parents were available for three cases and only the mother for five cases). Parents of proband 6-0282-03 who also had a clinical microarray were tested by fluorescence in situ hybridization (FISH) in the Genome Diagnostics Cytogenetics Laboratory at SickKids. Other parental samples were not available or individuals declined to participate.

For the 15 probands from the clinical genetics cohort, five were analyzed on the Affymetrix CytoScan HD and 10 samples on the ISCA 4×180K Oligonucleotide array (Oxford Gene Technologies, Begbroke, UK) according to the manufacturer's instructions. Chromosome Analysis Suite (Thermo Fisher Scientific, Waltham, MA, USA) was used for the detection of CNVs in the samples analyzed on the CytoScan HD. CytoSure Interpret (Oxford Gene

Technologies, Begbroke, UK) was used for identification of CNVs in the group analyzed on the 4×180K array (Supplementary Table 2). For our analysis, we obtained all CNVs that were categorized as being pathogenic or a variant of unknown significance. Rare CNVs were determined as those being in fewer than two independent studies in the database of genomic variants (DGV) and less than 1% in the in-house database.

Prioritization of rare CNVs

We prioritized rare CNVs as clinically relevant if they corresponded with a genomic disorder that has obesity as a known phenotype and as potentially clinically relevant if they satisfied at least one of the following criteria: locus previously reported in obesity studies (i.e., CNV, GWAS), a genomic locus identified in two or more unrelated probands, genomic locus overlapping CNV reported in DECIPHER [17] database in patients with obesity phenotype, and/or impact gene(s) found to be involved in energy homeostasis or related processes (Supplementary Table 3). We used the NCBI RefSeq numbering scheme for the numbering of exons (detailed information on exon boundaries provided in Supplementary Table 5). For the research microarray cohort, we validated selected CNVs using quantitative real-time PCR (qPCR) according to previously reported procedures [18] and tested parental samples when available. In the clinical genetics cohort, microarray testing of parental samples was conducted following a clinically relevant or variant of unknown significance (potentially clinically relevant) finding in the proband. All CNVs found to be clinically relevant or potentially clinically relevant were confirmed using one of these methods with the exception of the CNV finding in proband P0310 (from the clinical genetics cohort), as her father was unavailable to undergo testing and her mother declined.

Results

The clinical features of the 67 individuals tested are presented in Table 1. All probands included in this study presented with obesity; 22 also exhibited DD. A total of 128 rare CNVs (Supplementary Table 4) were identified in the 52 probands from the research microarray cohort. Of these, 48% (61/128) were duplications and 52% (67/128) were deletions. 34% (43/128) of identified rare CNVs impact coding regions. We found variants of clinical or potential clinical relevance in 15% (10/67) of the probands (Table 2). By combining both cohorts, this represents ~4% (3/67) of probands harboring clinically relevant variants, but 20% (3/15) of probands in the clinical genetics cohort alone. In ~10% (7/67) of the probands, we identified CNVs of

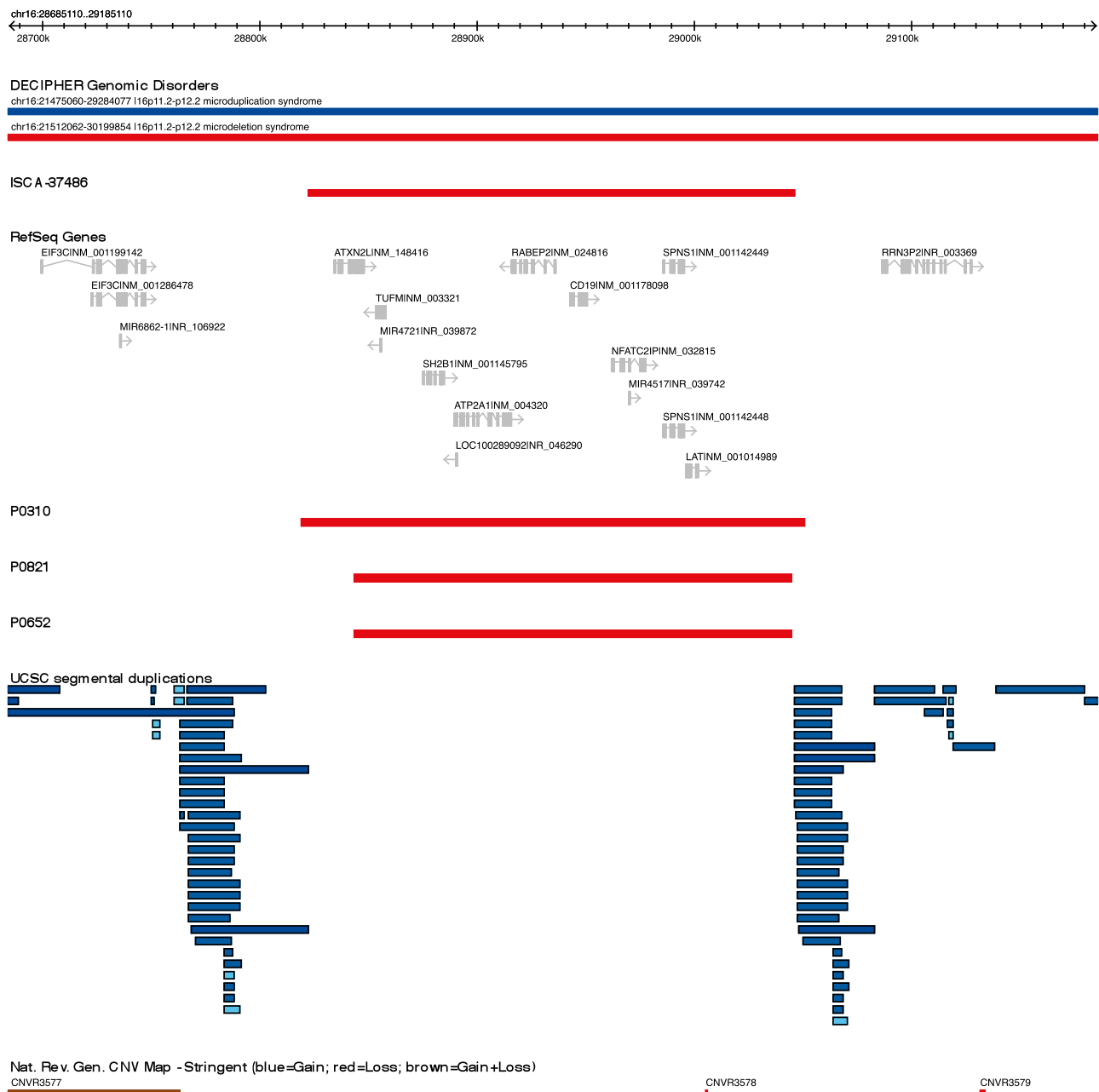


Fig. 1 CNVs impacting 16p11.2 locus. Female proband P0310 harbors a 232 kb deletion, and brothers P0821 and P0652 harbor a 201 kb deletion

potential clinical relevance using a stringent set of criteria, as described in Materials and Methods.

Clinically relevant CNVs

We prioritized rare CNVs as clinically relevant if they had previously been associated with a genomic disorder that has obesity as a known phenotype. We identified three individuals (including two siblings) from two families with pathogenic 16p11.2 deletions encompassing *SH2B1*, which is known to be the most common genomic variation

associated with early-onset obesity [11]. One female proband (P0310), who presented with early-onset obesity and DD, was found to have a 232 kb deletion in this region (Fig. 1). Her father, mother, and maternal grandmother are all reported to be obese. However, because her mother declined testing and her father was unavailable, we were unable to investigate the segregation of this deletion.

A 16p11.2 deletion encompassing 201 kb was also found in two brothers (P0821 and P0652) (Fig. 1). Both presented with early-onset obesity (≤ 3 years old) and DD. Microarray testing of parental DNA samples revealed the 16p11.2

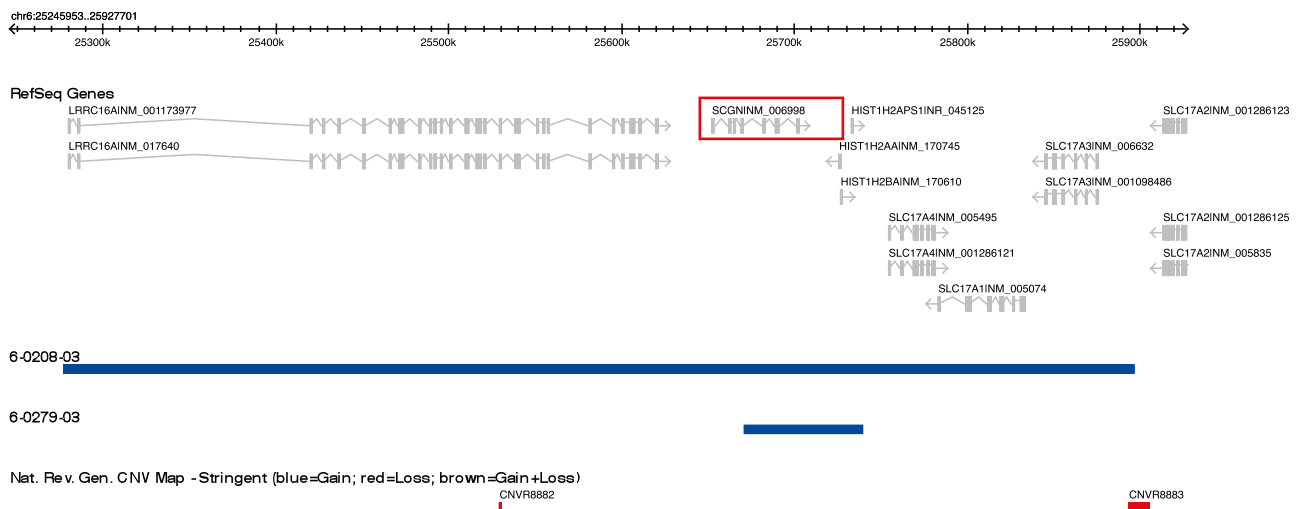


Fig. 2 CNVs impacting *SCGN* locus. Male proband 6-0208-03 harbors a 619 kb duplication, and female proband 6-0279-03 harbors a 68 kb duplication

deletion to be maternally inherited. Their mother's BMI ($\sim 22 \text{ kg/m}^2$) was within the normal range according to WHO, which is interesting considering the high penetrance of 16p11.2 deletions for the obesity trait [11]. However, the mother had a history of psychiatric illness (i.e., eating disorders, anxiety, and depression), which has been observed in some individuals with 16p11.2 deletions [19]. A third family that included two obese siblings (male and female) and an asymptomatic mother with a 16p11.2 deletion was previously reported [20]. These families reflect the clinically variable (obesity) phenotype amongst 16p11.2 deletion carriers.

Potentially clinically relevant CNVs

We used the following criteria to prioritize CNVs as potentially clinically significant: (1) locus previously reported in obesity studies (i.e., CNV and GWAS), (2) genomic locus identified in two or more unrelated probands, (3) genomic locus overlapping CNV reported in DECIPHER database in patients with obesity phenotype, or (4) genomic locus reported to be involved in energy homeostasis or related processes. We identified CNVs of potential clinical significance in 10% (7/67) of probands; the majority were found to be duplications (Table 2). In two unrelated male and female probands (6-0208-03 and 6-0279-03), we detected a duplication at 6p22.2 impacting *SCGN* (Fig. 2). Male proband 6-0208-03 harbors a 619 kb duplication of unknown inheritance encompassing seven genes including *SCGN*. His BMI plots at +5.85 SD and he has multiple obesity-associated comorbidities including type 2 diabetes, hypertension, obstructive sleep apnea, asthma, fatty liver, and dyslipidemia. Female proband 6-0279-03 carries a 68 kb duplication of unknown inheritance

also impacting *SCGN*. Her BMI plots at +2.96 SD and she has acanthosis nigricans. *SCGN* encodes secretagogin, a calcium-binding protein that is highly expressed in the neurons of the hypothalamus, neuroendocrine cells, and pancreatic β -cells. Recent studies have found that secretagogin is involved in the secretion of insulin and corticotropin-releasing hormone within pancreatic β -cells and the hypothalamus, respectively [21, 22]. Moreover, secretagogin was determined to be expressed in a fraction of NPY/AgRP neurons found in the leptin–melanocortin pathway of the hypothalamus [23]. This pathway regulates food intake and maintains energy homeostasis, supporting a role for secretagogin in energy balance. *SCGN* is copy number stable according to the CNV map [24] and the Database of Genomic Variants Gold Standard map [25].

Female proband 6-0197-03 (BMI +3.30 SD) with typical development carries a 730 kb duplication of unknown inheritance within the 22q11DS locus (2.44 Mb) encompassing 12 genes (Supplementary Figure 1). We prioritized this CNV as potentially clinically relevant as obesity has been reported in individuals with 22q11.21 deletion and duplication syndromes (ISCA: 37446) [26]. One study investigating the clinical features of adults with 22q11DS found obesity to be a prevalent phenotype [27]. A subsequent study found that an increased prevalence of obesity in an adult cohort of 22q11DS is independent of psychotropic medication use; however, the degree of obesity is increased in individuals on these medications [28].

Female proband 6-0282-03 (BMI +4.03 SD) with typical development carries a 4.4 Mb deletion of at 10q21.2–21.3 encompassing 10 genes (Supplementary Figure 2), which was determined by FISH to be de novo in origin. She also presented with renal tubular acidosis, growth hormone deficiency, delayed puberty, and autoimmune hepatitis. This

CNV satisfied our potentially clinically relevant criteria as *ARID5B*, one of the genes within this deletion is highly expressed in adipose tissue and is reported to be involved in adipogenesis and lipid metabolism in mice [29]. Further, a larger 15 Mb deletion encompassing the same locus is present in a patient presenting with obesity, cognitive impairment, delayed speech and language development, sparse hair, and supernumerary nipples, from the DECIPHER database [17].

We identified a maternally inherited 634 kb duplication at 2q21.2 encompassing the promoter region and exon 1 of *GPR39* (NM_001508.2) in male proband 6-0210-03 (BMI +2.41 SD), who developed obesity at an early age (<5 years) and has fatty liver and high triglycerides (Supplementary Figure 3). We classified this CNV as potentially clinically relevant as *GPR39*, which is widely expressed throughout the body with high levels of expression in adipose tissue, stomach, intestine, pancreas, liver, and kidney, has been shown to be involved in energy homeostasis in mice [30]. A positive family history of obesity is reported in the maternal grandfather.

A 65 kb duplication of unknown inheritance at 7q36.3 impacting exons 17–22 of *PTPRN2* (NM_001308267.1) was detected in female proband 6-0248-03 (BMI +4.16 SD) (Supplementary Figure 4). She developed cerebral palsy (spastic diplegia) following complications from interferon treatment for a hemangioma during infancy. We considered this CNV to be potentially clinically relevant because *PTPRN2* has been reported to be involved in insulin secretion processes [31]. *PTPRN2* encodes a tyrosine phosphatase receptor that is found in neuroendocrine cells and is associated with cell growth and differentiation. Moreover, a recent study investigating the role of rare CNVs in obsessive-compulsive disorder implicates a duplication encompassing *PTPRN2* [15].

Lastly, male proband 6-0191-03 (BMI +4.82 SD) with typical development carries a maternally inherited 15 kb duplication of 8q21.11 encompassing the last exon of *HNF4G* (NM_004133.4) (Supplementary Figure 5). Knockout mouse studies suggest that this gene likely plays a role in the regulation of energy balance [32]. Thus, we considered this CNV to be of potential clinical relevance. In addition, *HNF4G* was also identified as a novel obesity-associated locus in a meta-analysis of GWAS studies investigating anthropometric traits [33].

pLI

We obtained the probability of truncating loss-of-function intolerance (pLI) [34] for all genes within the 14 rare deletions impacting coding regions. Only 4 of these 27 genes were likely to be intolerant of loss-of-function variants (pLI \geq 0.9), two of which, (*ARID5B* (pLI = 1) and

JMJD1C (pLI = 1)) were within the 10q21.2–21.3 deletion in proband 6-0282-03, which was considered likely pathogenic. The other two genes were *PPM1A* (pLI = 0.99), which has not previously been implicated in disease and *LARGE* (pLI = 0.96), which is associated with autosomal recessive congenital muscular dystrophy type 1D. Neither *PPM1A* nor *LARGE* have been implicated in obesity to date.

Discussion

We describe the identification of CNVs affecting known obesity risk genes/loci or new functional candidate genes in 15% (10/67) of individuals in our study cohorts. In our research microarray cohort, we have a higher ratio of females (63%) to males (37%) and only 15% of these individuals have DD. In contrast, we have a greater proportion of males (67%) in our clinical genetics cohort who comprised of probands with obesity as well as DD, which is consistent with the literature where a greater frequency of males are reportedly affected by DD [35]. In 47/51 (92%) of probands from the research microarray cohort and 83% (10/12) of probands from the clinical genetics cohort, a positive family history of obesity was observed. However, we were unable to investigate CNV segregation patterns in all families due to limited availability of parental samples.

Previously, Vuillaume et al. investigated the contribution of CNVs in 100 individuals (40 females, 60 males) who were 1–18 years old with syndromic obesity, which is generally defined as the accompaniment of obesity with congenital malformations and/or DD [36]. They reported clinically relevant or potentially clinically relevant CNVs in 22% of their patients [10]. This is a higher frequency than we found in our cohort, a difference that could be attributed to the variation in clinical features among cohorts. Their study was comprised of patients with obesity accompanied by mainly DD (85%) and/or dysmorphic facial features (78%). Rare CNVs are well recognized to significantly contribute to the pathogenesis of DD [37]. When we restrict our cohort to only those probands with comorbid DD, we find clinically relevant or potentially clinically relevant CNVs in 14% (3/21) of unrelated cases.

We identified two families with pathogenic 16p11.2 deletions. This highly penetrant deletion is the most frequently reported (~0.7%) genomic alteration in individuals with early-onset obesity. This CNV was only identified in our clinical genetics cohort, where 14% (2/14) of unrelated probands carry this deletion. This higher than expected frequency likely reflects our small sample size and/or sample ascertainment bias. Our list of potentially clinically relevant CNVs were all identified in our research microarray cohort with 13% (7/52) of these individuals harboring CNVs.

We identified multiple candidate genes that are likely biologically relevant to obesity. We found a 10q21.2–21.3 de novo deletion (4.4 Mb) in proband 6-0282-03. *ARID5B*, a gene found in this region is essential for the formation of histone H3K9me2 demethylase complexes involved in the regulation of genes required for adipogenesis [29, 38]. Interestingly, a recent study found that the obesity-associated *FTO* SNP rs1421085 affects a conserved binding motif of *ARID5B*, leading to disrupted regulation of two genes (i.e., *IRX3* and *IRX5*) involved in adipocyte thermogenesis. This disruption, which leads to decreased thermogenesis and an increased storage of lipids in primary preadipocytes of *FTO*-risk allele carriers, has been postulated to be responsible for the development of obesity [39].

In another proband, we identified a duplication partially encompassing *GPR39*. There is compelling data for the involvement of *GPR39* in lipolysis via the regulation of hormone-sensitive lipase (HSL) and adipose triglyceride lipase (ATGL). Mice deficient for *GPR39* have been shown to have significantly increased body weight compared with wild-type littermates. The authors concluded from this study that the observed body weight increase was the result of decreased lipolytic capacity rather than increased food intake [30].

Another proband was identified with a duplication encompassing the last exon of *HNF4G*. This gene encodes a nuclear hormone receptor that has been identified through GWAS to be associated with BMI [33]. As well, targeted disruption of *Hnf4g* in mice leads to increased body weight as a result of decreased physical activity and reduced energy expenditure [32].

Last, our finding of CNVs impacting *SCGN* in two unrelated probands is potentially the most interesting discovery. *SCGN* encodes a protein containing six calcium-binding domains. *SCGN* resides in a copy stable region of the genome, i.e., CNVs are rarely found at this locus in control populations [24]. There is intriguing evidence of the possible role of *SCGN* in the regulation of energy homeostasis [23] and these observations together make *SCGN* an attractive candidate gene for obesity. One of the duplications (in proband 6-0208-03) spans the whole *SCGN* gene as well as seven other genes (619 kb). The second duplication (in proband 6-0279-03) is 68 kb in size and spans the last five exons of *SCGN* (NM_006998.3). This partial duplication encompasses the last three calcium-binding domains of *SCGN*. Given that *SCGN* has a low loss-of-function intolerance score (pLI: 0.04) [34], it is likely not sensitive to haploinsufficiency; so the impact of these CNVs may be due to other mechanisms, such as dosage or gain of function. Additional studies are needed to ascertain the pathogenicity of CNVs involving the *SCGN* locus with respect to human obesity.

Here, we identify important candidate genomic regions for obesity using a CNV scanning approach. The novel

candidate genes identified do not function in the same biological pathways or processes, adding additional support for significant heterogeneity involved in pediatric obesity.

In the current approach, microarrays were used because they are the first-tier test in our clinical laboratories providing the best-understood data [40]. However, microarrays are able to only accurately detect deletions and duplications of size 10 kb or greater. Further, as not all probands underwent genetic testing for known pediatric obesity syndromes (e.g., obesity gene panel and Prader–Willi syndrome testing), it is also possible that some may have a pediatric obesity syndrome due to sequence-level variant(s) in known disease-associated genes. We anticipate that the application of whole-genome sequencing [41] to detect all classes of genetic variants will lead to additional genetic factors and loci being identified in pediatric obesity, and potentially add further support for the novel loci we have identified.

Acknowledgments The authors would like to thank all the families who participated in this study. Special thanks to the centre for applied genomics for assistance with laboratory and bioinformatics analysis; in particular Bhooma Thiruvahindrapuram and Dr. John Wei. We also thank Drs. Mehdi Zarrei and Sanaa Choufani for helpful comments and suggestions. This project was supported by grants from the Canada Foundation for Innovation, Government of Ontario, the Canadian Institutes of Health Research, and the University of Toronto McLaughlin Centre. SWS holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at The Hospital for Sick Children and the University of Toronto.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Spiegelman BM, Flier JS. Obesity and the regulation of energy balance. *Cell*. 2001;104:531–43.
2. De Onis M, Onyango AW. WHO child growth standards. *Lancet*. 2006;95:76–85.
3. Maes HH, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behav Genet*. 1997;27:325–51.
4. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature*. 2015; 518: 197–206.
5. Lee C, Scherer SW. The clinical context of copy number variation in the human genome. *Expert Rev Mol Med*. 2010;12:e8.
6. Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nat Genet*. 2017; 49: 27–35.
7. Wellcome Trust Case Control Consortium. Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3000 shared controls. *Nature*. 2010;464:713–20.
8. Richards AA, Garg V. Genetics of congenital heart disease. *Curr Cardiol Rev*. 2010;6:91–7.

9. Baskin B, Choufani S, Chen YA, Shuman C, Parkinson N, Lemyre E, et al. High frequency of copy number variations (CNVs) in the chromosome 11p15 region in patients with Beckwith-Wiedemann syndrome. *Hum Genet.* 2014;133:321–30.
10. Vuillaume ML, Naudion S, Banneau G, Diene G, Cartault A, Cailley D, et al. New candidate loci identified by array-CGH in a cohort of 100 children presenting with syndromic obesity. *Am J Med Genet Part A.* 2014;164:1965–75.
11. 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.* 2010;463:671–5.
12. WHO. WHO AnthroPlus software. 2015. <http://www.who.int/growthref/tools/en/>.
13. Oskoui M, Gazzellone MJ, Thiruvahindrapuram B, Zarrei M, Andersen J, Wei J, et al. Clinically relevant copy number variations detected in cerebral palsy. *Nat Commun.* 2015;6:7949.
14. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, et al. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. *Sci Transl Med.* 2011;3:95ra75.
15. Gazzellone MJ, Zarrei M, Burton CL, Walker S, Uddin M, Shaheen SM, et al. Uncovering obsessive-compulsive disorder risk genes in a pediatric cohort by high-resolution analysis of copy number variation. *J Neurodev Disord.* 2016;8:36.
16. Zarrei M, Fehlings DL, Mawjee K, Switzer L, Thiruvahindrapuram B, Walker S, et al. De novo and rare inherited copy number variations in the hemiplegic form of cerebral palsy. *Genet Med.* 2018;20:172–80.
17. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources. *Am J Hum Genet.* 2009;84:524–33.
18. Gazzellone MJ, Zhou X, Lionel AC, Uddin M, Thiruvahindrapuram B, Liang S, et al. Copy number variation in Han Chinese individuals with autism spectrum disorder. *J Neurodev Disord.* 2014;6:34.
19. Bachmann-Gagescu R, Mefford HC, Cowan C, Glew GM, Hing AV, Wallace S, et al. Recurrent 200-kb deletions of 16p11.2 that include the SH2B1 gene are associated with developmental delay and obesity. *Genet Med.* 2010;12:641–7.
20. Dasouki MJ, Youngs EL, Hovanes K. Structural chromosome abnormalities associated with obesity: report of four new subjects and review of literature. *Curr Genom.* 2011;12:190–203.
21. Yang SY, Lee JJ, Lee JH, Lee K, Oh SH, Lim YM, et al. Secretagogin affects insulin secretion in pancreatic β -cells by regulating actin dynamics and focal adhesion. *Biochem J.* 2016;473:1791–803.
22. Romanov RA, Alpár A, Zhang MD, Zeisel A, Calas A, Landry M, et al. A secretagogin locus of the mammalian hypothalamus controls stress hormone release. *EMBO J.* 2014;34:36–54.
23. Gyengesi E, Paxinos G, Andrews ZB. Oxidative stress in the hypothalamus: the importance of calcium signaling and mitochondrial ROS in body weight regulation. *Curr Neuropharmacol.* 2012;10:344–53.
24. Zarrei M, Macdonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. *Nat Rev Genet.* 2015;16:172–83.
25. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The database of genomic variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res.* 2014;42:D986–92.
26. D'Angelo CS, Varela MC, de Castro CI, Kim CA, Bertola DR, Lourenço CM, et al. Investigation of selected genomic deletions and duplications in a cohort of 338 patients presenting with syndromic obesity by multiplex ligation-dependent probe amplification using synthetic probes. *Mol Cytogenet.* 2014;7:75.
27. Bassett AS, Chow EW, Husted J, Weksberg R, Caluseriu O, Webb GD, et al. Clinical features of 78 adults with 22q11 deletion syndrome. *Am J Med Genet Part A.* 2005;138:307–13.
28. Voll SL, Boot E, Butcher NJ, Cooper S, Heung T, Chow EWC, et al. Obesity in adults with 22q11.2 deletion syndrome. *Genet Med.* 2017;19:204–8.
29. Yamakawa T, Sugimoto K, Whitson RH, Itakura K. Modulator recognition factor-2 regulates triglyceride metabolism in adipocytes. *Biochem Biophys Res Commun.* 2010;391:277–81.
30. Petersen PS, Jin C, Madsen AN, Rasmussen M, Kuhre R, Egerod KL, et al. Deficiency of the GPR39 receptor is associated with obesity and altered adipocyte metabolism. *FASEB J.* 2011;25:3803–14.
31. Cai T, Hirai H, Zhang G, Zhang M, Takahashi N, Kasai H, et al. Deletion of *Ia-2* and/or *Ia-2 β* in mice decreases insulin secretion by reducing the number of dense core vesicles. *Diabetologia.* 2011;54:2347–57.
32. Gerdin AK, Surve VV, Jönsson M, Bjursell M, Björkman M, Edenro A, et al. Phenotypic screening of hepatocyte nuclear factor (HNF) 4- γ receptor knockout mice. *Biochem Biophys Res Commun.* 2006;349:825–32.
33. Berndt SI, Gustafsson S, Mägi R, Ganna A, Wheeler E, Feitosa MF, et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat Genet.* 2013;45:501–12.
34. Lek M, Karczewski K, Minikel E, Samocha K, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536:285–91.
35. Boyle CA, Boulet S, Schieve LA, Cohen RA, Blumberg SJ, Yeargin-Allsopp M, et al. Trends in the prevalence of developmental disabilities in US children, 1997–2008. *Pediatrics.* 2011;127:1034–42.
36. Bray GA. Definition, measurement, and classification of the syndromes of obesity. *Int J Obes.* 1978;2:99–112.
37. 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.* 2011;43:838–46.
38. Baba A, Ohtake F, Okuno Y, Yokota K, Okada M, Imai Y, et al. PKA-dependent regulation of the histone lysine demethylase complex PHF2-ARID5B. *Nat Cell Biol.* 2011;13:668–75.
39. Claussnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C, et al. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med.* 2015;373:895–907.
40. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet.* 2010;86:749–64.
41. Stavropoulos DJ, Merico D, Jobling R, Bowdin S, Monfared N, Thiruvahindrapuram B, et al. Whole-genome sequencing expands diagnostic utility and improves clinical management in paediatric medicine. *NPJ Genom Med.* 2016;1:15012.