Genetic analysis of very obese children with autism spectrum disorder

Herman D. Cortes and Rachel Wevrick*

Dept. of Medical Genetics, University of Alberta, Edmonton, Alberta Canada

*Corresponding author.

Authors:

Rachel Wevrick, Dept. of Medical Genetics, University of Alberta, Edmonton, Alberta Canada,
rwevrick@ualberta.ca

Herman D. Cortes, Dept. of Medical Genetics, University of Alberta, Edmonton, Alberta Canada,
hdc2@ualberta.ca

Running title: Syndromic autism with obesity in the SSC

Key words: SSC, autism spectrum disorder, obesity, Wnt, Gene Ontology

Abstract: Autism spectrum disorder (ASD) is defined by the triad of deficits in social interactions, deficits in communication, and repetitive behaviors. Common co-morbidities in syndromic forms of ASD include intellectual disability, seizures, and obesity. We asked whether very obese children with ASD had different behavioral, physical and genetic characteristics compared to children with ASD who were not obese. We found that very obese children with ASD had significantly poorer scores on standardized behavioral tests. Very obese boys with ASD had lower full scale IQ and increased impairments with respect to stereotypies, communication and social skills. Very obese girls with ASD had increased impairments with respect to irritability and oppositional defiant behavior. We identified genetic lesions in a subset of the children with ASD and obesity and attempted to identify enriched biological pathways. Our study demonstrates the value of identifying co-morbidities in children with ASD as we move forward towards understanding the biological processes that contribute to this complex disorder and prepare to design customized treatments that target the diverse genetic lesions present in individuals with ASD.

Introduction

Autism spectrum disorder (ASD) has complex origins with contributions from genetic and environmental factors. As more genes are identified in which sequence variants contribute to ASD predisposition, it can be useful to determine whether mutations in certain genes are responsible not only for ASD, but also for co-morbid conditions that define syndromic forms of ASD [1]. Fragile X syndrome, Angelman syndrome, Rett syndrome and a syndrome associated with *POGZ* mutations [2] are examples of syndromes with defined genetics that have high rates of behavioral phenotypes either fitting the definition of ASD or closely mimicking ASD [3]. Other syndromes with defined genetics and with elevated risks of ASD include Down syndrome [4], del16p11 syndrome [5], Bardet-Biedl syndrome [6], and Prader-Willi syndrome (PWS) [7]. However, in the latter syndromes, how the loss of specific genes confers increased risk of ASD is still unclear. For example, the expression of multiple contiguous genes is disrupted in PWS. However, mutations in only one of the inactivated genes, namely *MAGEL2*, cause a PWS-like syndrome called Schaaf-Yang syndrome [8-10]. Inactivating mutations in *MAGEL2* also carry a very high risk of ASD, suggesting that loss of *MAGEL2* also contributes to the increased incidence of ASD in children with PWS.

Children with ASD have an increased risk of overweight and obesity [11-14]. However, the genetic contribution to the co-morbidity of obesity in ASD has not been systematically studied [15, 16]. While rising rates of obesity in adults have been attributed to societal changes in physical activity and food choices, severe childhood obesity more often involves changes in the central neural circuits that regulate food intake and energy expenditure [17]. In principle, ASD and obesity could co-occur in an individual because of a single inherited or sporadic genetic mutation (e.g. *MAGEL2* deficiency), two inherited or sporadic genetic mutations (one gene responsible for ASD and a different gene responsible for obesity), or a combination of genetic mutations and multifactorial contributions. Indeed, a variety of circumstances such as reduced activity, access to health care, and interrupted sleep patterns may contribute to obesity in ASD probands [12, 14]. Some probands could have an inherited predisposition to obesity that is unrelated to their ASD diagnosis, such as variation in the melanocortin-4 receptor or *FTO* genes [18].

ASD- and obesity-predetermining genes participate in a large variety of biological processes, many of which affect the function of central neural circuits. Proteins important for ion transport and chromatin remodeling are common sources of mutations in autism gene sets [19]. The planar cell polarity (PCP) and Wnt-beta catenin pathways that control cell polarity and adhesion have been implicated in both autism [20-22] and in the hypothalamic regulation of energy balance [23]. A key player is the ASD-associated gene *CTNNB1*, encoding beta catenin, in which mutations cause

intellectual disability (ID)/ASD with neonatal hypotonia and respiratory problems [24, 25]. Mutations in a set of genes important for PCP pathways and ciliary function cause Bardet-Biedl syndrome, a neurodevelopmental disorder associated with intellectual disability and obesity [26].

An excess of *de novo* copy number variants (dnCNVs) in probands (0.05 per individual) compared to siblings (0.02 per individual) and a larger number of genes contained within proband dnCNVs compared to sibling dnCNVs are found in autism cohorts, suggesting that dnCNVs are associated with risk of ASD [27]. Whole exome sequencing studies have estimated an average of 0.17 *de novo* likely gene disrupting (dnLGD) and 0.94 *de novo* non-synonymous coding mutations (dnNSC) per affected child, also in excess over the *de novo* events in the control siblings [27, 28]. Most recent estimates have implicated at least 8 CNVs and 65 genes with increased risk for ASD [29]. We hypothesized that candidate genes for ASD with severe obesity could be identified by analyzing genetic variants found in very obese children with ASD from the Simons Simplex Collection (SSC) [30].

Methods

Probands and siblings: At the time of study, the SSC Version 15 Phenotype Data Set 9 contained data on 2873 probands with ASD, their parents, and one or more siblings [31]. The data include psychological and IQ testing, medical history, dysmorphology, birth order, and anthropometric information. Details of phenotypic data are available from SFARI Base (sfari.org/resources/autism-cohorts/simons-simplex-collection). Height Z-score, body mass index (BMI), and BMI Z-score were drawn from the SSC dataset for each proband and designated unaffected sibling. As with height, normal ranges of BMI vary with age in children, so the use of Z-scores rather than absolute values allowed standardized comparison of children from 3 to 18 years of age. Height-Z and BMI-Z scores were pre-calculated in the SSC dataset using the World Health Organization growth standards and the age at evaluation from the SSC “Height, Weight, and Head Circumference” form.

Genes: Genetic mutations in SSC probands were derived from data in GPF: Genotype and Phenotype in Families [32]. Constraint scores defining tolerance to genetic mutation were derived from Exome Aggregation Consortium (ExAC) data [33]. Intolerance to mutation is assumed when the observed number of mutations (missense or loss of function) is significantly less than the number of mutations expected.

Statistics: Comparisons between groups of probands were performed by Student t-test followed by post-hoc comparisons with Bonferroni corrections and a value of $P < 0.05$ for significance. Data in Table 1 were analyzed by Student t-test. Data in Table 2 were analyzed by the PANTHER Overrepresentation Test (release 20170413) using a Bonferroni post hoc correction for multiple testing

[34]. The correlation between birth order and full scale IQ was calculated by the Pearson method. Fisher's exact test was used to compare the mean BMI-Z and height-Z scores of probands vs unaffected siblings.

Results

Relationships between body mass index and anthropometric measurements, IQ, and neonatal history in SSC probands. We assessed the relationships between BMI and other anthropometric measurements, IQ, behavior, and neonatal history in the SSC SFARI Base, which contains data on probands with ASD, their parents, and siblings when available [31]. Age, height-Z, and BMI-Z scores were available for 2246 male probands and 338 female probands (Fig. 1). The mean BMI-Z of probands (0.66 ± 1.38 , mean $\pm$ SD) was slightly larger than that of unaffected siblings (0.55 ± 1.27 , Student t-test, $P=0.01$). More probands than siblings had a BMI-Z ≥ 2 (18% vs. 13%, Fisher's exact test, $P<0.0001$). These data are consistent with previous reports of increased risk of obesity in children with ASD [11-14]. Mean height-Z of probands (0.4 ± 1.1 for males and 0.2 ± 1.2 for females) was similar to that of unaffected siblings (Student t-test).

Mother's BMIs (mean 27 ± 7 kg/m²) and father's BMIs (mean 29 ± 6 kg/m²) were calculated from their weight and height when available. Both weight and height have large genetic and environmental components, so we expected that probands with higher BMI would have parent(s) with higher BMI. Consistent with this prediction, 7% of obese probands with BMI-Z ≥ 2 had one or two severely obese parents (BMI ≥ 35), whereas 2.5% of probands with $-2 \leq \text{BMI-Z} \leq 2$ had one or two severely obese parents. This suggests that some probands with high BMI for age have an inherited predisposition to obesity, while others, particularly those with normal weight parents and siblings, may also carry deleterious mutations in obesity-predisposing genes.

We next tested whether there is a relationship between BMI-Z, ASD symptoms and full scale IQ (FSIQ) among SSC probands. In females, BMI-Z and FSIQ were not correlated. However, compared to typical weight males ($-2 \leq \text{BMI-Z} \leq 2$), mean FSIQ was 12 points lower among very obese males (BMI-Z ≥ 3) (Table 1). Obese male probands (BMI-Z ≥ 2) and underweight male probands (BMI-Z ≤ -2) also had significantly lower FSIQ than typical weight males. Verbal IQ was 13 points lower in very obese males compared to typical weight males. Consistent with a previous study [35], mean scores for some behaviors were worse in very obese compared to typical weight children with ASD: very obese males had increased impairments with respect to stereotypies on the Aberrant Behavior Checklist (ABC), overall language on the Autism Diagnostic Interview (ADI), social problems, somatic problems

and complaints, and total competence on the Child Behavior Checklist (CBCL 6-18), awareness and communication on the teacher version of the Social Responsiveness Scale, and in communication, daily living skills, and socialization on the Vineland Adaptive Behavior Rating Scales (VABS) (Fig. 2). Very obese females with ASD had larger impairments with respect to irritability (ABC) and aggressive and oppositional defiant behaviors (CBCL) compared to typical weight females (Fig. 2). It is important to note that many of these measures are correlated with each other: for example, lower IQ is correlated with increased stereotypies [36]. There were no behaviors that were less impaired in the very obese group in any of the scales used.

Table 1. FSIQ (mean±SD) and BMI-Z in male probands.

Group	BMI-Z	FSIQ	Verbal IQ
Group 1 (typical, n=1778)	-2<BMI-Z<2	83±28	80±31
Group 2 (very obese, n=110)	BMI-Z≥3	72±29 ^{\$}	67±33 ^{\$}
Group 3 (obese, n=412)	BMI-Z≥2	79±27 [*]	75±31
Group 4 (underweight, n=61)	BMI-Z≤-2	76±28 [#]	73±30

*P=0.002 different from Group 1

#P=0.05 different from Group 1

\$P<0.00002 different from Group 1

In humans, increased paternal age correlates with an increased number of *de novo* mutations in offspring at about 2 mutations per year [37]. However, Sanders *et al.* found no evidence for a correlation between the number of *de novo* mutations per proband and parental age in the SSC. We examined paternal age in very obese SSC probands, using birth order as a surrogate index for parental age. Very obese male probands had a higher average birth order (2.0±0.9) compared to typical weight male probands (1.7±0.8, $P<0.0001$), and so are expected to have slightly older parents. However (see below), there was no evidence that very obese probands had more *de novo* mutations than probands of typical weight.

We next examined SSC phenotypic data to determine whether obese or very obese SSC probands had other clinical findings that could suggest a syndromic form of autism. We were particularly interested in neonatal history, given that neonatal hypotonia or difficulty feeding are found in many neurodevelopmental disorders. Of 2591 SSC probands, 19% reported some neonatal complication (Fig. 3). The most common neonatal complications were difficulty feeding and respiratory problems (in 19%), poor suck (10%), or decreased muscle tone (7%). APGAR scores at 1 minute were reported to be below 7 in 5% of infants, and 7% were admitted to the NICU. These data

represent minimum prevalence, as the absence of a report does not necessarily mean that the phenotype was not present. The prevalence of most neonatal complications was similar for probands who later presented as underweight ($\text{BMI-Z} \leq -2$, $n=66$), obese ($\text{BMI-Z} \geq 2$, $n=465$), or very obese ($\text{BMI-Z} \geq 3$, $n=119$). There was a trend towards a higher incidence of “stiff infant” among the underweight class and a trend towards higher incidence of broken bones reported in the very obese class, but these trends were not statistically significant after accounting for multiple comparisons (Fig. 3).

Genetic analysis of very obese probands in the SSC.

We used GPF: Genotype and Phenotype in Families to identify *de novo* mutations in 133 probands with $\text{BMI-Z} \geq 3$ (10 F, 123 M) [32]. We limited the search to variant types sub, ins, del, or CNV with effect types coding (Nonsense, Frame-shift, Splice-site, Missense, Non-frame-shift, noStart, noEnd) or CNV (+ or -). There were 120 such *de novo* mutations, including 17 dnCNVs, 20 dnLGD, Frame-shift, Nonsense, or Splice-site) mutations and 83 mutations of the remaining varieties (Suppl. Table 1). The number of *de novo* mutations per very obese proband was not different from the number of *de novo* mutations per typical weight proband. Some probands had *de novo* mutations in more than one category. Three dnCNVs in very obese probands had already been associated with ASD risk [38-41] (dup7q11, dup15q12, del16p11). Both the dup15q12 and del16p11 CNVs are also associated with obesity [42].

Analysis of loss of function mutations. Of the 20 genes with dnLGD mutations in very obese probands, 11 had already been associated with ASD in SFARI Gene [31] (*ANK2*, *DNMT3A*, *DSCAML1*, *MYH10*, *POGZ*, *PPM1D*, *PSD3*, *PTK7*), OMIM (*VCP*) or in neurodevelopmental disorders [2, 43] (*SPEN*, *STAM*). Only two of these genes have been associated with overgrowth: *DNMT3A* and *POGZ*. The proband identified here is the only one in the SSC with a dnLGD mutation in *DNMT3A* (DNA-methyltransferase 3 alpha). *DNMT3A* is mutated in children with Tatton-Brown-Rahman syndrome, which is associated with tall stature, a distinctive facial appearance, and intellectual disability. *De novo* inactivating mutations in *POGZ*, encoding the chromatin factor Pogo transposable element with ZNF domain, cause syndromic autism with hypotonia and obesity (White-Sutton syndrome [2]). Only two probands in the SSC have dnLGD mutations in *POGZ*: the one identified here with $\text{BMI-Z}=3.6$ and another with $\text{BMI-Z}=2$, both consistent with White-Sutton syndrome. Inactivating mutations in *PTK7* (encoding a tyrosine-protein kinase), and inherited or *de novo* mutations in *PPM1D* (encoding a protein phosphatase) have been found in children with ASD [27, 44]. The female with a *PPM1D* mutation also carries a dnNSC in *SCN1A*, encoding a sodium channel associated with epilepsy, and a dnLGD mutation in *TNFRSF8*. dnLGD mutations in *MYH10* (encoding a non-muscle myosin) have been found in children with moderate to severe ID [44-47]. Heterozygous LGD

mutations in *VCP* (valosin containing protein) cause amyotrophic lateral sclerosis with dementia or myopathy [48]. The male with a *VCP* dnLGD mutation also carried a dnNSC in *SCN4A*, a gene mutated in autosomal dominant and autosomal recessive muscular conditions. The other 10 genes in which dnLGD mutations were found have not been associated with neurodevelopmental genetic disorders to date. There were 83 dnNSC mutations, of which 20 are in genes plausibly associated with autism in SFARI gene (*ANK3*, *CSMD1*, *DISC1*, *GLRA2*, *GRM7*, *KMT2C*, *MCC*, *NLGN2*, *PRKDC*, *PTK7*, *SCN1A*, *SCN4A*, *SLCO1B3*, *TRIP12*, *TTN*, *TUBGCP5*, *ZC3H4*) or other neurodevelopmental disorders (*ANO3*, *CNTN6*, *C9orf156*, *RYR1*, *SCN11A*, *SPEN*, *TAF6*, *TRPM6*). Variants in *ZC3H4* are also associated with uncontrolled eating [49], but otherwise none of the genes with dnNSC are associated with obesity.

Consideration of tolerance to mutation for ASD/obesity candidate genes. We next considered the tolerance of the candidate genes to either loss of function (LoF) or missense mutations using the ExAC interface [33]. Reduced tolerance would be expected of a gene for which haploinsufficiency reduces fertility, and would increase the likelihood that the *de novo* mutation is pathogenic. Of the genes with dnLGD mutations, ten previously identified ASD/ID genes were intolerant to LoF, missense or both (*ANK2*, *DNMT3A*, *DSCAML1*, *MYH10*, *POGZ*, *PPMID*, *PTK7*, *SPEN*, *STAM*, and *VCP*). Of the remaining dnLGD genes, some were highly intolerant to LoF mutations (*CDC23*, *SPAG9*) or to missense mutations (*CDC23*) while the remaining were tolerant to both LoF and missense mutations. Of the genes in which there were dnNSC mutations in very obese probands, some were intolerant to either missense or LoF mutations, including genes associated with neurodevelopmental disorders (*GLRA2*, *GRM7*, *KMT2C*, *MEGF8*, *PTK7*, *RYR1*, *SCN1A*, *TRIP12*). Other genes not yet associated with disorders were also deemed intolerant to either missense or LoF mutations, while the remaining genes were either tolerant or of unknown tolerance (Suppl. Table 1).

Gene ontology pathways identifies proteins important for neuronal action potentials as contributors to ASD in very obese probands. We used Gene Ontology (GO) to identify pathways enriched in the *de novo* mutated gene set compared to expected for all GO-annotated genes in the genome [50]. This analysis is constrained by the fact that the very obese group differs from the typical weight group not only in high BMI-Z but also in lower IQ (Table 1) and more severe behavioral scores (Fig. 2), so that the genes mutated in this group may contribute to ASD or any one of these co-morbid phenotypes. The combined set of mutated genes (dnLGD and dnNSC) is significantly enriched in genes involved in neuronal membrane depolarization during action potential and voltage-gated ion channel activity (*ANK2*, *ANK3*, *ANO9*, *ELK1*, *GLRA2*, *GRM7*, *NLGN2*, *RYR1*, *SCN1A*, *SCN3A*, *SCN4A*,

SCN11A, *SLC17A3*, *TRPM6*, *VCP*) (Table 2). Analysis of missense mutations alone produced a similar profile of disruption of neuronal action potential.

Table 2. Gene Ontology Consortium Enrichment Analysis.

GO biological process complete	Fold Enrichment	P value
1: neuronal action potential	36	0.0003
1a: regulation of membrane potential	6	0.003
2: membrane depolarization during action potential	25	0.01
GO molecular function complete		
3: voltage-gated sodium channel activity	36	0.02
3a: sodium channel activity	30	0.004
3b: voltage-gated ion channel activity involved in regulation of postsynaptic membrane potential	58	0.003
3bi: gated channel activity	8	0.02

Data were derived from PANTHER Overrepresentation Test (release 20170413), using the GO Ontology database released 2017-06-29, and a Bonferroni correction for multiple testing.

Planar cell polarity / Wnt signaling genes are also mutated in obese probands. We next investigated whether any of the mutated genes were previously implicated in obesity pathways. Intriguingly, all three SSC probands who carry *de novo* mutations in the PCP/Wnt pathway *PTK7* gene [51] are very obese (Suppl. Table 2). PCP pathways intersect with phosphoinositide 3-kinase (PI3K) pathways, and 3 additional very obese probands carry *de novo* mutations in the PI3K pathway genes *PSD3*, *MYH10*, and *ANK3* respectively. We further investigated PCP/Wnt pathway genes in obese probands in the SSC (i.e. BMI-Z $\geq$ 2). An obese SSC proband (BMI-Z=2.5) carries a dnLGD mutation in *DVL3*, one of the three human genes encoding forms of the Wnt relay protein Dishevelled [52]. This proband had neonatal feeding difficulties and physical anomalies including genital anomalies. Another obese SSC proband (BMI-Z=2.5) carries a dnLGD mutation in *NOTCH1*, encoding a Dishevelled antagonist. This proband had neonatal feeding difficulties, and was stiff at birth, phenotypes also seen in infants with Adams-Oliver syndrome caused by *NOTCH1* mutations [53]. *CNTN6* encodes a NOTCH signaling pathway protein (contactin) and is mutated in a very obese proband [54]. Yet another obese SSC proband (BMI-Z=2.7) carries a dnLGD mutation in *MED13L*, a gene associated with ID with transposition of the great arteries, and implicated in Wnt signaling [55]. This proband had neonatal respiratory distress, feeding difficulty, and other unspecified defects. Last, we looked at the *DYRK1A* gene, encoding a kinase that inactivates GSK3-beta and is implicated in neurodevelopmental disorders with co-morbid autism [56]. Of four probands in the SSC with *DYRK1A* mutations, three

were obese and two had multiple neonatal complications. The Wnt/PCP genes *PTK7*, *RAPGEF2*, *CTNNB1*, *DVL3*, *MED13L*, *DYRK1A*, and *NOTCH1* are all highly intolerant to LoF mutations, supporting their essential role in development. There were no loss of function mutations in the SSC collection in genes responsible for Bardet-Biedl syndrome, which is also a disorder of planar cell polarity. This is likely because children with Bardet-Biedl syndrome typically develop blindness and deafness, which would exclude them from the SSC.

We next considered a potential role for the *MUC* genes in ASD with high BMI. The encoded mucin proteins are aberrantly regulated through Wnt-dependent pathways in tumors, particularly in cancers of the gastrointestinal tract. Variants in *MUC5B* are associated with type 2 diabetes [57]. Recurrent *de novo* nonsynonymous mutations in the *MUC5B* gene have previously been noted in cohorts of patients with autism or intellectual disability. However, *MUC5B* has been previously excluded as an ASD candidate gene because of high variation in controls [58]. In contrast, ExAC analysis suggests that *MUC5B* is highly constrained for both missense and LoF mutations despite a high absolute number of variants in controls. All five probands in the SSC collection carrying *MUC5B* mutations had neonatal complications, and two were obese with BMI-Z=3.0 (missense mutation) and BMI-Z=2.9 (nonsense mutation) (Supplementary Table 3).

Last, genes encoding proteins important for endosomal sorting, protein ubiquitination and autophagy, pathways recently implicated in neurodevelopmental and neurodegenerative disorders [59] were also analyzed further. This analysis was motivated by the finding that mutations in the *MAGEL2* gene cause ASD in Schaaf-Yang syndrome and contribute to obesity and ASD in Prader-Willi syndrome through disruption of endosomal protein trafficking [8-10, 60]. Genes associated with these pathways and mutated in very obese probands include *VCP* (valosin containing protein, mutations associated with ubiquitinated inclusions and protein aggregates in cells, frontotemporal dementia with myopathy, vesicle transport and fusion), *STAM* (component of the endosomal sorting complex), and *SPAG9/JIP4* (kinesin-mediated transport in neurons).

Discussion

The Simons Simplex collection is a well-characterized cohort of individuals with autism spectrum disorder, for whom a rich variety of clinical data has also been collected. By its nature, individuals with ASD but also a diagnosis of a syndromic genetic disorder, such as Prader-Willi syndrome or Bardet-Biedl syndrome, are underrepresented in or excluded from SSC. As observed in other autism cohorts, probands in the SSC are overweight compared to unaffected siblings. We postulated that probands in the SSC at extreme ends of the phenotypic ranges for various traits could

inform us about as yet unrecognized subtypes of ASD. Indeed, we found that very obese male SSC probands had lower full scale and verbal IQ compared to typical weight probands, and also had significantly worse scores in the areas of communication, daily living skills, and socialization. More severe behaviors were also seen in very obese female probands, but the smaller number of girls compared to boys represented in the SSC limited this analysis. We cannot exclude the possibility that being very obese in and of itself impairs social communication because of exclusion from activities with peers. Conversely, more severe behaviors could contribute to obesity because the child may be excluded from activities that promote fitness. Our study is also compromised by the considerable environmental contribution to the obesity phenotype, particularly in a country that has seen childhood obesity rates increase substantially in the last few decades. Limiting the analysis to very obese probands with two lean parents might have mitigated some of this confounding variable, but would have also reduced the sample size to an unmanageably small number of probands. We did not find that neonatal complications were associated with obesity in this study, although the lack of reporting of neonatal histories limited this analysis.

We proposed that genetic investigation of individuals with dual diagnoses of ASD and obesity could point us towards pathways that are essential for normal neurodevelopment and normal development of energy-regulating pathways. Our focus on *de novo* mutations in very obese probands correctly identified obesity predisposing lesions such as del16p11 and loss of function mutations in *DNMT3A* and *POGZ* and support a role for mutations in *PTK7* in autism [40] and also in obesity. Additional very obese probands carried mutations in genes implicated in ASD or other neurological disorders, such as those important for neuronal action potentials, as predicted from previous autism analyses [61, 62]. Obese probands do not appear to have more pathogenic mutations compared to typical weight probands with ASD. Despite analyzing over 2000 probands, we could not determine whether some probands had mutations in two different genes or environmental factors that independently contributed to ASD and obesity. We also only examined *de novo* mutations, and so do not have information about other genetic mechanisms, such as recessive mutations that may contribute to either the obesity or ASD phenotypes.

The PCP/Wnt pathway has been implicated in neurodevelopmental disorders associated with neonatal complications and obesity, in part through the identification of key genes such as *CTNNB1*, *MIB1*, *RAB23*, *MEGF8*, *PTK7*, and the *BBS* genes in syndromic forms of intellectual disability. *PTK7* encodes protein tyrosine kinase 7 and is implicated in Wnt signaling pathway. *PTK7* (CCK4) binds intracellular scaffolding proteins of the Dishevelled family that act downstream of WNT receptors. *PTK7* has mostly been studied in the context of cancer cell motility and in epithelial to mesenchymal

transitions during development. A recessive *Ptk7* mutation in mice (*chuzhoi*) causes severe developmental defects consistent with dysregulation of PCP, but no phenotype has been described in *chuzhoi* heterozygotes. Loss of *ptk7* function in zebrafish causes ciliary and PCP abnormalities. In humans, we would not expect to find homozygous mutations, considering that this state is embryonic lethal in mice. It would however be interesting to determine whether *chuzhoi* heterozygous mice have behavioral abnormalities recapitulating ASD. The endosomal sorting/ubiquitination pathways have also been implicated in syndromic intellectual disability, with mutations in *MAGEL2*, *VPS13B*, and *VCP* causing syndromic neurological disorders. Our data support the previously noted involvement of both the PCP/Wnt pathway and the endosomal sorting /ubiquitination pathway in ASD. There are too few probands with obesity and ASD and *de novo* mutations to determine whether this pathway is indeed enriched in the very obese subgroup of children with ASD. Further investigation of these two pathways in ASD may be warranted. In particular, genes that are mutated in obese probands with ASD and in many cases also intolerant to mutation (*BTBD9*, *DVL3*, *MED13L*, *RAPGEF2*, *RYK*, *STAM*, and *SPAG9*) may be good candidates for involvement in syndromic forms of ASD.

Conclusions.

Our study provides a model for gene discovery using additional phenotypic information to categorize subsets of ASD probands. For example, probands with dysmorphic features could have a different set of genetic lesions compared to those without dysmorphic features. Extensive phenotypic characterization and collection of a larger cohort of children with ASD, as is now ongoing in the SPARK initiative [63] will enrich our understanding of the multitude of genes associated with the many flavors of autism spectrum disorder. This understanding is essential for the design of customized therapies that target the underlying genetic lesion that contributes to the diagnosis of ASD in each individual.

List of Abbreviations

- Aberrant Behavior Checklist (ABC)
- Autism Diagnostic Interview (ADI)
- Autism spectrum disorder (ASD)
- Body mass index (BMI),
- Child Behavior Checklist (CBCL 6-18)
- De novo copy number variants (dnCNVs)
- De novo likely gene disrupting (dnLGD)

De novo non-synonymous coding mutations (dnNSC)

Exome Aggregation Consortium (ExAC)

Full scale IQ (FSIQ)

Gene Ontology (GO)

Intellectual disability (ID)/

Loss of function (LoF)

Planar cell polarity (PCP)

Prader-Willi syndrome (PWS)

Simons Simplex Collection (SSC)

Vineland Adaptive Behavior Rating Scales (VABS)

Declarations

Ethics approval and consent to participate: Use of SSC data from human subjects was approved by the Health Research Ethics Board-Biomedical Panel of the University of Alberta, Edmonton, Alberta Canada.

Availability of data and material: The data that support the findings of this study are available from the Simons Foundation but restrictions apply to the availability of these data, which were used under license for the current study. Data are however available from the authors upon reasonable request and with permission of the Simons Foundation.

Funding: This work was supported by a grant from the Simons Foundation for Autism Research (SFARIPA304216, to RW).

Authors' contributions: HDC and RW analyzed and interpreted the mutation data and wrote the manuscript. Both authors read and approved the final manuscript.

Acknowledgements: We are grateful to all of the families at the participating Simons Simplex Collection (SSC) sites, as well as the principal investigators (A. Beaudet, R. Bernier, J. Constantino, E. Cook, E. Fombonne, D. Geschwind, R. Goin-Kochel, E. Hanson, D. Grice, A. Klin, D. Ledbetter, C. Lord, C. Martin, D. Martin, R. Maxim, J. Miles, O. Ousley, K. Pelphrey, B. Peterson, J. Piggot, C. Saulnier, M. State, W. Stone, J. Sutcliffe, C. Walsh, Z. Warren, E. Wijsman). We appreciate obtaining access to phenotypic data on SFARI Base. Approved researchers can obtain the SSC population dataset described in this study by applying at <https://base.sfari.org>. We are grateful to the Prader-Willi Association of Alberta for their donation in support of our research on autism spectrum disorders.

Figure Legends.

Figure 1. Distribution of BMI-Z versus age in the SSC.

Figure 2. Behavioral characteristics of very obese probands on the A) Child Behavior Checklist (CBCL), B) autism behavior checklist (ABC), C) Social Responsiveness Scale (SRS), and D) Vineland Adaptive Behavior Scales (VABS). Values that are significantly different from same sex typical weight probands are indicated ($*P \leq 0.01$, $\#P \leq 0.001$). Note that higher scores on the CBCL(6-18), ABC and SRS but lower scores on the VABS are indicative of behavioral deficits. Error bars are not shown for clarity.

Figure 3. Frequency of neonatal findings in SSC probands, divided into groups by BMI-Z.

Supplementary Tables.

Supplementary Table 1. *De novo* genetic events detected in very obese probands (BMI-Z ≥ 3).

Supplementary Table 2. Phenotypes associated with *de novo* mutations in *PTK7*, *MEGF8* and *DVL3*.

Supplementary Table 3. Phenotypes associated with *de novo* *MUC5B* mutations.

References

1. Miles JH: Autism spectrum disorders--a genetics review. *Genet Med* 2011, 13(4):278-294.
2. Stessman HA, Willemsen MH, Fenckova M, Penn O, Hoischen A, Xiong B, Wang T, Hoekzema K, Vives L, Vogel I *et al*: Disruption of POGZ Is Associated with Intellectual Disability and Autism Spectrum Disorders. *Am J Hum Genet* 2016, 98(3):541-552.
3. Casanova EL, Sharp JL, Chakraborty H, Sumi NS, Casanova MF: Genes with high penetrance for syndromic and non-syndromic autism typically function within the nucleus and regulate gene expression. *Molecular autism* 2016, 7:18.
4. Channell MM, Phillips BA, Loveall SJ, Conners FA, Bussanich PM, Klinger LG: Patterns of autism spectrum symptomatology in individuals with Down syndrome without comorbid autism spectrum disorder. *Journal of neurodevelopmental disorders* 2015, 7(1):5.
5. Maillard AM, Hippolyte L, Rodriguez-Herreros B, Chawner SJ, Dremmel D, Aguera Z, Fagundo AB, Pain A, Martin-Brevet S, Hilbert A *et al*: 16p11.2 Locus modulates response to satiety before the onset of obesity. *Int J Obes (Lond)* 2016, 40(5):870-876.
6. Kerr EN, Bhan A, Heon E: Exploration of the cognitive, adaptive and behavioral functioning of patients affected with Bardet-Biedl syndrome. *Clin Genet* 2015.
7. Dimitropoulos A, Ho A, Feldman B: Social responsiveness and competence in Prader-Willi syndrome: direct comparison to autism spectrum disorder. *Journal of autism and developmental disorders* 2013, 43(1):103-113.
8. Fountain MD, Aten E, Cho MT, Juusola J, Walkiewicz MA, Ray JW, Xia F, Yang Y, Graham BH, Bacino CA *et al*: The phenotypic spectrum of Schaaf-Yang syndrome: 18 new affected individuals from 14 families. *Genet Med* 2017, 19(1):45-52.
9. Soden SE, Saunders CJ, Willig LK, Farrow EG, Smith LD, Petrikin JE, LePichon JB, Miller NA, Thiffault I, Dinwiddie DL *et al*: Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Science translational medicine* 2014, 6(265):265ra168.
10. Schaaf CP, Gonzalez-Garay ML, Xia F, Potocki L, Gripp KW, Zhang B, Peters BA, McElwain MA, Drmanac R, Beaudet AL *et al*: Truncating mutations of MAGEL2 cause Prader-Willi phenotypes and autism. *Nat Genet* 2013, 45(11):1405-1408.
11. Dreyer Gillette ML, Borner KB, Nadler CB, Poppert KM, Odar Stough C, Swinburne Romine R, Davis AM: Prevalence and Health Correlates of Overweight and Obesity in Children with Autism Spectrum Disorder. *Journal of developmental and behavioral pediatrics : JDBP* 2015, 36(7):489-496.
12. Hill AP, Zuckerman KE, Fombonne E: Obesity and Autism. *Pediatrics* 2015, 136(6):1051-1061.
13. Jones KB, Cottle K, Bakian A, Farley M, Bilder D, Coon H, McMahon WM: A description of medical conditions in adults with autism spectrum disorder: A follow-up of the 1980s Utah/UCLA Autism Epidemiologic Study. *Autism : the international journal of research and practice* 2016, 20(5):551-561.
14. Broder-Fingert S, Brazauskas K, Lindgren K, Iannuzzi D, Van Cleave J: Prevalence of overweight and obesity in a large clinical sample of children with autism. *Academic pediatrics* 2014, 14(4):408-414.
15. D'Angelo CS, Kohl I, Varela MC, de Castro CI, Kim CA, Bertola DR, Lourenco CM, Perez AB, Koiffmann CP: Obesity with associated developmental delay and/or learning disability in patients exhibiting additional features: report of novel pathogenic copy number variants. *Am J Med Genet A* 2013, 161A(3):479-486.

16. Ramachandrapa S, Raimondo A, Cali AM, Keogh JM, Henning E, Saeed S, Thompson A, Garg S, Bochukova EG, Brage S *et al*: Rare variants in single-minded 1 (SIM1) are associated with severe obesity. *J Clin Invest* 2013, 123(7):3042-3050.
17. Farooqi IS: Defining the neural basis of appetite and obesity: from genes to behaviour. *Clinical medicine* 2014, 14(3):286-289.
18. Choquet H, Meyre D: Genetics of Obesity: What have we Learned? *Current genomics* 2011, 12(3):169-179.
19. Schmunk G, Gargus JJ: Channelopathy pathogenesis in autism spectrum disorders. *Front Genet* 2013, 4:222.
20. De Ferrari GV, Moon RT: The ups and downs of Wnt signaling in prevalent neurological disorders. *Oncogene* 2006, 25(57):7545-7553.
21. Hormozdiari F, Penn O, Borenstein E, Eichler EE: The discovery of integrated gene networks for autism and related disorders. *Genome research* 2015, 25(1):142-154.
22. Caracci MO, Avila ME, De Ferrari GV: Synaptic Wnt/GSK3beta Signaling Hub in Autism. *Neural plasticity* 2016, 2016:9603751.
23. Helfer G, Tups A: Hypothalamic Wnt Signalling and its Role in Energy Balance Regulation. *Journal of neuroendocrinology* 2016, 28(3):12368.
24. Ye T, Ip JP, Fu AK, Ip NY: Cdk5-mediated phosphorylation of RapGEF2 controls neuronal migration in the developing cerebral cortex. *Nature communications* 2014, 5:4826.
25. Krumm N, O'Roak BJ, Shendure J, Eichler EE: A de novo convergence of autism genetics and molecular neuroscience. *Trends Neurosci* 2014, 37(2):95-105.
26. Ross AJ, May-Simera H, Eichers ER, Kai M, Hill J, Jagger DJ, Leitch CC, Chapple JP, Munro PM, Fisher S *et al*: Disruption of Bardet-Biedl syndrome ciliary proteins perturbs planar cell polarity in vertebrates. *Nat Genet* 2005, 37(10):1135-1140.
27. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL *et al*: De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* 2012, 485(7397):237-241.
28. Iossifov I, Ronemus M, Levy D, Wang Z, Hakker I, Rosenbaum J, Yamrom B, Lee YH, Narzisi G, Leotta A *et al*: De novo gene disruptions in children on the autistic spectrum. *Neuron* 2012, 74(2):285-299.
29. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A, Volfovsky N, Packer A, Lash A, Troyanskaya OG: Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nature neuroscience* 2016, 19(11):1454-1462.
30. Fischbach GD, Lord C: The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron* 2010, 68(2):192-195.
31. Simons Simplex Collection [<https://sfari.org/>]
32. GPF: Genotype and Phenotype in Families
33. Lek M, Consortium EA, al. e: Analysis of protein-coding genetic variation in 60,706 humans. *bioRxiv Preprint server* 2015, doi: <http://dx.doi.org/10.1101/030338>.
34. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD: PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. *Nucleic Acids Res* 2017, 45(D1):D183-D189.
35. Kerekes N, Tajnia A, Lichtenstein P, Lundstrom S, Anckarsater H, Nilsson T, Rastam M: Neurodevelopmental problems and extremes in BMI. *PeerJ* 2015, 3:e1024.
36. Goldman S, Wang C, Salgado MW, Greene PE, Kim M, Rapin I: Motor stereotypies in children with autism and other developmental disorders. *Dev Med Child Neurol* 2009, 51(1):30-38.
37. Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A *et al*: Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 2012, 488(7412):471-475.

38. DiStefano C, Gulsrud A, Huberty S, Kasari C, Cook E, Reiter LT, Thibert R, Jeste SS: Identification of a distinct developmental and behavioral profile in children with Dup15q syndrome. *Journal of neurodevelopmental disorders* 2016, 8:19.
39. Mervis CB, Klein-Tasman BP, Huffman MJ, Velleman SL, Pitts CH, Henderson DR, Woodruff-Borden J, Morris CA, Osborne LR: Children with 7q11.23 duplication syndrome: psychological characteristics. *Am J Med Genet A* 2015, 167(7):1436-1450.
40. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, Murtha MT, Bal VH, Bishop SL, Dong S *et al*: Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron* 2015, 87(6):1215-1233.
41. Hippolyte L, Maillard AM, Rodriguez-Herreros B, Pain A, Martin-Brevet S, Ferrari C, Conus P, Mace A, Hadjikhani N, Metspalu A *et al*: The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. *Biological psychiatry* 2016, 80(2):129-139.
42. Walters RG, Jacquemont S, Valsesia A, de Smith AJ, Martinet D, Andersson J, Falchi M, Chen F, Andrieux J, Lobbens S *et al*: A new highly penetrant form of obesity due to deletions on chromosome 16p11.2. *Nature* 2010, 463(7281):671-675.
43. Guo H, Peng Y, Hu Z, Li Y, Xun G, Ou J, Sun L, Xiong Z, Liu Y, Wang T *et al*: Genome-wide copy number variation analysis in a Chinese autism spectrum disorder cohort. *Sci Rep* 2017, 7:44155.
44. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Cicek AE, Kou Y, Liu L, Fromer M, Walker S *et al*: Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* 2014, 515(7526):209-215.
45. Hamdan FF, Srouf M, Capo-Chichi JM, Daoud H, Nassif C, Patry L, Massicotte C, Ambalavanan A, Spiegelman D, Diallo O *et al*: De novo mutations in moderate or severe intellectual disability. *PLoS Genet* 2014, 10(10):e1004772.
46. Dong S, Walker MF, Carriero NJ, DiCola M, Willsey AJ, Ye AY, Waqar Z, Gonzalez LE, Overton JD, Frahm S *et al*: De novo insertions and deletions of predominantly paternal origin are associated with autism spectrum disorder. *Cell reports* 2014, 9(1):16-23.
47. Iglesias A, Anyane-Yeboah K, Wynn J, Wilson A, Truitt Cho M, Guzman E, Sisson R, Egan C, Chung WK: The usefulness of whole-exome sequencing in routine clinical practice. *Genet Med* 2014, 16(12):922-931.
48. Wehl CC, Pestronk A, Kimonis VE: Valosin-containing protein disease: inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia. *Neuromuscular disorders : NMD* 2009, 19(5):308-315.
49. Cornelis MC, Rimm EB, Curhan GC, Kraft P, Hunter DJ, Hu FB, van Dam RM: Obesity susceptibility loci and uncontrolled eating, emotional eating and cognitive restraint behaviors in men and women. *Obesity* 2014, 22(5):E135-141.
50. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT *et al*: Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000, 25(1):25-29.
51. Martinez S, Scerbo P, Giordano M, Daulat AM, Lhoumeau AC, Thome V, Kodjabachian L, Borg JP: The PTK7 and ROR2 Protein Receptors Interact in the Vertebrate WNT/Planar Cell Polarity (PCP) Pathway. *J Biol Chem* 2015, 290(51):30562-30572.
52. Bunn KJ, Daniel P, Rosken HS, O'Neill AC, Cameron-Christie SR, Morgan T, Brunner HG, Lai A, Kunst HP, Markie DM *et al*: Mutations in DVL1 cause an osteosclerotic form of Robinow syndrome. *Am J Hum Genet* 2015, 96(4):623-630.
53. Stittrich AB, Lehman A, Bodian DL, Ashworth J, Zong Z, Li H, Lam P, Khromykh A, Iyer RK, Vockley JG *et al*: Mutations in NOTCH1 cause Adams-Oliver syndrome. *Am J Hum Genet* 2014, 95(3):275-284.

54. Kashevarova AA, Nazarenko LP, Schultz-Pedersen S, Skryabin NA, Salyukova OA, Chechetkina NN, Tolmacheva EN, Rudko AA, Magini P, Graziano C *et al*: Single gene microdeletions and microduplication of 3p26.3 in three unrelated families: CNTN6 as a new candidate gene for intellectual disability. *Mol Cytogenet* 2014, 7(1):97.
55. Utami KH, Winata CL, Hillmer AM, Aksoy I, Long HT, Liany H, Chew EG, Mathavan S, Tay SK, Korzh V *et al*: Impaired development of neural-crest cell-derived organs and intellectual disability caused by MED13L haploinsufficiency. *Human mutation* 2014, 35(11):1311-1320.
56. Song WJ, Song EA, Jung MS, Choi SH, Baik HH, Jin BK, Kim JH, Chung SH: Phosphorylation and inactivation of glycogen synthase kinase 3beta (GSK3beta) by dual-specificity tyrosine phosphorylation-regulated kinase 1A (Dyrk1A). *J Biol Chem* 2015, 290(4):2321-2333.
57. Chen G, Zhang Z, Adebamowo SN, Liu G, Adeyemo A, Zhou Y, Doumatey AP, Wang C, Zhou J, Yan W *et al*: Common and rare exonic MUC5B variants associated with type 2 diabetes in Han Chinese. *PLoS One* 2017, 12(3):e0173784.
58. Hoischen A, Krumm N, Eichler EE: Prioritization of neurodevelopmental disease genes by discovery of new mutations. *Nature neuroscience* 2014, 17(6):764-772.
59. Tang G, Gudsnuk K, Kuo SH, Cotrina ML, Rosoklija G, Sosunov A, Sonders MS, Kanter E, Castagna C, Yamamoto A *et al*: Loss of mTOR-dependent macroautophagy causes autistic-like synaptic pruning deficits. *Neuron* 2014, 83(5):1131-1143.
60. Hao YH, Doyle JM, Ramanathan S, Gomez TS, Jia D, Xu M, Chen ZJ, Billadeau DD, Rosen MK, Potts PR: Regulation of WASH-dependent actin polymerization and protein trafficking by ubiquitination. *Cell* 2013, 152(5):1051-1064.
61. Glessner JT, Wang K, Cai G, Korvatska O, Kim CE, Wood S, Zhang H, Estes A, Brune CW, Bradfield JP *et al*: Autism genome-wide copy number variation reveals ubiquitin and neuronal genes. *Nature* 2009, 459(7246):569-573.
62. Chahrour MH, Yu TW, Lim ET, Ataman B, Coulter ME, Hill RS, Stevens CR, Schubert CR, Collaboration AAS, Greenberg ME *et al*: Whole-exome sequencing and homozygosity analysis implicate depolarization-regulated neuronal genes in autism. *PLoS Genet* 2012, 8(4):e1002635.
63. Simons Foundation Powering Autism Research for Knowledge [<https://sparkforautism.org/>]

Fig. 1 Cortes

Fig. 2 Cortes

Fig. 3 Cortes

family id	location	variant	in child	worst request	genes	proband verbal iq	proband non-verbal iq
14031	2:38956697-1	CNV+	prbM	CNV+	GALM	29	53
14114	2:38956697-1	CNV+	prbM	CNV+	GALM	20	34
13367	4:45410-354	CNV+	prbF	CNV+	MIR943;FGFRL1;SLBP;ZNF732;LOC402160;ADD1;LOC100130872;NELFA;LETM1;LOC	46	57
13000	5:814843-82	CNV-	prbM	CNV-	ZDHHC11	48	29
13866	5:98767783-1	CNV-	prbM	CNV-		67	92
12594	7:72722981-1	CNV+	prbF	CNV+	STX1A;WBSCR27;WBSCR22;LAT2;LIMK1;WBSCR28;MIR4284;ABHD11;FKBP6;MIR59	101	100
12134	7:73978801-1	CNV+	prbM	CNV+	NCF1;PMS2P5;WBSCR16;STAG3L2;GTF2IRD2;GTF2IRD1;GTF2I	121	127
13876	8:35859685-1	CNV+	prbM	CNV+	IDO2;TM2D2;IDO1;STAR;BAG4;ASH2L;SMIM19;AP3M2;LETM2;IKBKB;HTRA4;WHSC	35	49
13646	8:89252843-1	CNV-	prbM	CNV-	MMP16	101	78
11718	10:12075111	CNV+	prbM	CNV+	FAM45A;FAM45B;SNORA19;NANOS1;EIF3A;GRK5;PRDX3;SFXN4	108	93
12437	11:55370916	CNV-	prbM	CNV-	OR4P4;OR4C11;OR4S2	72	78
12976	11:55370916	CNV-	prbM	CNV-	OR4P4;OR4C11;OR4S2	82	85
12343	13:41104610	CNV+	prbF	CNV+	KBTBD7;KBTBD6;MTRF1;MRPS31;NAA16;MIR621;FOXO1;MIR320D1;SUGT1P3;TPT	18	42
13646	14:22538628	CNV-	prbM	CNV-		101	78
11501	14:22591454	CNV-	prbM	CNV-		71	79
13355	15:26971834	CNV+	prbM	CNV+	GABRA5;GABRG3;GABRB3	17	43
11540	16:29647342	CNV-	prbM	CNV-	MVP;ASPHD1;CORO1A;TBX6;SLX1B-SULT1A4;LOC388242;PRRT2;CDIPT;QPRT;YPEL	55	67

family id	location	variant	in child	worst request	genes	all effects	effect details	SSCfreq	EVSfreq	E65freq	proband verbal iq	proband non-verbal iq	SFARI gene	neurodevelopmental	tolerance to LoF	tolerance to missense
13768	4:114277600	del(1)	prbM	frame-shift	ANK2	ANK2:fr	2609/3957 3	nan	nan	nan	100		94	yes	intolerant	
14114	11:428174	ins(T)	prbM	frame-shift	ANO9	ANO9:fr	416/782	nan	nan	nan	20		34			
12610	17:41197776	sub(C->T)	prbM	nonsense	BRCA1	BRCA1:nonse	1858/1884(T	nan	nan	nan	71		99			
11256	11:68029417	del(10)	prbM	frame-shift	C11orf24	C11orf24:frai	346/449	nan	nan	nan	98		127			
13809	5:137534409	del(2)	prbM	frame-shift	CDC23	CDC23:fr	239/597	nan	nan	nan	68		58		intolerant	intolerant
12311	2:25505357	ins(T)	prbM	frame-shift	DNMT3A	DNMT3A:frar	134/166;134	nan	nan	nan	89		71	yes		intolerant
13582	11:11730143	sub(A->G)	prbM	splice-site	DSCAML1	DSCAML1:spl	1956/2114	nan	nan	nan	79		88	yes	intolerant	intolerant
13590	17:28400679	del(1)	prbM	frame-shift	EFCAB5	EFCAB5:fram	848/856 14/	0.01	nan	nan	100		80			
14287	19:48255803	del(4)	prbM	frame-shift	GLTSCR2	GLTSCR2:frar	235/478	nan	nan	nan	36		43			
13000	17:8424205	del(1)	prbM	frame-shift	MYH10	MYH10:fram	733/1985;72	nan	nan	nan	48		29	yes	intolerant	intolerant
13627	1:151378489	sub(G->A)	prbM	nonsense	POGZ	POGZ:nonser	955/1357(Ar	nan	nan	nan	91		76	yes	intolerant	intolerant
12340	17:58740376	sub(G->A)	prbF	nonsense	PPM1D	PPM1D:non	427/605(Trp	nan	nan	nan	19		29	yes		intolerant
11842	8:18730165	del(1)	prbM	frame-shift	PSD3	PSD3:fr	70/1047	nan	nan	nan	15		38	yes		
11256	6:43109471	sub(C->T)	prbM	nonsense	PTK7	PTK7:nonse	562/1070(Ar	nan	nan	0	98		127	yes	intolerant	
14165	17:49072430	del(5)	prbM	frame-shift	SPAG9	SPAG9:fr	727/1311;73	nan	nan	nan	108		85		intolerant	
11160	1:16255763	ins(A)	prbF	frame-shift	SPEN	SPEN:fr	1010/3664	nan	nan	nan	99		119	yes	intolerant	
14147	10:17738825	del(1)	prbM	frame-shift	STAM	STAM:fr	260/540 37C	nan	nan	nan	96		93		intolerant	
12340	1:12175787	sub(G->A)	prbF	splice-site	TNFRSF8	TNFRSF8:spli	205/484;316	nan	nan	nan	19		29			
13054	19:57641997	del(1)	prbM	frame-shift	USP29	USP29:fr	652/922	nan	nan	nan	29		51			
13646	9:35060456	del(5)	prbM	frame-shift	VCP	VCP:fr	515/806	nan	nan	nan	101		78	yes	intolerant	intolerant

family id	location	variant	in child	worst req	genes	all effects	effect details	alternative	al number	of al	SSCfreq	EVSfreq	E65freq	proband	vert	proband	non	SFARI	gene	neurodevelopmental	tolerance to LoF	tolerance to missense	
14242	1:179077107	sub(C->T)	prbM	missense	ABL2	ABL2:missen:1078/1161(Ala->Thr);1099/1182(Ala->T	nan	nan	nan	nan	nan	nan	nan	84	99								
13834	4:74318320	sub(C->T)	prbM	missense	AFP	AFP:missense544/609(Ala->Val)	nan	nan	nan	nan	nan	0.02	0.01	63	85								
11540	X:49955614	sub(C->T)	prbM	missense	AKAP4	AKAP4:missen 843/845(Ala->Thr);852/854(Ala->Thr)	0.153	nan	nan	nan	nan	0.45	0.13	55	67						intolerant		
11540	10:6183821	sub(A->G)	prbM	missense	ANK3	ANK3:missen 2271/4377(Met->Thr) 13/19[6060];36/nan	nan	nan	nan	nan	nan	nan	0	55	67	yes					intolerant		
13834	11:26484618	sub(G->A)	prbM	missense	ANO3	ANO3:missen 119/981(Ala->Thr)	nan	nan	nan	nan	nan	nan	0	63	85				yes				
13834	16:28509202	sub(G->A)	prbM	missense	APOBR	APOBR:missen 947/1097(Glu->Gly)	nan	nan	nan	nan	nan	nan	nan	63	85								
12340	X:129149138	sub(G->A)	prbF	missense	BCORL1	BCORL1:miss 797/1711(Cys->Tyr)	nan	nan	nan	nan	nan	nan	nan	19	29						intolerant		
11263	9:136899892	sub(T->C)	prbM	missense	BRD3	BRD3:missen 666/726(Lys->Glu)	0.01	nan	nan	nan	nan	0.02	0.01	92	100							intolerant	
12661	6:36168202	sub(G->A)	prbF	missense	BRPF3	BRPF3:missen 35/1205(Ala->Thr)	nan	nan	nan	nan	nan	nan	nan	33	59						intolerant		
14242	19:5678611	sub(C->T)	prbM	missense	C19orf70	C19orf70:miss103/118(Arg->His)	0.01	nan	nan	nan	nan	nan	nan	84	99								
14287	9:100672811	sub(G->A)	prbM	missense	C9orf156	C9orf156:miss166/441(Pro->Leu)	nan	nan	nan	nan	nan	0.01	0	36	43				yes				
13809	2:241534627	sub(G->A)	prbM	missense	CAPN10	CAPN10:miss395/672(Arg->Gln);395/517(Arg->Gln)	0.01	nan	nan	nan	nan	nan	0	68	58								
11094	3:123663756	sub(A->G)	prbM	missense	CDC14	CDC14:miss 435/912(Ile->Thr)	nan	nan	nan	nan	nan	nan	nan	86	105								
12742	17:80086416	sub(G->A)	prbM	missense	CDC57	CDC57:miss 767/915(Arg->Cys)	0.01	nan	nan	nan	nan	nan	0	106	103								
11540	X:50051731	sub(T->A)	prbM	missense	CENB3	CENB3:missen 188/1395(Ser->Thr) 2/6[19844]	0.109	nan	nan	nan	nan	0.44	0.12	55	67								
11055	1:179983048	sub(G->A)	prbM	missense	CEP350	CEP350:miss487/3117(Asn->Thr)	nan	nan	nan	nan	nan	nan	nan	19	32						intolerant		
12385	3:1371543	sub(G->A)	prbM	missense	CNTN6	CNTN6:missen 430/1028(Gly->Arg)	nan	nan	nan	nan	nan	0.01	0	58	55								
13535	12:53553913	sub(A->A)	prbM	missense	CSAD	CSAD:missen 384/493(Ala->Val);151/260(Ala->Val);4/nan	nan	nan	nan	nan	nan	nan	nan	12	29								
14165	8:3165969	sub(G->A)	prbM	missense	CSMD1	CSMD1:missen 1230/3564(Arg->Cys)	nan	nan	nan	nan	nan	nan	nan	108	85	yes							
14183	1:155706848	sub(G->A)	prbM	missense	DAP3	DAP3:missen 369/398(Glu->Lys);369/398(Glu->Lys);3/nan	nan	nan	nan	nan	nan	nan	nan	92	104								
11055	1:112308784	sub(T->G)	prbM	missense	DDX20	DDX20:missen 580/824(Ser->Ala)	nan	nan	nan	nan	nan	nan	0	19	32								
12343	1:231829616	sub(C->T)	prbF	missense	DISC1	DISC1:missen 38/832(Arg->Trp);38/854(Arg->Trp);38/nan	nan	nan	nan	nan	nan	0.04	0.01	18	42	yes							
12667	7:108213622	sub(T->G)	prbM	missense	DNAJB9	DNAJB9:miss 166/223(Phe->Cys)	nan	nan	nan	nan	nan	nan	nan	19	49								
11726	11:6580460	sub(A->G)	prbM	missense	DNHD1	DNHD1:miss3156/4753(Thr->Ala)	nan	nan	nan	nan	nan	nan	nan	90	96								
11540	X:47497511	sub(C->T)	prbM	missense	ELK1	ELK1:missen 242/428(Arg->Gln);242/428(Arg->Gln) 1/nan	nan	nan	nan	nan	nan	0.02	0	55	67								
11842	X:14599441	sub(A->G)	prbM	missense	GLRA2	GLRA2:missen 136/452(Asn->Ser);136/452(Asn->Ser);1/nan	nan	nan	nan	nan	nan	nan	nan	15	38	yes					intolerant	intolerant	
13000	X:101912889	sub(A->G)	prbM	missense	GRASP1	GRASP1:miss1350/1395(Lys->Glu);1350/1395(Lys->G	nan	nan	nan	nan	nan	nan	nan	48	29								
11470	3:7620458	sub(G->A)	prbM	missense	GRM7	GRM7:missen 622/922(Arg->Glu);622/915(Arg->Gln)	nan	nan	nan	nan	nan	nan	0	76	95	yes					intolerant		
13646	16:27540006	sub(A->G)	prbM	missense	GTFC31	GTFC31:missen 296/2109(Tyr->His)	nan	nan	nan	nan	nan	nan	nan	101	78							intolerant	
12343	5:175816383	sub(G->A)	prbF	missense	HIGD2A	HIGD2A:miss 69/106(Arg->Gln)	nan	nan	nan	nan	nan	nan	nan	18	42							intolerant	
14554	11:62490296	sub(C->T)	prbM	missense	HNRNPUL2	HNRNPUL2:n324/747(Arg->His) 4014	0.01	nan	nan	nan	nan	0	70	87							intolerant		
12029	19:14161680	sub(C->T)	prbM	missense	IL27RA	IL27RA:missen 505/636(Arg->Trp)	nan	nan	nan	nan	nan	nan	nan	0	94	108							
11584	11:66999299	sub(G->A)	prbM	missense	KDM2A	KDM2A:missa449/1162(Arg->Lys) 6925	nan	nan	nan	nan	nan	nan	nan	73	75						intolerant	intolerant	
12661	9:5763642	sub(A->C)	prbF	missense	KIAA1432	KIAA1432:mi 835/1386(Glu->Ala);872/1423(Glu->Ala);nan	nan	nan	nan	nan	nan	0.01	0.01	33	59								
11470	16:57805151	sub(C->T)	prbM	missense	KIFC3	KIFC3:missen 242/833(Glu->Lys);242/826(Glu->Lys);1/nan	nan	nan	nan	nan	nan	nan	0	76	95								
12742	7:151845714	sub(G->A)	prbM	missense	KMT2C	KMT2C:missen 4433/4911(Ala->Val)	nan	nan	nan	nan	nan	nan	nan	106	103	yes					intolerant		
13315	11:1606038	sub(A->G)	prbM	missense	KRTAP5-1	KRTAP5-1:mi 148/278(Ser->Pro) 1/5[42406] 1/1[117]nan	nan	nan	nan	nan	nan	nan	nan	123	113								
11765	5:112824034	ins(GCTGCCGprbM	no-frame-shi	MCC	MCC:no-fran 26/1019	nan	nan	nan	nan	nan	nan	nan	nan	24	39	yes							
14088	11:12225938	sub(G->A)	prbM	missense	MICAL2	MICAL2:miss 136/112(Asp->Asn)	nan	nan	nan	nan	nan	nan	nan	107	88								
13071	16:58071464	sub(C->G)	prbM	missense	MMP15	MMP15:missen 84/669(Ala->Gly)	nan	nan	nan	nan	nan	nan	nan	108	124								
13590	15:80137554	ins(A)	prbM	noEnd	MTNFS5,ST20	MTNFS5: noEn 147,204,180 2365	nan	nan	nan	nan	nan	nan	nan	100	80								
13196	3:195510958	sub(G->C)	prbM	missense	MUC4	MUC4:missen 2498/5412(Pro->Arg) 1/22[9782];1/23[nan	nan	nan	nan	nan	nan	nan	nan	95	91								
14183	11:1263122	sub(G->A)	prbM	missense	MUC5B	MUC5B:missen 1671/5762(Thr->Met)	nan	nan	nan	nan	nan	nan	nan	92	104						intolerant	intolerant	
11470	6:135507151	sub(G->A)	prbM	missense	MYB	MYB:missen 45/758(Arg->Gln);45/555(Arg->Gln);45/nan	nan	nan	nan	nan	nan	nan	nan	76	95								
1367	17:7318418	sub(C->T)	prbF	missense	NLGN2	NLGN2:missen 330/835(Arg->Cys)	nan	nan	nan	nan	nan	nan	nan	46	57	yes					intolerant	intolerant	
11597	19:49167998	sub(G->A)	prbM	missense	NTNS	NTNS:missen 220/489(Arg->Cys) 2/4[1474]	nan	nan	nan	nan	nan	nan	nan	128	118								
13939	1:248569659	sub(A->G)	prbM	missense	OR2T1	OR2T1:missen 122/369(Met->Val)	nan	nan	nan	nan	nan	nan	nan	92	93								
11726	16:90128549	sub(G->A)	prbM	missense	PRDM7	PRDM7:missen 221/492(Pro->Leu);15/171(Pro->Leu)	nan	nan	nan	nan	nan	nan	nan	90	96								
12266	8:48827944	sub(G->A)	prbM	missense	PRKDC	PRKDC:missen 854/4128(Arg->Ile);854/4097(Arg->Ile)	nan	nan	nan	nan	nan	nan	0	54	65	yes					intolerant		
13907	7:18067224	sub(C->T)	prbM	missense	PRPS1L1	PRPS1L1:miss61/318(Gly->Asp)	nan	nan	nan	nan	nan	nan	nan	62	71								
14180	5:40692019	sub(G->T)	prbM	missense	PTGERA	PTGERA:missen 336/488(Val->Leu)	nan	nan	nan	nan	nan	nan	nan	134	112						intolerant		
14242	1:117484681	sub(T->C)	prbM	missense	PTGFRN	PTGFRN:missen 132/879(Tyr->His)	nan	nan	nan	nan	nan	nan	nan	84	99								
11726	6:43109472	sub(G->A)	prbM	missense	PTK7	PTK7:missen 562/1070(Arg->Gln);522/1030(Arg->Gln	nan	nan	nan	nan	nan	nan	nan	90	96	yes					intolerant		
13808	6:43099768	sub(G->A)	prbM	missense	PTK7	PTK7:missen 276/1070(Arg->His);276/1030(Arg->His)	0.01	nan	nan	nan	nan	nan	nan	19	48	yes					intolerant		
12437	12:13135912	sub(G->T)	prbM	missense	RAN	RAN:missen 93/216(Thr->Ile)	nan	nan	nan	nan	nan	nan	nan	72	78						intolerant		
11263	1:84945109	sub(C->A)	prbM	missense	RPF1	RPF1:missen 49/349(Pro->Thr)	nan	nan	nan	nan	nan	nan	nan	92	100								
13464	19:39019316	sub(G->A)	prbM	missense	RYR1	RYR1:missen 3667/5033(Arg->His);3672/5038(Arg->t	nan	nan	nan	nan	nan	nan	0	103	105				yes			intolerant	
12306	3:38892193	sub(T->G)	prbM	missense	SCN11A	SCN11A:missen 1369/1791(Asp->Ala)	nan	nan	nan	nan	nan	nan	nan	84	57				yes				
12340	2:166848006	sub(T->C)	prbF	missense	SCN1A	SCN1A:missen 1927/2009(Arg->Gly);1916/1998(Arg->C	nan	nan	nan	nan	nan	nan	nan	19	29	yes					intolerant	intolerant	
12271	2:166011075	sub(C->T)	prbM	missense	SCN3A	SCN3A:missen 423/2000(Val->Met);423/1951(Val->M	nan	nan	nan	nan	nan	nan	nan	40	41						intolerant	intolerant	
13646	17:62049125	sub(G->A)	prbM	missense	SCN4A	SCN4A:missen 190/1836(Arg->Trp)	0.01	nan	nan	nan	nan	0	101	78	yes								
13000	6:25862544	sub(A->G)	prbM	missense	SLC17A3	SLC17A3:miss 74/420(Ser->Pro);74/498(Ser->Pro)	nan	nan	nan	nan	nan	nan	nan	48	29								
13130	19:47935439	sub(C->G)	prbM	missense	SLC8A2	SLC8A2:miss 792/921(Gly->Arg)	nan	nan	nan	nan	nan	nan	nan	113	102								
14676	12:21069032	sub(A->G)	prbM	missense	SLCO1B3	SLCO1B3:miss 654/702(Lys->Glu)	nan	nan	nan	nan	nan	nan	nan	98	94	yes							
14031	20:10603465	sub(C->A)	prbM	missense	SLX4IP	SLX4IP:missen 222/408(Ala->Glu)	nan	nan	nan	nan	nan	nan	nan	29	53								
14287	1:183502868	sub(C->T)	prbM	missense	SMG7	SMG7:missen 313/1145(Ser->Phe);355/109																	

Supplementary Table 2. Mutations in *PTK7*, *MEGF8* and *DVL3* in the SSC (rows 1-5, proband IDs are masked) or the Database of Genomic Variants (rows 6-8, Kamisnky et al., Genet Med. 2011:777-84.). The mouse mutant *chuzhoi* is listed in the bottom row for comparison.

Proband	mutation	BMI-Z	height-Z	neonatal and other phenotypes	FSIQ
11256.p1 (M)	PTK7 nonsense	3.2	1.8	greater than 8 ear infections vision corrected clumsy	102
11726.p1 (M)	PTK7 R562Q	4.3	.8	respiratory difficulty frequent otitis media normal hearing vision	92
13808.p1 (M)	PTK7 R276H	3.0	.8	neonatal genital abnormality, broken bones neonatal heart problems, abnormal eeg frequent strep throat normal hearing vision	40
11798.p1 (M)	MEGF8 Q1962G	2.0	-0.2	poor suck, lethargic, floppy, feeding difficulties, respiratory difficulties	114
11242.p1 (M)	DVL3 frameshift	2.5	0.24	poor suck, physical anomalies, feeding difficulties, genital abnormalities	94
nsv530893	PTK7 CNVdel (7.8 Mb) de novo			abnormality of cardiac morphology; Abnormality of the skeletal system; Dandy-Walker malformation; seizure	
nsv529573	PTK7 CNVdel 2.2Mb			intellectual disability	
nsv530894	PTK7 CNVdel 4.9 Mb			developmental delay and/or other significant developmental or morphological phenotypes	
mouse	<i>chuzhoi</i> p.L491_Q 492insAN P			heterozygote overtly normal, multiple congenital anomalies in homozygote	

Supplementary Table 3. Phenotypes associated with *de novo* *MUC5B* mutations in the Simons Simplex Collection.

Proband (sex)	Mutation	BMI-Z	height-Z	neonatal phenotype	FSIQ
11078.p1 (M)	nonsense	2.93	.99	stiff infant lethargic/overly sleepy large head circ	29
14183.p1 (M)	T1671M	3.01	3.47	feeding clumsy broken bones >8 otitis	100
12779.p1 (M)	P4997L	1.91	1.15	feeding vision corrected	99
13983.p1 (F)	P3887T	-1.75	.28	feeding	40
11653.p1 (M)	del(15) no frame shift V2574A T3883M	-.43	-.77	stiff infant	36
11247.p1 (M)	P4997L	-0.6	+0.4	kidney or urinary	128
11957.p1 (M)	C4174F	1.6	-2.0	physical anomalies, respiratory	77
12447.p1 (M)	H4050Q	0.4	-0.5	lethargic	53