

# Prevalence and Phenotypic Impact of Rare Likely Pathogenic Variants in Autism Spectrum Disorder

**Behrang Mahjani**

Icahn School of Medicine at Mount Sinai Seaver Autism Center for Research and Treatment <https://orcid.org/0000-0001-6087-9537>

**Silvia De Rubeis**

Icahn School of Medicine at Mount Sinai

**Christina Gustavsson Mahjani**

Icahn School of Medicine at Mount Sinai

**Maureen Mulhern**

Icahn School of Medicine at Mount Sinai

**Xinyi Xu**

Icahn School of Medicine at Mount Sinai

**Lambertus Klei**

University of Pittsburg

**F. Kyle Satterstrom**

Broad Institute

**Jack Fu**

Broad Institute

**Michael E. Talkowski**

Broad Institute

**Abraham Reichenberg**

Icahn School of Medicine at Mount Sinai

**Sven Sandin**

Karolinska Institutet

**Christina M. Hultman**

Karolinska Institutet

**Dorothy E. Grice**

Icahn School of Medicine at Mount Sinai

**Kathryn Roeder**

Carnegie Mellon University

**Bernie Devlin**

University of Pittsburg

**Joseph D. Buxbaum** (✉ [joseph.buxbaum@mssm.edu](mailto:joseph.buxbaum@mssm.edu))

Icahn School of Medicine at Mount Sinai Seaver Autism Center for Research and Treatment <https://orcid.org/0000-0001-8898-8313>

---

**Research**

**Keywords:** autism spectrum disorder, copy number variant, single nucleotide variant, intellectual disability, PAGES

**Posted Date:** November 17th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-104532/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

**Background:** Copy number variants (CNVs) and single nucleotide variants (SNVs) are sources of risk for autism spectrum disorder (ASD). The distribution of such pathogenic variants in individuals with ASD and the characterization of those who carry such variants versus those who do not are understudied at the population level. We describe a population sample from Sweden, evaluating the distribution of likely pathogenic variants and their impact on medical, neurological, and psychiatric phenotypes.

**Methods:** The genotyped sample consisted of 1,236 children born in Sweden with autistic disorder, a severe form of ASD (International Classification of Diseases, Tenth Revision, code F84.0.) Of these individuals, CNVs were called from 997, while SNVs were called from 808.

**Results:** Out of 997 individuals from whom CNVs were called, 104 (10.4 %) carried one or more likely pathogenic CNV, including 15q11q13 (n=8), 15q13.3 (n=5), 16p13.11 (n=5), 16p11.11 (n=5), 22q11.2 (n=5). Of 808 individuals assessed by whole-exome sequencing, 69 (8.5%) had a likely pathogenic SNV, including in *GRIN2B* (n = 6), *POGZ* (n = 5), *SATB1* (n=4), *DYNC1H1* (n=4), and *CREBBP* (n=3). Fourteen individuals carried two likely pathogenic CNVs, and 5 carried a likely pathogenic CNV and SNV. Carriers of likely pathogenic CNVs or SNVs were more likely to have intellectual disability (ID), scholastic skill disorders, and epilepsy, with odds ratios of 2.31 (95%CI, 1.55,3.47), 1.98 (95%CI, 1.19,3.21), and 1.63 (95%CI, 1.08,2.44) respectively. Carriers of likely pathogenic CNVs also showed significant increased rates of congenital anomalies. We compared rates of likely pathogenic CNVs, SNVs, and phenotypes from genotyped AD subjects with and without ID: rates were not significantly different between groups.

**Limitations:** As a case-control cohort, we did not have *de novo* information to aid in classification. More broadly, there were judgment calls involved in identifying likely pathogenic variants. For these reasons, some misclassification is possible. In addition, phenotypes are defined from medical registers, which may lead to underestimates of milder findings.

**Conclusions:** People with ASD who carry likely pathogenic CNVs or SNVs show increased rates of various comorbidities, most prominently ID. Despite the strong association with ID, conditioning on its presence explains little of the variation for other comorbidities and physical traits.

## Background

Autism spectrum disorder (ASD) is a childhood-onset neurological and developmental disorder that affects more than 1% of the population (1). The affected individuals can have lifelong impairments in social interaction, communication, and adaptive functioning. The complexity of the autism spectrum involves different symptoms and severities according to the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition (DSM-IV) (2), ranging from a mild form called Asperger's syndrome, to the severest form called Autistic Disorder (AD). In 2013, in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) (3), AD, Asperger's syndrome and additional pervasive developmental disorder diagnoses were replaced with the umbrella diagnosis of autism spectrum disorder (ASD), with severity specifiers for social communication and for restricted interests and repetitive behaviors.

ASD has a complex genetic architecture with both rare and common variation contributing to risk. While common variation accounts for the majority of genetic liability for autism, rare variation, often *de novo*, accounts for substantial individual liability (5). For example, while only 14% of individuals with ASD in the Simons Simplex Collection dataset carried a *de novo* risk CNV or loss-of-function (LoF) SNV, 80% and 57% of these probands would not have ASD if they did not carry the *de novo* CNV or LoF mutation, respectively (5). As a result, identification of rare genetic variants in individuals with ASD can be clinically beneficial, both for diagnostic and genetic counseling purposes.

Numerous studies have identified *de novo* and inherited CNVs and SNVs associated with ASD (6–14). Microdeletions and microduplications in 1q21.1, 7q11.23, 15q11-q13, 16p11.2, 22q11.2, and 22q13.3 regions are among the most commonly reported CNVs (11). In the most extensive genetic sequencing study to date, researchers have identified 102 genes that, when mutated, are associated with risk for autism (4).

Chromosome microarray (CMA) and whole-exome sequencing (WES) are being increasingly used as part of genetic evaluation for ASD. In a study of 258 probands with ASD, 9.3% received a diagnosis by CMA and 8.4% by WES, for a total diagnostic yield of 15.8% with both. This diagnosis rate varied based on the presence of co-morbid major congenital abnormalities and minor physical anomalies, ranging from as low as 6.3% for the least complex group to 37.5% for the most complex group (15).

Despite these findings, which are either from clinic- or cohort-based studies, the prevalence and characterization of CNVs and SNVs in populations is understudied. The objective of this work is to examine and test for possible associations between phenotypes of individuals diagnosed with ASD with, or without, likely pathogenic genetic findings. By systematically curating rare variants based on pathogenicity, we are able to describe the genetic architecture of rare variants in the population. By incorporating robust and relatively unbiased phenotype data obtained from the Swedish national registers, we compare the phenotypes of those with ASD with a known genetic cause and those with ASD without a known genetic cause. In the companion study (Klei et al), the inter-related role of common variation in ASD risk is explored in the PAGES sample.

## Methods

### Study population

In this study, we used data collected from study participants in PAGES (Population-based Autism Genetics and Environmental Study), a large ongoing population-based cohort study in Sweden that started in 2012 with the overall aim to identify possible genetic and environmental risk factors for ASD (5). The PAGES study is a collaboration between researchers at Karolinska Institutet in Stockholm, Sweden, and the Icahn School of Medicine at Mount Sinai, New York, USA. The study was approved by the Regional Ethical Review Board in Stockholm, Sweden, and the Institutional Review Board (IRB) at the Icahn School of Medicine at Mount Sinai, New York, USA. All individuals with a diagnosis of ASD with ICD-9 and -10 (International Classification of Diseases) were identified in the Swedish National Patient Register (NPR). Our focus here is on Autistic Disorder, defined by ICD-9 codes 299.A/B/X and ICD-10 code F84.0. The eligible individuals were born in Sweden between 1960–1996 and followed up through 2011.

In PAGES, once a potential case was identified in the NPR, and the ASD diagnosis confirmed by the clinic, a team of research nurses informed the family with a letter about the genetic study, followed up with a phone call to schedule an appointment, and then visited the family to collect informed consent, summary phenotypic information through a questionnaire, and the biomaterial (blood, in most cases). Information about sex, age at the time of diagnosis, date of admission and discharge, and diagnostic codes for intellectual functioning and psychiatric comorbidities were extracted from the NPR after the consent form was signed. The date of the first registered ASD diagnosis was used as the diagnosis date. Individuals with a diagnosis of Down syndrome were excluded from this study.

In addition to the NPR, the Multi-generation Register was also accessed, which allowed for the identification of family relations, and the Swedish Medical Birth Register (MBR), which contained birth characteristics of all Swedish-born children since 1973 (including prenatal, perinatal and neonatal variables).

As of March 2019, biological samples from 1726 individuals were genotyped: 1236 individuals with ASD, and 238 controls. DNA was genotyped from 1,154 ASD samples using Infinium Omni Express Exome and 82 ASD samples using Infinium Global Screening Array; furthermore, the DNA from 827 of these samples were whole exome sequenced.

### CNV Calling

CNV calls were generated from 1,154 ASD samples genotyped on the Infinium Omni Express Exome by PennCNV using hg19 genomic coordinates for autosomes. Data and calls were cleaned using standard procedures in PennCNV (BAF drift  $\leq 0.01$ ,  $|WF| \leq 0.05$ , LRR SD  $\leq 0.3$ ). We filtered CNVs with the number of SNPs  $\geq 20$ , and overlap with common regions  $\leq 50\%$  of the CNV's own length, leaving 997 high quality samples from which CNVs were called. All CNV calls that span 1) recurrent CNVs - recurrent loci within specific chromosomal regions that are found to cause genomic disorders (GDs) (for more details, see (4)); or 2) larger than 1 MB and include at least one gene were called likely pathogenic. For most GDs, there is information on whether

the deletion and/or duplication is likely pathogenic (Decipher, ClinVar). If a CNV has such information, we adhere to it; if a given CNV does not include such information, we assume that the deletion and duplication are GD.

Individuals with two or more large CNVs (> 1 MB) due to concerns about the quality of the sample; likewise, individuals with a CNV larger than 45 MB were removed from the analysis.

The GATK Team at Broad Institute has generated CNV calls for PAGES sequenced data using germline copy number variants (gCNVs) (16). We use these calls as a validation step in our analysis.

## SNV Calling

SNVs were called using the Genome Analysis Toolkit (17) HaplotypeCaller package version 3.4 (for more details, see (4)). We reviewed the SNVs that were identified in a known and candidate ASD risk gene based on Satterstrom et al. gene list (4). SNVs in these 102 genes were classified as likely pathogenic if there was either 1) a loss-of-function mutation (LoF) that was rare (defined as absent from the Genome Aggregation Database (gnomAD)), or 2) a missense rare (absent from gnomAD) mutation with an MPC score > 2.

## Exposure Covariates

We extracted information for the following variables from the NPR and the MBR: intellectual disability (ID), attention-deficit/hyperactivity disorder (ADHD), schizophrenia (SZ), obsessive-compulsive disorder (OCD), anxiety disorder, speech and language disorders, scholastic skill disorders, motor function disorders, epilepsy, sleeping disorders, hypotonia, birth defects, prenatal growth rate, gestational age in weeks, weight and height at birth, head circumference, and Apgar scores (Table S1).

Due to the limited number of individuals diagnosed with OCD and anxiety disorder, we grouped these diagnoses together to gain statistical power and ensure valid statistics. OCD was included in the anxiety disorder section in DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, fourth edition). In DSM-5, OCD is in a separate section from anxiety disorders, albeit with minimal changes in diagnostic criteria (2, 3).

The average head circumference of healthy newborns is 35 cm (13 ¾ inches). While the range depends on the height of the child, among other attributes (18), we used head circumference without adjustment, defining "HC-small" if the circumference was smaller than 32 cm and "HC-big" if it was larger than 38.

## Statistical analysis

To identify comorbidities and birth characteristics associated with ASD probands who carry damaging mutations, we used a logit model in which carrier status of the damaging variant type was the outcome (carrier or not) and predictors were sex, used as a covariate, and potential comorbidity or characteristic. Thus, a series of models were fit, one for each potentially associated feature. We reported the resulting odds ratio (OR), P values, and 95% confidence intervals (CIs) for the OR after adjusting for sex.

## Results

### Demographic Data:

For quality control, we removed twelve individuals with a diagnosis of Down syndrome, five individuals with more than two large CNVs, and one individual with a CNV larger than 45 MB. After quality control, genotype data was available for 997 ASD individuals and WES data for 808 ASD individuals (Table 1). Of these probands, 70% were male, and males tended to have a lower age of ASD diagnosis (Table 1).

Table 1  
Genetic characterization of probands with ASD

	Genotyped Probands	Sequenced Probands	Intersection
<b>Total (n)</b>	997	808	689
Average diagnosis age (SD)	8.2 (8.5)	7.9 (8.3)	8 (8.5)
<b>Females (n, (%))</b>	299 (30%)	229 (28%)	189 (28%)
Average diagnosis age (SD)	8.9 (9.3)	8.6 (8.9)	9 (9.4)
<b>Males (n, (%))</b>	698 (70%)	579 (72%)	500 (72%)
Average diagnosis age (SD)	7.9 (8.1)	7.6 (8)	7.6 (8.1)

Of the comorbidities and birth characteristics for the population of ASD probands (Table 2), ID was most common (46%), and epilepsy was second (31%). Individuals with congenital anomalies had the lowest age of ASD diagnosis, while individuals with SZ had the highest age of ASD diagnosis. Comorbidities and birth characteristics of the probands which were not genotyped or sequenced are presented in Table S4.

Table 2  
Demographics of Comorbidities and Birth Characteristics

Comorbidities and Birth Characteristics	Genotyped Probands (N = 997)		Sequenced Probands (N = 808)	
	n (%)	Average diagnosis age for AD (SD)	n (%)	Average diagnosis age for AD (SD)
ID	474 (47%)	6.5 (7)	381 (47%)	6.5 (7.1)
Epilepsy	309 (31%)	5.5 (7.2)	225 (31%)	5.5 (7.2)
ADHD	203 (20%)	8.4 (7.4)	160 (20%)	7.7 (7.2)
Scholastic skill disorders	155 (15%)	5.1 (5.6)	121 (15%)	5.2 (6.2)
Speech/language disorders	133 (13%)	4.8 (5.8)	102 (13%)	4.6 (5.1)
OCD + anxiety disorder	100 (10%)	9.2 (7.3)	85 (10%)	9.7 (7.7)
HC – small	87 (11%)	4.6 (5.9)	73 (11%)	4 (4.4)
SZ	86 (9%)	13.9(10.2)	67 (8%)	11.7 (9.8)
Congenital anomalies	74 (11%)	2.9 (4.9)	57 (10%)	3.1 (4.7)
Motor function disorders	55 (5%)	4.8 (5)	43 (5%)	4.6 (4.8)
HC – large	40 (5%)	7.9 (7.3)	30 (5%)	8.4 (7.7)
Large for gestational age	38 (5%)	7.6 (7.7)	30 (4%)	5.2 (5.7)
Small for gestational age	37 (5%)	4.1 (6.2)	35 (5%)	4.3 (6.3)

Missing values for congenital anomalies (Genotyped Probands /Sequenced Probands) 336/260, large for gestational age 207/139, small for gestational age 207/139, head – small 220/153, head – large 220/153 individuals. ADHD: attention-deficit/hyperactivity disorder, SZ: schizophrenia (SZ), OCD: obsessive-compulsive disorder, ID: intellectual disability.

## Genetic findings

Of the 997 individuals who were genotyped, 104 (10.4%) carried one or more likely pathogenic CNVs (Table 3, Table S2). Fourteen probands had two likely pathogenic CNVs each, for a total of 118 likely pathogenic CNVs in the population sample; 65 of these were deletions and 53 were duplications, and they ranged in size from 217 kb to 43 Mb (median 3.6 Mb). There were 53

CNVs considered to be recurrent genomic disorders. Those loci with more than one overlapping likely pathogenic CNV/genomic disorder in different probands are described in Table 3; see Table S2 for full results.

Table 3  
Known genomic disorders with more than one case identified, based on CNV findings

Chromosomal Disorder	Number Identified
15q11q13 PWS BP1-BP3 class 1	8 (6 dup, 2 del)
15q13.3 microdeletion syndrome BP4 - BP5	5 (2 dup, 3 del)
16p13.11 syndrome MYH11	5 (2 dup, 3 del)
22q11.2 DGS typical AD 3 MB	5 (1 dup, 4 del)
16p11.2 typical BP5-BP5 TBX6	4 (1 dup, 3 del)
1q21 recurrent region	3 (dup)
2q37 monosomy	3 (del)
Phelan-McDermid Syndrome (22q13 deletion syndrome)	2 (del)
16p12.2 recurrent region (EEF2K)	2 (del)
17q12 RCAD syndrome (HNF1B)	2 (dup)

Of the 808 individuals for whom WES was performed, 69 (8.8%) had a likely pathogenic variant, and no individuals had more than one (Table S3). Of these SNVs, 34 were predicted loss of function (frameshift, nonsense, splice acceptor or donor), and the remaining 35 were deleterious missense variants. Genes with the highest frequency of suspected likely pathogenic findings were *GRIN2B* (n = 6), *POGZ* (n = 5), *SATB1* (n = 4), *DYNC1H1* (n = 4), and *CREBBP* (n = 3). Two individuals had likely significant variants in each of the following genes: *CACNA1E*, *CHD8*, *DIP2A*, *FOXP1*, *RORB*, *SETD5*, *STXBP1*, *SUV420H1* (also referred to as *KMT5B*), and *SYNGAP1*. One individual had likely significant variants in each of the following genes: *ADNP*, *ANK2*, *ARID1B*, *ASXL3*, *CELF4*, *CHD2*, *DNMT3A*, *DSCAM*, *FOXP2*, *GABRB2*, *GIGYF1*, *HECTD4*, *KCNMA1*, *KDM6B*, *LRR4C*, *MYT1L*, *NACC1*, *NR3C2*, *NUP155*, *PHF21A*, *PPP1R9B*, *PPP2R5D*, *SCN1A*, *SCN2A*, *SHANK3*, *SIN3A*, *TCF4*, *TEK*, and *ZMYND8*. In total, 168 individuals had a likely pathogenic CNV or SNV (5 individuals had both a CNV and SNV).

## Comorbidities and Birth Characteristics of the Probands

Evaluating medical and psychiatric comorbidities among individuals with ASD (Tables 4 and 5), the likely pathogenic CNV and SNV groups showed slightly different average ages of ASD diagnosis by group and by sex, although none of these differences were significant (P value > 0.05). Of the phenotypes of individuals with likely pathogenic CNVs and SNVs (Table 5), ID was the most common disorder.

Table 4  
 Characteristics of individuals with significant CNVs and SNVs

	<b>Genotyped Probands Individuals with likely pathogenic CNV n (%)</b>	<b>Sequenced Probands Individuals with likely pathogenic SNV n (%)</b>	<b>Intersection Individuals with likely pathogenic SNV or CNV n (%)</b>
<b>Total (%)</b>	104 (10.4%)	69 (8.5%)	121 (17.6%)
Average diagnosis age (SD)	7.1 (8.1)	7.3 (7.3)	6.7 (7.4)
<b>Females (n, (%))</b>	37 (35%)	20 (30%)	35 (29%)
Average diagnosis age (SD)	7.8 (9.2)	8.2 (7)	6.2 (6.6)
<b>Males (n, (%))</b>	67 (65%)	49 (70%)	86 (71%)
Average diagnosis age (SD)	6.8 (7.6)	6.9 (7.5)	7 (7.7)

Table 5  
 Comorbidities and birth characteristics of individuals with ASD

<b>Phenotypes</b>	<b>Genotyped Probands Individuals with likely pathogenic CNV (N = 104) n (%)</b>	<b>Sequenced Probands Individuals with likely pathogenic SNV (N = 69) n (%)</b>	<b>Intersection Individuals with likely pathogenic CNV/SNV (N = 121) n (%)</b>
ID	58 (55%)	47 (68%)	75 (50%)
Epilepsy	34 (32%)	30 (43%)	49 (32%)
Scholastic skill disorders	25 (24%)	11 (16%)	33 (11%)
ADHD	21 (20%)	11 (16%)	23 (15%)
Speech/language disorders	18 (17%)	7 (10%)	16 (7%)
Congenital anomalies	18 (17%)	3 (4%)	15 (10%)
HC - small	13 (12%)	4 (5.7%)	14 (9%)
SZ	12 (11%)	4 (6%)	14 (9%)
OCD + Anxiety	10 (10%)	4 (6%)	10 (8%)
Motor function disorders	8 (8%)	2 (3%)	7 (5%)
Small for gestational age	7 (7%)	0 (0%)	5 (3%)
HC - big	4 (4%)	4 (5.7%)	6 (4%)
Large for gestational age	2 (2%)	3 (2%)	1 (1%)

Missing values for congenital anomalies (Genotyped Probands /Sequenced Probands) 336/260, large for gestational age 207/139, small for gestational age 207/139, head – small 220/153, head – large 220/153 individuals. ADHD: attention-deficit/hyperactivity disorder, SZ: schizophrenia (SZ), OCD: obsessive-compulsive disorder, ID: intellectual disability.

Congenital anomalies and scholastic skill disorders were associated with having likely pathogenic CNVs (Table 6) with an odds ratio (OR) of 1.85 (95%CI, 1.12,2.97), and 3.05 (95%CI, 1.64,5.47), respectively (Table 6). For probands carrying a damaging SNV, versus those who did not, ID and epilepsy showed a significant positive association (Table 6). In general, these patterns were similar when carrier status for either a CNV or SNV was assessed (Table 6).

Table 6  
Odds ratios for comorbidities and birth characteristics of individuals with ASD

Phenotypes	Genotyped Probands			Sequenced Probands			Intersection		
	Individuals with Likely pathogenic CNV			Individuals with Likely pathogenic SNV			Individuals with Likely pathogenic CNV/SNV		
	OR	95% CI	P value	OR	95% CI	P value	OR	95% CI	P value
Congenital anomalies	3.05	(1.64,5.47)	0.0002*	0.57	(0.14,1.64)	0.36	1.63	(0.84,3.03)	0.13
Small for gestational age	2.10	(0.82,4.71)	0.09	0	-	-	1.01	(0.33,2.52)	0.98
Scholastic skill disorders	1.85	(1.12,2.97)	0.01*	1.08	(0.52,2.05)	0.82	1.98	(1.19,3.21)	0.007*
HC - small	1.63	(0.82,3.01)	0.14	0.67	(0.20,1.73)	0.46	1.36	(0.70,2.52)	0.34
Motor function disorders	1.51	(0.64,3.12)	0.3	0.51	(0.08,1.70)	0.36	1.33	(0.52,3.00)	0.51
ID	1.45	(0.96,2.18)	0.07	2.6	(1.55,4.48)	0.0003*	2.31	(1.55,3.47)	5e-05*
SZ	1.42	(0.71,2.62)	0.29	0.66	(0.20,1.66)	0.43	1.48	(0.76,2.72)	0.23
Speech/language disorders	1.39	(0.79,2.35)	0.23	0.77	(0.31,1.62)	0.52	1.14	(0.62,2.00)	0.66
Epilepsy	1.08	(0.69,1.65)	0.73	1.76	(1.06,2.89)	0.03*	1.63	(1.08,2.44)	0.02*
HC - big	1.01	(0.30,2.62)	0.99	0.86	(0.53,5.02)	0.27	1.39	(0.50,3.36)	0.49
ADHD	0.99	(0.58,1.61)	0.97	0.75	(0.37,1.41)	0.4	1.05	(0.62,1.70)	0.86
OCD + Anxiety	0.93	(0.44,1.77)	0.84	0.5	(0.15,1.25)	0.19	0.74	(0.35,1.43)	0.4
Large for Gestational age	0.47	(0.08,1.58)	0.31	0	-	-	0.16	(0.01,0.77)	0.07

Missing values for congenital anomalies (Genotyped Probands /Sequenced Probands) 336/260, large for gestational age 207/139, small for gestational age 207/139, head – small 220/153, head – large 220/153 individuals. ADHD: attention-deficit/hyperactivity disorder, SZ: schizophrenia (SZ), OCD: obsessive-compulsive disorder, ID: intellectual disability.

We next compared the effect of carrier status for ASD subjects who do or do not manifest ID for the largest group of genetically characterized subjects, those who were genotyped (Table 7). Although ASD subjects who had a likely pathogenic CNV were more likely to have ID, ASD subjects with and without ID showed no significant differences in the association of carrier status with other, potentially associated phenotypes (P value > 0.05 for all tests). Thus, conditioning on ID status does not explain much of the variation for other CNV-related associations. For instance, the risk for congenital abnormalities are similar for likely pathogenic CNV carriers whether or not they meet criteria for ID, 2.46 versus 3.87 (Table 7).

Table 7  
Comparison of individuals with and without ID

Phenotypes	Genotyped Probands <i>without ID (N = 523)</i>			Genotyped Probands <i>with ID (N = 474)</i>		
	Individuals with Likely pathogenic CNV (N = 46, 8.8%)			Individuals with Likely pathogenic CNV (N = 58, 12.2%)		
	OR	95% CI	P value	OR	95% CI	P value
Congenital anomalies	3.87	(1.41,9.66)	0.005*	2.46	(1.09,5.26)	0.02*
Speech/language disorders	2.19	(0.94,4.66)	0.05*	0.92	(0.41,1.89)	0.83
Small for gestational age	2.17	(0.48,7.08)	0.24	2.01	(0.55,5.81)	0.23
Scholastic skill disorders	2.06	(0.92,4.25)	0.06	1.61	(0.83,3.00)	0.14
HC - big	1.89	(0.43,5.99)	0.33	0.42	(0.02,2.11)	0.4
OCD + Anxiety	1.58	(0.65,3.41)	0.27	0.39	(0.06,1.33)	0.2
SZ	1.24	(0.41,3.09)	0.66	1.63	(0.63,3.68)	0.27
Epilepsy	1.23	(0.58,2.45)	0.57	0.86	(0.49,1.51)	0.61
ADHD	1.14	(0.54,2.26)	0.72	0.89	(0.41,1.78)	0.76
Large for gestational age	0.92	(0.14,3.31)	0.91	0	-	-
Motor function disorders	0.86	(0.13,3.02)	0.84	1.99	(0.71,4.86)	0.15
HC - small	0.26	(0.01,1.25)	0.19	2.85	(1.31,5.94)	0.006*

Data for sleeping disorders, hypotonia, birth defects, prenatal growth rate, gestational age in weeks, weight and height at birth, and Apgar scores were underpowered due to a high number of missing values. ADHD: attention-deficit/hyperactivity disorder, SZ: schizophrenia (SZ), OCD: obsessive-compulsive disorder, ID: intellectual disability.

## Discussion

In this study, we investigated the characterization of CNVs and SNVs in a population sample of individuals from Sweden identified with autistic disorder. Of this population sample, 17.6% had a likely pathogenic CNV or SNV. Carriers of likely pathogenic CNV make up 10.4% of the individuals with AD in the Swedish population, similar to that reported for European ancestry in other studies (19, 20). Among individuals with AD, we observed a significant association between likely pathogenic genetic findings and intellectual disability, scholastic skills disorders, and epilepsy. Individuals with CNVs had elevated risk for congenital anomalies, a relevant risk factor for autism.

Eight individuals had 15q11q13 duplication syndrome, which contributes to delayed development and impairment of motor functions. Six individuals had *GRIN2B* mutations (n = 6), which is reported in individuals with muscle tone abnormalities, epilepsy, and ASD (21). One individual had a *SHANK3* SNV and two individuals had a CNV including *SHANK3*; *SHANK3* encodes a protein that is essential for proper functioning of the synapse, the junction between neurons.

PAGES is a population sample, and we did not exclude subjects with complex presentations, and therefore is likely to be more representative of the population. Importantly, the results do agree with previous literature. For example, 15q11-q13 duplication syndrome is often cited as a prominent genomic disorder in ASD (22, 23); ASD in the presence of CNV or SNV in *SHANK3* Phelan-McDermid syndrome is estimated to account for 0.5-2% of ASD (24); and comorbid epilepsy and ASD are often traced to *GRIN2B* mutations. *POGZ* is emerging as a major gene in ASD (4), similar to what is observed here. Other ASD related high frequency mutations were *SATB1* (n = 4), *DYNC1H1* (n = 4), and *CREBBP* (n = 3).

## Limitations

The results of this study should be interpreted in the context of some limitations. First, not all variants were validated by a second method; therefore, some could be artifacts. Nonetheless, a substantial portion of the CNVs were independently validated by calling CNVs from the whole exome data (16) and the validation rate of SNV is similarly high, as documented by variant calls from whole-genome versus whole-exome sequencing (25). Second, judgment calls and empirically defined thresholds were used to identify likely pathogenic variants. Third, we did not have de novo information to aid in classification. Fourth, this study focused on AD and future studies on individuals with less profound ASD are warranted in order to draw a more comprehensive picture of the genetic architecture of the autism spectrum. Finally, phenotypes are defined from medical registers, which may lead to underestimates of milder findings.

## Conclusions

This population survey, with its characterization of rare potentially likely pathogenic variants, provides greater insight into the genetic architecture of ASD and associated comorbidities. Moreover, because many of the same subjects have been characterized for genotypes from common variants, we can explore the genetic architecture of ASD in far greater detail. Indeed, in an accompanying manuscript (Klei et al.), we explore the joint contributions of rare and common variation to liability for ASD, finding that they work together approximately additively.

## List Of Abbreviations

AD: Autistic Disorder, defined by the ICD-10, code F84.0

ADHD: Attention-Deficit/Hyperactivity Disorder

ASD: Autism Spectrum Disorder

CMA: Chromosomal Microarray

CNV: Copy Number Variant (CNV)

DSM: Diagnostic and Statistical Manual of Mental Disorders

ICD: International Classification of Diseases

ID: Intellectual Disability

LoF: Loss-of-function mutation

OCD: Obsessive-Compulsive Disorder

MPC: "Missense badness, Polyphen-2, Constraint" pathogenicity score

NPR: Swedish National Patient Register

PAGES: Population-based Autism Genetics and Environmental Study

SNV: Single Nucleotide Variants

SZ: Schizophrenia

WES: Whole-Exome Sequencing

## Declarations

**Ethics approval and consent to participate:** The study was approved by the Regional Ethical Review Board in Stockholm, Sweden, and the Institutional Review Board (IRB) at the Icahn School of Medicine at Mount Sinai, New York, USA.

**Consent for publication:** Not relevant

**Availability of data and materials:** The data that support the findings of this study are available from the corresponding author upon reasonable request.

**Competing interests:** The authors declare that they have no competing interests

**Funding:** This study was supported by the National Institute of Mental Health (NIMH) grant R01MH097849, R01MH097849-S1, and the Beatrice and Samuel A. Seaver Foundation.

**Author Contributions:** BM had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: JDB, BD, KR, SDR, BM

Acquisition, analysis, or interpretation of data: JDB, BD, SDR, BM, CGM, MM

Drafting of the manuscript: JDB, BD, SDR, BM, CGM, MM

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: BD, BM

Obtained funding: JDB, BD, KR

Study supervision: JDB, BD, KR

## References

1. Levy SE, Mandell DS, Schultz RT. Autism. Vol. 374, The Lancet. 2009. p. 1627–38.
2. Association American Psychiatric. Diagnostic and statistical manual of mental disorders (4th ed.). 2000;
3. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). 2013;
4. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An JY, et al. Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. *Cell*. 2020 Feb 6;180(3):568-584.e23.
5. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. *Nat Genet*. 2014;46(8):881–5.
6. Matsunami N, Hadley D, Hensel C, Christensen GB, Kim C, Frackelton E, et al. Identification of Rare Recurrent Copy Number Variants in High-Risk Autism Families and Their Prevalence in a Large ASD Population. *PLoS One*. 2013;8(1):e52239.
7. Leblond CS, Cliquet F, Carton C, Hugué G, Mathieu A, Kergrohen T, et al. Both rare and common genetic variants contribute to autism in the Faroe Islands. *npj Genomic Med*. 2019;4(1).
8. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, et al. A Discovery resource of rare copy number variations in individuals with autism spectrum disorder. *G3 Genes, Genomes, Genet*. 2012;2(12):1665–85.
9. Fan Y, Du X, Liu X, Wang L, Li F, Yu Y. Rare Copy Number Variations in a Chinese Cohort of Autism Spectrum Disorder. *Front Genet*. 2018;9.
10. Velinov M. Genomic copy number variations in the autism clinic—work in progress. *Front Cell Neurosci*. 2019;13.
11. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. *Neuron*. 2015;87(6):1215–33.

12. Bitar T, Hleihel W, Marouillat S, Vonwill S, Vuillaume ML, Soufia M, et al. Identification of rare copy number variations reveals PJA2, APCS, SYNPO, and TAC1 as novel candidate genes in Autism Spectrum Disorders. *Mol Genet Genomic Med*. 2019;7(8).
13. D'Abate L, Walker S, Yuen RKC, Tammimies K, Buchanan JA, Davies RW, et al. Predictive impact of rare genomic copy number variations in siblings of individuals with autism spectrum disorders. *Nat Commun*. 2019;10(1).
14. Griswold AJ, Ma D, Cukier HN, Nations LD, Schmidt MA, Chung RH, et al. Evaluation of copy number variations reveals novel candidate genes in autism spectrum disorder-associated pathways. *Hum Mol Genet*. 2012;21(15):3513–23.
15. Tammimies K, Marshall CR, Walker S, Kaur G, Thiruvahindrapuram B, Lionel AC, et al. Molecular diagnostic yield of chromosomal microarray analysis and whole-exome sequencing in children with autism spectrum disorder. *JAMA - J Am Med Assoc*. 2015;314(9):903–909.
16. Team G. (How to) Call common and rare germline copy number variants. Broad Institute. Available from: <https://gatk.broadinstitute.org/hc/en-us/articles/360035531152-How-to-Call-common-and-rare-germline-copy-number-variants>
17. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr Protoc Bioinforma*. 2013;(SUPL.43).
18. Chaste P, Klei L, Sanders SJ, Murtha MT, Hus V, Lowe JK, et al. Adjusting head circumference for covariates in autism: Clinical correlates of a highly heritable continuous trait. *Biol Psychiatry*. 2013;74(8):84–91.
19. Shen Y, Dies KA, Holm IA, Bridgemohan C, Sobeih MM, Caronna EB, et al. Clinical genetic testing for patients with autism spectrum disorders. *Pediatrics*. 2010;125(4).
20. Schaefer GB, Mendelsohn NJ. Erratum: Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions (*Genetics in Medicine* (2013) DOI: 10.1038/gim.2013.32). *Genet Med*. 2013;15(8):669.
21. Platzer K, Lemke JR. GRIN2B-Related Neurodevelopmental Disorder. *GeneReviews*®. 1993;
22. Reiter LT. Chapter 9 - Developmental disabilities, autism, and schizophrenia at a single locus: complex gene regulation and genomic instability of 15q11-q13 cause a range of neurodevelopmental disorders. *Neurodev Disord*. 2020;201–21.
23. Lu Y, Liang Y, Ning S, Deng G, Xie Y, Song J, et al. Rare partial trisomy and tetrasomy of 15q11-q13 associated with developmental delay and autism spectrum disorder. *Mol Cytogenet*. 2020;13(1).
24. Oberman LM, Boccuto L, Cascio L, Sarasua S, Kaufmann WE. Autism spectrum disorder in Phelan-McDermid syndrome: Initial characterization and genotype-phenotype correlations. *Orphanet J Rare Dis*. 2015;10(1).
25. An JY, Lin K, Zhu L, Werling DM, Dong S, Brand H, et al. Genome-wide de novo risk score implicates promoter variation in autism spectrum disorder. *Science* (80- ). 2018;362(6420).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [PAGESCSNVSupplement.docx](#)
- [PAGESCSNVSupplement.docx](#)