Copy-Number Variants in The Contactin-5 Gene Are a Potential Risk Factor for Autism Spectrum Disorder

Zoe Schmilovich, Guillaume Huguet, Qin He, Amélie Musa-Johnson, Elise Douard, Mor Absa Loum, Calwing Liao, Jay P Ross, Alexandre Dionne-Laporte, Dan Spiegelman, et al.

Copy-Number Variants in The Contactin-5 Gene Are a Potential Risk Factor for Autism Spectrum Disorder

Zoe Schmilovich
McGill University
Guillaume Huguet
CHU Sainte-Justine: Centre Hospitalier Universitaire Sainte-Justine
Qin He
Institut universitaire en santé mentale: Institut universitaire en sante mentale de Montreal
Amélie Musa-Johnson
Institut universitaire en santé mentale: Institut universitaire en sante mentale de Montreal
Elise Douard
CHU Sainte-Justine: Centre Hospitalier Universitaire Sainte-Justine
Mor Absa Loum
CHU Sainte-Justine: Centre Hospitalier Universitaire Sainte-Justine
Calwing Liao
McGill University
Jay P. Ross
McGill University
Alexandre Dionne-Laporte
McGill University
Dan Spiegelman
McGill University
Caroline Hayward
The University of Edinburgh MRC Human Genetics Unit
Tobias Banaschewski
Central Institute of Mental Health: Zentralinstitut fur Seelische Gesundheit
Arun Bokde
Trinity College Institute of Neuroscience
Sylvane Desrivieres
King's College London Institute of Psychiatry Psychology and Neuroscience
Herve Lemaitre
Universite Paris-Saclay
Gunter Schumann
King's College London Institute of Psychiatry Psychology and Neuroscience
Lan Xiong
Montreal Neurological Institute and Hospital
Patrick A. Dion
Montreal Neurological Institute and Hospital
Sébastien Jacquemont
CHU Sainte-Justine: Centre Hospitalier Universitaire Sainte-Justine
Boris Chaumette
IPNP: Institut de Psychiatrie et Neurosciences de Paris
Guy A. Rouleau (guy.rouleau@mcgill.ca)
McGill University https://orcid.org/0000-0001-8403-1418

Research Article

Keywords: ASD, CNTN5, CNV, intronic deletions, neurodevelopment, inherited

Posted Date: July 6th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-660740/v1

Abstract

Background

Contactin-5 (CNTN5) is a candidate risk gene for autism spectrum disorder (ASD), yet previous attempts to associate copy-number variants (CNVs) encompassing *CNTN5* with ASD-susceptibility were limited by insufficient statistical power. Here, we aim to clarify the putative association between *CNTN5* CNVs and ASD-risk using large samples.

Methods

First, we calculated the prevalence and transmission of *CNTN5* CNVs in ASD across three ASD cohorts (SSC, MSSNG, and SPARK), the cases reported in the Mercati *et al.* study, and the BBGRE database (n = 16,607). Second, we modelled their transmission in children with ASD compared to their unaffected siblings. Third, we assessed their frequency in cases with ASD compared to unselected population controls (n = 24,898) and replicated the findings in UK Biobank (UKBB), an independent general population cohort (n = 459,855). Finally, we evaluated the clinical impact of *CNTN5* CNVs by assessing their enrichment in a broad neurodevelopmental disorder (NDD) cohort, and the clinical profile of *CNTN5* CNV carriers in the DECIPHER database.

Results

The prevalence of *CNTN5* exonic deletions and duplications was stable across ASD and across unselected cohorts (0.042% and 0.020%, respectively). We found a significant enrichment of intronic *CNTN5* deletions CNVs in ASD compared to unselected controls (0.175% and 0.004%, respectively). CNVs in most cases with ASD (29 out of 30, 96.7%) were inherited. Parents transmitted the variants to their affected and unaffected children with the same frequency. No differences in exonic *CNTN5* CNVs enrichment between cases with ASD compared to individuals with NDDs was observed.

Limitations

The lack of phenotypic data available for unaffected family members of probands with ASD limits the potential to assess whether *CNTN5* CNVs segregate with other neuropsychiatric or sub-threshold autistic traits. Different genotyping or sequencing technologies may affect the differences in *CNTN5* CNV prevalence across cohorts.

Conclusion

CNTN5 CNVs are rare inherited ASD susceptibility variants. They may also confer risk for other neuropsychiatric disorders. We offer a powerful framework to investigate candidate susceptibility variants that may not be detected through small-scale approaches. This approach may reveal more intermediate effect-size variants that are implicated in the etiology of ASD.

Background

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with a high estimated overall heritability (0.7–0.8) and a complex genetic architecture consisting of rare and common variants [1, 2]. Rare *de novo* and inherited copy-number variants (CNVs), corresponding to deleted or duplicated regions of the genome, substantially increase ASD-risk and are present in 8–14% of individuals with ASD [3]. Variants of intermediate effect size may also contribute to ASD-susceptibility, but they remain to be discovered. Some of them are inherited from healthy parents due to the incomplete penetrance. In a multiplex Pakistani family with three brothers with autism and a fourth brother with a learning disorder, we identified a *CNTN5* deletion inherited from their unaffected father. Focusing on this CNV, we proposed a framework to confirm the relevance of inherited CNV with intermediate effect size by using large case-control studies. This approach may provide sufficient statistical power to characterize the role of candidate intermediate effect size CNVs in ASD risk.

Variants within the contactin gene family are interesting candidates for ASD-risk [4]. Contactins are a group of neuronal cell adhesion molecules, encoded by the *contactin-1* to *contactin-6* genes, which have been shown to play a key role in axonal guidance and organization, myelinisation, neurogenesis, neuronal development, synaptogenesis and axon-glia interactions [5]. The aberrant functioning and incorrect cellular localization of these proteins has been previously implicated in autism [6–8] and neurodevelopmental delay [9, 10]. Contactin-1 and contactin-2 interact with CNTNAP1 and CNTNAP2, and have been associated with demyelinating diseases [5]. Contactin-3 is involved in outgrowth and guidance of axons and dendrites [11], while contactin-4, 5, and 6

play a key role in central nervous system developmental processes [5]. While the association between ASD and variants in *CNTN3, CNTN4*, and *CNTN6* have been well characterized, the involvement of *CNTN5* variants in the etiology of ASD remains elusive.

Previous attempts to elucidate the role of *CNTN5* in ASD have been limited by small sample sizes and insufficient statistical power. Mercati *et al.* identified a maternally-inherited *CNTN5* deletion in an individual with ASD with hypersensitivity to sounds and abnormal motor coordination, and a *CNTN5* deletion in an individual with ASD from the Brain & Body genetic Resource Exchange (BBGRE version 3.0; https://bbgre.brc.iop.kcl.ac.uk/) [12]. CNVs in *CNTN5* have similarly been reported in other cohorts of multiplex and simplex families with ASD [8, 11].

Similar to other contactins, contactin-5 is involved in neuronal cell adhesion, and interacts with CNTNAP4 as a scaffold on inter-neurons to support the growth of ganglion cells [13–15]. Given the function of CNTN5 in neuronal maintenance and previous attempts to associate its variants with ASD-risk, a large-scale case-control study could clarify the role of *CNTN5* CNVs in ASD-risk.

In this study, we reported a Pakistani family with a *CNTN5* CNV and assessed its role in ASD-risk. Firstly, we evaluate the prevalence of *CNTN5* CNVs and how they arise in more than 16,000 individuals with ASD. Secondly, we evaluate whether *CNTN5* CNVs are over-transmitted in individuals with ASD in comparison to their unaffected siblings. Thirdly, we assess the differential enrichment of *CNTN5* CNVs between cases with ASD and unselected individuals from the general population ($n \approx 25{,}000$) and replicate the findings in UK Biobank, an independent control cohort ($n \approx 460{,}000$). Finally, we determine the clinical significance of *CNTN5* CNVs by comparing their prevalence in cases with ASD to patients with suspected neurodevelopmental disorders (NDDs) ($n \approx 16{,}500$), and assess their prevalence in DECIPHER, a rare disease genomic database.

Our study outlines an approach for investigating the role of intermediate effect size candidate variants using large-scale data. Our results implicate the possible role of *CNTN5* CNVs in ASD-susceptibility.

Methods

Exploration of a multiplex ASD family

A multiplex family with ASD, consisting of five brothers, a sister, and two parents was identified and recruited in Pakistan by a local psychiatrist (Dr. Brohi Qasim). Following a face-to-face standardized evaluation based on DSM-IV, three brothers were given a diagnosis of autism, and a fourth brother was diagnosed with a learning disorder (Supplementary **Figure A**). Following standard DNA extraction and genotyping protocol, CNVs across all family members were annotated and called (**Supplementary Note**). In the interest of identifying candidate ASD-risk variants, an in-house Python script, SV-Segregation, was used to identify CNVs that were shared by all three brothers with ASD.

Cohorts

ASD cohorts

We included individuals with ASD from the Simons Simplex Collection (SSC) (n = 2,585) [16], MSSNG (n = 3,289) [17], and Simons Foundation Powering Autism Research for Knowledge (SPARK) (n = 8,423) [18] cohorts as cases in this study. The SSC is a cohort of simplex families consisting of trios (one proband with ASD and unaffected parents) and quads (one proband with ASD, unaffected parents, and unaffected sibling(s)). The MSSNG database includes simplex and multiplex (multiple affected family members with ASD) from the Autism Speaks whole-genome sequencing (WGS) project. The SPARK database is a newly established genetic cohort that includes genotyping data of over 8,000 families or singletons with ASD. Parents diagnosed with ASD or half-siblings were excluded from the study. We also included individuals with ASD reported in Mercati *et al.* (n = 1,534) [12] and the BBGRE database (https://bbgre.brc.iop.kcl.ac.uk/) (n = 776) in our meta-analysis. In total, 16,607 cases with ASD were studied.

When available, the unaffected siblings and unaffected parents of probands with ASD from the SSC, MSSNG, and SPARK cohorts were used as intrafamilial control subjects. Parents diagnosed with ASD from the SPARK cohort ($n_{fathers} = 48$; $n_{mothers} = 101$) were excluded from the analyses. In total, we studied 5,290 unaffected siblings and 21,767 unaffected parents from the SSC, MSSNG, and SPARK cohorts.

Unselected control cohorts

In our discovery cohort, we included 1,802 individuals from the IMAGEN Study [19] and 14,160 individuals from Generation Scotland Scottish Family Health Study (GS) [20] to represent unselected controls from the general population. We also included the 8,936 unaffected controls reported in Mercati *et al*, which are detailed in their Supplementary Materials. In total, we used 24,898 individuals as independent unselected controls in our discovery cohort. We used the genetic data of 459,855 individuals from the UK Biobank (UKBB) [21], a large-scale population-based study, as an independent replication cohort to validate our findings.

Clinical cohorts

We included 16,586 individuals from the Centre Universitaire Sainte-Justine (CHU-SJ; Montreal, Canada) to represent a broad clinical cohort. These individuals were referred to the cytogenetic pediatric laboratory for suspected NDDs.

CNV Calling and Annotation

CNVs were detected, filtered, and annotated using the genotyping data from SSC, SPARK, IMAGEN, GS, UKBB and WGS data from MSSNG according to the criteria detailed in the Supplementary Methods of Douard *et al.* [22]. CNVs from chromosomal microarray data of individuals from CHU-SJ were called according to the Supplemental Methods of Huguet *et al.* [23].

Individuals carrying a CNV deletion or duplication in either the exonic or intronic region of the *CNTN5* gene were identified. As only exonic CNVs were reported in Mercati *et al.* and BBGRE, these cohorts were excluded from combined *CNTN5* CNV (intronic and exonic) carrier counts and intronic *CNTN5* CNV carrier counts.

Statistical Analysis

To assess whether *CNTN5* CNVs were over-transmitted in cases with ASD, in comparison to their unaffected siblings, we used a generalized linear mixed-effects model, accounting for relatedness as a random effect, with ASD diagnosis as the outcome variable and *CNTN5* CNV presence as the predictor variable. The analysis was done using the R package ("lme4") [24].

Enrichment of exonic and intronic *CNTN5* CNVs in cases compared to unselected controls and individuals from the clinical cohort was assessed using a Fisher's exact test ("fisher.test") in the R ("stats") package.

All analyses were performed on the R version 3.6.1 [25].

Results

Identification of a CNTN5 CNV deletion in a multiplex family with ASD

In the multiplex Pakistani family, a paternally-inherited 50.3 kB deletion in 11q22.1 (hg19 - chr11:99480175–99530431) shared by all three brothers with autism and the fourth brother with a learning disorder was identified. The CNV deletion encompassed the intronic region of the *CNTN5* gene between exon 2 and 3 (**Supplementary Figure B**).

Prevalence of CNTN5 CNVs across cohorts

The number of *CNTN5* CNV carriers across cohorts is detailed in Table 1. There were 32 out of 16,607 individuals with ASD carrying a *CNTN5* CNV (0.193%). The prevalence of exonic *CNTN5* CNVs ranged from 0.018% (duplications) to 0.024% (deletions). The prevalence of intronic *CNTN5* CNVs in ASD ranged from 0.007% (duplications) to 0.168% (deletions). Most of the intronic deletions in cases with ASD were identified from the SPARK cohort. This positions *CNTN5* CNVs as a rare to intermediate prevalence variant. Of the total 171 *CNTN5* carriers across cohorts, 32 individuals had a diagnosis of ASD (18.71%).

With the exception of one carrier from the SPARK cohort, *CNTN5* deletions and duplications in individuals with ASD were transmitted by an unaffected parent. Similarly, all *CNTN5* CNVs in cases with ASD from Mercati *et al.* were inherited. Of the 50 unaffected parents carrying a *CNTN5* CNV, 33 (66%) transmitted it to their child, with 7 (14%) parents transmitting it to their unaffected child only. As such, 17 (34%) parents carried a *CNTN5* CNV and did not transmit it at all.

Table 1
Prevalence of CNTN5 CNV carriers across cohorts.

			All CNTN5 CNVs (Exonic and Intronic)			Exonic			Intronic			
	Cohort	Total samples	DEL and DUP (%)		DEL	DUP	DEL and DUP (%)		DEL	DUP	DEL and DUP (%)		DEL	DUP
	Probands with ASD													
Cases	SSC	2585	4	0.155%	1	3	3	0.116%	1	2	1	0.039%	0	1
	MSSNG	3289	1	0.030%	1	0	1	0.030%	1	0	0	0.000%	0	0
	SPARK	8423	25	0.297%	24	1	1	0.012%	0	1	24	0.285%	24	0
	Mercati et al. (2017)[a,b]	1534	1	0.065%	1	0	1	0.065%	1	0	-	-	-	-
	BBGRE[b,c]	776	1	0.129%	1	0	1	0.129%	1	0	-	-	-	-
	Total cases with ASD	16607	32	0.193%	28	4	7	0.042%	4	3	25	0.175%	24	1
Controls	**Unaffected siblings of probands with ASD**													
	SSC	2425	3	0.124%	1	2	2	0.082%	1	1	1	0.041%	0	1
	MSSNG	131	0	0.000%	0	0	0	0.000%	0	0	0	0.000%	0	0
	SPARK	2734	10	0.366%	9	1	1	0.037%	0	1	9	0.329%	9	0
	Total unaffected siblings	5290	13	0.246%	10	3	3	0.057%	1	2	10	0.189%	9	1
	Unaffected parents of probands with ASD													
	SSC	5153	5	0.097%	2	3	4	0.078%	2	2	1	0.019%	0	1
	MSSNG	3600	1	0.028%	1	0	1	0.028%	1	0	0	0.000%	0	0
	SPARK	13014	44	0.338%	42	2	3	0.023%	1	2	41	0.315%	41	0
	Total unaffected parents	21767	50	0.230%	45	5	8	0.037%	4	4	42	0.193%	41	1
	Discovery: Unselected individuals from the general population													
	GS	14160	4	0.028%	4	0	1	0.007%	1	0	3	0.021%	3	0
	IMAGEN	1802	0	0.000%	0	0	0	0.000%	0	0	0	0.000%	0	0
	Mercati et al. (2017)[b]	8936	4	0.045%	1	3	4	0.045%	1	3	-	-	-	-
	Total discovery controls	24898	8	0.050%	5	3	5	0.020%	2	3	3	0.019%	3	0
	Replication: Unselected individuals from the general population													
	UKBB	459855	68	0.015%	46	22	48	0.010%	26	22	20	0.004%	20	0
Clinical pediatric cohort	**Individuals with suspected NDDs**													
	CHU-SJ	16586	2	0.012%	1	1	2	0.012%	1	1	0	0.000%	0	0

	All CNTN5 CNVs (Exonic and Intronic)	Exonic	Intronic
The number and proportion of *CNTN5* CNV carriers across ASD, control, and clinical pediatric cohorts.			
[a] Cases with ASD reported in Mercati *et al.* include 901 individuals from Pinto *et al* [26] and 633 individuals from their cohort.			
[b] Intronic CNV information for these cohorts unavailable – these individuals are excluded from total and exonic *CNTN5* CNV prevalence counts			
[c] The number of *CNTN5* CNV carriers in BBGRE is reported in Mercati *et al.*			

Probands with ASD and their unaffected siblings have a comparable prevalence of CNTN5 **CNVs**

Results from the generalized linear-mixed effects model, accounting for family relatedness, found no association between *CNTN5* CNVs in ASD-risk in cases with ASD compared to their unaffected siblings (Table 2). These results suggest that unaffected parents transmit *CNTN5* CNVs to their children with ASD and their unaffected children at the same rate.

Table 2
No association of CNTN5 CNVs with ASD-risk in probands with ASD compared to unaffected siblings

	All *CNTN5* CNVs (DEL and DUP)				DEL				DUP			
	Carrier		p-value	OR (95%CI)	Carrier		p-value	OR (95%CI)	Carrier		p-value	OR (95%CI)
All CNVs (Exonic and Intronic)												
Cases with ASD	30	0.21%	0.634	0.854	26	0.18%	0.917	0.962 (0.46-2.0)	4	0.03%	0.355	0.493
Unaffected siblings	13	0.25%		(0.5–1.6)	9	0.17%			3	0.06%		(0.1–2.2)
Exonic												
Cases with ASD	5	0.04%	0.508	0.617	2	0.01%	0.806	0.740 (0.07–8.2)	3	0.02%	0.519	0.56 (0.09–3.3)
Unaffected siblings	3	0.06%		(0.15–2.6)	1	0.02%			2	0.04%		
Intronic												
Cases with ASD	25	0.18%	0.835	0.925	24	0.17%	0.973	0.987	1	0.007%	0.482	0.370 (0.02–5.9)
Unaffected siblings	10	0.19%		(0.4–1.9)	9	0.17%		(0.5–2.1)	1	0.02%		
Results from a generalized linear-mixed effects model of the association between *CNTN5* CNVs and ASD-risk in probands with ASD compared to their unaffected siblings. Only probands and unaffected siblings from family-based cohorts (SSC, MSSNG, SPARK) are included in analyses. Cases with ASD from Mercati *et al.* and BBGRE are excluded.												

Frequency of CNTN5 **CNVs in cases with ASD and controls**

We first compared the frequency of all combined (exonic and intronic) *CNTN5* deletions and duplications in cases with ASD (30 out of 14,297; 0.201%) and controls (4 out of 15,962; 0.025%), which excluded samples from Mercati *et al.* and the BBGRE database, where intronic CNV information was unavailable (Table 3A). A significant enrichment of *CNTN5* deletion and duplication for all CNVs (exonic and intronic) in ASD was observed (Fisher's exact test; p = 0.0001; OR = 4.13; 95%CI = [0.85–10.44]). To determine which CNVs were driving the association signal, we assessed the differential enrichment across deletions or duplications, and CNVs localized in the exonic or intronic regions (Table 3 details the results of the nine tests). Intronic deletions were significantly enriched in cases with ASD (0.168%) compared to controls (0.019%) (Fisher's exact test; $p = 1.68 \times 10^{-5}$; OR = 8.82; 95%CI = [2.68-45.70]). Only three intronic deletions identified in controls, whereas 24 intronic deletions were identified in cases with ASD (all of which were found in cases from the SPARK cohort, involving the intronic region of the *CNTN5* gene between exons 2 and 3). No significant difference in the prevalence of exonic deletions and duplications between cases and controls was identified.

We replicated the case-control enrichment analyses in UKBB, an independent general population cohort (Table 3B). Combined exonic and intronic *CNTN5* deletions and duplications were enriched in cases with ASD in comparison to controls (68 out of 459,855; 0.015%) (Fisher's exact test; $p = 4.48 \times 10^{-22}$; OR = 14.21; 95%CI = [8.92-22.15]). Similar to the discovery control cohort, we assessed the enrichment across all CNV types, and found that intronic deletions were driving the association signal (Table 3B) (Fisher's exact test; $p = 1.45 \times 10^{-22}$; OR = 27.73; 95%CI = [15.07-51.05]). No significant difference in the prevalence of exonic deletions and duplications between cases and controls in the replication cohort was identified.

Table 3
Case-control CNTN5 CNV enrichment analyses.

	All CNVs (DEL and DUP)				**DEL**				**DUP**			
	Carrier		p-value	OR (95%CI)	Carrier		p-value	OR (95%CI)	Carrier		p-value	OR (95%CI)
A. Discovery set												
All CNVs (Exonic and Intronic)												
Cases[a]	30	0.210%	1.17E-04*	4.13 (0.9–10.4)	26	0.182%	4.05E-05*	5.73 (2.2–19.1)	4	0.028%	0.72	1.47 (0.25–10.02)
Controls[b]	4	0.025%			4	0.025%			0	0.000%		
Exonic												
Cases	7	0.042%	0.2439	2.07 (0.57–8.29)	4	0.024%	0.23	2.96 (0.42–32.74)	3	0.018%	0.31	4.44 (0.36–232.9)
Controls	5	0.020%			2	0.008%			3	0.012%		
Intronic												
Cases[a]	25	0.175%	4.90E-06*	9.18 (2.80-47.48)	24	0.168%	1.68E-05*	8.82 (2.68–45.69)	1	0.07%	0.48	Inf
Controls[b]	3	0.015%			3	0.019%			0	0.00%		
B. Replication set												
All CNVs (Exonic and Intronic)												
Cases[a]	30	0.210%	4.48E-22*	14.22 (8.93–22.15)	26	0.18%	2.03E-21*	18.21 (10.80-30.12)	4	0.03%	0.0072	5.85 (1.47–17.22)
Controls	68	0.015%			46	0.01%			22	0.00%		
Exonic												
Cases	7	0.042%	0.001	4.70 (1.80-10.43)	4	0.02%	0.01	4.95 (1.26–14.27)	3	0.02%	0.038	4.39 (0.84–14.61)
Controls	48	0.010%			26	0.01%			22	0.00%		
Intronic												
Cases[a]	25	0.175%	1.46E-24*	31.78 (17.3-58.63)	24	0.17%	1.45E-22*	27.73 (15.1-51.05)	1	0.01%	0.035	Inf
Controls	20	0.004%			20	0.004%			0	0.00%		

Fisher's exact test to identify an association between *CNTN5* CNVs and ASD-risk. A significant (*) enrichment threshold ($p \leq 0.006$) was determined according to a Bonferroni correction, adjusting for 9 independent tests. Intronic *CNTN5* CNV deletions were significantly associated with ASD-risk in the discovery (unselected controls from GS, IMAGEN, and Mercati *et al.*) and replication (unselected controls from the UKBB cohort) cohorts.

[a] Samples exclude cases with ASD from Mercati *et al.* and the BBGRE database, where intronic CNV information is unavailable.

[b] Samples exclude controls from Mercati *et al.* and the BBGRE database, where intronic CNV information is unavailable.

CNTN5 **CNVs are not specific to ASD**

Two out of 16,589 individuals referred to the pediatric hospital of CHU-SJ for suspected NDDs carried an exonic *CNTN5* CNV. No intronic *CNTN5* deletions or duplications were identified in the CHU-SJ cohort. The two carriers were screened postnatally for an indication of epilepsy. One carried a 95.3kb duplication (chr11:100173136–100268422) and the other a 53.6kb deletion (chr11:99675192–99728831), both restricted to *CNTN5*. They did not carry any other pathogenic CNV that could explain their clinical presentation. A carrier with a 29.52 Mb deletion encompassing 119 genes, including *CNTN5*, was excluded from the analysis. A Fisher's exact test revealed no significant enrichment of *CNTN5* CNVs in individuals with ASD in comparison to individuals with NDDs from the CHU-SJ cohort ($p \geq 0.05$).

In the DECIPHER database v10.3 (https://decipher.sanger.ac.uk/), 30 *CNTN5* CNV carriers (deletions and duplications) are listed among 36,530 patients (0.082%). The phenotype for autism is indicated 3 times out of 30 carriers (10%). Of the three individuals with autism, 1 individual has a duplication inherited from a deemed to be unaffected parent and the other two are deletion carriers with unknown transmission. Half (15 out 30) of the *CNTN5* deletion and duplication carriers in DECIPHER had either a cognitive impairment, intellectual disability, global developmental delay, learning disability, or autism (**Supplementary Table**).

Discussion

A paternally-inherited *CNTN5* CNV intronic deletion in a multiplex family we identified, with three affected brothers with ASD and a fourth brother with a learning disorder, provided further motivation to investigate its role in a large-scale case-control study. Based on this analysis, we identified an enrichment of intronic *CNTN5* deletion in ASD compared to an unselected control population.

Similar to previous studies, we position *CNTN5* CNVs as rare risk variants in ASD. Across individuals with ASD from SSC, MSSNG, and SPARK (n_{total} = 14,297), *CNTN5* CNVs occurred at a frequency of 0.210%. This rare characterization is slightly higher than the prevalence reported in Mercati *et al.* (0.065%) and BBGRE (0.129%) which reported exonic CNVs while we reported both intronic and exonic CNVs. We found that, except one case, all individuals with ASD had inherited the variant from an unaffected parent. These findings align with previous reports of *CNTN5* CNVs in ASD. Van Daalen *et al.* [8] identified a paternally-inherited *CNTN5* CNV deletion in three children with ASD from a multiplex family. While the father was not diagnosed with ASD, the variant co-segregated with a high Social Responsiveness Scale (SRS) score, reflecting a sub-clinical behavioural impairment. Following their initial screen, Mercati *et al.* [12] identified two additional families with *CNTN5* CNVs in which the children with ASD had inherited the variant. Though unaffected with ASD, one of the mothers who transmitted the variant had a language impairment, while the other mother had a specific learning disorder. We found that unaffected parents transmitted the *CNTN5* CNV to their children with ASD and their unaffected children at the same rate; we identified no significant association between *CNTN5* CNVs and ASD-risk in probands with ASD in comparison to their unaffected siblings, suggesting that the at-risk CNVs are running in the families, and could be associated with mild phenotype.

We observed a significant association of intronic deletions in ASD compared to unselected individuals from the general population. These findings replicated in another independent control cohort. Interestingly, the three brothers with ASD and the fourth brother with a learning disorder from the multiplex Pakistani family we originally identified shared an intronic *CNTN5* deletion as well. According to Ensembl, this intron is a retained intron, thus it may encode the protein depending on its alternative splicing. Intronic CNVs have been associated with significant difference in gene expression levels in the population and may be associated with expression differences of other genes through intron-promoter 3D interactions [27].

We investigated the clinical impact of *CNTN5* CNVs and observed no significant difference in *CNTN5* CNV frequency between individuals with ASD and those with suspected NDDs from the CHU-SJ pediatric cohort. Furthermore, only 10% of the *CNTN5* CNV carriers in DECIPHER had a diagnosis of autism, while half had a broad developmental disorder. These findings reveal a lack of ASD-specific clinical impact. This is not surprising, given that CNVs conferring risk for ASD are often associated with other neurodevelopmental phenotypes.

Taken together, this study characterizes *CNTN5* CNVs as inherited rare frequency variants in ASD. We show that *CNTN5* CNVs are not fully penetrant, but might confer risk for ASD in specific genetic backgrounds through possible additive or epistatic interactions, as previously described [5, 8, 12]. Our findings likely position *CNTN5* CNVs as risk factors for broad NDDs. We demonstrate the potential of rare inherited CNVs that are otherwise difficult to characterize.

Limitations

Despite the large-scale case-control study design, the rareness of the CNV, as well as their incomplete penetrance, limits the significance of the results for clinical practice.

There is a noticeable difference in *CNTN5* CNV frequency across cohorts which may be attributed to difference in genotyping or sequencing technologies used. Most (24 out of 25) of the intronic *CNTN5* deletions in the ASD cohort came from carriers in SPARK. While all SPARK CNVs were manually visualized by plotting their Log R ratio (LRR) and B-allele frequencies (BAF) to ensure their validity, the ASD-risk association signal is mostly driven by this cohort. Furthermore, different CNV calling, annotation, and filtering parameters in Mercati *et al.*, the BBGRE database, and the DECIPHER database may limit the ability to comprehensively compare the prevalence of *CNTN5* CNVs across cohorts.

Van Daalen *et al.* [8] and Mercati *et al.* [12] describe multiplex families with ASD where *CNTN5* CNVs are transmitted to affected children with ASD from parents who, while unaffected with ASD, have behavioural impairments, learning disorders, and language impairments. We lacked detailed phenotypic information of the unaffected family members of the probands with ASD to study the potential role of *CNTN5* CNVs in other traits. While the unaffected family members do not have an ASD diagnosis, they could have neuropsychiatric traits or sub-threshold autistic phenotypes that may segregate with the variant, but do not meet the criteria for ASD diagnosis. Familial risk for neuropsychiatric disease in ASD etiology and the potential role of increased phenotypic characterization has been demonstrated [28]. Detailed behavioural and clinical information of the ASD family members could explain the comparable frequency of *CNTN5* CNVs between probands with ASD and their unaffected siblings. Similarly, this may explain the parental transmission of *CNTN5* CNVs in ASD.

Conclusions

Investigating the role of rare intermediate effect-size CNVs is important to understanding the complete genetic etiology of ASD but requires sufficient statistical power through large-scale cohorts. Here, we observe that *CNTN5* CNVs in ASD are rare and inherited intermediate effect-size variants. Parents transmit the variant to their children with ASD and unaffected children at the same rate. While intronic *CNTN5* CNV deletions significantly increase ASD-risk compared to the general population, we could not demonstrate ASD-specificity when compared to a broad NDD cohort. We position *CNTN5* CNVs as potential inherited risk factors for neurodevelopmental disorders. Moreover, we demonstrate a framework for investigating the role of candidate variants in complex heterogenous disorders. Given the heterogenous nature of ASD, identifying more susceptibility variants may elucidate a clearer picture of the genetic architecture of the disorder.

Abbreviations

ASD: Autism Spectrum Disorder

CNTN5: Contactin-5

CNV: Copy Number Variant

CNS: Central Nervous System

CHU-SJ: Centre Hospitalier Universitaire Sainte Justine

SSC: Simons Simplex Collection

SPARK: Simons Foundation Powering Autism Research for Knowledge

UKBB: UK Biobank

WGS: Whole-Genome Sequencing

SRS: Social Responsiveness Scale

Ethics approval and consent to participate

All individuals or their legal representatives gave their written informed consent. All procedures were carried out according to the Declaration of Helsinki. Regarding the access to the cohorts and databases, approvals were obtained from the involved institutions. All data were deidentified.

Consent for publication

Not applicable.

Availability of data and materials

Data from the SSC, Generation Scotland and Imagen are available upon request to the promoter of the respective study. Data from the MSSNG database can be accessed through Google Genomics (for access, see http://mss.ng/researchers).

Access to the genetic and phenotypic data for SPARK was obtained from the SFARI Base. Approved researchers may obtain this population dataset (further details available at https://www.sfari.org/resource/spark/) by applying at https://base. sfari.org.

Access to the individual-level genotype and phenotype data for UKBB are available by application (http://www.ukbiobank.ac.uk/)

Competing interests

All the authors declare they have no competing interests.

Funding

The analysis of the Pakistani family was supported by a CIHR operating grant to Lan Xiong, and Guy A. Rouleau. Bioinformatic analyses were supported by the Canadian Institutes of Health Research (CIHR). ZS is funded by the Transforming Autism Care Consortium, a thematic network supported by the Fonds de Recherche Québec-Santé. CL is funded by the Vanier Graduate Scholarship. JPR is funded by a Canadian Institutes of Health Research (CIHR) Frederick Banting & Charles Best Canada Graduate Scholarship (FRN159279). CH was supported by an MRC Human Genetics Unit programme grant 'Quantitative traits in health and disease' (U. MC_UU_00007/10). GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences and is supported by a Canadian Institutes of Health Research Foundation Scheme Grant #332971. BC has received a postdoctoral fellowship from the Healthy Brains for Healthy Lives project (Talent program) and a grant from Fondation Bettencourt-Schueller.

Authors' contributions

ZS, GH, SJ and BC designed the study. SJ, PAD, LX, GAR and BC obtained the funding and supervised the study. ZS, GH, QH, AMJ, SJ, and LX collected the data. CH curated and contributed the Generation Scotland data. HL, AB, SD, AH, and GH curated and contributed to the IMAGEN data. ZS, GH, QH, AMJ, MAB, ED, CL, JPR, ADL, DS and BC analyzed the data. ZS, GH, MAL, PAD, GAR and BC interpreted the data. ZS, GH, QH, PAD, GAR and BC drafted the report. All authors contributed to and have approved the final manuscript.

Acknowledgements

We would like to thank Brohi Qasim who recruited the Pakistani family and Lydia Werhli, who contributed to the analyses about the Pakistani family during her IFMSA Exchange program.

This study makes use of data generated by the DECIPHER community. A full list of centres who contributed to the generation of the data is available from http://decipher.sanger.ac.uk and via email from decipher@sanger.ac.uk. Funding for the project was provided by

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z).

IMAGEN received support from the following sources: the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behaviour in normal brain function and psychopathology) (LSHM-CT- 2007-037286), the Horizon 2020 funded ERC Advanced Grant 'STRATIFY' (Brain network based stratification of reinforcement-related disorders) (695313), Human Brain Project (HBP SGA 2, 785907, and HBP SGA 3, 945539), the Medical Research Council Grant 'c-VEDA' (Consortium on Vulnerability to Externalizing Disorders and Addictions) (MR/N000390/1), the National Institute of Health (NIH) (R01DA049238, A decentralized macro and micro gene-by-environment interaction analysis of substance use behavior and its brain biomarkers), the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, the Bundesministeriumfür Bildung und Forschung (BMBF grants 01GS08152; 01EV0711; Forschungsnetz AERIAL 01EE1406A, 01EE1406B; Forschungsnetz IMAC-Mind 01GL1745B), the Deutsche Forschungsgemeinschaft (DFG grants SM 80/7-2, SFB 940, TRR 265, NE 1383/14-1), the Medical Research Foundation and Medical Research Council (grants MR/R00465X/1 and MR/S020306/1), the National Institutes of Health (NIH) funded ENIGMA (grants 5U54EB020403-05 and 1R56AG058854-01). Further support was provided by grants from: – the ANR (ANR-12-SAMA-0004, AAPG2019 – GeBra), the Eranet Neuron (AF12-NEUR0008-01 – WM2NA; and ANR-18-NEUR00002-01 – ADORe), the Fondation de France (00081242), the Fondation pour la Recherche Médicale (DPA20140629802), the Mission Interministérielle de Lutte-contre-les-Drogues-et-les-Conduites-Addictives (MILDECA), the Assistance-Publique-Hôpitaux-de-Paris and INSERM (interface grant), Paris Sud University IDEX 2012, the Fondation de l'Avenir (grant AP-RM-17-013), the Fédération pour la Recherche sur le Cerveau; the National Institutes of Health, Science Foundation Ireland (16/ERCD/3797), U.S.A. (Axon, Testosterone and Mental Health during Adolescence; RO1 MH085772-01A1), and by NIH Consortium grant U54 EB020403, supported by a cross-NIH alliance that funds Big Data to Knowledge Centres of Excellence.

References

1. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016;57:585–95.
2. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017;49:978.
3. Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu T, Baker C, et al. A Copy Number Variation Morbidity Map of Developmental Delay. Nat Genet. 2011;43:838–46.
4. Zuko A, Kleijer KTE, Oguro-Ando A, Kas MJH, van Daalen E, van der Zwaag B, et al. Contactins in the neurobiology of autism. Eur J Pharmacol. 2013;719:63–74.
5. Chatterjee M, Schild D, Teunissen CE. Contactins in the central nervous system: role in health and disease. Neural Regen Res. 2019;14:206–16.
6. Burbach JPH, van der Zwaag B. Contact in the genetics of autism and schizophrenia. Trends Neurosci. 2009;32:69–72.
7. Kumar RA, Christian SL. Genetics of autism spectrum disorders. Curr Neurol Neurosci Rep. 2009;9:188–97.
8. van Daalen E, Kemner C, Verbeek NE, van der Zwaag B, Dijkhuizen T, Rump P, et al. Social responsiveness scale-aided analysis of the clinical impact of copy number variations in autism. Neurogenetics. 2011;12:315–23.
9. Fernandez T, Morgan T, Davis N, Klin A, Morris A, Farhi A, et al. Disruption of contactin 4 (CNTN4) results in developmental delay and other features of 3p deletion syndrome. Am J Hum Genet. 2004;74:1286–93.
10. Roohi J, Montagna C, Tegay DH, Palmer LE, DeVincent C, Pomeroy JC, et al. Disruption of contactin 4 in three subjects with autism spectrum disorder. J Med Genet. 2009;46:176–82.
11. Hussman JP, Chung R-H, Griswold AJ, Jaworski JM, Salyakina D, Ma D, et al. A noise-reduction GWAS analysis implicates altered regulation of neurite outgrowth and guidance in autism. Mol Autism. 2011;2:1.
12. Mercati O, Huguet G, Danckaert A, André-Leroux G, Maruani A, Bellinzoni M, et al. CNTN6 mutations are risk factors for abnormal auditory sensory perception in autism spectrum disorders. Mol Psychiatry. 2017;22:625–33.

13. Ashrafi S, Betley JN, Comer JD, Brenner-Morton S, Bar V, Shimoda Y, et al. Neuronal Ig/Caspr Recognition Promotes the Formation of Axoaxonic Synapses in Mouse Spinal Cord. Neuron. 2014;81:120–9.
14. Peng Y-R, Tran NM, Krishnaswamy A, Kostadinov D, Martersteck EM, Sanes JR. Satb1 regulates Contactin 5 to pattern dendrites of a mammalian retinal ganglion cell. Neuron. 2017;95:869–83.e6.
15. Karayannis T, Au E, Patel JC, Kruglikov I, Markx S, Delorme R, et al. Cntnap4 differentially contributes to GABAergic and dopaminergic synaptic transmission. Nature. 2014;511:236–40.
16. 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;87:1215–33.
17. Trost B, Walker S, Wang Z, Thiruvahindrapuram B, MacDonald JR, Sung WWL, et al. A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. Am J Hum Genet. 2018;102:142–55.
18. Feliciano P, Daniels AM, Green Snyder L, Beaumont A, Camba A, Esler A, et al. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. Neuron. 2018;97:488–93.
19. Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Büchel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry. 2010;15:1128–39.
20. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;42:689–700.
21. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLOS Medicine Public Library of Science. 2015;12:e1001779.
22. Douard E, Zeribi A, Schramm C, Tamer P, Loum MA, Nowak S, et al. Effect Sizes of Deletions and Duplications on Autism Risk Across the Genome. AJP. American Psychiatric Publishing; 2020.;appi.ajp.2020.19080834.
23. Huguet G, Schramm C, Douard E, Jiang L, Labbe A, Tihy F, et al. Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry. 2018;75:447–57.
24. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw. 2015;67:1–48.
25. Core Team R. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. Available.; 2013.
26. Pinto D, Delaby E, Merico D, Barbosa M, Merikangas A, Klei L, et al. Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. The American Journal of Human Genetics. 2014;94:677–94.
27. Rigau M, Juan D, Valencia A, Rico D. Intronic CNVs and gene expression variation in human populations. PLOS Genetics Public Library of Science. 2019;15:e1007902.
28. Robinson EB, Samocha KE, Kosmicki JA, McGrath L, Neale BM, Perlis RH, et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. PNAS National Academy of Sciences. 2014;111:15161–5.

Supplementary Files

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

- SupplementaryFiles.docx
- SupplementaryTable.xlsx