Université de Montréal

Differential effect of deletions and duplications on general intelligence and social responsiveness

Par

Petra Tamer

Département de neurosciences
Faculté de médecine

Mémoire présenté en vue de l'obtention du grade de Maîtrise ès Sciences (Msc.)

en neurosciences

Novembre 2019

Université de Montréal

Unité académique : département de neurosciences, Faculté de médecine

Ce mémoire intitulé:

Differential effect of deletions and duplications on general intelligence and social responsiveness

Présenté par

Petra Tamer

A été évalué par un jury composé des personnes suivantes

Alexis Beauchamp-Châtel
Président-rapporteur

Sébastien Jacquemont
Directeur de recherche

Sherif Karama
Membre du jury

RÉSUMÉ

Les délétions et les duplications délétères (Variations de nombre de copies, CNV) sont identifiés dans environ 11% des individus référés dans des cliniques du neurodéveloppement pédiatrique. Certains CNVs récurrents ont été formellement associés avec des troubles du neurodéveloppement, mais la majorité des CNVs sont non-récurrents et donc trop rares pour être évalués par des études d'association. Dans cette optique, nous avons récemment développé une nouvelle approche pour estimer l'effet des CNVs non-documentés sur le quotient intellectuel non-verbal (QINV) et nous visons étendre cette approche pour l'appliquer sur une mesure de traits autistiques.

Nous avons identifié les CNVs dans deux cohortes d'autisme du Simons Simplex Collection (SSC) et du MSSNG, dans leurs apparentés de premier-degré, dans une cohorte du neurodéveloppement et dans une population générale. Des modèles statistiques intégrant les scores des gènes inclus dans les CNVs ont été utilisés pour expliquer leur effet sur l'intelligence générale et sur la réciprocité sociale.

Les délétions et les duplications diminuent le QINV et l'effet des duplications est 3 fois inférieur à celui des délétions. L'effet différentiel est aussi observé pour la réciprocité sociale avec un ratio d'altération de 2:1 pour les délétions et les duplications et cet effet est principalement expliqué par le QINV. Les estimés de notre modèle pour l'intelligence générale et la réciprocité sociale concordent bien avec des observations déjà publiés.

Nos modèles entraînés sur des CNVs couvrant >4,500 gènes suggèrent que l'effet des CNVs sur la cognition et la réciprocité sociale est dû à leurs propriétés polygéniques. Ces modèles pourront aider dans l'interprétation des CNVs en clinique.

Mots-clés : Variations de nombre de copies, autisme, intelligence générale, QI, réciprocité sociale, SRS, modèle statistique, score génétique.

ABSTRACT

Deleterious deletions and duplications (copy number variations, CNVs) are identified in up to 11% of individuals referred to neurodevelopmental pediatric clinics. However, only few recurrent CNVs have been formally associated with neurodevelopmental disorders because the majority are too rare to perform individual association studies. We recently developed a new framework to estimate the effect size of undocumented CNVs on non-verbal intelligence quotient (NVIQ) and sought to extend this approach to another score measuring autistic traits.

We identified CNVs in an autism sample from the Simons Simplex Collection (SSC) and MSSNG, in their first-degree relatives, in a neurodevelopmental cohort and in individuals from an unselected population. Statistical models integrating scores of the genes encompassed in the CNVs were used to explain their effect on general intelligence and on social responsiveness.

Deletions and duplications decreased NVIQ and the effect of duplications was three-fold smaller than deletions. There was also a differential effect on social responsiveness: the ratio of the impairment conferred by deletions and duplications was 2:1 and this effect was mainly driven by NVIQ. Models estimates for general intelligence and social responsiveness were consistent with previously published observations.

Our models, trained on CNVs encompassing >4,500 genes, suggest highly polygenic properties of CNVs with respect to cognition and social responsiveness. These models will help interpreting CNVs identified in the clinic.

Keywords: Copy-number variants, autism, general intelligence, IQ, Social responsiveness, SRS, statistical models, genetic scores.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADDM: Autism and Developmental Disabilities Monitoring

ADHD: Attention-deficit/hyperactivity disorder

ASD: Autism Spectrum Disorder

BRCA1: Breast cancer 1

CGH: Comparative genomic hybridization

CHRNA7: Cholinergic receptor nicotinic alpha 7

CI: Confidence interval

CNV: Copy number variants

DAS: Differential ability scales

DSM-IV-TR: Diagnostic and Statistical manual of Mental Disorders - 4th version

DSM-5: Diagnostic and Statistical manual of Mental Disorders - 5th version

FISH: Fluorescent in situ hybridization

ICC: Intraclass correlation coefficient

ID: Intellectual disability

IQ: Intelligence quotient

Kb: Kilobase

LEITER-R: Leiter International Performance Scale – Revised

MAF: minor allele frequency

MAPK3: Mitogen-activated protein kinase 3

Mb: Megabase

MSEL: Mullen Scales of Early Learning

NLGN: Neuroligins

NDD : Neurodevelopmental disorder

NRXN: Neurexin

NVIQ: Non-verbal intelligence quotient

OR: Odds ratio

PIQ: Performance Intelligence quotient

pLI: probability of loss-of-function intolerance

pLI DEL/DUP: Sum of pLI scores by individuals for deletions or for duplications

RhoA: Ras homolog family member A

SD: Standard deviation

SHANK: SH3 and multiple ankyrin repeat domains

SNP: Single nucleotide polymorphism

SNV: Single nucleotide variant

SRS: Social Responsiveness Scale

SSC: Simons Simplex Collection

TAOK2: TAO kinase 2

WASI: Wechsler Abbreviated Scale of Intelligence

WISC: Wechsler Intelligence Scale for Children

WPPSI: Wechsler Preschool and Primary Scale of Intelligence

DEDICATION

To all the children with neurodevelopmental disorders and their families

ACKNOWLEDGMENTS

I would like to give my most sincere thanks

To Dr. Jacquemont for giving me the opportunity to do my master's in his multidisciplinary laboratory and for his continuous support. His dedication to research is definitely inspiring.

To all the students and employees in the laboratory. Every single person in the laboratory has somehow contributed to my progress and helped me through the process especially Miss Elise Douard, a PhD candidate, who supervised me since I started as a research intern. She invested a lot of her time and effort in training me. She was always there when I needed help and always answered my questions.

To Dr. Huguet, Dr. Schramm and Dr. Absa for their supervision and their availability.

To the research assistants Dr. Martin and Miss Nadine Younis who were managing the cohort of St-Justine and to all the students of the laboratory of Dr. Sarah Lippé who guided me during the neuropsychological assessments.

I would like to extend my acknowledgement,

To the Canadian Institutes of Health Research (CIHR) and to the department of neurosciences for the scholarships that supported me financially.

And lastly to my family and friends, my source of motivation and encouragement.

CHAPTER 1: INTRODUCTION

INTRODUCTION

Autism spectrum disorder (ASD) is one of the many neurodevelopmental disorders described in the Diagnostic and Statistical manual of Mental Disorders (5th version) (DSM-5) (1). Constant changes and refinements have been made to the diagnosis of autism since the documented observations of Kanner in 1943 up until the publication of the DSM-5 in 2013 (1,2). This disorder is now widely accepted as complex, pervasive, heterogeneous with multiple aetiologies, sub-types and developmental trajectories (3). Nowadays, the diagnosis is based on the observation of the two domains of core symptoms: Deficits in social communication and interaction and stereotyped, restricted and repetitive behaviors or interests (1) (Figure 1).

The prevalence of this disorder is in constant increase (Figure 2, Figure 4). According to a surveillance study done by the Autism and Developmental Disabilities Monitoring network (ADDM) in 2014 among children aged 8 years in 11 sites in the United States, the estimated prevalence of autism is 16.8 per 1000 (one in 59) (4). This prevalence is higher than previously reported estimates from the ADDM network and this is not solely due to the extension of the diagnostic criteria by the DSM-5 since there is an 86% overlap between the DSM-IV-TR (2002) and DSM-5 (2013) case counts (1,4,5).

The causes and mechanisms underlying ASD are still not fully known but several epidemiological studies have firmly established a genetic component underlying ASD with a heritability ranging from 50-90% (6–8) as well as a complex interaction between genetic and environmental factors (9,10). Furthermore the implementation of advanced technology for chromosomal microarray-based analysis in clinic has rapidly expanded the number of genes associated with ASD (e.g. SFARI genes) through the identification of deleterious Copy-Number Variants (CNVs) (11,12) (Figure 4). The unveiling of the genetic contribution to ASD could be an approach that would ultimately lead to developing specific molecular diagnostics and targeted therapeutics.

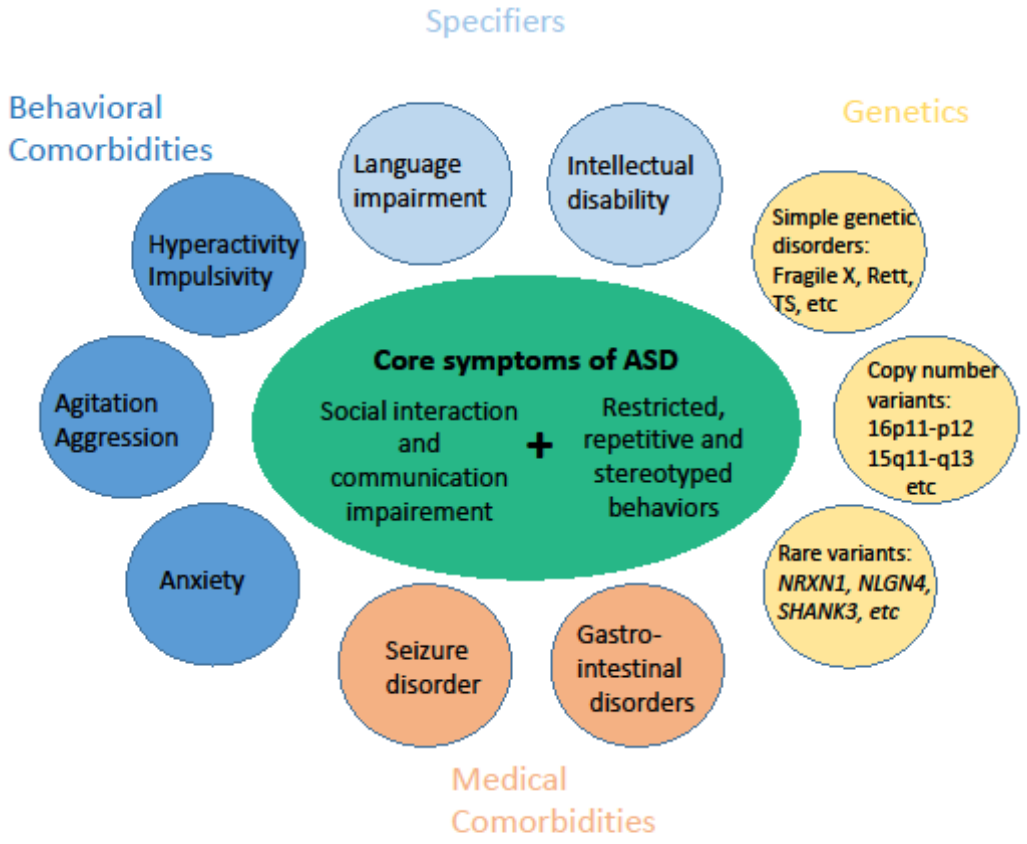


Figure 1: Representation of the phenotypical heterogeneity of ASD based on DSM-5 criteria for ASD. The core symptoms of ASD are represented in the center and represent the common features required to receive a diagnosis. Comorbidities spanning behavior, cognition and genetic disorders are represented around the periphery of the figure. Adapted from Veenstra-VanderWeele & Blakely (2012) (13).

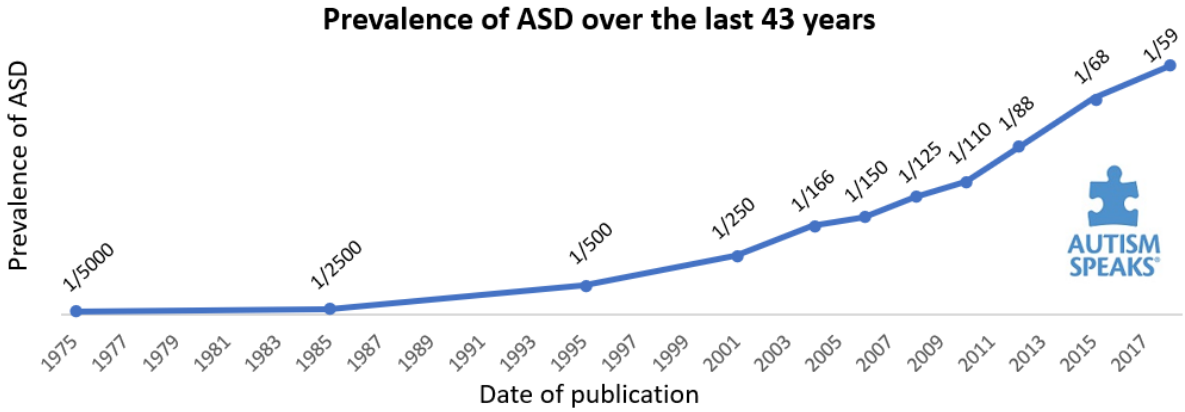


Figure 2: Prevalence of ASD reported by the Center for Disease Control and Prevention since 1975 (4)**.**

Phenotypical heterogeneity of autism:

The heterogeneity of the clinical presentation is a hallmark of autism (3). Different specifiers and many psychiatric and medical comorbidities are associated with this disorder.

Specifiers of autism

The DSM-5 focuses on a dimensional assessment to examine the core and associated features of ASD which led to the inclusion of "specifiers". The specifiers are dimensions provided to describe the heterogeneity of the presentation of ASD and they indicate the presence of intellectual and/or language impairment as well as the severity level of the core ASD symptoms (14). Intellectual disability (ID) is a developmental disorder characterized by intellectual and adaptive functioning deficits (1). Studies published so far have reported highly variable rates of ID prevalence in ASD, ranging from 16.7% to 84% (15,16). However, some individuals with autism have above-average intelligence quotient (IQ) and high levels of academic and occupational functioning (17). Language impairment is also specifier of ASD with up to 76% of occurrence in children with autism (18–20). Thus, some children with ASD fail to acquire spoken language skills beyond a basic or minimal level, which may range from no spoken words to fewer than 20–30 words (21,22); about 30% of children with autism fall into this group (23). Within the group of children who are verbal, some have a notable language deficit, including difficulties with the understanding and use of grammar (24,25). Furthermore, the specifiers of autism include the recording of any known genetic or medical disorder and other co-occurring neurodevelopmental, mental, or behavioral disorder (Figure 3).

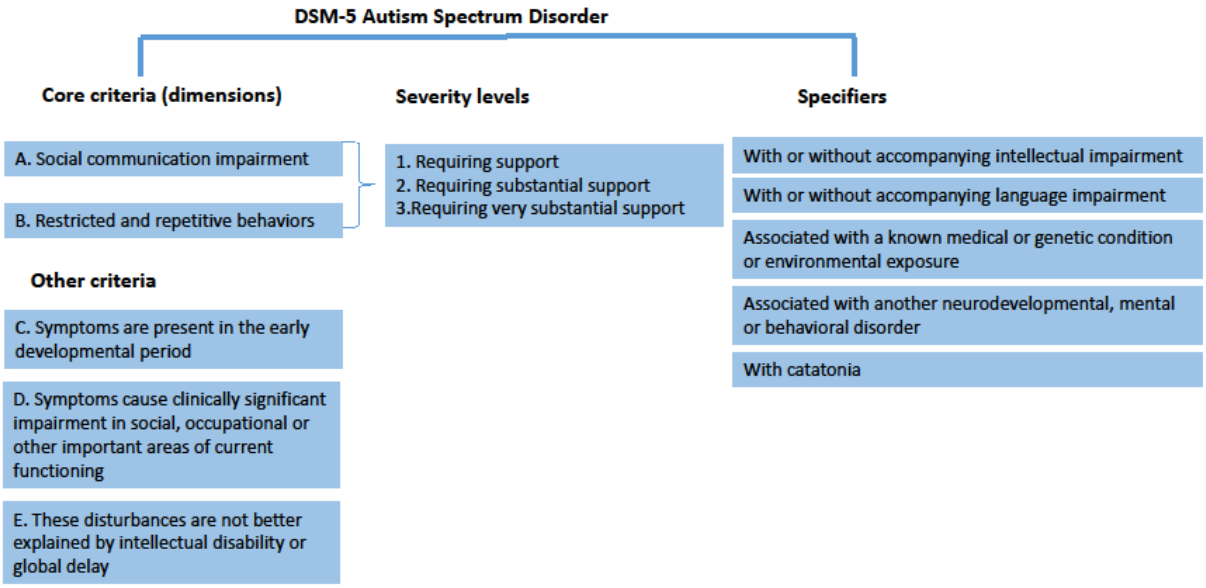


Figure 3: Representation of the phenotypical heterogeneity of ASD based on DSM-5 criteria for ASD. Adapted from Ousley& Cermak (2013) (26).

Psychiatric comorbidities

Neurodevelopmental and psychiatric comorbidities occur in up to 70% of children with autism and the most common are social anxiety disorder, oppositional defiant disorder and attention-deficit/hyperactivity disorder (ADHD) (27). The social anxiety disorder, also known as social phobia, occurs during specific social situations and leads to an avoidance reaction and constant fear that substantially affects social life, academic performance and professional success (28).

This comorbidity is seen in approximately 29.2% of individuals with autism (29). The oppositional defiant disorder is a behavioral disorder that has a prevalence of 28.1% among autistic individuals (29). This disorder is characterized by uncooperative, defiant, negativistic and irritable behaviors toward parents, peers, teachers and other authority figures and often interferes with learning, school adjustment and with the child's relationships with others (28). ADHD is outlined by symptoms of inattention, hyperactivity and impulsivity and it frequently co-occurs with ASD (30) with a prevalence ranging from 30% to 80% (31). Symptoms such as poor social skills, emotional dysregulation, and oppositional behavior were found in both diagnoses, but these may be qualitatively distinct.

Medical comorbidities

Some medical conditions are more frequently observed in individuals with autism compared to the general population. These conditions include epilepsy, gastrointestinal disorder, immune system abnormalities, sleep disorders and motor disorders (13).

A meta-analysis study shows that the prevalence of epilepsy among autistic individuals is 8% compared to 2-3% in the general population (32). This prevalence is of 21.4% among patients who have simultaneously autism and ID. Epilepsy is also more frequent among autistic children with poor verbal abilities, among those suffering from a neurological impairment such as cerebral palsy and among females (33,34). Individuals who have autism and epilepsy tend to have more severe impairments affecting their adaptive and social functioning domains and their fine and gross motor skills which leads to more challenging behaviors (35,36). This severity can be the result of either the physiological complications caused by the recurrent seizures or the manifestations of the genetic mutation underlying these two disorders.

A systematic review reported that the prevalence of gastrointestinal disorders among individuals with autism ranges from 9 to 91% (37). The 11 studies included in this review, except for the study of Black et *al*. (38), agree on the fact that this prevalence is higher in the autistic population compared to the general population with symptoms including constipation, feeding issues/food selectivity and diarrhea (39). Patients with autism who have gastro-intestinal problems are at a higher risk of anxiety, sensory over-responsivity, irritability, social withdrawal and language regression (40–42).

A case-control study found that autoimmune diseases were diagnosed significantly more often among children with ASD than controls and psoriasis was the most frequently diagnosed autoimmune condition; it occurred over twice as often in cases than in controls (43). A different study reported that allergic manifestations were 5 times as prevalent in children with ASD than in controls (52% vs. 10%) (44). Other studies of biological markers of immune function in individuals with ASD have found neuroinflammation in brain tissues (45,46) and imbalances in cytokine/chemokine levels and other abnormalities (47).

Children with ASD are more likely to have sleep problems including dyssomnias and parasomnias (48). Dyssomnias include problems such as sleep onset delay, night awakenings (49), early morning awakening (50) and periodic limb movements (51). Parasomnias consist of problems such as sleep talking, sleep walking, sleep terrors and nightmares (52). These sleep problems intensify the symptoms of autism as some research identified that fewer hours of sleep per night predicted overall autism scores, social skills deficits and stereotypic behaviors (53).

Movement disorders have also been frequently identified in individuals with ASD, with ataxia (54) reported in a number of studies, as well as akinesia, dyskinesia, bradykinesia and catatonic-like symptoms (55) among others, with cerebellum and basal ganglia dysfunction being implicated (56).

Genetic heterogeneity of autism

The genetic architecture of autism is very complex. It involves a lot of variants with numerous characteristics such as size, frequency and inheritance. First, the size depends on the number of nucleotides altered which divides the variants into two classes: single nucleotide variants (SNV) and structural variants including the CNVs (57). Second, the frequency divides the genetic variants into four categories: A very common variant is a variant with the minor allele frequency (MAF) between 5 and 50% , a less common variant has a MAF between 1 and 5%, a rare variant has a MAF of less than 1% but still polymorphic in one or more major human populations, and a private variant is restricted to probands (the individual through whom a family with a genetic disorder is ascertained) and their immediate relatives (58). Third, the patterns of inheritance include autosomal recessive, autosomal dominant and X-linked through which the variant is transmitted from one parent to the descendant, whereas a *de novo* variant is only seen in the proband (59). These different characteristics determine the severity of the variants which leads to their clinical classification into benign, pathogenic or variant of uncertain significance defined by the American College of Medical Genetics (60).

Copy-number variations

In 2004, two landmark studies (61,62) have demonstrated that submicroscopic variations (<500 Kilobases (kb) in size - 1 Kb =1000 base pairs) are widespread in the human genome. These variations are known as CNVs and are defined as a genomic deletion or duplication of over 1000 base pairs. On average, each individual has more than 1000 CNVs in the genome accounting for 4 million base pairs of genomic difference and contributing to 0.1% of the genetic variation at the structural level (63).

Technological and methodological advances in genetics and genomics such as the Fluorescence in situ hybridization (FISH) and comparative genomic hybridization (CGH) have permitted the identification of pathogenic CNVs in 11% of individuals with autism but only 12 recurrent CNVs have been formally associated to ASD (Table 1) (64). Recurrent CNVs share a common size, show clustering of fixed breakpoints, and recur in multiple individuals (65). They are mainly caused by a non-allelic homologous recombination between two low-copy repeats. Low-copy repeats are "hotspots" meaning that they are unstable regions of the genome and are subject to high rates of structural mutation. On the other hand, non-recurrent CNVs have distinct breakpoints and different sizes and consequently are less common. The major mechanism underlying these CNVs is the non-homologous end-joining (65).

- *16p11.2 deletion*

Since the frequency of recurrent CNVs allows the collection of large samples of carriers, many of them have been studied and characterized such as the 16p11.2 deletion.

The 16p11.2 deletion is a recurrent CNV between breakpoints 4 and 5 on the 16th chromosome (29.5–30.1 Megabases - 1 Megabase: 1 million base pairs) that encompasses 29 genes (66). It is associated with a broad range of neurodevelopmental and neuropsychiatric diagnoses including developmental delay, ID and autism (66). The overrepresentation of this deletion has been demonstrated in ASD cohorts with an odds ratio (OR) of 10 (64). This would translate into a risk for ASD of 15% based on the ASD population prevalence of 1.5%.

Many studies have attempted to quantify the effect size of this deletion on cognition and other phenotypes notably a study by Moreno-De-Luca et *al.* conducted on 56 probands with the *de novo* CNV and their non-carrier parents and siblings (66). When comparing the probands to their family members, they scored 1.7 standard-deviation (SD) lower in cognitive abilities, 2.2 SD in social behavior and 1.3 SD in neuromotor performance. Another study assessing phonology found that probands have a decrease of 2 SD which reflects their language impairments (67). Of note, these phenotypes are not specific to the deletion 16p11.2 but they are observed in other recurrent CNVs such as the duplication 1q21.1 (68).

Several studies have attempted to identify causative driver genes (altered by the CNV) of these above-mentioned neurodevelopmental phenotypes. Manipulations of mouse models found that *TAOK2* heterozygous and knockout mice have gene dosage-dependent impairments in cognition, anxiety and social interaction (69). They also have dosage-dependent abnormalities in brain size and neural connectivity in multiple regions, deficits in cortical layering, dendrite and synapse formation, and reduced excitatory neurotransmission (69). The purported mechanism of synaptic development impairment could be that the loss of TAOK2 causes a reduction in the activation of RhoA implicated in the reorganization of the actin cytoskeleton and the regulation of the cell shape, attachment and motility (69,70) .

Other studies suggest that *MAPK3* being the most topologically important hub in protein-protein interaction networks could be the driver gene in this CNV (71).

However, studies of individual genes do not always replicate across publications. This suggests that there is not one clear major candidate, but multiple genes within the 16p11.2 are responsible for the observed phenotype by additive effects or interactions (72).

Table 1: Recurrent CNVs formally associated to ASD.

	Odds ratio in autism [95%CI]	
Regions	**Deletions**	**Duplications**
1q21.1	3 [1-9]	5 [3-10]
3q29	19 [5-81]	-
5q35	∞	-
7q11.23	32 [2-517]	32 [10-112]
15q13.3	15 [5-42]	-
16p11.2 (proximal)	14 [8-25]	14 [6-19]
17p11.2	-	32 [2-517]
17q12	97 [10-933]	-
22q11.2	32 [9-112]	-

95% CI: 95% Confidence interval. Table adapted from Sanders et *al* (2019) (64).

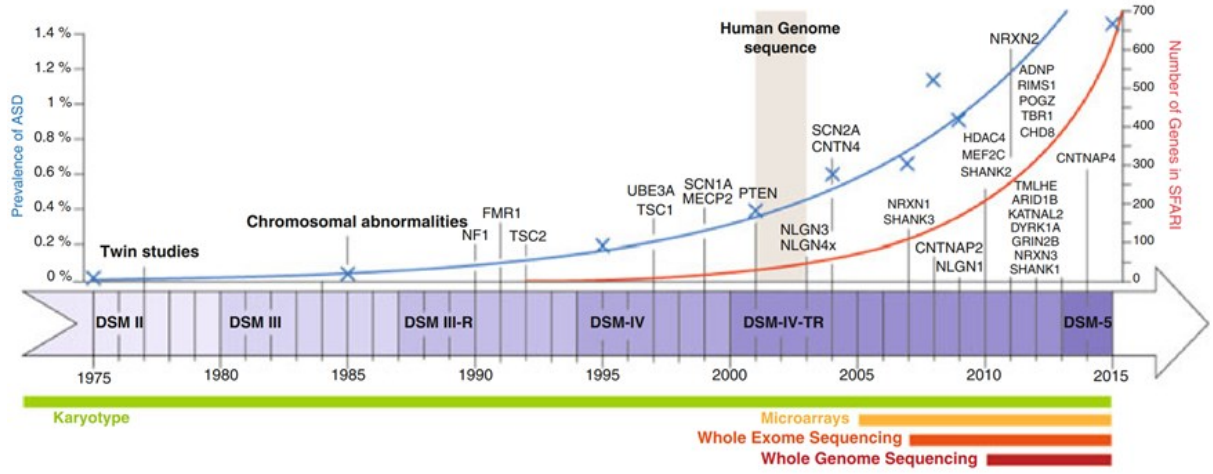


Figure 4: The history of the genetics of autism from 1975 to 2015.

The increase in the identified genes associated with ASD (SFARI) (11) is represented together with the prevalence of ASD reported by the Center for Disease Control and Prevention (4), the different versions of the Diagnostic Statistical Manual (from DSM II to DSM 5.0) and the advance in genetics technology. From Huguet et *al.* (2016) (73).

Single nucleotide variations

The identification of CNVs through microarrays and the implementation of new technologies such as the whole-exome and the whole-genome sequencing have led to a large and rapidly growing number of genes associated with autism susceptibility. It is now estimated that 400-1000 genes are associated with autism (74). These susceptibility genes contain thousands of variants in particular SNVs which are a substitution of one nucleotide. A SNV mutation can be synonymous or non-synonymous: a synonymous SNV does not alter the protein sequence while a non-synonymous SNV changes the protein either by changing the sequence of amino acids (missense mutation) or by changing an original amino acid to a stop codon which leads the protein coded by the gene to terminate prematurely (nonsense mutation).

Synaptic genes are highly penetrant for autism and this is based on recurrent findings of rare, *de novo*, damaging variants of these genes in probands (75). Neuroligins (*NLGN*), neurexin families (*NRXN*) and SH3 and multiple ankyrin repeat domains (*SHANK)* harbour some of the most consistently reported genetic abnormalities that are associated with autism (75). *NRXN*s and *NLGN*s are trans-synaptic cell-adhesion molecules that mediate essential signalling between presynaptic (*NRXN*) and postsynaptic (*NLGN*) specializations (76). Seven point mutations (including SNVs) in *NRXN1*, five missense mutations in *NLGN4* and one missense mutation in *NLGN3* were detected in patients with autism (76). In addition to the NRXN–NLGN complex, mutations in the gene encoding SHANK3- a molecular scaffolding protein in the postsynaptic density of excitatory synapses that binds indirectly to NLGNs- may also occur frequently in autism (77). An astounding 18 point mutations (among which 8 non-synonymous mutations) were detected in the *SHANK3* gene in patients with autism, in addition to several cases containing CNVs that cover this gene (e.g. 22q deletion syndrome) (75,78,79). Mutations in synaptic genes are not specific to ASD but are also found in other neuropsychiatric disorders, such as schizophrenia (80). Interestingly, neuropsychiatric conditions share common features such as cognitive dysfunction, limited emotional expression and lack of social reciprocity

suggesting that synaptic dysfunction is the common pathway of these major, chronic neuropsychiatric illnesses (81).

Although rare *de novo* mutations in genes expressed in the brain are identified in approximately 5–14% of individuals with idiopathic autism (82,83), they only contribute to autism liability by 2.6% (7). Conversely, common variants which are shared with more than 5% of the human population contribute to this liability by almost 50% (7). This contribution is therefore important, but unfortunately the causative SNVs still remain unknown since they are numerous (>1000) and each is associated with a low risk. So far, the largest genome-wide association studies performed on <5000 families with autism were underpowered to identify a single SNV with genome-wide significance (73). The recruitment of larger cohorts of patients is required to identify these common variants that explain most of the genetic variance in autism.

Overall, the genetic architecture in autism varies substantially from a single penetrant mutation being enough to cause autism (*NRXN1*, *SHANK3*…) to an accumulation of over one thousand low-risk alleles (common variants) (3,84).

Prediction models

As mentioned before, CNVs are identified in 11% of autistic individuals but only 12 recurrent CNVs have been formally associated to autism (Table 1) (64) and since the majority of these CNVs are rare or even private, the possibility of studying them through association studies is ruled out. Therefore, their effect on neurodevelopment and cognition is poorly understood.

In line with this, the research team of Pr. Jacquemont has recently published a new framework to estimate the effect of CNVs on general intelligence as measured by NVIQ (85) because 1) ID is one of the specifiers of autism and 2) the CNVs associated to autism have an impact on cognition. This framework corresponds to a statistical model trained on two cohorts of the general population where the majority of CNVs are benign. It consists of linear regressions including functional annotations of genes encompassed by the CNVs used to identify features that explain their association with IQ. Among the 10 functional annotation of genes, a stepwise

procedure identified that the probability of loss-of-function intolerance score (pLI) (86), which is a haploinsufficiency score, is what best explains the association of any deletion CNV with Non-verbal IQ (NVIQ). This score separates the genes tolerant to a loss-of-function (pLI≤ 0.1) from the genes that are intolerant to a loss-of-function (pLI≥0.9). Note that a pLI≥0.9 points to the genes that would have a significant impact on survival and reproduction (fitness) if altered by a heterozygous loss-of-function. However, not all genes associated with diseases have high scores of pLI (e.g. *BRCA1* is associated to breast cancer but does not affect fertility and survival). This model was validated with a concordance analysis using published measures of 15 recurrent pathogenic CNVs. The estimates of the model were 75% (95% CI, 39-91%) concordant with the loss of IQ measured in published case-control studies (Figure 5). This means that the effect of deleterious CNVs can be precisely estimated using a model trained on mostly benign CNVs (with a low pLI score). In addition, the application of this model in autistic cohorts shows similar results to what was found in the general population (Douard et *al*., manuscript under review).

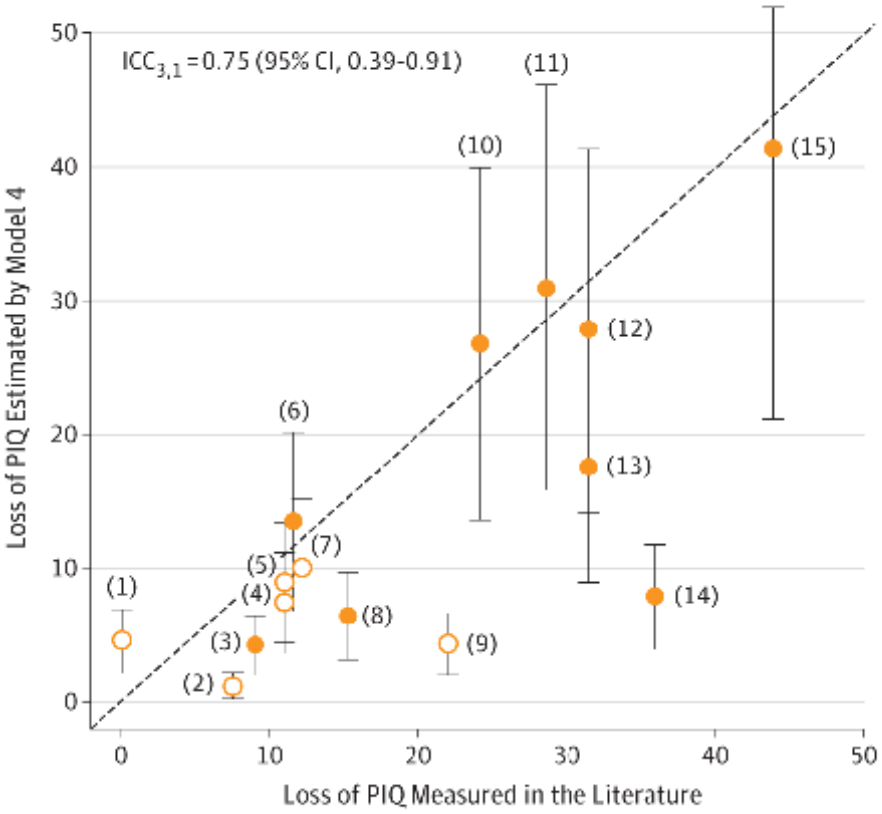


Figure 5: Concordance between loss of PIQ estimated by the first model (y-axis) and loss of PIQ measured by previously published studies (x-axis) for 15 recurrent CNVs. Each point corresponds to a known recurrent CNV: (1) 17p12_(HNPP), (2) 16p12.1, (3) 15q11.2, (4) 16p13.11, (5) 1q21.1 TAR, (6) 17q12, (7) 16p11.2 Distal (SH2B1), (8) 1q21.1 Distal (Class I), (9) 15q13.3 (BP4-BP5), (10) 16p11.2 proximal (BP4-BP5), (11) 22q11.2, (12) 7q11.23 (William-Beuren), (13) 3q29 (DLG1), (14) 8p23.1, and (15) 17p11.2 (Smith-Magenis). The diagonal dashed line represents exact concordance. When loss of IQ was not directly measured in a previous study, we derived the loss of IQ from the published OR measuring the enrichment of a CNV in the neurodevelopmental clinic (open circles). From Huguet et *al.* (2018) (85).

In the neurodevelopment clinic, CNVs are reported daily but the effect size of 90% of these CNVs on autism risk remains undocumented. And it is still unknown if the enrichment of these CNVs is related to the core symptoms or to the specifiers and comorbidities of autism.

We hypothesize that the effect of CNVs on autism risk and comorbidities follows a general principal and can be estimated using statistical models.

The general aim of this study is to quantify the effect of CNVs on general intelligence and on social responsiveness which is a domain impaired in autism.

The specific aims:

1) Replicate the model assessing the effect of CNVs on general intelligence in a cohort of heterogeneous neurodevelopmental disorders.

2) Validate the previous model through correlation and concordance analysis

3) Test the model on a measure associated to a core symptom of autism (social responsiveness) in different cohorts: cohorts of autism, an unselected population and in a cohort of heterogeneous neurodevelopmental disorders.

CHAPTER 2: METHODOLOGY

METHODOLOGY

Cohorts

Autism cohorts

Two autism cohorts (Figure 6) were used: The Simons Simplex Collection (SSC) (87) , a cohort of 2,569 simplex families, including 2,074 quads (one proband, unaffected parents, one unaffected sibling) and 495 trios (one autistic proband and unaffected parents); and the MSSNG database (88), which was used as an independent replication cohort and includes 845 probands.

Neurodevelopment cohort of Sainte-Justine

This cohort is designed by the project of Brain Canada and it includes 106 families with at least one child carrying a large or a recurrent CNV and diagnosed with a neurodevelopmental disorder (autism, ADHD, ID…). All 282 participants were assessed with an IQ test (Table 2) and have filled the SRS and other questionnaires that are not used in this study (e.g. Children behavioral checklist, Conners Comprehensive Behavior Rating Scale…).

Unselected cohort

One community-based cohort was used: the IMAGEN (89) (Figure 6) which includes 1802 adolescents.

Genetic data - CNV detection, annotation and filtering

Genotyping and whole genome sequencing

- *Genotyping data*

CNV detections and standard filtering strategies were previously published (85). CNV calling was performed using the same pipeline for individuals from the SSC (87) and IMAGEN (89) to obtain a harmonized dataset.

In the IMAGEN cohort (89), 2,090 individuals were genotyped using a combination of the Illumina 610Kq (N probes=620,901; N arrays=708) and 660Wq (N probes=657,366; N arrays=1,385). The genotyping was performed at the Centre National de Genotypage (Paris,

France). In the SSC (87), 10,032 individuals were genotyped at Yale University using Illumina single nucleotide polymorphism (SNP) genotyping arrays 1Mv1, 1Mv3 Duo, or Omni2.5M.

- *Whole genome sequencing data*

In the MSSNG database (88), 7,231 individuals were sequenced at multiple sites using Illumina sequencing HiSeq, HiSeq 2,500, or HiSeqX.

Next generation sequencing data were analysed using Broad Institute Genome Analysis Toolkit best practices (90).

- *CGH and FISH data*

In the neurodevelopment cohort of Sainte-Justine, the detection of CNVs was done mainly by CGH. In some cases, FISH was done for family members to confirm the inheritance of the CNV carried by the proband (*de novo* or inherited) or to confirm if a sibling carries the same CNV as the proband. These genetic tests were performed in the laboratory of Pediatric Genetics at CHU Sainte-Justine (Montreal, Canada). The calling and filtering of CNVs were also completed by the same laboratory. All the CNVs of this cohort are variants of uncertain significance or pathogenic.

CNV Calling

CNVs from SSC and IMAGEN were called using PennCNV (91) and QuantiSNP (92) with the following parameters:

- Number of consecutive probes for CNV detection ≥ 3
- CNV size $\geq$ 1Kb
- Confidence scores ≥ 15.

Then, we merged detected CNVs from both algorithms with CNVision (82).

For MSSNG, read alignment data were used to compute CNV calling following the workflow of Trost et *al*. (2018) (93).

Filtering of microarrays

To ensure good quality of CNVs, we kept only microarrays without too much noise.

- For IMAGEN:
  - Wave Factor < |0.05|
  - Standard deviation of the Log-R-Ratio < 0.35
  - Standard deviation of the B allele frequency < 0.08
  - Call Rate > 0.99
- For the SSC cohort: all microarrays detecting ≥ 200 CNVs were considered as noisy and were removed from the analysis.

CNV coordinates

The CNVs coordinates were updated from hg18 to hg19 using Illumina information and the liftover tool from the genome browser (https://genome.ucsc.edu/cgi-bin/hgLiftOver and http://grch37.ensembl.org/Homo_sapiens/Tools/AssemblyConverter).

Concatenation of CNVs

In a subsequent step, using an in-house algorithm (Pasteur, Paris, France) (85) followed by visual inspection (SnipPeep, http://snippeep.sourceforge.net), we stitched CNVs that appeared to be incorrectly split by the calling algorithms, and we removed any CNVs (size of ≥ 500 Kb and ≥ 100 SNPs) that spanned known large assembly gaps (greater than 150 Kb).

CNV filtering

CNVs with the following criteria were selected for analysis:

- Size ≥ 50 Kb
- Autosomal
- Unambiguous type: deletions or duplications
- Confidence score ≥ 30 with at least one of both detection algorithms
- Cross array criteria: CNVs overlapping ≥ 10 probes in each of the array technologies used in the study

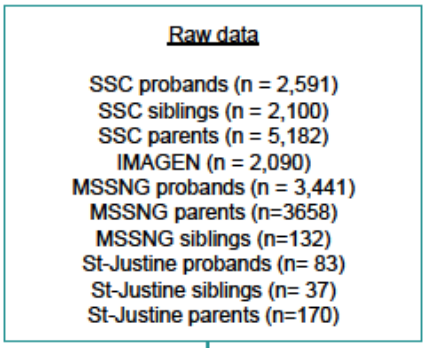
- Additional filters were applied for CNVs which are not 40% overlapping with recurrent CNVs : overlap with segmental duplicates or centromeric regions < 50%

CNV annotation and scoring:

The genetic annotation was based on RefSeq genes (https://genome.ucsc.edu/) using ANNOVAR (94). Each gene was annotated using functional scores such that CNV scores are the sum of scores of genes with all isoforms fully contained in the CNV. If an individual carried more than one CNV, CNV scores for this individual were summed (Figure 6). The default value associated to a gene without available scores was 0. Functional annotations of CNVs were performed using a home-made R package grouping several information about genes included in the CNVs and obtained using RefSeq genes. Genes were annotated using different scores and transformations, but in this study, we were only interested in the pLI because the previous study identified this score as the best variable explaining CNVs effects (85).

Genetic analysis of pairwise relatedness and population stratification

Relatedness was computed using Huguet et *al*. (2018) methodology (85) in IMAGEN and SSC cohorts. Ancestry was estimated using Admixture (95) (http://www.genetics.ucla.edu/software/admixture) with reference populations from HapMap3 (96) allowing for 4 ancestry components (Africa, Asia, Europe and India). Results show a strong European ancestry component in both datasets with 1,630 individuals from IMAGEN and 9,799 individuals from the SSC being estimated to have more than 80% of European ancestry. We then performed a principal components analysis based on the variance-standardized relationship matrix. For the analysis including IMAGEN and SSC, we only used the first three components (C1-C3) as covariates.



Samples after microarray and phenotypic quality control

Cohorts	No. individuals (N_{total}=13,949)	Age, mean (SD)	Males No. (%)	N NVIQ	NVIQ, mean (SD)	N raw SRS	Raw SRS mean (SD)	N SRS T-score	SRS T-score (SD)
SSC	2,569 probands	9.0 (3.6)	2,227 (86.7)	2569	84.5 (26.3)	2557	97.9 (26.9)	2560	79.5 (10.5)
	2,092 siblings	10.1 (4.3)	969 (46.3)	-	-	2079	18.4 (13.8)	2008	43.8 (7.0)
	5,138 parents	41.5 (6.2)	2,569 (50.0)	-	-	5138	29.5 (21.3)	-	-
MSSNG	845 probands	9.7 (4.7)	1,106 (80.1)	600	93.0 (23.7)	845	98.4 (32.3)	475	81.7 (16.1)
	628 parents	41.9 (6.5)	310 (49.4)	65	113.8 (10.6)	628	34.1 (26.3)	-	-
	28 siblings	8.1 (7.4)	16 (57.1)	10	108.8 (13.9)	28	39.4 (34.8)	25	53.7 (17.8)
IMAGEN	1,802 unselected	14.4 (0.3)	880 (48.9)	1744	106.6 (14.8)	1010	25.4 (17.0)	626	46.7 (9.0)
St-Justine	75 probands	7.5 (4.4)	50 (66.6)	75	80.19 (14.5)	57	80.5 (32.4)	57	68.4 (12.9)
	37 siblings	10.1 (6.6)	20 (54.1)	37	96.0 (15.9)	26	42.6 (30.7)	26	53.5 (12.9)
	170 parents	37.7 (6.9)	70 (41.2)	170	98.2 (16.9)	127	35.8 (27.7)	127	48.4 (9.7)

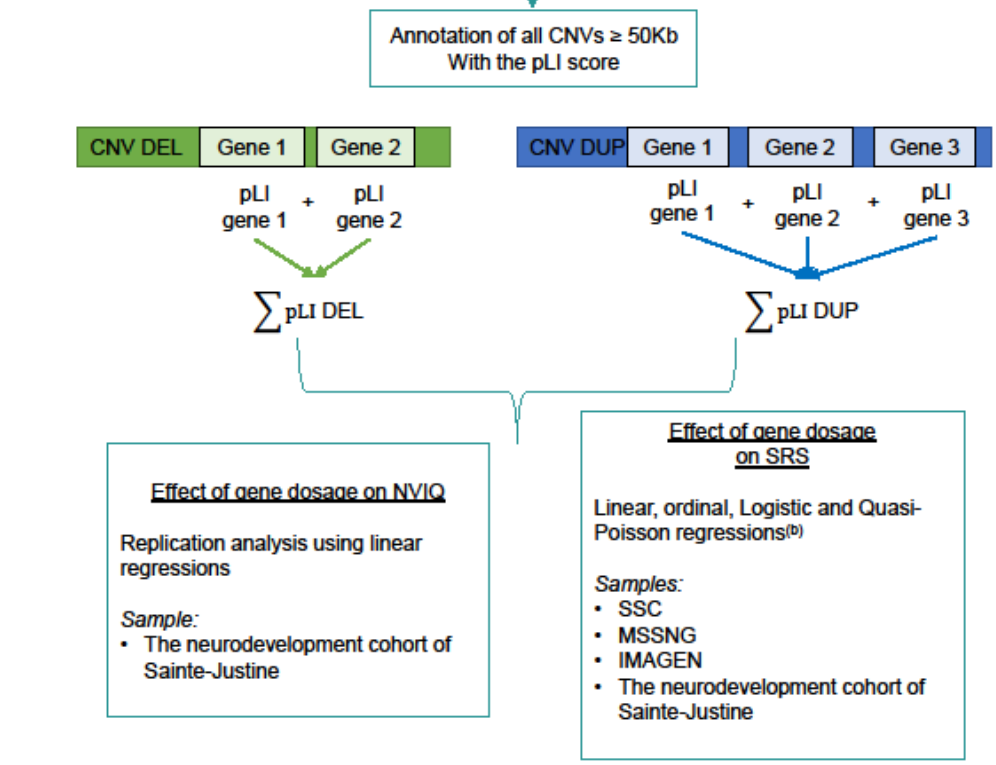


Figure 6: Methodological pipeline. Microarray quality control and CNV selection and annotation were performed as previously published in Huguet et *al.* (2018) (85) (Methods); The model used and available data for each phenotype are detailed in the statistical analysis section of the methods. SSC: Simon Simplex Collection; CNV: copy number variants; SD: standard deviation; N.A.: Not applicable; NVIQ: non-verbal intelligence quotient; SRS: Social Responsiveness Scale.

Clinical assessments

NVIQ

Intellectual abilities were measured using standardized tests according to the cognitive level of the participant (Table 2). We used NVIQ because individuals with autism or neurodevelopmental disorders are either non-verbal or have language impairment and consequently are not able to complete the verbal tasks in the cognitive tests. Then, the NVIQ was z-scored using the mean (mean=100) and the SD (SD=15) of the general population as follows: (NVIQ-100)/15. Norm-referenced standard scores (deviation IQ) were available for most of the participants (85.10%). However, for individuals from SSC who were not able to obtain a deviation IQ due to their age and/or developmental level, ratio IQ were derived by dividing mental age by chronological age and multiplying by 100. This was only done for the kids who underwent the Mullen Scale of Early Learning (MSEL) test. See Bishop et *al*., (2011) for more details concerning convergence between ratio and deviation IQ (98).

Table 2: NVIQ available from the different cohorts.

Cohorts	N available NVIQ	Age		Males		NVIQ		
		mean	SD	N	%	Test used	Mean	SD
IMAGEN	1,744	14.45	0.37	880	48.91	WISC-IV	106.62	14.77
SSC probands	2,564	9.03	3.58	2,227	86.66	DAS-II (N= 2244) MSEL(N=213), WASI-I (N=63), WISC-IV (N=45) Leiter (N=73)	84.47	26.27
MSSNG (probands, parents, siblings)	673 (Probands=598; Parents=65; Siblings=10)	12.8	10.2	511	75.7	Raven (N=147)(97), Stanford-Binet(N=89), WASI-I,WASI-II(N=218), WISC-IV (N=5), WPPSI-IV (N=69), Leiter (N=13), MSEL (N=18),	94.62	22.61
St-Justine (probands, parents, siblings)	282 (Probands= 75; Parents= 170; Siblings=37)	26.3	16.0	143	50.7	WPPSI-IV (N=31), WISC-V (N=36), WASI-II (N=12), WAIS-IV short version (N=172),	93.42	18.04

SD: Standard deviation; NVIQ: Non-verbal intelligence quotient; DAS-II: Differential Ability Scales - Second Edition (99); MSEL: Mullen Scale of Early Learning (100); WASI-I or II: Wechsler Abbreviated Scale of Intelligence – First

or Second Edition (101,102); WISC-IV: Wechsler Intelligence Scale for Children, Fourth Edition (103); Leiter: Leiter international performance scale – Original and revised (104,105); WPPSI-IV: Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (106).

Social Responsiveness Scale (SRS)

For all the individuals from the neurodevelopment cohort of Sainte-Justine, SSC, MSSNG and for the unselected population from IMAGEN, severity of social deficits was ascertained with scores from the SRS (107,108). Parents completed the SRS – an extensively validated quantitative measure of characterizing traits and symptoms of autism– about their offspring. Parents also had the questionnaire SRS filled either by a self-report or by a relative or a spouse to evaluate their social responsiveness as well.

The SRS is a 65 items questionnaire rated on a 4-point Likert-type scale and it contains 5 treatment subscales: Social Awareness, Social Cognition, Social Communication, Social Motivation, and Restricted Interests and Repetitive Behavior (109). It generates a total raw score that serves as an index of severity of social deficits in the autism spectrum. Note that this score can be raw (not corrected) or a T-score (corrected for sex and the type of the questionnaire used (preschool-form, school-age form, adult form and adult self-report)).

Higher scores on the SRS indicate greater severity of social impairment. In the SSC, probands have a mean of 97.9 (SD=26.9) for the total raw score which is much higher than the means of their parents (mean: 29.5 (SD=21.3)) and their unaffected siblings (mean: 18.4 (SD=13.8)) (Figure 6). Similar observations are seen in MSSNG and in the neurodevelopment cohort of Sainte-Justine (Figure 6). However, in the neurodevelopment cohort of Sainte-Justine, this mean is slightly lower for the probands (mean: 80.5 (SD=32.4)) since they don't all have a diagnosis of autism (heterogeneous neurodevelopmental disorders) (See distribution of SRS raw score in Figure 8 and Figure 9 and distribution of SRS T-score in Figure 10).

Statistical analyses

Effect of gene dosage on general intelligence in the neurodevelopment cohort of St-Justine

The model assessing the effect of CNVs on general intelligence applied in the unselected population sample (85), in ASD cohorts (Douard et *al*., manuscript under review) and in the

meta-analysis of 7 cohorts including unselected and autism populations (Huguet et *al*., manuscript in preparation) was also replicated in the neurodevelopment cohort of Sainte-Justine. It is a linear mixed-effect model, explaining the z-NVIQ according to total pLI measured for deletions and duplications. These analyses were performed with the function lmekin() from the R package "coxme" (110). The random effect takes into account a kinship matrix generated to model the genetic covariance between related individuals using the kinship() function from the R package "kinship2" (111).

This model could be written as:

NVIQ z-score ~ $\boldsymbol{\alpha}X + \gamma Z + \beta_1$ pLI $_{DEL} + \beta_2$pLI $_{DUP}$

where *X* represents the adjustments covariates (NVIQ test used and sex) and *Z* is the familial relatedness; pLI $_{DUP/DEL}$: sum of pLI scores for deletions or for duplications. α, β_1, β_2 and γ are respectively the vectors of coefficients for fixed and random effects.

Correlation and concordance analysis of the meta-analysis model in the neurodevelopment cohort of St-Justine

For this analysis, we calculated the correlation and concordance between the loss of NVIQ predicted by the model of the meta-analysis and the loss of NVIQ calculated by comparison to the biparental mean. This was performed only in the sample of probands of the neurodevelopment cohort of Sainte-Justine after separating them in two groups: probands with an inherited CNV and probands with a *de novo* CNV. The loss of NVIQ compared to the parents was calculated by subtracting the mean of the NVIQ z-score of the parents from the NVIQ z-score of the proband. The loss of NVIQ estimated by the model is calculated based on the sum of pLI for deletions and duplications of each individual.

Effect of gene dosage on social responsiveness

- *SRS as a continuous variable*

We used a linear mixed effect model to quantify the effect of gene dosage measured by pLI scores on the SRS total raw score after pooling SSC, MSSNG and Imagen. A kinship matrix was generated to model the genetic covariance between related individuals using the kinship()

function from the R package "kinship2" (111) and this covariance was used as a random effect in the model performed with the function lmekin() from the R package "coxme" (110).

This model could be written as:

SRS raw score ~$\boldsymbol{\alpha}X + \gamma Z$ + β_1 pLI $_{DEL}$ + β_2 pLI $_{DUP}$

where *X* represents the adjustments covariates (age, sex, NVIQ, ASD diagnosis) and *Z* is the familial relatedness; pLI $_{DUP/DEL}$: the sum of pLI score for deletions or for duplications. α, β_1, β_2 and γ are respectively the vectors of coefficients for fixed and random effects.

We further explored a potential effect of gene dosage on the SRS within the different groups (probands, unaffected siblings and parents, unselected population) separately using a linear regression (with the function lm ()) and adjusting for the abnormal distributions with a square root transformation of the SRS scores when necessary (Table 5).

Finally, we investigated the effect of gene dosage on SRS in the neurodevelopment cohort of Sainte-Justine. The aim of this analysis was to explore if there is a significant effect of gene dosage on social communication in a cohort with heterogeneous neurodevelopmental disorders and not only autism (Replication analysis). To do so, we used a Quasi-Poisson mixed-effect model fitted with the function glmmPQL() from the R package “MASS” (112).

This model could be written as:

Y= Poisson (μ (X, Z), θ)

With log μ (X,Z)= $\boldsymbol{\alpha}X + \gamma Z$ + β_1 pLI $_{DEL}$ + β_2 pLI $_{DU}$

Where θ: overdispersion parameter $_{P;}$ and *X* represents the covariates used in this model (age, sex and NVIQ). α, β_1, β_2 and γ are respectively the vectors of coefficients for fixed and random effects.

- *SRS as a categorical variable*

We also investigated the SRS scores based on the previously published *T*-score categorization (109) as follow:

- ***T*-scores of 76 or higher**: Clinically significant deficits in social functioning that interfere with interactions with others;
- **66 <*T*-scores< 75**: Moderate, signaling some clinically significant social deficits;
- **60 <*T*-scores< 65** : Mild to moderate deficits in social behavior;
- ***T*-scores< 59**: Indicate an individual probably does not have social difficulties indicative of a possible autism diagnosis.

A logistic regression was applied in this pooled dataset (autistic probands; unaffected siblings and unselected population) to investigate the effect of gene dosage on binary categories of the SRS: clinical (obtained after merging the moderate, mild and clinically significant categories) and normal (Table 7, Figure 10). This logistic regression model took into account the family relatedness as random factor using the MCMCglmm () function from the R package "MCMCglmm" (113).

An ordinal regression model was also performed on SRS coding for 4 different levels of social deficits (normal, moderate, mild and clinically significant) (Table 7, Figure 10). This model was applied using the function MCMCglmm () from the R package "MCMCglmm" (113).

This model could be written as:

$$LogP\ (Y>k)/P(Y<=k)) \sim \alpha_k X + \gamma_k Z + \beta_1\ pLI_{DEL} + \beta_2\ pLI_{DUP;}$$

Where X represents the adjustment covariates used in this model (age, NVIQ and ASD diagnosis). α, β_1, β_2 and γ are respectively the vectors of coefficients for fixed and random effects.

CHAPTER 3: RESULTS

RESULTS

Effect of gene dosage on general intelligence in the neurodevelopment cohort of St-Justine

In the neurodevelopment cohort of St-Justine, each point of pLI for deletions decreases z-scored NVIQ by 0.14 points (β=-0.14, SD=0.03, $p=2.4\times10^{-7}$) (equivalent to 2.09 points of NVIQ). These results are very similar to what has been found in the previous analyses conducted in unselected populations (β=-0.19) (85), in the autism cohort of SSC (β=-0.17) (Douard et *al*., manuscript under review) and in the meta-analysis combining unselected populations with autism cohorts (β=-0.19) (Huguet et *al*., manuscript in preparation).

On the other side, duplications have an effect size on z-scored NVIQ that's 2-3 fold smaller than deletions and this effect size is the same in the different cohorts: in the neurodevelopment cohort of St-Justine (β=-0.06, SD=0.02, $p=3\times10^{-3}$ per one unit of pLI) (equivalent to a loss of 0.9 points of NVIQ), in SSC cohort (β=-0.06) (Douard et *al*., manuscript under review) and in the meta-analysis (β=-0.06) (Huguet et *al*., manuscript in preparation) (Table 3, Figure 7).

Table 3: Effect of gene dosage measured by pLI on NVIQ z-score in the neurodevelopment cohort of Saint-Justine..

Population (N)	Intercept	β pLI DEL [SD]	P value	β pLI DUP [SD]	P value
Probands (75)	-0.75	**-0.10 [0.03]**	**2.7×10^{-3}**	-0.04 [0.02]	0.14
Probands+ Siblings (112)	-0.03	**-0.16 [0.03]**	**4.3×10^{-8}**	**-0.07 [0.02]**	**6.2×10^{-4}**
Parents (170)	-0.21	**-0.12 [0.06]**	**4.2×10^{-2}**	-0.01 [0.03]	0.75
Parents + Siblings (202)	-0.17	**-0.12 [0.05]**	**3.3×10^{-2}**	-0.01 [0.03]	0.68
All (289)	-0.58	**- 0.14 [0.03]**	**2.4×10^{-7}**	**-0.06 [0.02]**	**3.0×10^{-3}**

SD: Standard deviation; Significant results are in bold black and borderline results are in bold blue.

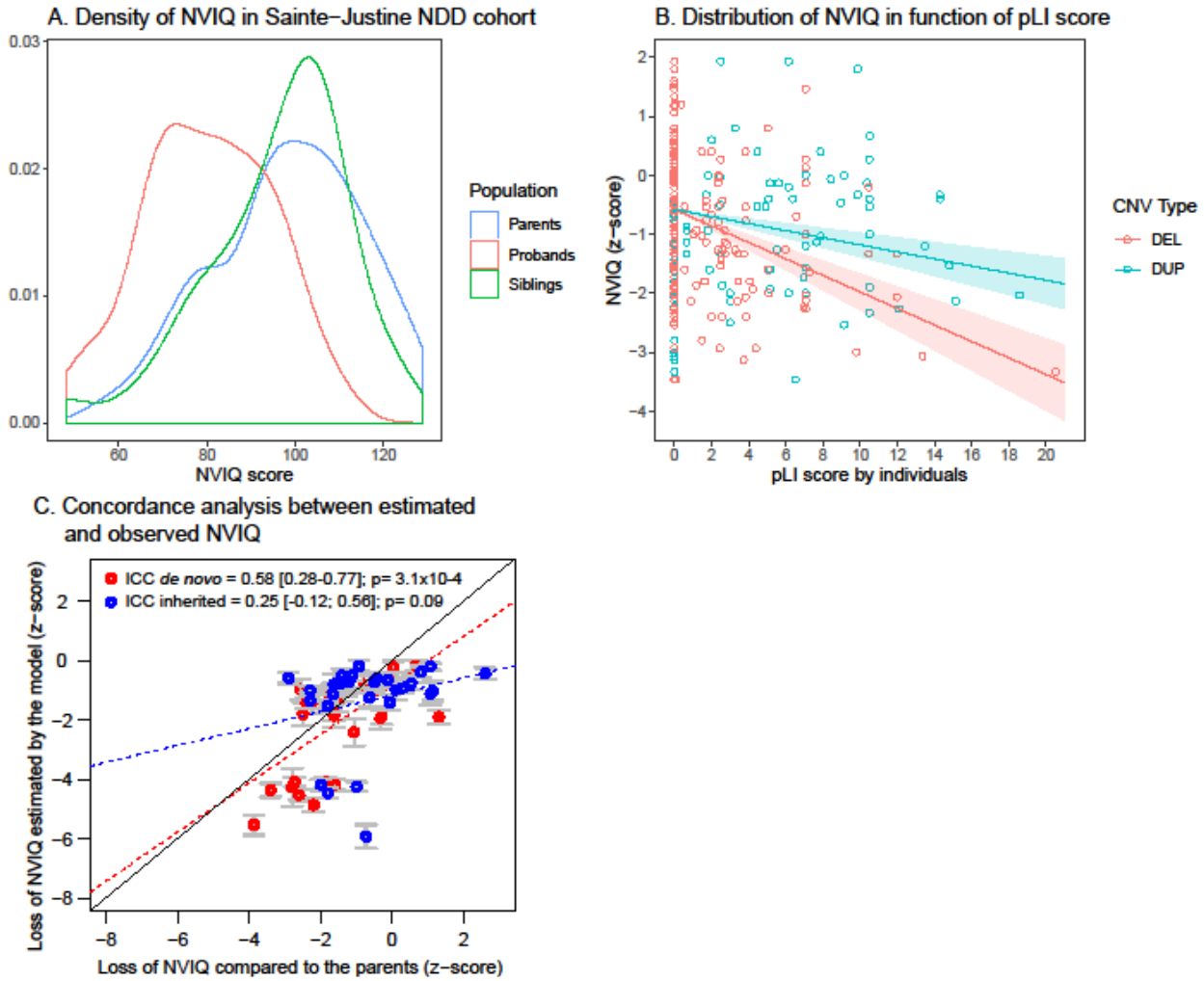


Figure 7: Effect of gene dosage on general intelligence in the neurodevelopment cohort of St-Justine.

(A) Density distribution of the NVIQ for the different kinships in Sainte-Justine cohort (probands in red, siblings in green, parents in blue). (B) Relationship between NVIQ (y-axis) and gene dosage measured by the pLI score (x-axis) (deletion: DEL in red, duplications: DUP in blue). (C) Concordance analysis between estimated loss of NVIQ computed in the meta-analysis (Huguet et al., manuscript in preparation) (y-axis) and observed loss of NVIQ in Sainte-Justine cohort (x-axis). Concordance for probands carriers of de novo CNVs are represented in red, and probands carriers of inherited CNVs are represented in blue. NDD: neurodevelopmental disorders; ICC: Intraclass correlation coefficient.

Correlation and concordance analysis of the meta-analysis model in the neurodevelopment cohort of St-Justine

In the sample of probands with a *de novo* CNV (N=30), the concordance, measured by intraclass correlation coefficient (ICC) between the loss of NVIQ z-score estimated by the model of the meta-analysis and the loss of NVIQ z-score observed by comparison to the parents is of 0.58

(95% CI: 0.28 to 0.77; p= 3.11×10^{-4}) and the Pearson correlation is of 0.61 (p=3.71×10^{-4}). However, in the sample of probands with an inherited CNV (N=29), the concordance (ICC= 0.25 (95% CI: -0.12 to 0.56)) and the Pearson correlation (0.25) are lower and not significant (p>.05). In the pooled sample (Inherited + *de novo* CNVs) (N=71), the concordance is of 0.35 (95% CI: 0.13 to 0.54; p=1.13×10^{-3}) and the Pearson correlation is of 0.36 (p=2×10^{-3}) (Table 4, Figure 7C).

Table 4: Concordance and correlation between the loss of NVIQ estimated by the model and the real loss compared to parents.

Inheritance	N	ICC [CI]	P valeur	Pearson correlation	P valeur
De novo	30	0.58 [0.28-0.77]	**3.1×10^{-4}**	0.61	**3.7×10^{-4}**
Inherited	29	0.25 [-0.12; 0.56]	0.09	0.25	0.19
De novo+ Inherited	71	0.35 [0.13-0.54]	**1.1×10^{-3}**	**0.36**	**2.0×10^{-3}**

ICC: Intraclass correlation coefficient; CI: Confidence interval. Significant results are in bold.

Effect of gene dosage on social responsiveness

SRS raw score in the autism cohorts and the unselected population

The pLI significantly increases the SRS raw score, with a 2:1 effect size ratio for deletions and duplications in the pooled SSC, MSSNG and IMAGEN dataset (deletions: β=3.37 points of raw SRS score per unit of pLI, SD=0.55, p= 7.8×10^{-10}; duplications: β=1.77 points of raw SRS score per unit of pLI, SD=0.40, p=9.7×10^{-6}). This effect of gene dosage on SRS is mainly accounted for by the NVIQ and the autism diagnosis (Table 5, Figure 8). Similar results were also observed when these analyses were done in the SSC cohort (deletions: β=3.47 points of raw SRS score per unit of pLI, SD=0.58, p= 2.4×10^{-9}; duplications: β=1.54 points of raw SRS score per unit of pLI, SD=0.44, p=5.2×10^{-4}) and in a pooled sample of SSC and IMAGEN (deletions: β=3.72 points of raw SRS score per unit of pLI, SD=0.57, p= 5.1×10^{-11}; duplications: β=1.87 points of raw SRS score per unit of pLI, SD=0.43, p=1.4×10^{-5}) (Table 5). However, once we adjust for the autism diagnosis, the effect of pLI on SRS disappears (Table 5)

Table 5: Effect of gene dosage measured by pLI on SRS raw score in two autism cohorts and IMAGEN.

Population	N	SRS-score	Model	CNV score	Effect size (β)	SE	p
SSC Probands	2 556	Total-raw	Linear not adjusted for NVIQ	pLI DEL	-0.21	0.40	0.60
				pLI DUP	-0.28	0.36	0.42
			Linear adjusted for NVIQ	pLI DEL	-0.31	0.41	0.45
				pLI DUP	-0.32	0.36	0.36
MSSNG Probands	845	Total-raw	Linear not adjusted for NVIQ	pLI DEL	-1.79	1.51	0.24
				pLI DUP	0.14	0.79	0.86
	600		Linear adjusted for NVIQ	pLI DEL	-0.72	1.76	0.68
				pLI DUP	-0.46	0.94	0.62
SSC+ MSSNG Probands	3 403	Total-raw	Linear not adjusted for NVIQ	pLI DEL	0.05	0.51	0.91
				pLI DUP	-0.27	0.36	0.47
	3 157		Linear adjusted for NVIQ	pLI DEL	-0.41	0.50	0.41
				pLI DUP	-0.65	0.36	0.07
SSC Unaffected Siblings	2,078	√Total-raw[(a)]	Linear not adjusted for NVIQ	pLI DEL	0.05	0.08	0.47
				pLI DUP	0.001	0.06	0.99
SSC Unaffected Parents	4,838	√Total-raw[(a)]	Linear not adjusted for NVIQ	pLI DEL	0.07	0.09	0.43
				pLI DUP	0.01	0.04	0.83
IMAGEN	1,010	√Total-raw[(a)]	Linear not adjusted for NVIQ	pLI DEL	-0.06	0.15	0.66
				pLI DUP	0.03	0.09	0.71
			Linear adjusted for NVIQ	pLI DEL	-0.09	0.15	0.56
				pLI DUP	0.02	0.09	0.81
SSC Unaffected siblings and parents+ IMAGEN	7,926	Total-raw	Linear mixed-effect	pLI DEL	0.62	0.62	0.32
				pLI DUP	0.08	0.36	0.82
SSC Probands + IMAGEN	3,567	Total-raw	Linear not adjusted for NVIQ or autism diagnosis	pLI DEL	**2.68**	**0.67**	**$7.15x10^{-5}$**
				pLI DUP	1.29	0.58	0.03
			Linear adjusted for autism diagnosis	pLI DEL	0.41	0.47	0.38
				pLI DUP	-0.19	0.40	0.63
			Linear adjusted for NVIQ	pLI DEL	-0.31	0.46	0.50
				pLI DUP	-0.44	0.39	0.27
				NVIQ	**-0.54**	**0.02**	**$< 1x10^{-7}$**
SSC (Probands and unaffected parents and siblings)	9,473	Total-raw	Linear mixed-effect not adjusted for autism diagnosis	pLI DEL	**3.47**	**0.58**	**$2.40x10^{-9}$**
				pLI DUP	**1.54**	**0.44**	**$5.20x10^{-4}$**
			Linear mixed-effect adjusted for autism diagnosis	pLI DEL	0.75	0.37	$4.30x10^{-2}$
				pLI DUP	-0.003	0.29	0.99
SSC + IMAGEN	10,483	Total-raw	Linear mixed-effect not adjusted for autism diagnosis	pLI DEL	**3.72**	**0.57**	**$5.10x10^{-11}$**
				pLI DUP	**1.87**	**0.43**	**$1.40x10^{-5}$**
			Linear mixed-effect adjusted for autism diagnosis	pLI DEL	0.55	0.36	0.13
				pLI DUP	-0.10	0.27	0.72
SSC + MSSNG + IMAGEN	11,979	Total-raw	Linear mixed-effect not adjusted for autism diagnosis nor NVIQ	pLI DEL	**3.37**	**0.55**	**$7.80x10^{-10}$**
				pLI DUP	**1.77**	**0.40**	**$9.70 x10^{-6}$**
			Linear mixed-effect adjusted for autism diagnosis	pLI DEL	0.30	0.35	0.39
				pLI DUP	-0.04	0.26	0.88
	4,210		Linear mixed-effect adjusted for NVIQ	pLI DEL	0.29	0.60	0.62
				pLI DUP	0.28	0.50	0.57

SD: Standard deviation; NVIQ: Non-verbal intelligence quotient; DEL: deletion; DUP: duplication; pLI: probability of being Loss-of-function Intolerant; pLI DEL/pLI DUP: pLI score for deletions or for duplications; √Total-raw: square root transformation of the total SRS raw. All models used were adjusted for age, sex and ancestry. Models take into account family as random-effect when including related individuals (see methodology section). [(a)]Square root transformation of the total SRS raw score was performed to adjust for the non-Gaussian distribution or bimodality of SRS distribution (Figure 8). Significant results are in bold.

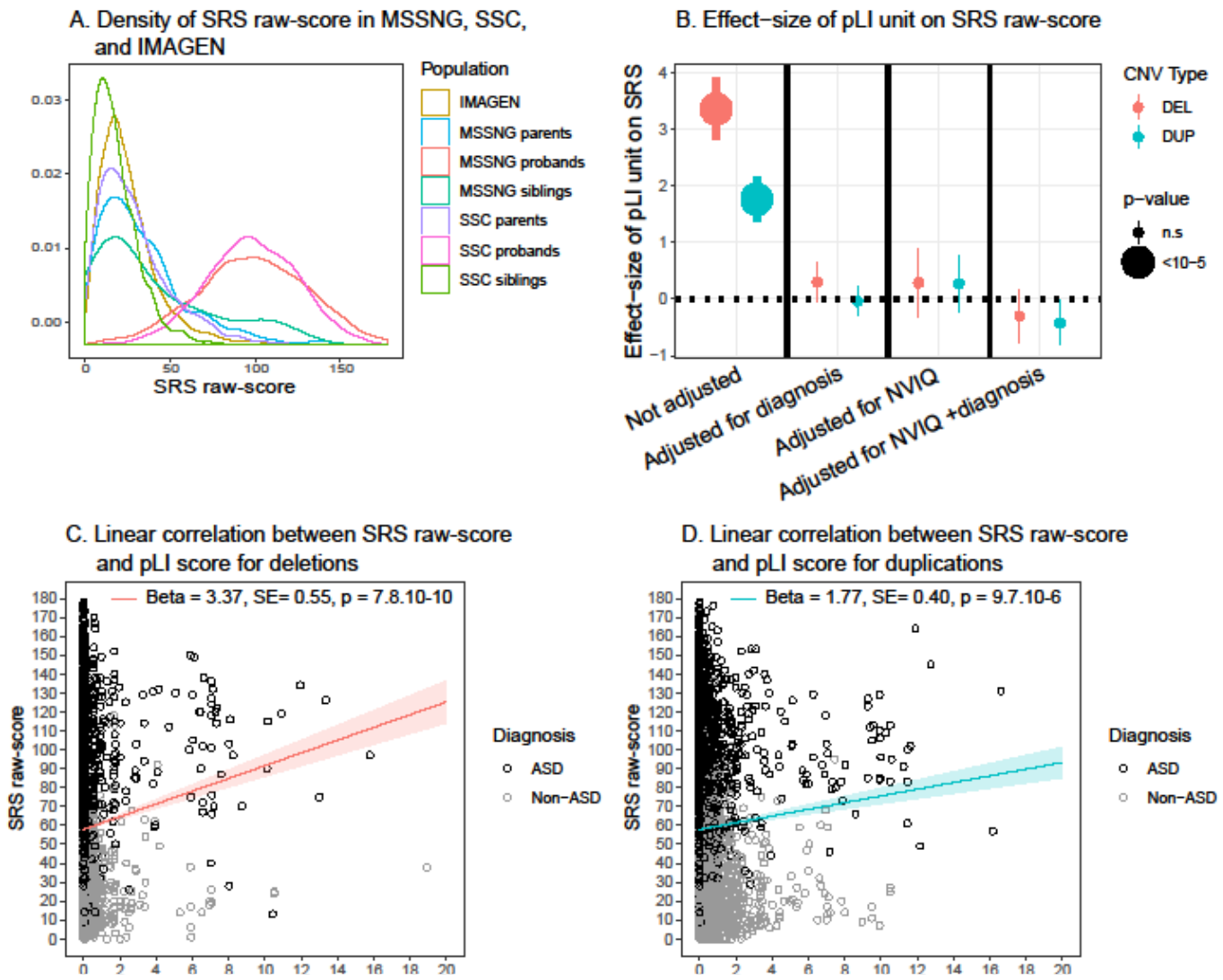


Figure 8: Effect of gene dosage on SRS raw score in the autism and unselected cohorts.

(A) Density distribution of the raw score of SRS in function of the cohort and kinship (SSC probands in pink, SSC siblings in dark green, SSC parents in violet, MSSNG probands in red, MSSNG siblings in light green, MSSNG parents in light blue, IMAGEN in brown). (B) Effect size of a unit of pLI on raw score of SRS (y-axis) for deletions (DEL, red) or duplications (DUP, blue) in function of the covariates used (x-axis), in the pooled dataset (SSC + MSSNG + IMAGEN). Effects were measured with and without adjustment for the diagnosis of autism or the NVIQ. The Y-axis represents the estimated effect of pLI on the SRS raw score. (C) Linear relationship between raw score of SRS (y-axis) and gene dosage measured by pLI for deletions (x axis). Individuals with a diagnosis of ASD are represented in black and unaffected individuals are in grey. (D) Linear relationship between raw score of SRS (y-axis) and gene dosage measured by pLI for duplications (x axis). Individuals with a diagnosis of ASD are represented in black and unaffected individuals are in grey.

SRS raw score in the neurodevelopment cohort of St-Justine

The relationship between gene dosage and SRS raw score (deletion: OR=1.05, 95% CI=1.02-1.08, $p=1.5\times10^{-3}$; duplication: OR=1.05, 95% CI=1.02-1.07, $p=5.3\times10^{-5}$) estimated by a Quasi-Poisson model translates into a gain of 3.0 points of SRS per one unit of pLI for deletions and a gain of 2.7 points of SRS per one unit of pLI for duplications. And since this model is a non-linear model, this also translates into an increase of a mean of 38.1 and 33.9 points of SRS for a deletion or a duplication encompassing 10 units of pLI, respectively (Table 6, Figure 9).

This effect of deletions and duplications on SRS remains significant after adjusting for NVIQ (deletion: OR=1.04, 95% CI=1.02-1.06, $p=2.3\times10^{-2}$; duplication: OR=1.05, 95% CI=1.01-1.08, $p=3.0\times10^{-3}$) and it translates into an increase of a mean of 3.8 and 4.3 points of SRS for a deletion or a duplication encompassing one unit of pLI respectively.

Table 6: Effect of gene dosage and NVIQ on SRS raw score in the neurodevelopment cohort of St-Justine.

Population (N)	OR DEL	95% CI	P value	OR DUP	95% CI	P value	OR NVIQ	95% CI	P value
NDD cohort (195)	1.05	1.02-1.08	**1.54×10^{-3}**	1.05	1.02-1.07	**5.33×10^{-5}**	0.99	0.99	**0.03**

OR: Odds ratio; DEL: deletions; DUP: Duplications; 95% CI: 95% confidence interval; NVIQ: Non-verbal intelligence quotient. Significant results are in bold and borderline results are in bold blue.

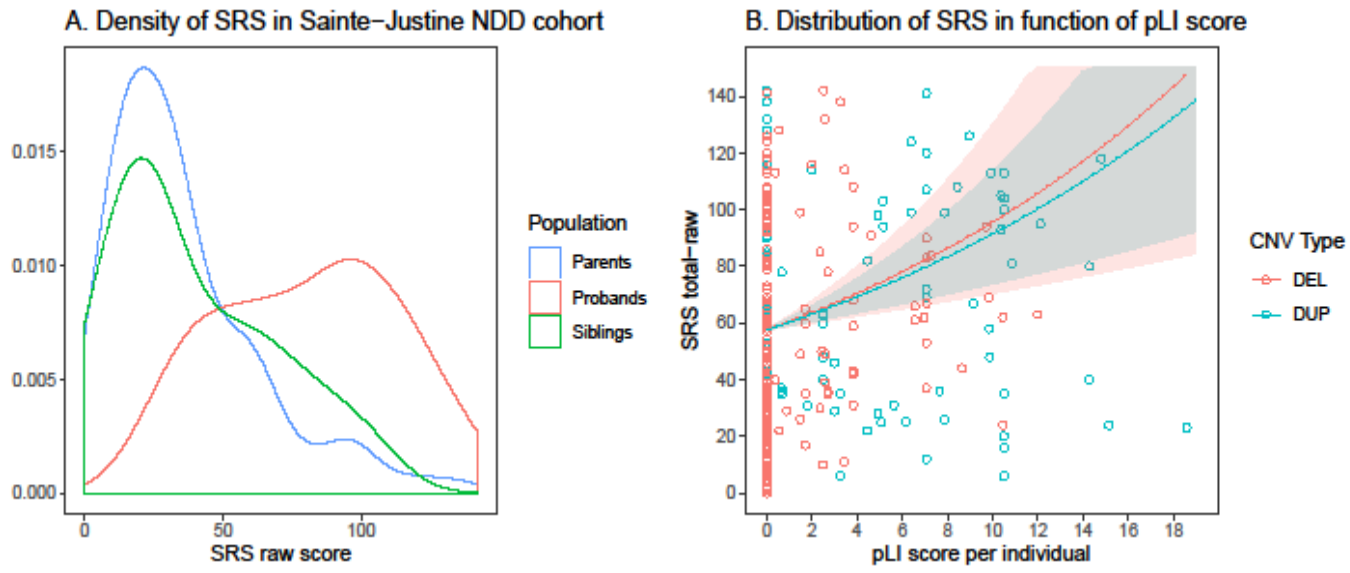


Figure 9: Effect of gene dosage on SRS raw score in the neurodevelopment cohort of St-Justine. (A) Density distribution of SRS raw score for the different kinships (probands in red, siblings in green, parents in blue). (B) Effect size of pLI (x-axis) on raw score of SRS (y-axis) for deletions (DEL, red) or duplications (DUP, blue) in Sainte-Justine cohort. The Y-axis represents the estimated effect of pLI on the SRS raw score using a quasi-poisson regression model. NDD: Neurodevelopmental disorders.

SRS categories in the autism cohorts and the unselected population

In the pooled dataset of SSC, MSSNG and IMAGEN, the ordinal regression performed on the 4 categories of SRS (normal, mild, moderate and clinically significant) shows that an increase of pLI for deletions or duplications increase the risk of being in a category with higher deficits in social functioning (deletions: OR= 1.15; 95% CI, 1.07-1.23; $p< 1.0\times10^{-3}$; duplications: OR= 1.10; 95% CI, 1.04-1.16; $p< 1.0\times10^{-3}$). This effect disappears once we adjust for the presence of an autism diagnosis or for NVIQ (Table 7, Figure 10). Similar results were obtained when using a pooled sample of SSC and IMAGEN (deletions: OR= 1.32; 95% CI, 1.12-1.60; $p< 1.0\times10^{-3}$; duplications: OR= 1.18; 95% CI, 1.04-1.35; $p< 1.0\times10^{-3}$) (Table 7).

The logistic regression performed on 2 categories of SRS (normal vs clinical) in the pooled dataset (SSC, MSSNG and IMAGEN) also shows that increasing the pLI leads to a higher probability of having social deficits that are clinically significant (deletions: OR= 1.21; 95% CI,

1.12-1.31; p< 1.0×10^{-3}; duplications: OR= 1.13; 95% CI, 1.07-1.21; p< 1.0×10^{-3}) and adjustments with the NVIQ or with the autism diagnosis make the effect of gene dosage on SRS insignificant. Similar results were obtained in the pooled dataset of SSC and IMAGEN (deletions: OR= 1.22; 95% CI, 1.12-1.34; p< 1.0×10^{-3}; duplications: OR= 1.14; 95% CI, 1.06-1.23; p< 1.0×10^{-3}) (Table7, Figure 10).

Table 7: Effect of gene dosage measured by pLI on SRS categories.

Population	N	SRS-score	Model	CNV score	Effect size (OR)	SD	p
SSC Probands+ Unaffected siblings+ IMAGEN	5,188	SRS categories (normal, clinical) [(a)]	Logistic regression not adjusted	pLI DEL pLI DUP	**1.22** **1.14**	**1.12-1.34** **1.06-1.23**	**<1.0 x10^{-3}** **<1.0 x10^{-3}**
			Logistic regression adjusted for autism diagnosis	pLI DEL pLI DUP	0.93 0.98	0.83- 1.05 0.88-1.05	0.22 0.69
SSC Probands+ Unaffected siblings+ IMAGEN	5,188	SRS categories (normal, moderate, mild, clinically significant) [(a)]	Ordinal not adjusted	pLI DEL pLI DUP	**1.32** **1.18**	**1.12- 1.60** **1.04-1.35**	**<1.0 x10^{-3}** **<1.0 x10^{-3}**
			Ordinal adjusted for autism diagnosis	pLI DEL pLI DUP	1.04 0.99	0.97-1.13 0.94-1.05	0.35 0.87
SSC and MSSNG Probands+ Unaffected siblings+ IMAGEN	5,688	SRS categories (normal, clinical) [(a)]	Logistic regression not adjusted	pLI DEL pLI DUP	**1.21** **1.13**	**1.12-1.31** **1.07-1.21**	**<1.0 x10^{-3}** **<1.0 x10^{-3}**
			Logistic regression adjusted for autism diagnosis	pLI DEL pLI DUP	0.94 0.97	0.86-1.06 0.87-1.08	0.41 0.61
	3,523		Logistic regression adjusted for NVIQ	pLI DEL pLI DUP	1.03 1.03	0.97-1.15 0.99-1.08	0.73 0.41
SSC and MSSNG Probands+ Unaffected siblings+ IMAGEN	5,688	SRS categories (normal, moderate, mild, clinically significant) [(a)]	Ordinal not adjusted	pLI DEL pLI DUP	**1.15** **1.10**	**1.07-1.23** **1.04-1.16**	**<1.0 x10^{-3}** **1.0 x10^{-3}**
			Ordinal adjusted for autism diagnosis	pLI DEL pLI DUP	1.02 1.00	0.96-1.09 0.99-1.05	0.53 0.89
	3,523		Ordinal adjusted for NVIQ	pLI DEL pLI DUP	1.00 1.01	0.95-1.06 0.96-1.06	0.90 0.55

SD: Standard deviation; NVIQ: Non-verbal intelligence quotient; DEL: deletion; DUP: duplication; pLI: probability of being Loss-of-function Intolerant; pLI DEL/pLI DUP: pLI score for deletions or for duplications. All logistic or ordinal regression models used were adjusted for age. Models take into account family as random-effect when including related individuals (Methods). [(a)] Based on the previously published *T*-score categorization (109) (Methods). Significant results are in bold.

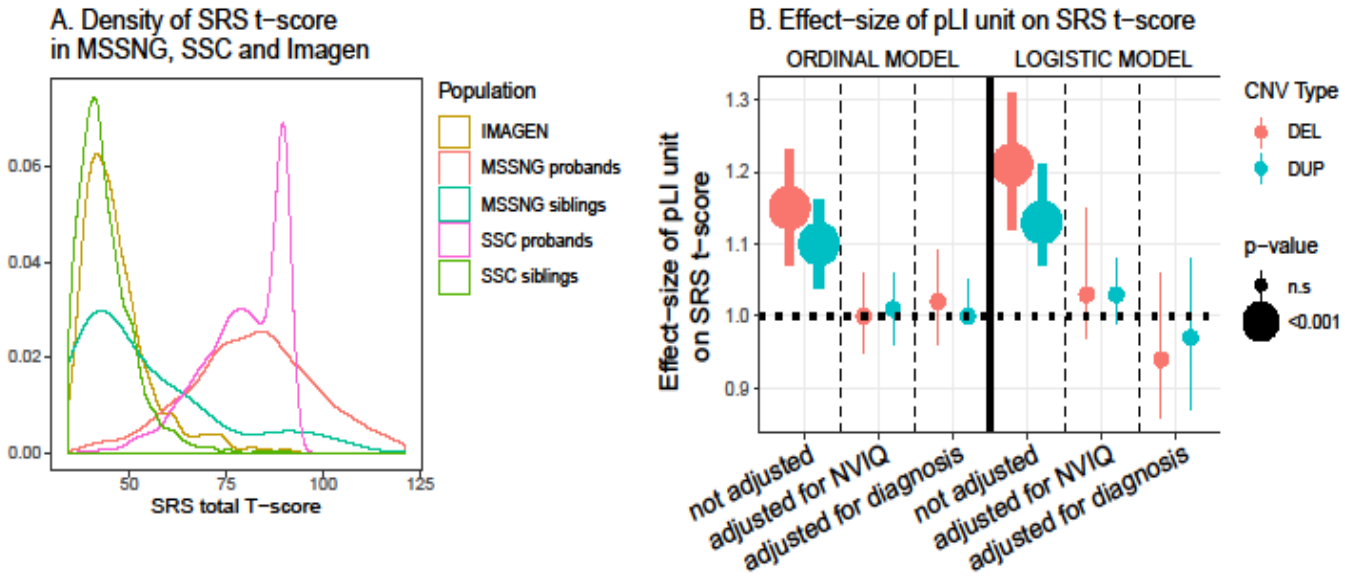


Figure 10: Effect of gene dosage on T-score of SRS in the pooled autism and general populations. (A) Density distribution of the T-score of SRS in function of the cohort and kinship (SSC probands in pink, SSC siblings in dark green, MSSNG probands in red, MSSNG siblings in light green, IMAGEN in brown). (B) Effect size of pLI on categories of SRS (y-axis) for deletions (DEL, red) or duplications (DUP, blue) in function of the covariates used (x-axis), in the pooled dataset (SSC + MSSNG + IMAGEN). The Y-axis represents the estimated OR for the clinical category conferred by one unit of pLI. Effects-size were measured as an OR by ordinal regressions on 4 categories (normal, mild, moderate, clinical) or by logistic regressions on binary categorical SRS (clinical and normal). These analyses were done with and without adjustment for the diagnosis of autism or the NVIQ.

CHAPTER 4: DISCUSSION AND CONCLUSION

DISCUSSION

Effect of gene dosage on general intelligence

In the first part, we aimed to replicate the model assessing the effect of CNVs on general intelligence (measured by NVIQ) in a cohort of heterogeneous neurodevelopmental disorders. The results showed that the effect of deletions is similar across the different populations (unselected, autistic and neurodevelopmental) and that the effect size of deletions is 2-3 fold higher than the one of duplications. When the model was first tested within the unselected population, duplications had no effect on general intelligence (85). However, when testing their effect in the autistic cohorts, the neurodevelopment cohort and the meta-analysis, we could quantify the impact of duplications on NVIQ and the effect size was the same across the different cohorts (a loss of 0.9 points of NVIQ per one unit of pLI). These results confirm that models trained on non-pathogenic CNVs in the general population reliably estimate the effect size of pathogenic CNVs and suggest "omnigenic" associations of haploinsufficiency with IQ. In fact, an omnigenic model is one of the many genetic models in the literature that aim to explain the genetic contribution to diseases; and this model supports the hypothesis that for complex traits such as autism, association signals tend to be spread across most of the genome—including many genes without an obvious connection to a specific disease (114). And since our model was trained in different cohorts, we covered CNVs encompassing over 4,000 genes. Furthermore, when large effect size *de novo* CNVs were excluded from the analyses, the effect of gene dosage on NVIQ remains unchanged (Douard et *al*., manuscript under review) supporting the fact that these results are not driven by highly pathogenic CNVs. In other words, these results suggest that autism risk is largely driven by genes with no direct relevance to autism (common variants) and is propagated through regulatory networks to a much smaller number of core genes with direct effects (e.g. synaptic genes).

The pLI, the genetic score used in our model, is a measure estimating the sensitivity of genes to haploinsufficiency based on the ratio of observed over expected Loss-of-Function mutations. This score quantified as well the effect of duplications on intelligence indicating

that it can assess intolerance to gene dosage regardless if the gene expression is increased (duplication) or decreased (deletion).

In the second part, we aimed to validate the model of the meta-analysis in the neurodevelopment cohort of St-Justine. To do so, we calculated the concordance and the correlation between the model estimations and the observed loss of NVIQ. In the sample of probands carrying *de novo* CNVs only, the concordance (0.58) and correlation (0.61) were higher than what was obtained in the sample of probands with inherited CNVs (ICC and Pearson correlation = 0.25). These results were expected since in the *de novo* sample, the difference of NVIQ between the biparental mean and the proband reflects the loss of NVIQ explained strictly by genetics and this can be well-estimated by the model. However, in the inherited sample, the probands share the pathogenic CNV with one of the parents, consequently the difference of NVIQ between the parents and the proband does not reflect the effect of the CNV that can be estimated by the model. Overall, our results fall between 0.2 and 0.7 which is the range of correlation between the biparental mean on psychometric and anthropometric quantitative traits and the individual's outcome (66,115).

Effect of gene dosage on social responsiveness

In the third part, we aimed at testing the model on a measure associated to a core symptom of autism. For that purpose, we applied the model on SRS which serves as an index of severity of social deficits in the autism spectrum. In the first place, we applied the model within samples of probands or unaffected individuals separately, but the results were not significant. Therefore, we attempted to pool them together to increase the statistical power needed to detect a signal and also to increase the variance of the SRS. Once this was done, we detected an effect of gene dosage on SRS with a 2:1 ratio for deletions and duplications in the different analyses when SRS was used as a continuous measure (SSC only, SSC+ Imagen, SSC+MSSNG+Imagen) (Table 5, Figure 8). When SRS was used as a categorical measure, the effect of gene dosage on SRS was also identified but without the 2:1 ratio for deletions and duplications. However, this effect disappears after adjusting for the presence of an autism diagnosis or for NVIQ. These results suggest that beyond its predictive value of a diagnosis, SRS

does not provide additional granularity such as the measure of ASD-trait severity implying that although SRS is a continuous measure, its use is mainly categorical. Also, the fact that the effect of gene dosage on SRS disappears after correction for NVIQ suggests that the latter mediates the effect between the gene dosage and SRS. In other words, haploinsufficiency causes a decrease in NVIQ which ultimately causes an increase in SRS scores. Taken together, our results indicate that the effect of CNVs on cognition is stronger than its effects on social behavior and autism core symptoms.

We also sought to replicate the previous analyses (done in a pooled sample of autistic and unaffected individuals) in a sample with neurodevelopmental disorders. The only result that differed is that the effect of gene dosage on SRS remained after adjusting for NVIQ. This can be due to the fact that in the neurodevelopment cohort of St-Justine, NVIQ has a very small effect on SRS and this effect is borderline significant (Table 6). In other words, since the probands in this cohort have different neurodevelopmental disorders and not only autism, they have low NVIQ but not necessarily high SRS scores. Whereas in the previous analyses, NVIQ is a mediator of the effect between gene dosage and SRS (Table 5) and since all the probands are autistic, low scores of NVIQ are mainly associated with high scores of SRS. Nevertheless, when NVIQ was accounted for, the effect size and the significance of gene dosage on SRS decreased in the neurodevelopment cohort suggesting that the impact of haploinsufficiency (measured by the pLI) on social behavior is largely driven by NVIQ.

Conclusion

This study highlights a differential effect of deletions and duplications on general intelligence and autism risk with a deletion: duplication effect size ratio of 2-3:1. This differential effect of gene dosage is a well-established characteristic in many psychiatric disorders where for example the 16p11.2 deletion is associated with ASD and obesity whereas the 16p11.2 duplication is associated with schizophrenia and leanness (116–120).

The results of this study corroborate the robustness of the model in predicting the effect of gene dosage on general intelligence by obtaining similar results across different cohorts and clinical diagnoses. Interestingly, when the model was extended to the prediction of social

communication by applying it on SRS measures, the estimates of the model overlapped with risk computed in previous studies. For example, our model estimates an increase of 37 and 42 points in the SRS raw score for the 16p11.2 and 22q11.2 deletion which is similar to the previously published effect of 44 and 55 points (66,121). The research team has developed a prediction tool available online (https://cnvprediction.urca.ca/) to estimate the effect size of undocumented deletions and duplications on IQ, autism risk and the SRS score. This tool will help clinicians evaluate quantitatively the contribution of a CNV to the patient's symptoms.

Limitations:

The predictions for some CNVs are discordant. Notably, deficits associated with the 15q13.3 deletion are underestimated by our model (Figure 5). This CNV may include genes for which the assigned pLI score does not capture the effects on psychiatric traits (*e.g.* gene dosage of *CHRNA7*, which has a pLI=0 may affect psychopathology without altering genetic fitness). Larger samples, novel functional annotations, and more refined models are required to improve our estimates of CNV effect sizes on cognitive and behavioral dimensions.

In conclusion, this study represents a new framework to study CNVs and can help in the interpretation of the effect size of undocumented CNVs identified in the neurodevelopmental clinic.

REFERENCES

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub; 2013. 1679 p.

2. Kanner L. A HISTORY OF THE CARE AND STUDY OF THE MENTALLY RETARDED. 1964.

3. Masi A, DeMayo MM, Glozier N, Guastella AJ. An Overview of Autism Spectrum Disorder, Heterogeneity and Treatment Options. Neurosci Bull. 2017 Feb 17;33(2):183–93.

4. Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. MMWR Surveill Summ. 2018 Apr 27;67(6):1–23.

5. First MB, Frances A, Pincus HA. DSM-IV-TR handbook of differential diagnosis. Arlington, VA, US: American Psychiatric Publishing, Inc.; 2002. xv, 247. (DSM-IV-TR handbook of differential diagnosis).

6. Sandin S, Lichtenstein P, Kuja-Halkola R, Larsson H, Hultman CM, Reichenberg A. THE FAMILIAL RISK OF AUTISM. JAMA. 2014 May 7;311(17):1770–7.

7. 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 Aug;46(8):881–5.

8. Hallmayer J, Cleveland S, Torres A, Phillips J, Cohen B, Torigoe T, et al. Genetic Heritability and Shared Environmental Factors Among Twin Pairs With Autism. Arch Gen Psychiatry. 2011 Nov;68(11):1095–102.

9. Kim YS, Leventhal BL, Koh Y-J, Fombonne E, Laska E, Lim E-C, et al. Prevalence of Autism Spectrum Disorders in a Total Population Sample. Am J Psychiatry. 2011 Sep 1;168(9):904–12.

10. Levy SE, Mandell DS, Schultz RT. Autism. The Lancet. 2009 Nov 7;374(9701):1627–38.

11. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. 2013 Oct 3;4(1):36.

12. Miller DT, Adam MP, Aradhya S, Biesecker LG, Brothman AR, Carter NP, et al. Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. Am J Hum Genet. 2010 May 14;86(5):749–64.

13. Veenstra-VanderWeele J, Blakely RD. Networking in Autism: Leveraging Genetic, Biomarker and Model System Findings in the Search for New Treatments. Neuropsychopharmacology. 2012 Jan;37(1):196–212.

14. Ousley O, Cermak T. Autism Spectrum Disorder: Defining Dimensions and Subgroups. Curr Dev Disord Rep. 2014 Mar 1;1(1):20–8.

15. Postorino V, Fatta LM, Sanges V, Giovagnoli G, De Peppo L, Vicari S, et al. Intellectual disability in Autism Spectrum Disorder: Investigation of prevalence in an Italian sample of children and adolescents. Res Dev Disabil. 2016 Jan;48:193–201.

16. Goin-Kochel RP, Peters SU, Treadwell-Deering D. Parental reports on the prevalence of co-occurring intellectual disability among children with autism spectrum disorders. Res Autism Spectr Disord. 2008 Jul 1;2(3):546–56.

17. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019 Mar;51(3):431.

18. Bishop DVM. Overlaps Between Autism and Language Impairment: Phenomimicry or Shared Etiology? Behav Genet. 2010 Sep;40(5):618–29.

19. Loucas T, Charman T, Pickles A, Simonoff E, Chandler S, Meldrum D, et al. Autistic symptomatology and language ability in autism spectrum disorder and specific language impairment. J Child Psychol Psychiatry. 2008 Nov;49(11):1184–92.

20. Kjelgaard MM, Tager-Flusberg H. An Investigation of Language Impairment in Autism: Implications for Genetic Subgroups. Lang Cogn Process. 2001 Apr 1;16(2–3):287–308.

21. Tager-Flusberg H, Rogers S, Cooper J, Landa R, Lord C, Paul R, et al. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. J Speech Lang Hear Res JSLHR. 2009 Jun;52(3):643–52.

22. Kasari C, Brady N, Lord C, Tager-Flusberg H. Assessing the minimally verbal school-aged child with autism spectrum disorder. Autism Res Off J Int Soc Autism Res. 2013 Dec;6(6):479–93.

23. Tager-Flusberg H, Kasari C. Minimally verbal school-aged children with autism spectrum disorder: the neglected end of the spectrum. Autism Res Off J Int Soc Autism Res. 2013 Dec;6(6):468–78.

24. Tager-Flusberg H, Joseph RM. Identifying neurocognitive phenotypes in autism. Philos Trans R Soc Lond B Biol Sci. 2003 Feb 28;358(1430):303–14.

25. Wittke K, Mastergeorge AM, Ozonoff S, Rogers SJ, Naigles LR. Grammatical Language Impairment in Autism Spectrum Disorder: Exploring Language Phenotypes Beyond Standardized Testing. Front Psychol. 2017;8.

26. Ousley O, Cermak T. Autism Spectrum Disorder: Defining Dimensions and Subgroups. Curr Dev Disord Rep. 2014 Mar 1;1(1):20–8.

27. Lai M-C, Lombardo MV, Baron-Cohen S. Autism. The Lancet. 2014 Mar 8;383(9920):896–910.

28. Glicken MD. Chapter 21 - Oppositional Defiant and Conduct Disorders Leading to Anti-Social Behavior and Violence. In: Glicken MD, editor. Evidence-Based Practice with Emotionally Troubled Children and Adolescents. San Diego: Academic Press; 2009. p. 357–76. (Practical Resources for the Mental Health Professional).

29. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. J Am Acad Child Adolesc Psychiatry. 2008 Aug;47(8):921–9.

30. Belardinelli C, Raza M, Taneli T. Comorbid Behavioral Problems and Psychiatric Disorders in Autism Spectrum Disorders. J Child Dev Disord. 2016 Mar 11;2(2).

31. van der Meer JMJ, Oerlemans AM, van Steijn DJ, Lappenschaar MGA, de Sonneville LMJ, Buitelaar JK, et al. Are Autism Spectrum Disorder and Attention-Deficit/Hyperactivity Disorder Different Manifestations of One Overarching Disorder? Cognitive and Symptom Evidence From a Clinical and Population-Based Sample. J Am Acad Child Adolesc Psychiatry. 2012 Nov 1;51(11):1160-1172.e3.

32. Amiet C, Gourfinkel-An I, Bouzamondo A, Tordjman S, Baulac M, Lechat P, et al. Epilepsy in Autism is Associated with Intellectual Disability and Gender: Evidence from a Meta-Analysis. Biol Psychiatry. 2008 Oct 1;64(7):577–82.

33. Bolton PF, Carcani-Rathwell I, Hutton J, Goode S, Howlin P, Rutter M. Epilepsy in autism: features and correlates. Br J Psychiatry. 2011 Apr;198(4):289–94.

34. Steffenburg S, Gillberg C. Autism spectrum disorders in children with active epilepsy and learning disability: comorbidity, pre- and perinatal background, and seizure characteristics. Dev Med Child Neurol. 2003;45(11):724–30.

35. Turk J, Bax M, Williams C, Amin P, Eriksson M, Gillberg C. Autism spectrum disorder in children with and without epilepsy: impact on social functioning and communication. Acta Paediatr. 2009;98(4):675–81.

36. Matson JL, Neal D, Hess JA, Mahan S, Fodstad JC. The effect of seizure disorder on symptom presentation in atypically developing children and children with autism spectrum disorders based on the BDI-2. Dev Neurorehabilitation. 2010 Oct 1;13(5):310–4.

37. Coury DL, Ashwood P, Fasano A, Fuchs G, Geraghty M, Kaul A, et al. Gastrointestinal Conditions in Children With Autism Spectrum Disorder: Developing a Research Agenda. Pediatrics. 2012 Nov 1;130(Supplement 2):S160–8.

38. Black C, Kaye JA, Jick H. Relation of childhood gastrointestinal disorders to autism: nested case-control study using data from the UK General Practice Research Database. BMJ. 2002 Aug 24;325(7361):419–21.

39. Ibrahim SH, Voigt RG, Katusic SK, Weaver AL, Barbaresi WJ. Incidence of Gastrointestinal Symptoms in Children With Autism: A Population-Based Study. Pediatrics. 2009 Aug 1;124(2):680–6.

40. Valicenti-McDermott MD, McVicar K, Cohen HJ, Wershil BK, Shinnar S. Gastrointestinal Symptoms in Children with an Autism Spectrum Disorder and Language Regression. Pediatr Neurol. 2008 Dec 1;39(6):392–8.

41. Nikolov RN, Bearss KE, Lettinga J, Erickson C, Rodowski M, Aman MG, et al. Gastrointestinal Symptoms in a Sample of Children with Pervasive Developmental Disorders. J Autism Dev Disord. 2009 Mar 1;39(3):405–13.

42. Mazurek MO, Vasa RA, Kalb LG, Kanne SM, Rosenberg D, Keefer A, et al. Anxiety, Sensory Over-Responsivity, and Gastrointestinal Problems in Children with Autism Spectrum Disorders. J Abnorm Child Psychol. 2013 Jan 1;41(1):165–76.

43. Zerbo O, Leong A, Barcellos L, Bernal P, Fireman B, Croen LA. Immune mediated conditions in autism spectrum disorders. Brain Behav Immun. 2015 May 1;46:232–6.

44. Mostafa G, Hamza R, El-Shahawi H. Allergic manifestations in autistic children: Relation to disease severity. J Pediatr Neurol. 2015 Jul 30;06(02):115–23.

45. Vargas DL, Nascimbene C, Krishnan C, Zimmerman AW, Pardo CA. Neuroglial activation and neuroinflammation in the brain of patients with autism. Ann Neurol. 2005;57(1):67–81.

46. Li X, Chauhan A, Sheikh AM, Patil S, Chauhan V, Li X-M, et al. Elevated immune response in the brain of autistic patients. J Neuroimmunol. 2009 Feb 15;207(1):111–6.

47. Han YMY, Cheung WKY, Wong CK, Sze SL, Cheng TWS, Yeung MK, et al. Distinct Cytokine and Chemokine Profiles in Autism Spectrum Disorders. Front Immunol. 2017 Jan 23;8.

48. Tsai F-J, Chiang H-L, Lee C-M, Gau SS-F, Lee W-T, Fan P-C, et al. Sleep problems in children with autism, attention-deficit hyperactivity disorder, and epilepsy. Res Autism Spectr Disord. 2012 Jan 1;6(1):413–21.

49. Goldman SE, Richdale AL, Clemons T, Malow BA. Parental Sleep Concerns in Autism Spectrum Disorders: Variations from Childhood to Adolescence. J Autism Dev Disord. 2012 Apr 1;42(4):531–8.

50. Wiggs L, Stores G. Sleep patterns and sleep disorders in children with autistic spectrum disorders: insights using parent report and actigraphy. Dev Med Child Neurol. 2004 Jun;46(06).

51. Godbout R, Bergeron C, Limoges É, Stip E, Mottron L. A laboratory study of sleep in Asperger's syndrome. NeuroReport. 2000 Jan;11(1):127.

52. Paavonen EJ, Vehkalahti K, Vanhala R, von Wendt L, Nieminen-von Wendt T, Aronen ET. Sleep in Children with Asperger Syndrome. J Autism Dev Disord. 2008 Jan 1;38(1):41–51.

53. Schreck KA, Mulick JA, Smith AF. Sleep problems as possible predictors of intensified symptoms of autism. Res Dev Disabil. 2004 Jan 1;25(1):57–66.

54. Fatemi SH, Folsom TD. Autism Spectrum Disorders and Ataxia. In: Manto M, Schmahmann JD, Rossi F, Gruol DL, Koibuchi N, editors. Handbook of the Cerebellum and Cerebellar Disorders. Dordrecht: Springer Netherlands; 2013. p. 1895–906.

55. Breen J, Hare DJ. The nature and prevalence of catatonic symptoms in young people with autism. J Intellect Disabil Res. 2017;61(6):580–93.

56. Nayate A, Bradshaw JL, Rinehart NJ. Autism and Asperger's disorder: Are they movement disorders involving the cerebellum and/or basal ganglia? Brain Res Bull. 2005 Oct 30;67(4):327–34.

57. Completing the map of human genetic variation. Nature. 2007 May;447(7141):161–5.

58. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet. 2010 Jun;11(6):415–25.

59. de la Torre-Ubieta L, Won H, Stein JL, Geschwind DH. Advancing the understanding of autism disease mechanisms through genetics. Nat Med. 2016 Apr;22(4):345–61.

60. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med Off J Am Coll Med Genet. 2015 May;17(5):405–24.

61. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, et al. Large-Scale Copy Number Polymorphism in the Human Genome. Science. 2004 Jul 23;305(5683):525–8.

62. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004 Sep;36(9):949–51.

63. Malhotra D, Sebat J. CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell. 2012 Mar 16;148(6):1223–41.

64. Sanders SJ, Sahin M, Hostyk J, Thurm A, Jacquemont S, Avillach P, et al. A framework for the investigation of rare genetic disorders in neuropsychiatry. Nat Med. 2019 Sep 23;

65. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. PathoGenetics. 2008 Nov 3;1:4.

66. Moreno-De-Luca A, Evans DW, Boomer KB, Hanson E, Bernier R, Goin-Kochel RP, et al. The role of parental cognitive, behavioral, and motor profiles in clinical variability in individuals with chromosome 16p11.2 deletions. JAMA Psychiatry. 2015 Feb;72(2):119–26.

67. Hippolyte L, Maillard AM, Rodriguez-Herreros B, Pain A, Martin-Brevet S, Ferrari C, et al. The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. Biol Psychiatry. 2016 15;80(2):129–39.

68. Bernier R, Steinman KJ, Reilly B, Wallace AS, Sherr EH, Pojman N, et al. Clinical phenotype of the recurrent 1q21.1 copy-number variant. Genet Med Off J Am Coll Med Genet. 2016 Apr;18(4):341–9.

69. Richter M, Murtaza N, Scharrenberg R, White SH, Johanns O, Walker S, et al. Altered TAOK2 activity causes autism-related neurodevelopmental and cognitive abnormalities through RhoA signaling. Mol Psychiatry. 2019 Sep;24(9):1329–50.

70. RHOA Gene - GeneCards | RHOA Protein | RHOA Antibody [Internet]. [cited 2019 Dec 11]. Available from: https://www.genecards.org/cgi-bin/carddisp.pl?gene=RHOA

71. Blizinsky KD, Diaz-Castro B, Forrest MP, Schürmann B, Bach AP, Martin-de-Saavedra MD, et al. Reversal of dendritic phenotypes in 16p11.2 microduplication mouse model neurons by pharmacological targeting of a network hub. Proc Natl Acad Sci U S A. 2016 26;113(30):8520–5.

72. Jensen M, Girirajan S. An interaction-based model for neuropsychiatric features of copy-number variants. PLOS Genet. 2019 Jan 17;15(1):e1007879.

73. Huguet G, Benabou M, Bourgeron T. The Genetics of Autism Spectrum Disorders. In: Sassone-Corsi P, Christen Y, editors. A Time for Metabolism and Hormones. Springer International Publishing; 2016. p. 101–29. (Research and Perspectives in Endocrine Interactions).

74. Krishnan A, Zhang R, Yao V, Theesfeld CL, Wong AK, Tadych A, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. Nat Neurosci. 2016;19(11):1454–62.

75. Yoo H. Genetics of Autism Spectrum Disorder: Current Status and Possible Clinical Applications. Exp Neurobiol. 2015 Dec;24(4):257–72.

76. Südhof TC. Neuroligins and neurexins link synaptic function to cognitive disease. Nature. 2008 Oct;455(7215):903–11.

77. Sheng M, Hoogenraad CC. The Postsynaptic Architecture of Excitatory Synapses: A More Quantitative View. Annu Rev Biochem. 2007;76(1):823–47.

78. Durand CM, Betancur C, Boeckers TM, Bockmann J, Chaste P, Fauchereau F, et al. Mutations in the gene encoding the synaptic scaffolding protein SHANK3 are associated with autism spectrum disorders. Nat Genet. 2007 Jan;39(1):25–7.

79. Wilson HL, Wong ACC, Shaw SR, Tse W-Y, Stapleton GA, Phelan MC, et al. Molecular characterisation of the 22q13 deletion syndrome supports the role of haploinsufficiency of SHANK3/PROSAP2 in the major neurological symptoms. J Med Genet. 2003 Aug 1;40(8):575–84.

80. Stephan KE, Friston KJ, Frith CD. Dysconnection in Schizophrenia: From Abnormal Synaptic Plasticity to Failures of Self-monitoring. Schizophr Bull. 2009 May 1;35(3):509–27.

81. Geschwind DH, State MW. Gene hunting in autism spectrum disorder: on the path to precision medicine. Lancet Neurol. 2015 Nov 1;14(11):1109–20.

82. Sanders SJ, Ercan-Sencicek AG, Hus V, Luo R, Murtha MT, Moreno-De-Luca D, et al. Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. Neuron. 2011 Jun 9;70(5):863–85.

83. Iossifov I, O'Roak BJ, Sanders SJ, Ronemus M, Krumm N, Levy D, et al. The contribution of *de novo* coding mutations to autism spectrum disorder. Nature. 2014 Nov;515(7526):216–21.

84. Bourgeron T. From the genetic architecture to synaptic plasticity in autism spectrum disorder. Nat Rev Neurosci. 2015 Sep;16(9):551–63.

85. 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 Mar 21;

86. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016 18;536(7616):285–91.

87. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron. 2010 Oct 21;68(2):192–5.

88. Yuen RKC, Merico D, Bookman M, Howe JL, Thiruvahindrapuram B, Patel RV, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci. 2017 Apr;20(4):602–11.

89. 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 Dec;15(12):1128–39.

90. 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;43:11.10.1-33.

91. Wang K, Li M, Hadley D, Liu R, Glessner J, Grant SFA, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res. 2007 Nov;17(11):1665–74.

92. Colella S, Yau C, Taylor JM, Mirza G, Butler H, Clouston P, et al. QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data. Nucleic Acids Res. 2007;35(6):2013–25.

93. 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 Jan 4;102(1):142–55.

94. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164.

95. Taravella Oill AM, Deshpande AJ, Natri HM, Wilson MA. PopInf: An approach for reproducibly visualizing and assigning population affiliation in genomic samples of uncertain origin. Genomics; 2019 Oct.

96. Hapmap Consortium. The International HapMap Project. Nature. 2003 Dec 18;426(6968):789–96.

97. Raven JC, Court JH, Raven J. Raven's Progressive Matrices. Harcourt Assessment; 1998.

98. Bishop SL, Guthrie W, Coffing M, Lord C. Convergent Validity of the Mullen Scales of Early Learning and the Differential Ability Scales in Children With Autism Spectrum Disorders. Am J Intellect Dev Disabil. 2011 Sep 1;116(5):331–43.

99. Elliott CD, Salerno JD, Dumont R, Willis JO. The Differential Ability Scales—Second Edition. In: Contemporary intellectual assessment: Theories, tests, and issues, 4th ed. New York, NY, US: The Guilford Press; 2007. p. 360–82.

100. Mullen EM. Mullen Scales of Early Learning. American Guidance Service Inc.; 1995.

101. Wechsler D. Wechsler Abbreviated Scale of Intelligence. Pearson; 1999.

102. Wechsler D. Wechsler Abbreviated Scale of Intelligence - Second Edition. Pearson; 2011.

103. Wechsler D. Wechsler Intelligence Scale for Children - Fourth Edition. Pearson; 2003.

104. Leiter RG. Leiter international performance scale. Stoelting; 1979.

105. Gale H. Roid, Miller LJ. Leiter international performance scale-revised. Stoelting; 1997.

106. Wechsler D. Wechsler Preschool and Primary Scale of Intelligence - Fourth Edition. Pearson; 2012.

107. Constantino J, Gruber C. Social Responsiveness Scale (SRS). Los Angel West Psychol Serv. 2005;

108. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, et al. Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. J Autism Dev Disord. 2003 Aug 1;33(4):427–33.

109. Bruni TP. Test Review: Social Responsiveness Scale–Second Edition (SRS-2). J Psychoeduc Assess. 2014 Jul 1;32(4):365–9.

110. Therneau TM. coxme: Mixed Effects Cox Models [Internet]. 2018. (R package). Available from: https://CRAN.R-project.org/package=coxme

111. Therneau TM, Sinnwell J. kinship2: Pedigree Functions [Internet]. 2015. (R package). Available from: https://CRAN.R-project.org/package=kinship2

112. Ripley B, Venables B, Bates D, Hornik K, Gebhardt A, Firth D. MASS: Support Functions and Datasets for Venables and Ripley's MASS [Internet]. 2019. (R package). Available from: http://www.stats.ox.ac.uk/pub/MASS4/

113. Hadfield JD. MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. J Stat Softw. 2010 Feb 2;33(1):1–22.

114. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell. 2017 Jun 15;169(7):1177–86.

115. Devlin B, Daniels M, Roeder K. The heritability of IQ. Nature. 1997 Jul;388(6641):468–71.

116. Jacquemont S, Reymond A, Zufferey F, Harewood L, Walters RG, Kutalik Z, et al. Mirror extreme BMI phenotypes associated with gene dosage at the chromosome 16p11.2 locus. Nature. 2011 Aug 31;478(7367):97–102.

117. McCarthy SE, Makarov V, Kirov G, Addington AM, McClellan J, Yoon S, et al. Microduplications of 16p11.2 are associated with schizophrenia. Nat Genet. 2009 Nov;41(11):1223–7.

118. Weiss LA, Shen Y, Korn JM, Arking DE, Miller DT, Fossdal R, et al. Association between microdeletion and microduplication at 16p11.2 and autism. N Engl J Med. 2008 Feb 14;358(7):667–75.

119. Zufferey F, Sherr EH, Beckmann ND, Hanson E, Maillard AM, Hippolyte L, et al. A 600 kb deletion syndrome at 16p11.2 leads to energy imbalance and neuropsychiatric disorders. J Med Genet. 2012 Oct 1;49(10):660–8.

120. Green Snyder L, D'Angelo D, Chen Q, Bernier R, Goin-Kochel RP, Wallace AS, et al. Autism Spectrum Disorder, Developmental and Psychiatric Features in 16p11.2 Duplication. J Autism Dev Disord. 2016 Aug 1;46(8):2734–48.

121. Vorstman JAS, Breetvelt EJ, Thode KI, Chow EWC, Bassett AS. Expression of autism spectrum and schizophrenia in patients with a 22q11.2 deletion. Schizophr Res. 2013 Jan;143(1):55–9.