University of Montreal

Measuring and estimating the effect of copy number variants on autism spectrum disorder and early-onset psychosis risk

A model-based estimation of the effect of copy number variants

by

Elise Douard, M.Sc.

Neurosciences, Université de Montréal

Thesis presented for the grade of Ph.D. in Neurosciences

Neurosciences department, University of Montreal

This thesis is entitled

Measuring and estimating the effect of copy number variants on autism spectrum disorder and early-onset psychosis risk

A model-based estimation of the effect of copy number variants

Presented by
Elise Douard

Evaluated by the following member of the jury:

Sébastien Jacquemont
Research director

Marc Lanovaz
President of the jury

Kimberley Kendall
Jury member

Konrad Karczewski
External jury member

Marie-Pierre Dubé
Dean's representative

To my family

"*Cette science qui devait tout m'apprendre finit dans l'hypothèse, cette lucidité sombre dans la métaphore, cette incertitude se résout en œuvre d'art. Qu'avais-je besoin de tant d'effort ? [...] L'intelligence aussi me dit donc à sa manière que ce monde est absurde*."

Albert Camus

Résumé

Les variations du nombre de copies (*i.e.,* VNC, perte ou gain de matériel génétique de plus de 1 kilobase) figurent parmi les facteurs biologiques les plus associés aux troubles neurodéveloppementaux (TNDs), tels que les troubles du spectre autistique (TSAs) ou la psychose précoce. Les variants génétiques classés comme pathogéniques sont identifiés chez environ 20% des enfants avec des symptômes de TSA référés en génétique clinique. Actuellement, seules les VNCs les plus récurrentes (*i.e.,* plusieurs individus non apparentés ont le même variant) ont été associées avec les TSAs et leurs tailles d'effets ont pu être décrites avec précision grâce à des études d'associations (*i.e.,* cas-contrôles). Cependant, la plupart des VNCs identifiées dans les cliniques neurodéveloppementales et génétiques sont ultra-rares. À ma connaissance, aucune méthode n'a été développée afin d'estimer et de prédire de façon précise la contribution de tels variants aux phénotypes cliniques. De ce fait, l'impact de ces variants ultra-rares sur les risques d'avoir des TNDs, comme les TSAs ou la psychose précoce, reste incertain. Une étude récente de mon groupe de recherche a démontré que les tailles d'effet des délétions et duplications à travers le génome sur les capacités cognitives pouvaient être prédites statistiquement avec 78% de précision en utilisant des mesures d'intolérance à la perte de fonction. Le but de cette thèse est de développer des modèles similaires pour définir les tailles d'effet des VNCs à travers le génome sur les risques de TSA et de psychose précoce, ainsi que sur quelques traits cognitifs et comportementaux affectés dans ces troubles.

J'ai analysé tous les VNCs ≥ 50 kilobases identifiées via les données de puces de génotypage et de séquençage sur génome entier chez 137 enfants et adolescents avec une psychose précoce (Boston Children's hospital), 5,540 probands avec des TSAs (Simons Simplex Collection et MSSNG), et 17,471 personnes de la population générale (Lothian birth cohort, Generation Scotland, IMAGEN et Saguenay Youth Study). Les gènes codants totalement compris dans les VNCs ont été annotés avec neufs variables quantitatives, incluant le score d'intolérance à la perte de fonction et d'autres scores

fonctionnels et génétiques. Des modèles statistiques incluant ces scores ont été testés afin de sélectionner celui qui explique le mieux l'effet des VNCs à travers le génome sur le risque de TSA et le quotient intellectuel (QI). Le meilleur modèle a été utilisé par la suite pour investiguer les tailles d'effets des VNCs sur d'autres traits cognitifs et comportementaux liés aux TSAs, ainsi que sur le risque de psychose précoce.

Le score d'intolérance à la perte de fonction expliquait le mieux les effets des VNCs sur le risque de TSA et la cognition générale. Les modèles incluant ces scores ont démontré que les délétions et les duplications augmentaient les risques de psychose précoce et de TSA, même après ajustement pour le QI. Il n'y avait aucune différence de tailles d'effets des VNCs entre la psychose précoce et le TSA. La fréquence de loci associé précédemment avec des TNDs et des troubles neuropsychiatriques était également similaire entre dans les TSA et la psychose précoce, et le modèle estimait précisément la taille d'effet de la plupart de ces loci sur le risque de TSA en comparaison aux observation empiriques publiées précédemment. Les CNVs à travers le génome mesurés par le score d'intolérance à la perte de fonction diminuaient de façon similaire le QI dans les populations TSA et générale. Les effets des duplications étaient systématiquement plus faibles que les effets des délétions pour chacun de ces phénotypes, ce qui suggère un effet plus pathogénique des délétions. Les délétions et les duplications affectaient différentiellement la communication sociale, les comportements, et la mémoire phonologique, tandis qu'elles affectaient similairement les capacités motrices dans les populations TSA.

L'enrichissement similaire des VNCs à travers le génome dans la psychose précoce et le TSA suggère un effet pléiotropique des VNCs dans ces différentes symptomatologies. Le dépistage routinier pour les VNCs doit être accessible dans les soins cliniques standards des jeunes avec une psychose précoce, comme il est recommandé pour les TSAs. Une telle pratique contribue à établir une médecine personnalisée et peut apporter des bénéfices médicaux comme la détection de comorbidités, la prédiction de la progression de la maladie, et faciliter la communication avec les parents à propos de la nature biologique du trouble. Les modèles appliqués dans ce projet, entraînés sur des VNCs

incluant plus de 4,500 gènes, suggèrent des propriétés hautement polygéniques du dosage génique dans les TNDs. J'ai estimé que chaque VNC de 1 mégabase, incluant au moins un gène scorant pour l'intolérance à la perte de fonction, augmente le risque de TSA. La combinaison de ces résultats ouvre de nouvelles perspectives dans la compréhension des effets des VNCs à travers le génome sur les TNDs et les traits associés (*e.g.*, QI ou symptômes comportementaux). Ces modèles ont été implémentés dans un outil en ligne qui a pour but d'aider les cliniciens à estimer les tailles d'effet des VNCs identifiés en clinique et à interpréter leur contribution au phénotype du patient.

Mots clés : Troubles du spectre autistique, psychose précoce, variation du nombre de copies, pléiotropie, polygénicité.

Summary

Copy number variants (CNVs; *i.e.,* loss or gain of genetic material of over 1 kilobase) are robustly associated with neurodevelopmental disorders (NDDs), such as autism spectrum disorder (ASD) and early-onset psychosis (EOP). Genetic variants classified as pathogenic are identified in approximately 20% of children with ASD symptoms referred to genetic clinics. To date, only the most recurrent CNVs (*i.e.,* similar variants across multiple unrelated individuals) were associated with ASD and their effect-sizes were characterized through association studies (*i.e.,* case-controls). However, most of the CNVs routinely identified in neurodevelopmental and genetic clinics are ultra-rare. To my knowledge, no method was developed to accurately estimate and predict the contribution of such variants to clinical phenotypes. Therefore, the impact of these ultra-rare variants on risk for NDDs, such as ASD and EOP, remains undocumented. A recent study from my research group has shown that the effect-size of genome-wide deletions and duplications on cognitive ability can be statistically predicted with an 78% accuracy using measures of loss-of-function (LoF) intolerance. The aim of this thesis was to develop similar models to define the effect-size of genome-wide CNVs on ASD and EOP risk, as well as on several cognitive and behavioral traits altered in these disorders.

I analyzed all CNVs ≥ 50 kilobases called from genotyping arrays and whole genome sequencing data from 137 children and adolescents with EOP (Boston Children's hospital), 5,540 probands with ASD (Simons Simplex Collection and MSSNG), and 17,471 individuals from unselected populations (Lothian birth cohort, Generation Scotland, IMAGEN and Saguenay Youth Study). Coding genes fully encompassed by CNVs were annotated with nine quantitative variables, including the LoF intolerance score and other functional and genetic scores. Statistical models including these scores were tested to select the one that best explained the effects of genome-wide CNVs on ASD risk and IQ. The best model was subsequently used to investigate the effect-size of genome-wide CNVs on cognitive and behavioral domains related to ASD, as well as on EOP risk.

The LoF intolerance score best explained the effect-sizes of genome-wide CNVs on ASD-risk and general cognition. Models including such scores demonstrated that deletions or duplications increased risks for EOP and for ASD, even after adjusting for IQ. There was no difference in effect-sizes between EOP and ASD. The frequency of loci previously associated with NDDs or neuropsychiatric disorders was also similar between EOP and ASD, and the model accurately estimated the effect-size of most of these loci on the risk for ASD comparing to previously published empirical observations. Genome-wide CNVs measured by LoF intolerance score also similarly decreased IQ in both ASD and unselected populations. The effect of duplications was smaller than the effect of deletion for all phenotypes investigated, suggesting a higher pathogenicity of deletions. Deletions and duplications were found to differentially affect social communication, behavior, and phonological memory, whereas both equally affected motor skills in the ASD population.

The identical enrichment of genome-wide CNVs in EOP and ASD suggests a pleiotropic effect of CNVs in these different symptomatology. Routine screening for CNVs should be made available in the standard clinical care for EOP youth, as is recommended in ASD. Such practice contributes to the establishment of personalized medicine and may bring medical benefits as detecting medical comorbidities, prediction of the disease progression, and facilitating the communication with parents about the biological nature of the disorder. The models applied in this project, trained on CNVs encompassing more than 4,500 genes, suggest highly polygenic properties of gene dosage in NDDs. I estimated that any 1 megabase CNV, encompassing at least one gene scoring for intolerance to LoF, would increase ASD risk. Overall, these results open new avenues for understanding the effect of genome-wide CNVs on NDD risk and related traits (*e.g.*, IQ or behavioral symptoms). These models were implemented in an online tool which aims to help clinicians estimate the effect-size of CNVs identified in the clinic and interpret their contribution to the patient's phenotype.

Keywords: Autism spectrum disorders, early-onset psychosis, copy number variants, pleiotropy, polygenicity.

Table of content

Figures

Tables

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
ADI: autism diagnosis interview
ADOS: autism diagnosis observation schedule
ARID1B: AT-rich interaction domain 1B
ASD: autism spectrum disorder
BD: bipolar disorder
BMI: body mass index
BP: breakpoint
CHD8: chromodomain helicase DNA binding protein 8
CNTN5: contactin 5
CNV: copy number variant
CORO1A: coronin-1A
COS: childhood-onset schizophrenia
CT: cortical thickness
CTOPP: comprehensive test of phonological processing
DCD: developmental coordination disorder
DCDQ: developmental coordination disorder questionnaire
DGCR8: DiGeorge syndrome critical region gene 8
DSM: diagnostic and statistical manual
EOP: early-onset psychosis
FMR1: fragile X mental retardation 1
FOXP2: forkhead-box P2
GWAS: genome-wide association studies
HIC2: HIC ZBTB transcriptional repressor 2
HIRA: histone cell cycle regulator
ID: intellectual disability
iPSYCH: integrative psychiatric research
IQ: intellectual quotient
KCTD13: potassium channel tetramerization domain containing 13
LBC: Lothian birth cohort
LCR: low copy repeats
LGD: likely gene-disruptive
LOEUF: loss-of-function observed/expected upper bound fraction
LoF: loss-of-function
MAPK3: MAP kinase, ERK1

MAZ: MYC associated zinc finger protein
MED15: mediator complex subunit 15
MRI: magnetic resonance imagery
NDD: neurodevelopmental disorder
NLGN3/4: neuroligin 3/4
NRXN1: neurexin 1
OR: odds Ratio
PDD: pervasive developmental disorder
pLI: probability of being loss-of-function intolerant
PRS: polygenic risk score
RNA: ribonucleic acid
RTN4R: reticulon 4 receptor
SCARF2: scavenger receptor expressed by endothelial cells 2 protein (SREC2)
SCARF2: scavenger receptor class F member 2
SD: standard deviation
SFARI: Simons foundation autism research initiative
SHANK3: SH3 and multiple ankyrin repeat domains 3
SNP: single nucleotide polymorphism
SNV: single nucleotide variant
SPARK: Simons foundation powering autism research for knowledge
SRS: social responsiveness scale
SSC: Simons simplex collection
SYS: Saguenay youth study
TAOK2: thousand and one amino-acid kinase 2
TBX1: T-box transcription factor
UFD1L: ubiquitin fusion degradation protein 1-like
VABS: vineland adaptive behavior scales
VEOP: very early-onset psychosis

Remerciements

Merci à mon directeur de thèse, Sébastien Jacquemont, de m'avoir permis de vivre cette épopée canadienne formatrice en m'accueillant dans son jeune laboratoire Montréalais en 2016. Tu as réussi à créer un environnement stimulant tant par tes idées novatrices que par les personnes qui composent ton équipage sur place et au-delà des frontières. Merci de m'avoir équipée de pagaies de qualité pour surfer sur les vagues, parfois imprévisibles, de cette thèse, mais aussi de rivières escarpées !

Merci à mon ancienne directrice, Monica Zilbovicius, qui m'a donné l'opportunité de mettre les pieds en recherche clinique, au sein de cette belle famille qu'elle a constitué à Necker ; et qui a su me transmettre la passion pour cet environnement qui me tient tant à cœur. Cette entrée en matière n'aurait pas été aussi sympathique et humainement riche sans Ana Saitovitch, Elza Rechtman, Alice Vinçon-Leite et Pierre Augé.

À Thérèse Collins et Hannah Englitz, je vous remercie d'avoir cru en moi et de m'avoir ouvert les portes vers les sciences cognitives ; ainsi qu'à tout le personnel du Cogmaster grâce à qui ce programme est si enrichissant. Merci d'encourager vos étudiants à laisser place à leur créativité.

Thanks to my jury members Kimberely Kendall, Konrad Karczewski and Marc Lanovaz, who accepted to review this thesis, and whose work played a foundational role in it.

Merci à mon comité de *marrainage*, Sarah Lippé et Mayada Elsabbagh de m'avoir conseillé pour les premières années de thèse.

A huge thanks to David Glahn and Catherine Brownstein who received us at Boston, helped me to reach this goal by merging our projects, and who brought a new perspective to this thesis' experiment! Also, thanks for your encouragements in continuing in this field.

Merci à tous ceux qui contribuent au succès de ce laboratoire : nos admin' rayonnantes Laura et Marine, ainsi que Mor et l'équipe de biostat', Kristian, la super team BC. Also, to all the PhD students including Rackeb and Zoe (thank you for your help and kindness! You will ace your thesis!). Mais aussi aux anciens, incluant Abderrahim (pour le co-parenting et cet exemple de détermination et résilience lors de ton épopée en médecine !), Sabrina, Clara, Marie-Pier.

À celles et ceux qui sont devenus des ami(e)s et qui ont contribué à ensoleiller ce périple:

Catherine et Flavien Schramm-Alleaume, merci pour votre bienveillance, votre partage et votre gentillesse. Vous avez adouci le début de thèse et je suis infiniment reconnaissante de cet accueil chaleureux dans vos vies. Le savez-vous ? Vos missives sont gardées bien précieusement !!

Petra Tamer, que je ne saurais jamais assez remercier pour son soutien et sa sagesse à toute épreuve, et que j'aimerais soutenir tout autant. Tu es un exemple de maturité et de persévérance. One of the most kindhearted people I met in my whole life, thanks for being so inspiring.

À Guillaume Huguet, sans qui l'expérience n'aurait définitivement pas eu le même goût ! Merci pour ta gentillesse, ton support et ta patience tous trois sans failles, même dans les plus gros pics de la courbe émotionnelle de cette thèse. Merci d'avoir saupoudré le tout de paillettes d'humour et de magie; et d'avoir contribué à effacer ces taches persistantes que j'appelle doutes. Et merci à Alan et toi, pour vos encouragements et vos conseils plus qu'appréciés.

À Zohra Saci, qui a un cœur plus gros qu'elle ne saurait l'admettre, et qui, en plus d'effectuer un travail irréprochable et d'être une super-maman, a pris le temps d'encourager JCVD dans ses pertes de vitesse. Merci, pour m'avoir averti de mettre un peu de bon sens de côté avant qu'il ne disparaisse tout à fait ! But... normal is boring, isn't it?

Merci également à ces ami(e)s que je suis reconnaissante d'avoir croisés sur cette route, avec qui j'ai partagé des moments de joies et de peines, et qui m'ont changé les idées en m'ouvrant à d'autre horizons culturels, philosophiques et spirituels :

À Louis-Joseph, pour ta franchise, ta loyauté, ton empathie. Quelle chance d'avoir rejoint ta famille de cœur après les montagnes russes… Merci pour ton support sans faille. Hâte de poursuivre les prochaines aventures et de raviver ma flamme créative dans ton beau projet. Et aussi, merci à cette joyeuse bande de mauvais chaotiques pour des jeudis de pandémie moins mornes !

À Jeanne, qui a une place particulière dans mon cœur et qui m'a ouvert les yeux sur un monde magique. Merci pour les festivités et ta tendresse.

À Lucas, pour ton accueil chaleureux dans ton univers, les concerts et le brin de folie.

Aux colocs qui m'ont fait découvrir la vie montréalaise et qui m'ont fait grandir en me confiant leurs expériences de vie : Francis, pour sa jovialité, son écoute et ces heures de réflexions philosophiques; mais aussi Sabrina, Marie, et Thien.

Merci à mes amis parisiens, avec qui les retrouvailles sont toujours festives, qui ont un cœur généreux, et dont les parcours m'inspirent à devenir une meilleure humaine: Carine, Julia (bien hâte de rencontrer tes deux plus belles créations), Robin, Morgan, Nidal, Clara.

Merci à Nadège et Matthias qui me font sembler que le temps n'altère rien.

Surtout, merci à ma famille.

À celle qui m'a adopté, bras ouverts, avec Suzanne et Claude, ainsi que les êtres généreux et brillants qui les entourent.

À mes parents, Sylvaine et Maurice, qui acceptent mes décisions de vie les plus excentriques et me soutiennent coûte que coûte. Merci pour l'éducation que vous nous avez apportée, votre amour, et cet environnement si riche que vous avez créé et dans lequel vous nous avez baignés. Quel bonheur de retrouver cette liberté réconfortante dans l'Art et la nature, et ce, grâce à vous... Je vous aime.

À ceux qui ne sont plus, mais dont chacune de mes cellules sont imprégnées : Stan et Jeanine. L'un pour m'avoir inculqué cette curiosité pour les nano- et les macro-éléments de notre univers et l'importance de l'*otium*. L'autre pour m'avoir transmis cette interprétation hédoniste de la vie. Tous deux, pour leur altruisme.

À mon frère, Romain, qui est le plus gros pôle (magnétique!) de ma vie malgré lui. Merci de m'induire ces réflexions, de me faire reconsidérer mes points de vue et ma façon de vivre cette grande aventure. Tu es la personne la plus courageuse que je connaisse. Je t'admire pour ta persévérance à t'échapper des nids de coucous en transcendant ce système, et je te souhaite de voler, épanoui, le plus loin possible de ceux-ci.

Enfin, à Martin, que j'ai rencontré sur la route il y a cinq ans. Merci de m'avoir accompagné avec ta chaleur et ta lumière dans ces années de transformation et de réalisation intenses. Quelle aventure... j'ai hâte d’ouvrir le prochain tome et de discuter de son contenu avec toi.

I. Introduction

Autism spectrum disorder (ASD) and early-onset psychosis (EOP) are neurodevelopmental disorders (NDDs) that occur during the first 2 decades of life. (Rapoport, Giedd, and Gogtay 2012; American Psychiatric Association 2013) These disorders are clinically heterogeneous and frequently co-occur with other NDDs and psychiatric diagnoses, such as intellectual disabilities (ID), language disorders or motor disorders. (American Psychiatric Association 2013; Downs et al. 2017)

The heritability of NDDs ranges from 50 to 80%. (Colvert et al. 2015; Sandin et al. 2017; Hilker et al. 2018; Akingbuwa et al. 2021) During the last decade, chromosomal microarray (CMA) has been recommended as a first-tier clinical test for children with a broad range of NDDs including ASD. However, genetic screening is not systematically recommended for all NDDs. In particular, there are no recommendations for children with EOP.

At my institution (Ste-Justine hospital), 3500 CMAs are performed annually. After ID, ASD is the most frequent diagnosis referred for pediatric genetic services. The routine implementation of CMA in clinic has allowed genome-wide detection of copy number variants (CNVs), which are either deletion or duplication of genomic fragments larger than 1 kilobase (kb). (Feuk, Carson, and Scherer 2006) CMA rapidly expanded the known number of genomic loci associated with ASD and deleterious CNVs are identified in 10 to 15 % of children with NDDs. (Miller et al. 2010)

Most studies on CNVs in NDDs are focused on the recurrent ones (Malhotra and Sebat 2012; Sanders et al. 2019), because they are detected in multiple unrelated patients and can be described through association studies. However, most CNVs reported back to patients are ultra-rare, as they are observed only once or a few times in the patient population. Because there is limited data characterizing and quantifying the effect-size of these ultra-rare CNVs on risk for ASD and EOP, and related cognitive or behavior

disturbances, it is impossible to reach the statistical power required to investigate them in individual association studies. There is currently no strategy to study the effects of the extreme diversity of ultra-rare variants. As the effects of ultra-rare variants are undocumented, it is unlikely for clinicians to estimate the precise contribution of such CNVs to neurodevelopmental symptoms in a given patient.

In this thesis, I will propose a new framework to investigate the effect-size of genome-wide CNVs on ASD while taking into account the clinical heterogeneity instead of reducing it. I will study the contribution of genome-wide CNVs to ASD risk, as well as cognitive and behavioral traits that are thought to be related to this condition. Once this methodology is validated for the study of ASD risk, I will apply a similar approach to investigate the contribution of such genetic variants to the most severe and rare form of psychosis — the EOP.

Clinical characterization of ASD and EOP

The shifting clinical boundaries of autism spectrum disorders

Evolution of ASD classification

The classification of ASD has evolved over time; from the merging and splitting of clinical features first described as a core symptom of schizophrenia by Eugene Bleuler in 1911 (Bleuler 1911), to the clinically accepted definition today. In 1943, Leo Kanner borrowed the term "autism" from Bleuler and was the first to consider it as a distinct disorder from schizophrenia. (Kanner 1943) Since the fifth edition of the diagnostic and statistical manual (DSM-V) released in 2013, the autism spectrum unifies three previously separate but highly related diagnoses from the previous edition (DSM-IV): autistic disorder, Asperger disorder, and pervasive developmental disorders-not otherwise specified. (American Psychiatric Association 2013)

ASD is now characterized by persistent deficits in social communication and interaction across multiple contexts, as well as restricted and repetitive patterns of behaviors,

interests or activities (Figure 1). A formal diagnosis can be made only when these deficits appear before the age of 3.

Psychometric instruments delineating ASD diagnostic

To confirm the diagnosis, clinicians use several standardized neuropsychological tools (Figure 2), such as the autism diagnostic interview (ADI) (Le Couteur et al. 1989; Lord, Rutter, and Couteur 1994; Rutter, Le Couteur, and Lord 2003) – a structured interview of retrospective questions administered to parents.

The autism diagnostic observation schedule (ADOS) (Lord et al. 1989; Lord et al. 2000; Lord et al. 2012), which is a semi-structured cognitive assessment adapted to the developmental-level of the patient, is also used as a diagnostic tool.

Beyond these diagnostic tools, the social responsiveness scale (SRS) is also an extensively used questionnaire ascertaining the presence and severity of autistic traits (Constantino et al. 2003; Constantino and Gruber 2012).

Demographic characteristics of ASD

The prevalence of autism has drastically increased in the last 20 years, from 1 in 250 individuals in 2000, to 1 in 59 presently (Figure 2). (Christensen 2018) This increase was thought to be related to the extension of the diagnosis criteria over the DSM versions, but since there is an 86% overlap in case counts between the IVth-revised and the Vth editions, other factors may contribute to this phenomenon, such as the increased accessibility to mental health services, systematic early screening procedures for ASD diagnosis and public awareness campaigns since the last two decades. The sex ratio of males to females is of 4:1 in clinical sample and decreases to 3:1 in the general population. (Loomes, Hull, and Mandy 2017) Despite this difference, both genders seem to have similar progression of symptoms and prevalence of language regression. (Fountain, Winter, and Bearman 2012; Barger, Campbell, and McDonough 2013; Pearson et al. 2018)

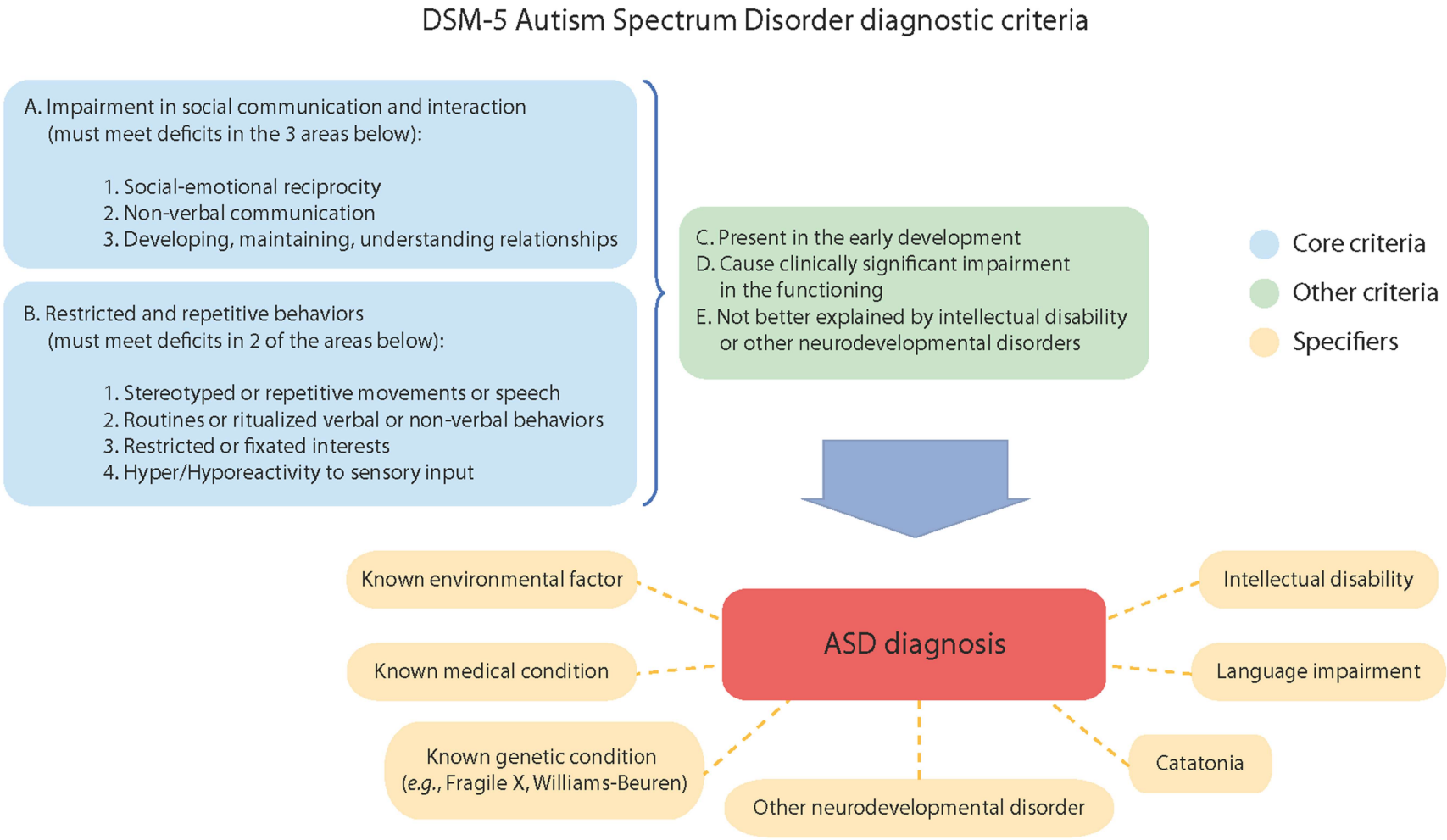


Figure 1: Definition of autism spectrum disorder from the DSM-5

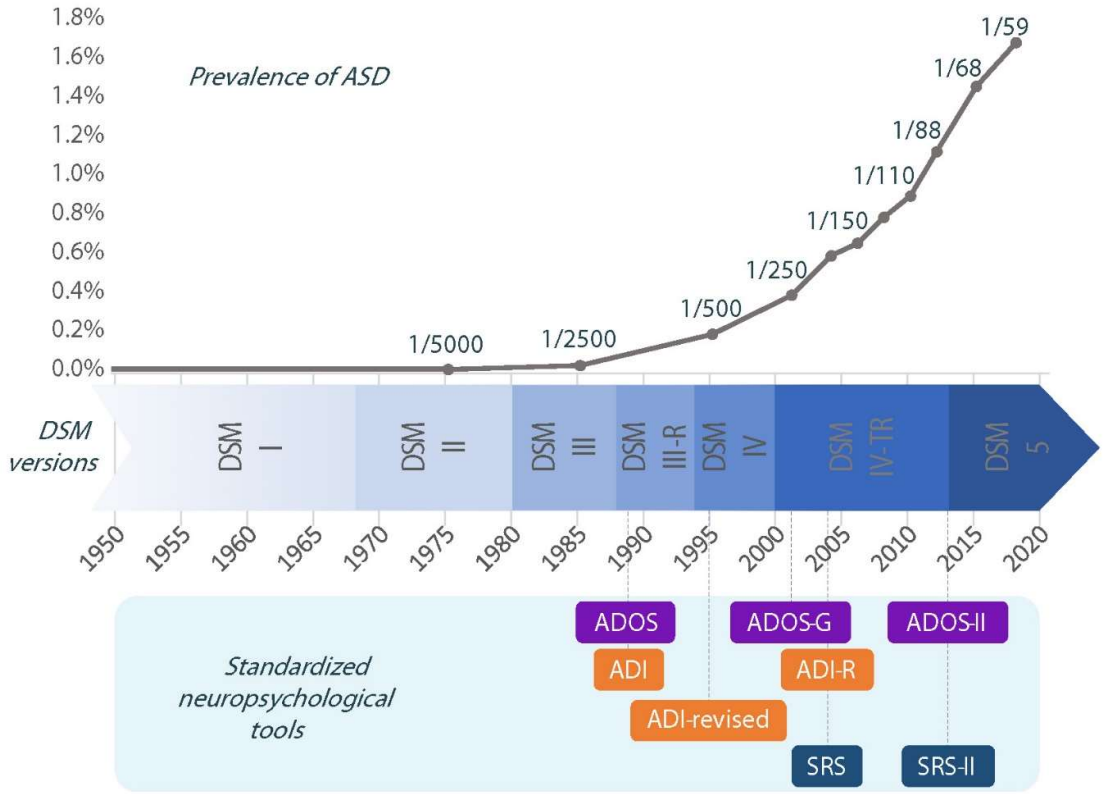


Figure 2: Evolution of the ASD prevalence, DSM-versions, and neuropsychological tools available to quantify ASD traits

The prevalence of ASD is based on Centers for Disease Control and Prevention data published since 1975. (Christensen 2018)

Comorbidities defined as specifiers of ASD

ASD occurs rarely in isolation and is characterized by an extreme clinical heterogeneity. To avoid the creation of subgroups and to take into account this heterogeneity in the diagnosis, the DSM-V recommends the use of "specifiers" to record the severity of cardinal symptoms, current language and intellectual ability, onset age and environmental/acquired conditions (Figure 1). (American Psychiatric Association 2013) The specifiers of ASD also include the recording of any known genetic syndromes (*e.g.*, Fragile X syndrome, DiGeorge syndrome, Williams-Beuren syndrome) or medical conditions, and other NDDs. The most prevalent specifiers are ID and language impairment. Out of the specifiers, motor disorders are also highly prevalent in ASD.

The clinical heterogeneity of ASD and the high prevalence of comorbidities are a challenge when studying its etiology, notably when investigating its genetic architecture.

NDD comorbidities increasing the challenge in ASD modelling

The complex relationship between intellectual disability, IQ, and ASD

ID is the main specifier of ASD, which also occurs during the developmental period. This disorder is diagnosed in 19% to 48% in ASD cases and ranges from mild to severe. (Braun et al. 2015; Croen et al. 2015; Postorino et al. 2016; Rydzewska et al. 2018) However, many individuals with ASD have above-average intelligence quotient (IQ), as well as high levels of academic and occupational functioning. (Duncan and Bishop 2015; Kim, Bal, and Lord 2018)

ID is characterized by deficits in both intellectual and adaptive functioning relative to peers of the same age, gender, linguistic and social background. (American Psychiatric Association 2013)

This disorder has a high heritability and the world-wide prevalence in the pediatric population has been estimated at up to 1.3% (Westerinen et al. 2017; Plomin and von Stumm 2018; McGuire et al. 2019), and the prevalence of severe ID is approximately 1 per 1000 births. (American Psychiatric Association 2013) Males are more likely to be diagnosed compared to females, with a sex ratio of 1.8:1. (McGuire et al. 2019) Sex-linked genetic factors and male vulnerability to brain injury may account for a small portion of these gender differences. (Maulik et al. 2011; Lubs, Stevenson, and Schwartz 2012; Printzlau, Wolstencroft, and Skuse 2017)

Commonly, IQ tests are used to measure intellectual functioning deficits, and the Vineland adaptive behavior scales (VABS) assesses adaptive functioning deficits. (Sparrow, Balla, and Cicchetti 1984; Sparrow, Balla, and Cicchetti 2005; Boat et al. 2015)

Thus, a diagnosis of ID can be made when standardized IQ and Vineland scores are 2 standard deviations (SDs) below the population mean[1] (Figure 3).

The severity of ID is divided into four levels: Mild (85% of cases), moderate (10% of cases), severe (3.5% of cases), and profound (1.5% of cases) (Table 1). (Boat et al. 2015)

The bimodal or trimodal IQ distribution often observed in ASD populations (Figure 3) is not observed in other psychiatric conditions such as schizophrenia. This high distribution of cases in several distinct IQ categories may suggest different subtypes of ASD with different underlying biological mechanisms. For example, most cases of prodigies or high functioning children with ASD express exacerbated attention to details and problems in communication or social skills. (Chiang et al. 2014; Yahya 2020)

General intelligence may be an important dimension in ASD. Many of the core social communication deficits that characterize ASD may lead to failure in the acquisition of developmentally expected skills and, therefore, are expected to be present to some extent in individuals with ID. (Thurm et al. 2019) The DSM-V classification of ID severity is based on adaptive functioning across conceptual, social, and practical domains, and the social ability expectations defined as not being met, overlap with the social deficits that define ASD. (American Psychiatric Association 2013)

However, even the most comprehensive and well-researched tools used in the diagnosis of ASD, such as the ADI and the ADOS, are far less specific when used for individuals with very low intellectual functioning. (Thurm et al. 2019; Myers et al. 2020) An unintended consequence is that people with co-occurring severe ID may not be included in ASD research, which reduces the generalizability of results.

The relationship between ASD and intelligence complexifies the delineation of whether social communication and interaction deficits are beyond what can be attributed to ID.

[1] This corresponds to a score of less than 70 points (generally ±5 points) for both tests which share the same normal-shaped distribution.

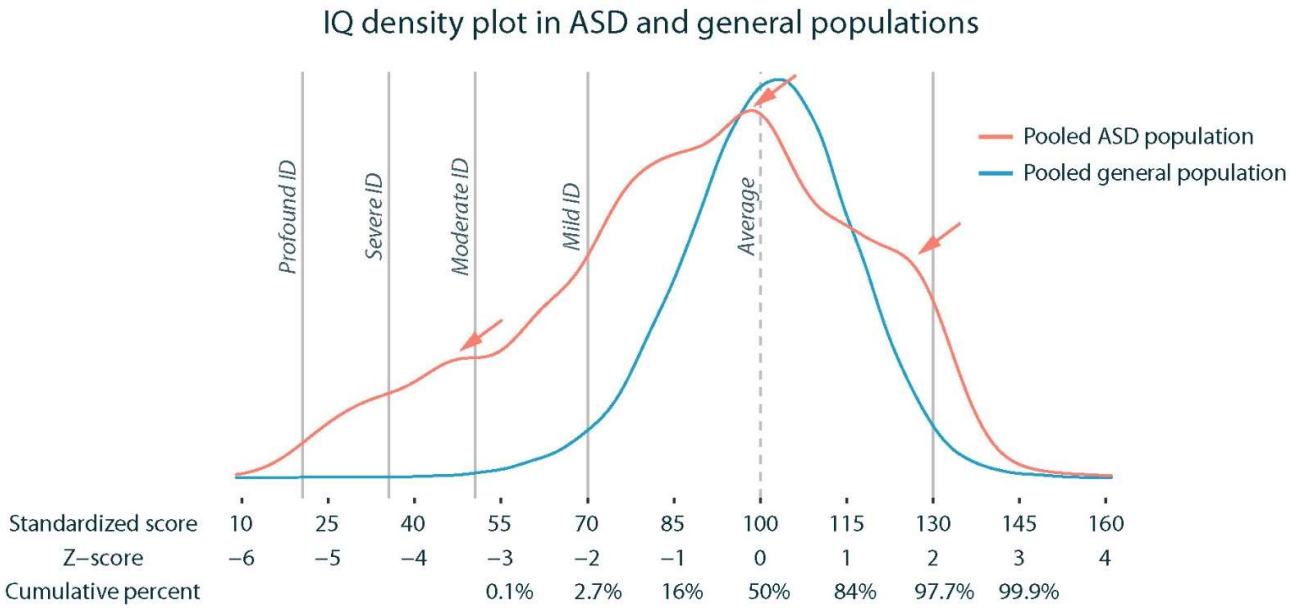


Figure 3: IQ distribution in the general population and in ASD populations

Distribution of IQ in large pooled ASD and general populations. The pooled ASD population (red) combines IQ data from Simons foundation powering autism research for knowledge (SPARK, N=2,111), Simons simplex collection (SSC, N=2,569) and MSSNG (N=1,394) datasets. The pooled general population group (blue) combines IQ data from Lothian birth cohort (LBC, N=504), Generation Scotland (N=13,745), Imagen (N=1,757), and Saguenay youth study (SYS, N=1,590) datasets. The mean IQ score defined with the reference populations is 100 points of IQ (dashed grey line), and the standard deviation is of 15 points of IQ. Intellectual deficiency (ID) is 2 standard deviations below the reference population mean, including a margin of error (generally ±5 points) The levels of ID severity (delimited by the grey lines) are based on the DSM-IV classification: IQs from 50 to 70 = mild ID; IQs from 36 to 49 = moderate ID; IQs from 20 to 35 = severe ID; and IQs below 20 are = profound ID. Red arrows point to the trimodality of the ASD curve observed in distinct IQ categories.

Table 1: Levels of ID severity

Severity Category	Distribution of Cases	DSM-V Criteria (severity classified on the basis of daily skills)	DSM-IV Criteria (severity levels based only on IQ)
Mild	85%	Can live independently with minimum levels of support.	IQ range 50–70
Moderate	10%	Independent living may be achieved with moderate levels of support, such as those available in group homes.	IQ range 36–49
Severe	3.5%	Requires daily assistance with self-care activities and safety supervision.	IQ range 20–35
Profound	1.5%	Requires 24-hour care.	IQ <20

Adapted from Boat *et al.* (2015). In the DSM-V, the levels of severity are defined based on adaptive functioning over IQ score, because it determines the level of support required. IQ measures are less valid in the lower IQ range, but in the DSM-IV they were classified in function of the IQ score. (Bell 1994)

Clinical overlaps between ASD and language disorders

Language disorders are the most prevalent communication disorders in ASD and are considered as one of the main specifiers of the diagnosis. These impairments are diagnosed in 10% to 30% of ASD cases. (D. V. M. Bishop, 2010; Bennett et al., 2014; T. W. Frazier, Georgiades, et al., 2014) Approximately 30% of children diagnosed with ASD are non-verbal (Tager-Flusberg, Paul, and Lord 2005; Boterberg et al. 2019). These children acquire no spoken language (*i.e.*, fewer than 30 words) or present a regression of language, which is an abrupt or gradual loss of previously acquired skills after apparently normal language development.

Language impairments affect the form, function, and use of a conventional system of symbols (*e.g.*, spoken words, sign language, written words, pictures), which are usually organized in a rule-governed manner for communication. (American Psychiatric Association 2013)

Epidemiological studies estimate that approximately 7% of children starting school have clinically significant language disorders of unknown origin (Law et al. 2000; Norbury et al. 2016), and the sex ratio in this population is of 1.8:1 in favor of males. (Pinborough-Zimmerman Judith et al. 2007; Rudolph and Leonard 2016) An additional 2.3% of children experiencing language impairments as part of a pervasive NDD such as ID or ASD. (Norbury et al. 2016)

Language impairments can be due to deficits in comprehension or the production of vocabulary, sentence structure and discourse, which are unrelated to sensory (*e.g.*, hearing), motor, or other medical conditions (*e.g.*, ASD or ID).

ASD diagnosis also involves impairments of communication, but these are much broader, affecting pragmatics (*i.e.,* the appropriate use of language in context), as well as nonverbal communication (*e.g.*, deficits in initiating and responding joint attention). Nevertheless, many children with ASD struggle with both structural and functional aspects of communication. (Bishop 2010)

Although ASD and language impairment have traditionally been regarded as distinct disorders, they often involve similar language deficits, raising the question of whether this is merely a superficial resemblance, or indicative of a deeper similarity, with overlap in etiology. Tager-Flusberg and Joseph argued that the existence of cases of ASD whose language features resembled those of language impairment suggested overlaps between these disorders at a deeper level, but the precise nature of this overlap has been the subject of debate. (Tager-Flusberg and Joseph 2003; Bishop 2010; Eicher and Gruen 2015; Nudel et al. 2020)

Motor disorders in ASD

Developmental coordination disorders (DCD) are the most common motor disorders in neurotypical children, as well as in children with ASD or ID. They occur in 25% to 80% of ASD cases. (Dewey, Cantell, and Crawford 2007; Emck et al. 2009; Green et al. 2009; Kopp, Beckung, and Gillberg 2010; Bishop et al. 2016; Bhat 2020)

The DCD, previously referred to as dyspraxia, are characterized by a substantial delay in reaching milestones of fine or gross motor skills development and execution in comparison to peers of the same age, gender and education. (American Psychiatric Association 2013) These impairments must be unrelated to other visual or neurological impairments. Among others, secondary consequences may be hypotonia, language impairment, dysgraphia, learning disorder, and sensory processing disorder.

The prevalence in school-aged children has been estimated at 5% to 6%, with a sex ratio of 1.7:1 to 7:1 in favor of males. (Lingam et al. 2010; Blank et al. 2012)

The most common motor impairments in ASD are hypotonia (25% to 51%) and apraxia (34% to 53%), followed by toe-walking (19%), and gross motor delay (9%). (Rapin 1996; Ming, Brimacombe, and Wagner 2007; Matson, Matson, and Beighley 2011)

Numerous studies show an enrichment of motor impairments in ASD compared to neurotypical (*e.g.*, an effect-size of +1.2 z-score to +2.9 z-score on motor skills scales (Fournier et al. 2010; Miyahara 2013)) or other NDD populations, such as attention-deficit/hyperactivity disorder (ADHD) and children with developmental delay. (Pan, Tsai, and Chu 2009; Fournier et al. 2010; Bhat, Landa, and Galloway 2011; Matson, Matson, and Beighley 2011)

Several studies demonstrate that the presence of DCD is a predictor of social impairment (Bhat, Landa, and Galloway 2011; Caeyenberghs et al. 2016; Sumner, Leonard, and Hill 2016; Hawks and Constantino 2020) and is associated with more severe forms of ASD. (Ming, Brimacombe, and Wagner 2007; Green et al. 2009; Kopp, Beckung, and Gillberg 2010; Hawks and Constantino 2020)

Of note, stereotyped and repetitive motor movements in ASD are distinct from the diagnosis of stereotypic movement disorder (which is one of DSM-V motor disorders). The latter is not associated with deficits of social communication and reciprocity that are usually present in ASD.

On the other hand, stereotyped and repetitive motor movements in ASD are diagnosed only when the symptoms are sufficiently severe to become a focus of treatment or if they cause self-injury. Self-injury occurs in 35.8% of ASD cases (Rattaz, Michelon, and Baghdadli 2015), yet there is insufficient data available regarding the prevalence of such stereotypic movement that lead to self-injury.

The clinical delineation and heterogeneity of early-onset psychosis

Definition of EOP symptoms and demographic characteristics

The early-onset variant of psychosis is a rare presentation occurring before the age of 18 and represents up to 19% of the psychotic cases reported in the clinic. (Schimmelmann et al. 2007) Very early-onset psychosis (VEOP) is rarer (1% of individual with psychosis) and refers to an onset before the age of 13 (Schimmelmann et al. 2007)

Psychosis is a common and functionally disruptive condition associated with many psychiatric and neurodevelopmental disorders (*e.g.*, schizophrenia, bipolar disorder (BD)), and is both a contributor to disability and a barrier to normal functioning (*e.g.*, in work, interpersonal relations or self-care).

Psychotic symptoms include hallucinations and delusions. (American Psychiatric Association 2013) Hallucinations are sensory perceptions occurring in the absence of corresponding external or somatic stimuli. (Arciniegas 2015) Delusions are beliefs based on incorrect inferences about reality external to, or about, oneself and maintained firmly despite the presentation of evidence that obviously and incontrovertibly contradicts the belief. (Arciniegas 2015) Psychosis is diagnosed in the patient referred in clinic when hallucinations and/or delusions interferes with the capacity to meet the ordinary demands of life. (Arciniegas 2015)

The category of primary psychotic spectrum disorders includes schizophrenia, affective psychosis (*i.e.*, schizoaffective disorders - bipolar type or depressed type, major depressive disorders with psychotic features and bipolar disorder with psychotic features) and other psychotic disorders (*i.e.*, schizophreniform disorders, brief psychotic disorders, other specified or unspecified schizophrenia spectrum and other psychotic disorder). (Arciniegas 2015)

EOP prevalence in the general population is 5.9 per 100 000 individuals and the sex ratio of male to female is around 1.6:1 in clinical populations. (Boeing et al. 2007; Stentebjerg-Olesen et al. 2016)

Comorbidities in EOP

Children and adolescents with psychosis are more likely to have severe outcomes and a greater number of comorbidities, such as NDDs and other medical disorders. Up to 57% of children with EOP may meet diagnostic criteria for childhood-onset schizophrenia (COS). (Downs et al. 2017)

The most prevalent comorbidities of EOP are ASD (18% to 38% of cases), ADHD (10 to 33% of cases) and ID (10% to 38% of cases) (Stentebjerg-Olesen et al. 2016; Downs et al. 2017; Petruzzelli et al. 2018; Downs et al. 2019). The mean IQ in the EOP population is 1.5 SD below the general population mean and among individuals with ID, 52% have mild ID, 38% have medium ID and 9% are in the severe range. (Petruzzelli et al. 2018; Smelror et al. 2021)

Tools assessing categorical diagnoses and measuring dimensional traits

As pointed out in the previous sections, co-morbidities in ASD and EOP are the rule rather than the exception. In clinical settings, most clinicians will typically extend their assessments beyond diagnostic items to evaluate other traits and to subsequently implement a management plan. This approach remains valuable, as patients recruited will be characterized beyond a single core diagnosis across different dimensions of symptoms, functioning, and social factors.

Over the last two decades, research has strongly advocated for investigating mental disorders along dimensional constructs which transcend diagnostic categories and integrate information across multiple measurement levels. (Thapar, Cooper, and Rutter 2017) To do so, a set of tools has been used in research to capture traits in affected individuals as well as those with subthreshold symptoms. However, the relevance of these dimensions remains debated and categorical diagnoses remain the norm in clinic.

Here, I will describe some of these assessment tools used for my thesis to validate the ASD and EOP diagnoses and to quantify the most common traits altered in ASD.

Assessing ASD diagnosis and quantifying autistic traits

Standardized instruments developed to establish a categorical diagnosis

The ADI consists of a structured interview of retrospective questions addressed to the parents or the caregivers. (Le Couteur et al. 1989; Lord, Rutter, and Couteur 1994) It was developed by Ann Le Couteur *et al*. in 1989 to help clinicians in the assessment of autism diagnosis. It yields a description of past and current functioning, in areas of development related to ASD. This tool allows clinicians to measure the severity of ASD impairment in the reciprocal interactions, the verbal and non-verbal communication, as well as the repetitive, restricted and stereotyped behaviors domains (Table 2). Each item of the main domains is rated on a 4-point severity scale from "not present" to "extremely invasive for the family and children functioning". Diagnosis of ASD is met when each of these three domains score under the cut-offs (Table 2).

Table 2: Domains assessed in the ADI and their corresponding cut-off for the ASD diagnosis

Domain assessed	N items	Cut-offs (points)
Reciprocal social interactions	16	10
Verbal and non-verbal communication	13	8 (7 if non-verbal)
Restricted, repetitive and stereotyped behaviors	8	3

Adapted from Rutter *et al.* (2003).

The ADOS is a semi-structured cognitive assessment adapted to the expressive language-level of the patient. (Lord et al. 1989; Lord et al. 2000; Lord et al. 2012) The first version was created by Catherine Lord *et al.* in 1989 and was introduced as a method of standardizing direct observations of social behavior, communication, and play in children suspected of having autism. It was proposed as a complementary instrument to the ADI. This tool is available in four age-adapted modules, and measures social affect, the severity of repetitive, restricted and stereotyped behaviors, and a total calibrated severity score (Table 3). Diagnosis of ASD is met when the total calibrated severity score is over the cut-off of 3 points.

Table 3: Organization of the four modules age-adapted versions of the ADOS and the corresponding domains assessed

Module	Language level of the patient	Calibrated severity score	Social communication	R.R.S.B.	Social affect
1	Non-verbal or one word	✓	✓	✓	✓
2	Phrase	✓	✓	✓	✓
3	Fluent	✓	✓	✓	✓
4	Fluent adolescents and adults	✗	✓	✓	✗

Domains can be either tested (✓) or not tested (✗) in function of the module used. R.R.S.B.: repetitive, restricted and stereotyped behaviors. Adapted from Lord et al. (2012).

Both tools were developed primarily for research on autism over a range of cognitive levels, from moderate ID to normal intelligence, with training required on each. Since their creation, they have been updated to better suit a wide age range, and for both verbal and non-verbal children. Nowadays, these instruments are complementary and broadly used in clinical and research settings to assess an ASD diagnosis.

Quantifying autistic traits in clinical and general populations

The SRS is an extensively validated quantitative measure used to ascertain the presence and severity of social communicative and repetitive behaviors that characterize ASD across the entire range of severity that occurs in nature. (Constantino et al. 2003; Constantino 2011; Constantino and Gruber 2012) It was developed in 2003 by John Constantino *et al.* for use in the general population, and in educational and clinical settings. This tool was created in response to the lack of established quantitative assessment tools. Since its creation, the SRS has been broadly used in behavior-genetic, epidemiological and intervention studies.

The SRS questionnaire can be filled either by a self-report, by a relative or a teacher. It contains 65 items rated on a 4-point Likert-type scale and is organized in 5 subscales: Social Awareness, Social Cognition, Social Communication, Social Motivation, and Restricted Interests and Repetitive Behavior. The generated total score serves as an index of severity of social deficits in the autism spectrum. This score can be either 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.

The SRS T-score allows for the creation of social impairment categories as follows (Bruni 2014):

- 75 < T-score: Clinically significant deficits in social functioning that interfere with interactions with others;

- 65 < T-score ≤ 75: Moderate, signaling some clinically significant social deficits;

- 60 < T-score ≤ 65: Mild to moderate deficiencies in social behavior;

- T-score ≤ 60: Probably does not have social difficulties indicative of a possible ASD diagnosis.

Instruments developed to establish a diagnosis of EOP

The structured clinical interview for DSM-V (SCID-V) is used to assess EOP diagnosis according to the DSM-V criteria for any current or lifetime axis I psychotic disorder[2] with onset prior to age 18. (First et al. 2015) This semi-structured interview is administered by a trained professional (clinician or researcher). EOP can be evaluated using module B in the core configuration of the SCID-V, which tests for the presence of psychotic and associated symptoms. This section includes the assessment of delusion and hallucination, as well as associated symptoms such as disorganized speech behavior, catatonic behavior, and negative symptoms.

The Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) is a semi-structured interview which can also help in the assessment of EOP. (Kaufman et al. 1996) It is administered by trained professionals to both the parent and child separately. This test serves to diagnose childhood mental disorders in school-aged children from 6 to 18. The first version of the K-SADS was developed by Puig-Antich and Chambers in 1978 and was modeled after the Schedule for Affective Disorders and Schizophrenia, an interview schedule for adults. (Endicott and Spitzer 1978; Puig-Antich and Chambers 1978) The interview is intended to assess both past and current episodes of psychopathology. It is focused on affective disorders, but also covers psychotic disorder. The structured sections cover about 50 psychiatric symptoms, each including 11 observational items. Each item is rated on a seven-point scale ranging from « *not at all »* to « *very extreme* ». The K-SADS has been found to be reliable and valid in multiple research and treatment settings. (Geller et al. 2001)

[2] Axis I disorders include substance use disorders and any mental health conditions other than personality disorders and mental retardation.

Measurement of general intelligence and daily adaptive functioning

General intelligence

IQ is commonly used to measure intellectual functioning and one of the most studied human traits throughout history. Its measurement has been used for controversial purposes before being largely implemented in the clinic.

The English statistician Francis Galton made the first attempt to rate intelligence with a standardized test in 1882. This test was based on measures of reaction time, sensory acuity, muscular power and body proportion. In his biological determinist theory, he believed that intelligence was mainly inherited. He was a pioneer in the research of genetic contribution to human traits through statistical methods. (Galton 1870) In 1905, when school became mandatory, two French psychologists, Alfred Binet and Theodore Simon, invented the Binet-Simon intelligence test under a government request. (Binet and Simon 1905) It was intended to identify children with mental retardation in school and provide them special care and education classes. This intelligence measure was based on mental age which was derived from the average performance for a particular age range in a reference population. In 1912, William Stern adapted this concept of mental age to create the concept of IQ. He developed the ratio IQ by dividing the mental age by the chronological age. (Stern and Whipple 1912) In 1916, Lewis Terman developed the Stanford-Binet test, which combined the method of Binet-Simon and William Stern. (Terman 1916) He tested this method in a longitudinal study including more than 1,400 high functioning individuals, from 1921 to his death in 1955 (the study continued until 1986). (Burks, Jensen, and Terman 1930) He believed in the heritability of intelligence and was a member of numerous eugenic organizations. In 1939, David Wechsler created the eponym tests of intelligence which measures ability in various features such as verbal comprehension, non-verbal reasoning, working memory, and information speed processing. (Wechsler 1939) This test measures the deviation IQ, which is the score of the individual compared to the average score of a reference

population in the same age range. The Wechsler tests have been adapted to multiple age ranges since its creation and are still largely used.

Nowadays, intellectual abilities are measured using standardized tests according to the cognitive level of the participant. Given the abstract nature of its concept, absolute measures of intelligence cannot be achieved. Consequently, IQ tests only allow an estimation of an individual's intelligence. The population mean is centered on 100 points and the SD is of 15 points (Figure 3).

Because individuals with ASD are either non-verbal or have a language impairment, and consequently are not able to complete the verbal tasks in the cognitive tests, non-verbal IQ was preferentially used in the analyses.

Daily adaptive functioning

Levels of adaptive functioning can be measured through the VABS, a psychological assessment developed by Sara S. Sparrow *et al*. in 1984. (Sparrow, Balla, and Cicchetti 1984; Sparrow, Balla, and Cicchetti 2005) They proposed to use this instrument in conjunction with standardized measures of cognitive development (*e.g.*, IQ) as a valuable clinical assessment tool. The test is organized in 5 domains: adaptive skills, communication, interaction, motor and daily living skills. The scores obtained in each domain allows for the measurement of a total scaled score of daily living skills. Similar to the IQ score, the population mean total score is centered on 100 points and the SD is of 15 points.

As aforementioned, both IQ and the VABS help in the diagnosis of ID. An ID is generally stated when standardized IQ and Vineland scores are 2 SD below the population mean, including a margin of error. This corresponds to a score of less than 70 (generally ±5 points) for both tests which share the same normal-shaped distribution.

Measuring language impairment

Language milestones

Delayed acquisition of early expressive language milestones, such as the age at first word or phrase, is the first indication of a NDD or language impairment that will persist throughout childhood, interfering with everyday communication and academic attainment. (Rescorla and Schwartz 1990; Dale Philip S. et al. 2003; Rudolph and Leonard 2016) Such delays can be diagnosed if the age reported is greater than 24 months for the first words and 33 months for the first phrase. (Rudolph and Leonard 2016) The age at first word or phrase are extensively used as they are generally well reported by the parents or in the follow-up of the pediatrician.

Phonological processing

The comprehensive test of phonological processing (CTOPP) measures phonological processing skills. (Wagner, Torgesen, and Rashotte 1999) This tool was developed by Richard K. Wagner *et al.* in 1999 to aid in the identification of individuals from nursery through to college who may profit from instructional activities to enhance their phonological skills. This test is organized into 13 subtests that assess phonological awareness, phonological memory, rapid symbolic naming, rapid non-symbolic naming, and the alternate phonological awareness composite. The subtest scaled scores have a mean of 10 and a standard deviation of 3. The composite score indexes have a mean of 100 and a standard deviation of 15.

Assessing the quality of motor skills development

Motor milestone

The age of onset for walking is an important milestone which is shifted in ID and ASD compared to neurotypical populations and has also been associated with more severe ASD symptoms. (Bishop et al. 2016; Reindal et al. 2019; Sumner, Leonard, and Hill 2016) Walking delay can be diagnosed if the age of first walking is greater than 18 months.

(WHO and de Onis 2006) Similar to the language milestones, this measure is broadly used as it is well reported by the parents or recorded during a pediatric follow-up.

Motor coordination

The developmental coordination disorder questionnaire (DCDQ) was created in 1998 by Branda N. Wilson *et al*. to respond to the need for a valid and reliable assessment tool in identifying children who have motor coordination problems. (Wilson et al. 2000) The last version consists of a 15-item parent questionnaire.

For each item, parents were asked to compare the degree of coordination of their child with other children of the same age, and to rate this on a 5-point Likert scale, ranging from “Not at all like this child” to “Extremely like this child.” This test allows three subscores: control during movement, fine motor/handwriting, and general coordination. The total raw score can be used as a quantitative measure or can be transformed in function of the age of the participant to indicate its DCD-level (“indication of/suspected DCD” or “probably not DCD”).

Recent studies reported that impairments identified with DCDQ score were associated with lower quality of life (Karras et al. 2019), atypical sensory processing (Delgado-Lobete et al. 2020) and persistent DCD in ASD adolescents. (Bhat 2020)

Overall, all these measures have been widely validated and adopted by both clinicians and researchers. Notably, they have been extensively used to investigate biomarkers associated with NDDs and associated traits.

In search of biomarkers and risk factors associated with ASD and EOP

Technological and methodological advances in imaging and genomics helped to formally associate various endophenotypes and biological risk factors with ASD and EOP. Such studies were driven by the needs to better understand the biological mechanisms underlying each disorder, and by extension to create adapted therapies or

medications, predict the progression and the severity of the disorder, as well as identifying the risk factors that may be encounter during the lifetime (*e.g.*, informing future parents of the risk factors during the peri-natal period).

Most recent findings on neuroimaging endophenotypes of ASD and EOP

Such advances in technologies and methodologies in magnetic resonance imagery (MRI) allowed to identify alterations of the central nervous system, which are now well-known biomarkers in several NDDs. These anatomical and functional alterations have been observed systematically in post-mortem, neuroimaging and electrophysiological studies. The following sections describe the most recent and robust neuroimaging endophenotype associated with ASD and EOP.

Neuroanatomical signatures and functional connectivity patterns associated with ASD

Structural MRI studies consistently report a higher total brain volume in children with ASD compared to controls during the first years of life, suggesting an early overgrowth of the brain in autism caused by abnormal cortical development and expansion. (Li, Karnath, and Xu 2017; Pagnozzi et al. 2018; van Rooij et al. 2018) In addition to this early biomarker, lower volumes of the cerebellum and corpus callosum in individual with ASD compared to controls are other robust biomarkers of ASD. (Li, Karnath, and Xu 2017; Pagnozzi et al. 2018; Valenti et al. 2020) Lower volume in these structures were associated with lower IQ, reduced integration of information and slower processing; and reflect the aberrant connectivity mediating the intra- and inter-hemispheric communications typical in autism. (Li, Karnath, and Xu 2017; Pagnozzi et al. 2018; Valenti et al. 2020) Another important early brain endophenotype of ASD is the increased cerebrospinal fluid volume compared to controls, which is also associated to an increased symptom severity measured by the ADOS. (Shen et al. 2013; Pagnozzi et al. 2018; van Rooij et al. 2018) Cerebrospinal fluid plays an important role in circulating nutrients and removing waste products from the brain and its alteration may reflect an

improper filtering and draining of waste particles, potentially leading to neuro-inflammation. (Shen et al. 2013; Pagnozzi et al. 2018)

Less consistent brain volumes alterations with smaller effect-sizes have also been identified in ASD individuals compared to controls, notably, in the putamen and amygdala, which are suspected to contribute to the repetitive and stereotyped and impaired social behaviors respectively. (Pagnozzi et al. 2018; van Rooij et al. 2018) Cortical thickness (CT) was observed as higher in the frontal and prefrontal cortex, as well as in the posterior cingulate gyri of ASD individuals compared to controls; and these modulations were positively correlated with the severity of ASD symptoms measured by the ADOS. (van Rooij et al. 2018; Bedford et al. 2020)

Multiple alterations in brain connectivity have also been highlighted by functional MRI studies. Resting state MRI studies have shown a widespread underconnectivity in ASD compared to controls, notably in the default mode network, the visual and auditory networks. (Di Martino et al. 2014; Holiga et al. 2019; Lau, Leung, and Lau 2019) A thalamocortical overconnectivity has also been observed, specifically between the thalamus and the somatosensory-motor network. (Woodward et al. 2017; Postema et al. 2019; Tomasi and Volkow 2019) However, some of these results have not been replicated, likely due to a lack of standardization in acquisition, ascertainment, or analytical strategies used by the different groups.

Interestingly, several of these alterations of brain structure and connectivity identified in ASD overlapped with other established psychiatric diagnoses, such as EOP or schizophrenia. (Goodkind et al. 2015; Cauda et al. 2017; Gurholt et al. 2020)

Neuroanatomical signature of EOP

Recent neuroimaging studies show a similar pattern of brain alterations in EOP compared to adult psychosis, schizophrenia and BD. (Hibar et al. 2016; van Erp et al. 2016) In 2020, Gurholt *et al*. demonstrated lower intracranial and hippocampal volumes and higher caudate and pallidum volumes in 263 patients with EOP in comparison to 359 individuals from the general population. (Gurholt et al. 2020) These results

replicated findings in much smaller EOP and COS datasets. (Frazier et al. 1996; Arango et al. 2008; El-Sayed et al. 2010) In contrast, while higher intracranial volume was reported in ASD in comparison to controls, the magnitude of alterations in EOP was greater. (van Rooij et al. 2018) Further, the directionality of effects on volume were similar for limbic structures, while opposing for basal ganglia structures between EOP and ASD, which may indicate differential neurodevelopmental mechanisms. Finally, given the relative stability of intracranial volume from mid-adolescence onwards, the decrease in volume found in EOP suggests an aberrant neurodevelopment, which is more severe and/or established earlier than in adult-onset psychosis. (Mills et al. 2016)

However, structural brain MRI studies of EOP are rare and most studies so far have focused on COS, which do not entirely reflect the heterogeneous population of people that first present to EOP intervention services. More generally, both EOP and COS studies are limited by their sample sizes.

Main environmental risk factors associated with ASD and EOP

ASD is multifactorial in origin with risk factors ranging from environmental to biological endogenous biological factors. The risk of ASD can be increased by exogenous factors (Figure 4) such as substance abuse during pregnancy (*e.g.*, alcohol, other drugs, teratogens), perinatal complications including labor and delivery-related events leading to neonatal encephalopathy, or postnatal causes. (Xie, Peltier, and Getahun 2016; Lyall et al. 2017; Modabbernia, Velthorst, and Reichenberg 2017) Postnatal factors include extreme malnutrition, hypoxic ischemic injury, traumatic brain injury, infections, seizure disorders (*e.g.*, infantile convulsions), social deprivation, and intoxications. Age of parents during the conception and maternal diseases (including placental disease) also contribute to these conditions.

Similarly to ASD, EOP risk factors can be environmental as well as endogenous. The major environmental exposures implicated in psychosis are represented in Figure 5. They can be classified based on the influence of exposure during the development. (Zwicker, Denovan-Wright, and Uher 2018)

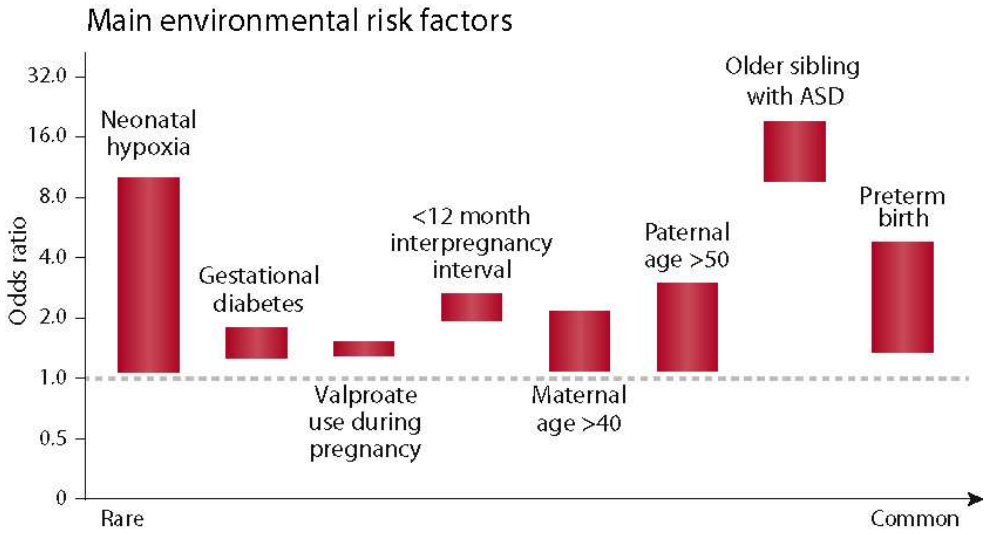


Figure 4: Main robust environmental risk factors for autism

Data can be broadly split into three categories: those with evidence supporting an association. Bars represent ranges. Having an older sibling with ASD is one of the most robust estimators for having a younger child with ASD (more than 10% (Xie, Peltier, and Getahun 2016; Lyall et al. 2017)). This Figure is adapted from Lord *et al.* 2020 (Lord et al. 2020). It combines findings from selected reviews and empirical papers aiming to identify risk factors for autism: neonatal hypoxia estimate (Modabbernia et al. 2016; Modabbernia, Velthorst, and Reichenberg 2017), gestational diabetes (Modabbernia, Velthorst, and Reichenberg 2017), valproate used during pregnancy (Christensen et al. 2013), interpregnancy interval (Cheslack-Postava, Liu, and Bearman 2011; Conde-Agudelo, Rosas-Bermudez, and Norton 2016; Xie, Peltier, and Getahun 2016; Lyall et al. 2017), parent age (Xie, Peltier, and Getahun 2016; Lyall et al. 2017; Modabbernia, Velthorst, and Reichenberg 2017), siblings (Conde-Agudelo, Rosas-Bermudez, and Norton 2016; Xie, Peltier, and Getahun 2016; Lyall et al. 2017), preterm birth (Schendel and Bhasin 2008; Guy et al. 2015; Lyall et al. 2017).

Exposure	Timing				Strength of evidence	Reference
	Prenatal	Perinatal	Childhood	Adolescence		
Maternal anemia					++	Nielsen *et al.*, 2016
Heavy metals					++	Attademo *et al.*, 2016
Infectious agents					+++	Brown & Patterson, 2011; Torrey *et al.*, 2012
Low socio-economic status					+++	Agerbo *et al.*, 2015
Obstetric complications					++	Byrne *et al.*, 2007
Preterm birth					++	Nosarti *et al.*, 2012
Migration					+++	Cantor-Graae, 2005
Urban residence					+++	Vassos *et al.*, 2012
Adversity					+++	Varese *et al.*, 2012; Trotta *et al.*, 2013, 2015
Head injury					++	Orlovska *et al.*, 2014
Stimulants					++	McKetin *et al.*, 2016; Gurillo *et al.*, 2015
Cannabis					+++	Marconi *et al.*, 2016; Kelley *et al.*, 2016

Figure 5: Environmental factors associated with psychosis

Number of plus signs denotes the strength of evidence for the association: +++ indicates consistent evidence from multiple large-scale studies or a meta-analysis; ++ indicates evidence from multiple smaller studies or a strong association in a high-quality study. The reference provided reflects meta-analyses or largest studies available. This Figure is adapted from Zwicker *et al.* (2018).

In pre- and perinatal periods, exposures that activate the immune system and subsequent inflammation (*e.g.*, infectious agents or maternal stress) are robustly associated with an increase of psychosis risk by contributing to abnormal neurodevelopment. (Buka et al. 2008; Malaspina et al. 2008; Brown and Patterson 2011; Torrey, Bartko, and Yolken 2012; Canetta et al. 2014; Fineberg et al. 2016) Interestingly, studies suggest that childhood trauma is associated with increased levels of inflammatory markers, which is in line with the biological pathway dysregulated by these exposure risk factor. (Baumeister et al. 2016)

Early adversity occurring in childhood (*e.g.*, physical, sexual, psychological abuse, neglect, and involvement in bullying) also consistently increase the risk of psychosis. (Varese et al. 2012; Wolke et al. 2014; Trotta et al. 2013; Trotta et al. 2015)

The risk factors with the highest effect-sizes during late development are the exposure to psychostimulants (*e.g.*, methamphetamine, tobacco, prescribed stimulant medication) and cannabis. (Gurillo et al. 2015; Kelley et al. 2016; Marconi et al. 2016; McGrath et al. 2016; McKetin et al. 2016; Hajebi et al. 2018) Recreational drug consumers with a family history of psychopathology are more vulnerable to persistent psychotic symptoms. (MacKenzie et al. 2016)

As most individuals are exposed to more than one of these risk factors, individual effects are difficult to investigate. Moreover, the effect of these factors differs between individuals likely to develop NDDs because genetic factors may modulate them. (Zwicker, Denovan-Wright, and Uher 2018)

The quest for genetic risk factors associated to NDDs such as ASD and EOP

The diagnostic-first approach: from a common clinical condition to associated genetic variants

It is twins! The concomitant birth of advanced statistics and heritability calculation

In the early XXth century, twin study was the first methodology used to investigate the genetic contribution to human complex traits. This method is based on the equal environments assumption, which is the assumption that environments for monozygotic twins are not more similar than the environments of dizygotic twins. Thus, the genetic contribution is estimated by comparing the similarity on a trait in monozygotic and dizygotic twins who grow up and experience the same degree of environmental similarity, but are of differing genetic relatedness. (Boomsma, Busjahn, and Peltonen 2002)

J.-C. Smith (1930), Kallmann and Roth (1956) and Suzan Folstein (1977) were the first to study pairs of twins with at least one with a diagnosis of ID, COS and autism, respectively, and to suggest the strong genetic contribution to the etiology of these conditions. (Smith 1930; Kallmann and Roth 1956; Folstein and Rutter 1977)

During the 1960's, the work of the geneticist Douglas Falconer on the computation of heritability[3] based on twin populations greatly contributed to a better understanding of the inheritance of liability to human complex diseases. (Falconer 1967)

Since, measurement of heritability evolved, and larger samples allowed to compute accurate and reliable estimations of the genetic contributions to several NDDs. Throughout the last three decades, studies converge on the high contribution of genetics to the etiology of NDDs. The larger up to date studies estimated an heritability of ASD

[3] The heritability estimates the relative contribution of genetics versus the environment to variation in a particular trait of interest.

ranging from 51% to 97%. (Bai et al. 2019; Taylor et al. 2020) The heritability of psychosis, schizophrenia and COS were estimated at up to 90%, 80% and 84%, respectively. (Kallmann and Roth 1956; Rijsdijk et al. 2011; Hilker et al. 2018; Sullivan and Geschwind 2019)

However, these heritability estimates from twin studies are based on models that do not take into account the interaction between genetics and the environment. (Zwicker, Denovan-Wright, and Uher 2018) Therefore, these values may be inflated by mechanisms such as assortative mating or dynastic effect[4]. (Morris et al. 2020)

Studies have demonstrated a complex genetic architecture of NDDs with contributions from both rare[5] and common[6] variants to the etiology of NDDs. Notably, common variants contributed at 14% to 59% to the heritability of hallucinations and paranoia in adolescents. (Zavos et al. 2014; Sieradzka et al. 2015; Pain et al. 2018) In ASD, studies have demonstrated that common variants accounted for at least 41% of the heritability and are enriched in regulatory elements, as well as in genes expressed in central nervous system cells. (Gaugler et al. 2014; Grove et al. 2019) This functional enrichment is similar to findings for rare structural variants and the genetic architecture of other psychopathologies such as schizophrenia, COS, BD and major depression disorder. (Bulik-Sullivan et al. 2015; Wray, Ripke, et al. 2018; Satterstrom et al. 2020) Rare variants account for less than 3% of the heritability of ASD (Gaugler et al. 2014), but historically, they were the first to be detected in individuals with ASD and other NDDs due to their large effect-size.

[4] A dynastic effect occurs when offspring inherit both phenotype-associated variants and phenotype-associated environments from parents, leading to biased genetic associations. It refers to the "inheritance" of environment in addition to genotype.

[5] Rare variants are defined as genomic variation with a minor allele frequency < 1% but still polymorphic in one or more major human populations.

[6] Common variants are defined as genomic variation with a minor allele frequency > 1% in the general population.

The identification of rare variants contributing to NDDs

The pace of rare variant discovery in NDDs is tightly related to the rapidly evolving genotyping and sequencing technologies. The Figure 6 represents the genetic discovery timeline in ID and ASD.

The first genetic syndromes were discovered through the investigation of groups of individuals with recognizable clinical features. In 1959, trisomy 21 was the first genetic diagnosis discovered under the microscope in a population of individuals with the Down syndrome. (Lejeune, Turpin, and Gautier 1959). Traditional thinking held that ASD was rare in individuals with trisomy 21, but the prevalence is up to 20%. (DiGuiseppi et al. 2010; Moss et al. 2013)

This discovery started the implementation of cytogenetics in the research on NDDs which allowed the discovery of rare CNVs. A CNV is defined as a deletion or duplication of a stretch of DNA as compared with the reference human genome. CNVs may range in size from a kilobase to several megabases (Mb) or even an entire chromosome (trisomies and monosomies) and can involve one or more genes. 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. (Malhotra and Sebat 2012) CNVs are either recurrent or ultra-rare (non-recurrent). Recurrent CNVs are genomic loci flanked by low copy repeats (LCR)[7] sequences, also called breakpoints (BPs), that greatly increase the risk of homologous recombination. These non-allelic homologous recombinations arise at meiosis and result in similar or identical CNVs in unrelated individuals. The non-recurrent CNVs are structural variants with non-recurrent end-point. On the other hand, the non-recurrent junctions do not coincide with LCRs, but tend to occur in the vicinity of regions that are rich in LCRs resulting in complex regional genomic architecture.

[7] Low copy repeats, or breakpoints, are highly homologous sequences which can cause a misalignment and an unequal crossing-over. These regions are susceptible to chromosomal rearrangements such as non-allelic homologous recombination during meiosis.

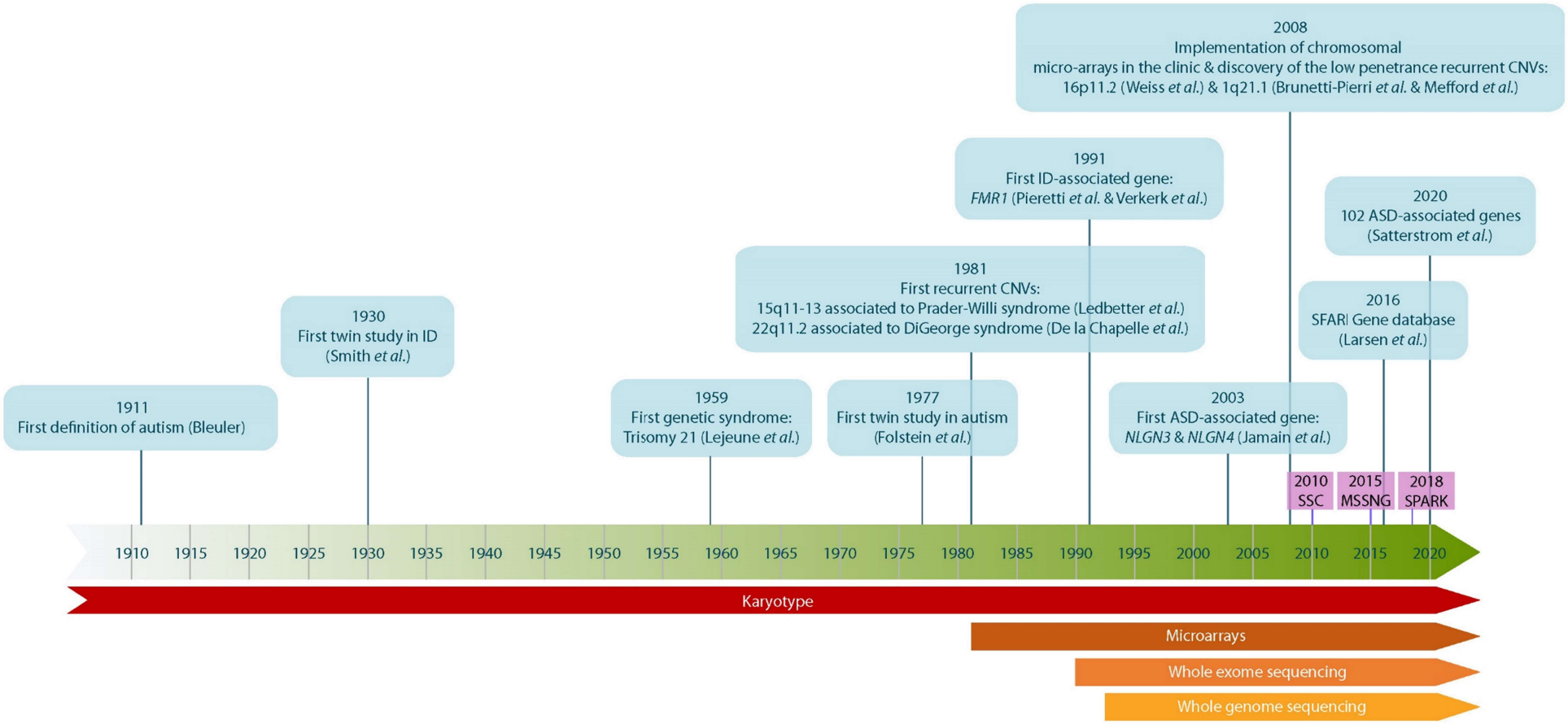


Figure 6: Evolution of the genotyping technologies and the discoveries of rare variants associated with ASD

ASD: autism spectrum disorder, ID: intellectual disability, FMR1: fragile X mental retardation 1, NLGN3/4: Neuroligin 3/4, SSC: Simons simplex collection, SPARK: Simons foundation powering autism research for knowledge.

Ten years after the discovery of trisomy 21, a smaller marker on chromosome X was discovered for the fragile X syndrome. (Lubs 1969) In 1991, an expansion of a CGG repeated nucleotide sequence in the *FMR1* (fragile X mental retardation 1) gene was identified as the cause of fragile X syndrome. (Pieretti et al. 1991; Verkerk et al. 1991) This was the first rare gene-disrupting variant associated with ID. More recently, this variant was associated with ASD, with 20% to 50% of cases in individuals with fragile X. (Kaufmann et al. 2017)

Rare pathogenic copy number variants

In the 80's and 90's, cytogenetic studies revealed large (above the 3 megabase (Mb) resolution of a karyotype) rare recurrent autosomal[8] deletions and duplications associated with clinically recognizable neurodevelopmental syndromes. Thus, Prader-Willi syndrome was linked to a deletion of the paternal copy in the 15q11-13[9] locus (Ledbetter et al. 1981), Angelman syndrome to a deletion of the maternal copy of the 15q11-13 locus (Magenis et al. 1987), the 22q11.2 deletion was associated with DiGeorge syndrome (de la Chapelle et al. 1981), the 17p11.2 deletion with the Smith-Magenis syndrome (Smith et al. 1986), and the 7q11.23 deletion was associated with Williams-Beuren syndrome (Robinson et al. 1996).

Around 2008, the implementation of CMA analysis in the clinic was a diagnostic revolution and allowed for the genome-wide detection of deleterious CNVs. This triggered a wave of discovery and rapidly expanded the number of genomic loci associated with NDDs. A handful of deleterious recurrent CNVs were rapidly identified due to their relatively high frequency and large effect-sizes: 3q29 (Willatt et al. 2005), 16p11.2 (Weiss et al. 2008), 1q21.2 (Brunetti-Pierri et al. 2008; Mefford et al. 2008), 15q13.3 (Sharp et al. 2008), and 17q12 (Moreno-De-Luca et al. 2010) deletions and duplications.

---

[8] Autosomal variants are genetic variants occurring on the non-sexual chromosome (from chromosome 1 to 22).

[9] This nomenclature can be read as follow: For the locus 15q11-13, this CNV is located in the chromosome 15, "q" indicates the larger arm of the chromosome (it will be "p" if it is the small one) and at the position 11 to 13.

Since their discovery, these recurrent pathogenic CNVs have been extensively investigated in a genetic-first approach described later in this thesis. Thus, the ASD and schizophrenia risks associated with these rare variants are well-established (Table 4).

Rare recurrent CNVs associated with COS have mainly been identified in one cohort and the majority were already associated with ASD or adult-onset schizophrenia (Ahn et al. 2014; Fernandez et al. 2019) Studies of EOP are even scarcer. Case series describing EOP have been reported in individuals with 22q11.2 deletion (Ivanov et al. 2003; Vorstman et al. 2006), also well known to be more enriched in schizophrenia than in ASD (Sanders et al. 2019), the 16p13.11 deletion and duplication (Brownstein et al. 2016), and the 3q29 deletion (Sagar et al. 2013). These limited findings in EOP are in part linked to the lack of genomic screening in the clinical care of such patients, which is a major constraint to identify the disease mechanisms and drug targets underlying this disorder.

Rare structural variants are identified in up to 28% of individuals with COS versus 15% in adult-onset schizophrenia, suggesting a greater genetic component in the early-onset form of the disorder. (Walsh et al. 2008; Ahn et al. 2014) Nevertheless, the limited number of individuals included in COS and EOP genetic studies do not allow to accurately measure the effect-size of CNVs identified.

Thanks to large ASD samples, such as the Simons simplex collection (SSC) (Fischbach and Lord 2010), MSSNG database (Yuen et al. 2015) or the Simons foundation powering autism research for knowledge (SPARK) (Feliciano et al. 2018), an increased burden of rare and large pathogenic CNVs in ASD has been clearly demonstrated. (Sebat et al. 2007; Girirajan et al. 2011; Girirajan et al. 2013; Sanders et al. 2015; Zarrei et al. 2019) Girirajan *et al*. found that the severity of ASD was positively correlated to the size of deletions and duplications. (Girirajan et al. 2013) The positive correlation observed for deletions was also associated with a decrease of non-verbal IQ. (Girirajan et al. 2013)

Table 4: Recurrent pathogenic CNVs, prevalence, de novo rate, and their odds ratio for ASD, schizophrenia and NDDs

Locus	Syndrome	Type	Position (Mb, Hg19)		Size (Mb)	Prevalence (%)	*de novo* rate (%)	OR for ASD	OR for SZ	OR for NDD
			Start	Stop						
1q21.1	-	deletion	146.6	147.5	0.9	0.027	16	3	6	11
	-	duplication				0.044	26	5	3	5
3q29	-	deletion	195.7	197.3	1.6	0.003	82	19	23	19
	-	duplication				0.001	25	n.s.	n.s.	42
7q11.23	Williams-Beuren	deletion	72.7	74.1	1.4	0.001	95	32	n.s.	∞
	-	duplication				0.003	71	32	23	36
15q11.2-13	Prader-Willi, Angelman	deletion	23.6	28.4	4.8	0.000	78	43	5	18
	-	duplication				0.006	59	50	12	42
15q13.3	-	deletion	30.9	32.5	1.6	0.007	26	15	18	36
	-	duplication				0.060	16	n.s.	n.s.	n.s.
16p11.2	-	deletion	29.6	30.3	0.7	0.028	42	14	n.s.	14
	-	duplication				0.028	24	11	12	4
17p11.2	Smith-Magenis	deletion	16.6	20.3	3.7	0.000	97	∞	n.s.	∞
	Potocki Lupski	duplication				0.001	92	32	n.s.	∞
17q12	Renal cysts and diabetes	deletion	34.8	36.2	1.4	0.001	69	97	54	∞
	-	duplication				0.023	15	n.s.	n.s.	5
22q11.2	Di George	deletion	18.9	21.9	3	0.003	90	32	92	∞
	-	duplication				0.064	34	n.s.	0.2	4

Non-exhaustive list of large effect-size reciprocal recurrent loci previously associated with NDDs. (Sanders et al., 2019) The genomic coordinates of the locus are under the human genome reference build version 19 (Hg19). The prevalence is based on the occurrence in the general population of UK Biobank dataset (N total individual = 151,619) (Kendall et al., 2017). The *de novo* rate was computed with the DECIPHER database, when including cases reported with an overlap >50% with the locus (*DECIPHER v9.17: Mapping the Clinical Genome*, n.d.). The odds ratio for ASD, schizophrenia and NDDs are based on those reported by Sanders *et al.*, (2019). Mb: mega base, OR: odds ratio, ASD: autism spectrum disorder, SZ: schizophrenia, NDDs:

However, individual effect-size of non-recurrent CNVs on ASD risk remain undocumented and can't be computed using case-control association studies due to their scarcity.

Relationship between de novo frequency and effect-size

Most of the rare variants previously discussed occur *de novo*[10] in over 20% of patients (Table 4). (DECIPHER v9.17: Mapping the Clinical Genome n.d.) The *de novo* frequency of a variant is highly correlated to fitness[11] and reflects its pathogenicity and effect-size: the more a variant occurs *de novo*, the more it is submitted to negative selection pressure[12]. Variants with effects on intelligence equal or above 2 SD are observed *de novo* in close to 100% of the cases (Figure 7). (cf. co-authored publications: Huguet et al. 2018; Sanders et al. 2019) Even variants with mild effects on general intelligence are under negative selection.

For example, carriers of a 1q21.2 deletion or s 15q11.2 deletion from the general population have a mean decrease of 0.30 and 0.15 IQ z-scores respectively, and these CNVs occurs *de novo* in 16 and 6% of cases. (DECIPHER v9.17: Mapping the Clinical Genome n.d.; Jønch et al. 2019; Kendall et al. 2019) Overall, *de novo* variants have been hypothesized to represent an important cause of severe early-onset NDDs, and the occurrence is estimated around 0.5 % of live births, explaining why severe NDDs under negative pressure remain frequent in the population. *De novo* variants were reported in COS at a similar rate as in ASD and adult-onset schizophrenia, but the small sample size (N = 17 trios) did not allow to establish a clear prevalence. (Ambalavanan et al. 2016)

[10] De novo variants are genetic alterations that are present for the first time in one family member as a result of a structural variation in the germ cell of one of the parents, or in the fertilized egg itself.

[11] The fitness describes individual fecundity/reproductive success.

[12] Meaning that the structural variation will be under strong purifying selection and quickly eliminated from the population.

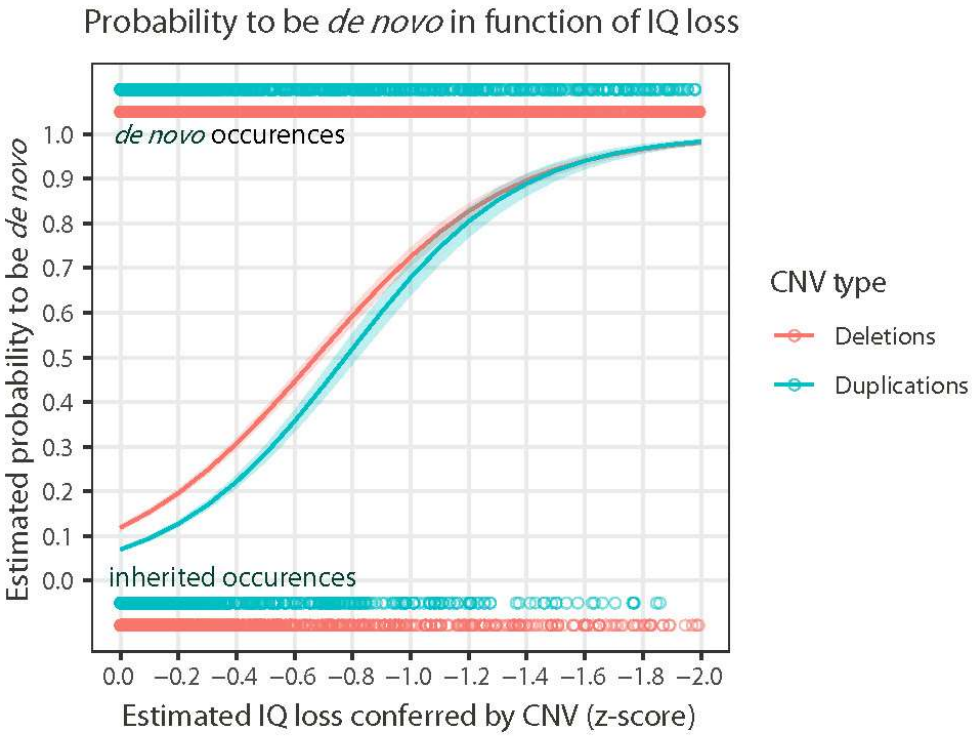


Figure 7: Probability to be *de novo* in function of the estimated IQ loss conferred by deletions and duplications (in z-score)

Probability of *de novo* (Y-axis) estimated by a logistic regression model according to the loss of IQ estimated by a model using pLI for deletions and duplications (X-axis). The *de novo* model was fitted on 13,114 deletions (red) and 13,323 duplications (blue) with available inheritance information observed in DECIPHER, CHU Sainte-Justine cytogenetic database, Simons simplex collection, MSSNG, Saguenay youth study and Generation-Scotland datasets. Circles represent the *de novo* or inherited occurrences in the pooled dataset. This Figure is based on analyses that I conducted for Huguet *et al.*, 2021. (Previous versions describing the estimated probability to be *de novo* for deletions were also published in Huguet *et al.*, 2018 and Sanders *et al.*, 2019).

De novo CNVs and single nucleotide variants (SNVs) associated with ASD and identified in COS are enriched in loss-of-function (LoF) intolerant gene[13]. (Iossifov et al. 2014; Ambalavanan et al. 2016; An et al. 2018; Satterstrom et al. 2020) *De novo* variants were also associated with a negative impact on cognition and the *de novo* burden is higher in individuals with ASD or schizophrenia which have a lower IQ. (Bishop et al. 2017; Singh et al. 2017; Weiner et al. 2017) Coe *et al.* demonstrated that the proportion of patients

[13] Loss-of-function (LoF) variants, also called inactivating variants, result in the gene product having less or no function.

with a *de novo* variations is significantly higher for ID compared to ASD (Odds Ratio (OR) = 3.70). (Coe et al. 2019) They are over-represented in females with ASD and can contribute to 45% of females diagnosed. (Jacquemont et al. 2014; Iossifov et al. 2015; Bishop et al. 2017; Satterstrom et al. 2020)

However, *de novo* variants including CNVs, missense[14] and likely gene-disruptive (LGD)[15] variants collectively explain less than 5% of the overall liability of ASD, and far less of the heritability. (Gaugler et al. 2014; Iossifov et al. 2014) There is evidence that ASD risk can be conferred by both rare and common inherited variants of smaller effect-size that disrupt genes intolerant to variations. (Krumm et al. 2015; Niemi et al. 2018; Constantino 2019) The transmission of such variants may often originate from the mother, who carry a variant without experiencing severe consequences. (Iossifov et al. 2015) These inherited variants are also seen more often in ASD offspring with lower IQ. Both inherited and *de novo* mutations also contribute to risks for schizophrenia and COS. (Singh et al. 2017; Fernandez et al. 2019; Forsyth and Asarnow 2020)

In search of causative genes: example of the elusive "ASD-genes"

In the past 10 years, next generation sequencing[16] technologies, such as exome[17] and whole-genome sequencing, as well as the inception of large cohorts, have accelerated the identification of genes associated with NDDs. (Boycott et al. 2013; Vissers, Gilissen, and Veltman 2016) Genes associated with psychopathologies were defined as those with an excess of SNVs in patients with overlapping clinically defined disorders, such as ASD, psychosis or schizophrenia. To identify such disease-causing genes, efforts were

[14] Missense variants are single base pair substitution which alter the genetic code (codon) in a way that produces an amino acid in the protein that is different from the usual amino acid at that position.

[15] Likely gene-disruptive are disruptive variations, either a frameshift, nonsense, or destroying a consensus splice site.

[16] Next generation sequencing is a collective term to describe the modern high-throughput sequencing technologies in the post-Sanger sequencing era.

[17] Exome sequencing is a genomic technique for sequencing all of the protein-coding regions of genes in a genome, also called exome.

focused on rare variants of large effect-size that are thought to account for the observed phenotype in participants.

The first genes associated with ASD were *NLGN3* (neuroligin 3) and *NLGN4* (neuroligin 4). (Jamain et al. 2003) Subsequently, *SHANK3* (SH3 and multiple ankyrin repeat domains 3) was associated with Phelan-McDermid syndrome (Phelan, Rogers, and Boccuto 1993; Durand et al. 2007), *NRXN1* (neurexin 1) with Pitt-Hopkins-like syndrome 2 (Bourgeron 2007; Zweier et al. 2009), and *ARID1B* (AT-rich interaction domain 1B) associated with Coffin-Siris syndrome (Santen et al. 2012; Iossifov et al. 2015). Other genes associated with ASD, but without or less syndromic clinical manifestation, were also identified, such as *FOXP2* (forkhead-box P2) (Wassink et al. 2002), *CHD8* (chromodomain helicase DNA binding protein 8) (O'Roak et al. 2012), or *CNTN5* (contactin 5). (van Daalen et al. 2011)

From the accumulated literature on ASD-associated variants, Larsen and colleagues developed a scoring algorithm to rank the candidate genes based on the strength of the evidence linking it to the development of ASD. (Larsen et al. 2016) This algorithm was implemented to create the Simons foundation autism research initiative (SFARI) Gene database, with currently 1,231 scored genes out of which 206 are ranked as high confidence and 221 as strong candidates' genes contributing to ASD. (https://gene.sfari.org/database/gene-scoring/) These genes are enriched in two major functional groups: gene expression regulation, including chromatin regulators and transcription factors, and neuronal communication, including synaptic functions. (De Rubeis et al. 2014; Satterstrom et al. 2020) They are expressed at high levels in the human cortex and cerebellum, and early in the development. (Chang et al. 2015; Satterstrom et al. 2020)

Several recent studies have attempted to identify genes that are "ASD-predominant" genes which are preferentially associated with ASD compared to other NDDs – by comparing the frequency of *de novo* LGD variants between cohorts of individuals with ASD and/or other NDD. Using this methodology, Satterstrom *et al.* have recently

identified 102 genes implicated in ASD risk. (Satterstrom et al. 2020) Based on the ratio of the frequency of such variants in 11,986 individuals with ASD and in 5,264 individuals with severe ID, 53 of the 102 genes were relatively ASD-predominant, and the others were associated with more global developmental impairment. Previously, Coe *et al*. also compared the frequency of *de novo* LGD variants between 5,624 individuals with ASD and 5,303 individuals with ID, but did not find evidence of ASD specificity for any of the 253 genes they identified as candidates for NDD. (Coe et al. 2019) In fact, after subsequent analyses, 72% of these genes showed evidence of excess of *de novo* variants in both ASD and ID cohorts, and the ID cohorts evaluated in these two studies were overlapping at 99%. (Myers et al. 2020)

However, potential bias introduced by the lack of uniform phenotyping across studies and phenotypic overlap between the groups is unclear. Furthermore, other studies demonstrate that such pathogenic ASD-predominant variants in the same genes are identified in individuals with a variety of different clinically defined brain disorders, including schizophrenia, COS and ID. (Singh et al. 2017; Fernandez et al. 2019; Wang, Corominas, and Lin 2019; Zarrei et al. 2019; Forsyth and Asarnow 2020; Myers et al. 2020; Ambalavanan et al. 2016) There is a clear lack of sufficient evidence to establish disorder specificity, as well as the possibility that other sources of genetic variation, such as common variation, may contribute to NDD susceptibility.

The contribution of common variants and the emergence of polygenic models in NDDs

Numerous evidences point toward a polygenic theory of NDDs – meaning that many genetic loci, mostly with small effect-sizes, contribute to NDD susceptibility – and that most of the genetic contribution is due to common variants. (Gaugler et al. 2014; Wray, Wijmenga, et al. 2018; Grove et al. 2019; Sullivan and Geschwind 2019; Myers et al. 2020) Collectively, rare pathogenic variants are identified in 15 to 28% of cases referred in the genetic clinic. (Walsh et al. 2008; Tammimies et al. 2015; Huguet et al. 2018) Conversely, common variants have been estimated to account for a major portion of

NDD heritability, and recent genome-wide association studies (GWAS) reveal a handful of disease-associated loci, such as in ASD, schizophrenia and COS. (Sieradzka et al. 2014; Sieradzka et al. 2015; Anney et al. 2017; Grove et al. 2019; Wang, Corominas, and Lin 2019; Akingbuwa et al. 2021; Guo et al. 2021)

In 2019, Grove *et al*. published the first common risk variants robustly associated with ASD. (Grove et al. 2019) They identified single nucleotide polymorphisms (SNPs)[18] from 18,381 ASD cases and 27,969 controls from the combination of the iPSYCH sample[19] and five family-based trio samples of European ancestry from the Psychiatric Genomics Consortium. In this study, SNP-based heritability of Asperger syndrome (h_G^2 = 0.097) was estimated to be twice the heritability of both autism (h_G^2 = 0.049) and a group of other/unspecified PDD (h_G^2 = 0.045). The heritability of ASD without ID (h_G^2 = 0.086) was also three times higher than that for ASD with ID (h_G^2 = 0.029). These results suggest that the contribution of common variants may be more prominent in high-functioning ASD, such as Asperger's syndrome, than in autism with lower functioning, for which rare variants may contribute more to their heritability. These results align well with the observation that de novo variants are more frequently observed in ASD cases with ID than in cases without comorbid ID. (Robinson et al. 2014) A total of 93 SNPs at three separate loci were significantly associated with ASD after Bonferroni correction ($p < 5 \times 10^{-8}$). These results were confirmed by five other studies. (Akingbuwa et al. 2021) Grove *et al*. identified significant enrichment of common variations in conserved DNA regions, in histones, as well as in genes expressed in central nervous system cells, in line with observations for common and rare variants in other NDDs. (Wray, Ripke, et al. 2018; Collins et al. 2019; Wang et al. 2019; Smeland et al. 2020; Satterstrom et al. 2020) These genes were specifically expressed in brain cells during the development. The common variations were located in regions that are highly enriched in regulatory elements

[18] SNPs are the most common type of genetic variation among people (almost once in every 1,000 nucleotides on average). Each SNP represents a substitution in a single nucleotide in the genome.

[19] The integrative psychiatric research (iPSYCH) dataset is a Danish nationwide population-based case–cohort sample including nearly all individuals born in Denmark between 1981 and 2005 and diagnosed with ASD before 2014.

predicted to be active in human corticogenesis, such as enhancer marks in the fetal brain. Interestingly, both common and rare variation associated with ASD preferentially affects elements regulating genes expressed during brain development. (Chang et al. 2015; Satterstrom et al. 2020)

The largest GWAS on multi-stage schizophrenia to date included 36,989 cases and 113,075 controls and found 108 genome-wide significant distinct risk loci. (Ripke et al. 2014) In these loci, there was an enrichment of coding genes expressed in brain, among which some were encoding calcium channels, as well as proteins involved in glutamatergic neurotransmission and synaptic plasticity. Previous studies on rare genetic variations reported that such gene functions were independently implicated in schizophrenia. (Kirov et al. 2012; Fromer et al. 2014; Purcell et al. 2014) These results suggests convergence at a broad functional level between studies of common and rare genetic variations. None of the coding genes highlighted in this study were overlapping with those found in Grove *et al.* 2019.

A recent GWAS on 2,159 children with COS and 6,561 controls identified 4 significant risk loci (*MTHFR*, *TDGF1*, *ANGPTL2*, and *RALGPS1*) in genes which are also crucial for the brain development, contributing to the regulation of genes expressed in the brain and previously associated with other brain disorders. (Guo et al. 2021) Yet, none of these genes were overlapping with those highlighted in the previous multi-stage schizophrenia GWAS.

The concordance of results in different symptomatology highlights potential spatiotemporal convergence of genetic risk across different psychiatric diagnoses, on this specific developmental epoch, despite the profound genetic heterogeneity.

Genetic overlaps and genetic correlations between ASD, schizophrenia and phenotypic traits

Through the last decade, multiple evidence demonstrated strong clinical and genetic correlations across NDDs and complex traits, confirming shared etiology and pleiotropic effects[20] of variants contributing to NDDs.

Cross-disorder genetic correlations

Rare and common variants individually associated to ASD, schizophrenia and COS were overlapping across disorders and alter similar expression pathways. (Ahn et al. 2016; Anney et al. 2017; Marshall et al. 2017; Gandal et al. 2018; Fernandez et al. 2019; Lee et al. 2019; Forsyth and Asarnow 2020; Guo et al. 2021)

For the analysis performed on common variants, the linkage disequilibrium score[21] and the polygenic risk score (PRS)[22] have been increasingly used to measure genetic overlaps between different disorders. With these methods, multiple studies found a significant association between ASD diagnosis and schizophrenia. (Weiner et al. 2017; Savage et al. 2018; St Pourcain et al. 2018; Grove et al. 2019) The rare studies on common variations in COS also reported a significant correlation with ASD PRS as well as schizophrenia PRS. (Ahn et al. 2016; Guo et al. 2021) This is in line with other studies on psychotic experiences in adolescents reporting overlap in polygenic susceptibility with ASD (mediated by social impairment) and schizophrenia. (Pain et al. 2018; Velthorst et al. 2018; Perkins et al. 2020)

[20] A pleiotropic effect of a variant is defined as a gene variation associated to multiple seemingly unrelated disorder or phenotype.

[21] LD occurs when single nucleotide polymorphisms (SNPs) are non-randomly correlated with other SNPs at different loci *i.e.* they are more or less frequently associated than would be expected at random. LD score, is estimated for each SNP by taking the sum of correlations between that SNP and all nearby SNPs.

[22] A polygenic risk score is based on a set of trait-related single nucleotide polymorphism, detected by GWAS, that may not achieve significance at the individual level, but collectively may explain a substantial portion of the trait variance. It captures the aggregated effect of common variants in a disorder.

Genetic correlations between disorders and related phenotypic traits

Genetic correlations between disorders and IQ

Rare large effect-size variants are well-known to be associated with lower IQ in ASD and schizophrenia populations. (Bishop et al. 2017; Singh et al. 2017; Myers et al. 2020; Satterstrom et al. 2020) Interestingly, a consistent positive correlation with ASD risk and negative correlation with schizophrenia risk with IQ and educational attainment was found in multiple studies, and all ASD, schizophrenia and general intellectual functioning share contributing genes and hubs. (Weiner et al. 2017; Mistry et al. 2018; Niemi et al. 2018; Plomin and von Stumm 2018; Savage et al. 2018; Grove et al. 2019; Yahya 2020; Smeland et al. 2020) The contrasting observations in ASD suggest that common variant-associated risk may load on cognitive dimensions that are distinct from those affected by rare variants and may be more specific to ASD core symptoms. The possibility that, for a given effect-size on IQ, rare variants in some genes may have a substantially greater or lesser effect-size on core ASD features is worth exploring. It may facilitate elucidation of the genomics and neurobiology of social communication and interaction. (Myers et al. 2020)

Genetic correlations between ASD and other related traits

In a study based on genetic model fitting in twins, a negative genetic correlation between the severity of ASD and IQ was mainly explained by autistic trait items assessing communication difficulties. (Hoekstra et al. 2010) In line with these results, a genetic correlation between ASD and impairment in social communication was also observed in other studies. (St Pourcain et al. 2018; Grove et al. 2019) Several contributing genes are shared between ASD and language impairment. (Eicher and Gruen 2015) Nevertheless, rare large effect-size variants have been previously associated with an atypical ASD profile characterized by less impairment in social communication and language, as well as greater motor delay. (Bishop et al. 2017; Buja et al. 2018) Satterstrom *et al*. observed that ASD individuals carrying disruptive *de novo* variants in one of the 102 genes associated with ASD, which are implicated either in expression regulation or neuronal

communication, showed delayed age of walking and reduced IQ compared with those with no variants in these genes. (Satterstrom et al. 2020) Yet, carriers of disruptive variants in expression regulation genes showed significantly greater delays in age of walking compared with those with disruptive variants in neuronal communication genes. Other studies suggested an overlap between genes contributing to ASD and motor skills. (Bishop et al. 2017; Buja et al. 2018; Yahya 2020) Recently, ASD PRS was negatively correlated with empathy and positively correlated with systemizing. (Warrier et al. 2019) This suggests that social and non-social core ASD symptoms are partially genetically dissociable, and it is possible that such distinct genetic backgrounds might influence the diagnostic classification and clinical trajectory of individuals

The genetic-first approach: understanding the impact of structural variations on risks for NDDs and on other cognitive and behavioral traits

Some results emerging from diagnostic-first approach highlighted knowledge gaps, such as the unknown effect-size of genomic variants on core symptoms or comorbidities, the understudied genomic overlap in several diagnoses, or the hypothesis of a potential impact on a single dimension contributing to all diagnoses. The relevance of conducting genetic-first studies emerged from these knowledge gaps.

Investigating pathogenic recurrent CNVs to measure a gene-dosage contribution to ASD, schizophrenia, and other cognitive traits

Pathogenic CNVs are amongst the most frequently identified rare variants in over 10% of cases from the NDD clinic (Huguet et al. 2018; Zarrei et al. 2019), but limited progress has been made in identifying phenotype-genotype relationships. Indeed, the effect-size of CNVs on cognitive and behavioral dimensions are only characterized for a handful of recurrent CNVs (*e.g.*, 22q11.2 and 16p11.2 loci). These CNVs show reproducible effect-sizes on cognition, language, socio-communication, and brain structure, suggesting that these alterations drive their over-representation in neurodevelopmental and psychiatric conditions, such as ID, ASD and schizophrenia (Table 4). (Moreno-De-Luca et al. 2015; Bernier et al. 2016; D'Angelo et al. 2016; Sanders et al. 2019) The interest of

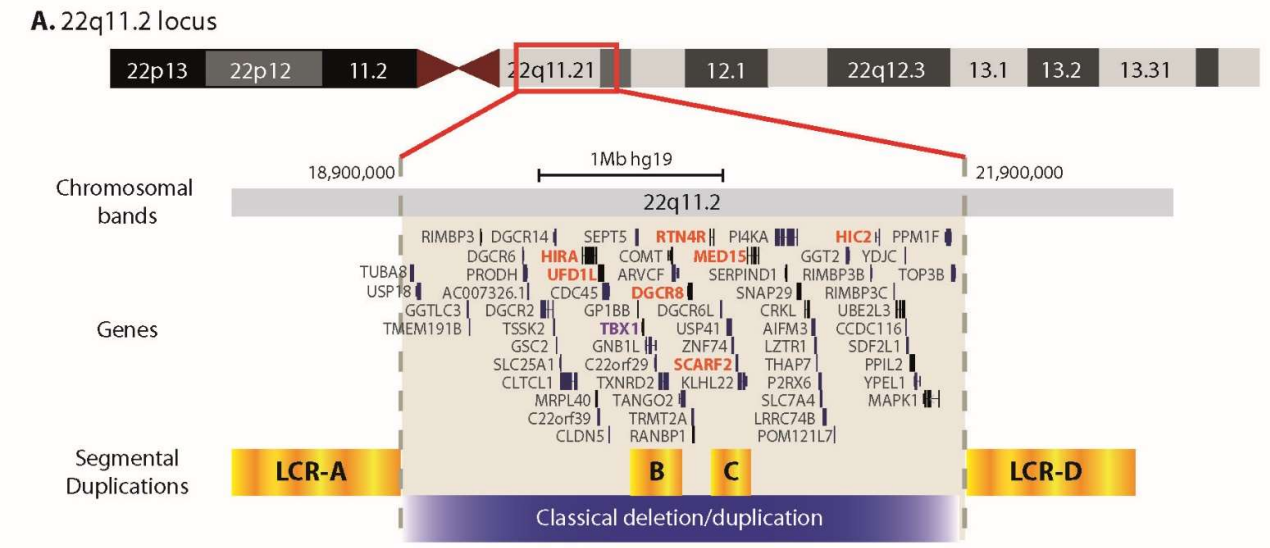
a genetic-first recruitment based on the presence of a common genetic risk factor is that it allows the investigation of pathways related to a particular biological risk for psychiatry irrespective of any clinical phenotype.

The 22q11.2 deletions and duplications

The 22q11.2 deletion was the first rare recurrent CNV associated with a distinct neurodevelopmental syndrome: the DiGeorge or velocardiofacial syndrome. (de la Chapelle et al. 1981) The reciprocal duplication was reported eighteen years later. (Edelmann, Pandita, and Morrow 1999) The micro-arrangement results in a hemizygous deletion or duplication of the long arm (q) of chromosome 22. This region is highly prone to non-allelic homologous recombination during meiosis due to the presence of 8 LCR, also named breakpoints (BPs), designated as LCR A to H. About 85–90% of micro-arrangement in this region are a 3 Mb deletion or duplication between LCR A and D (18.9–21.9 Mb, hg19[23]), which is referred to as the common 22q11.2 region including 46 protein-coding genes (Figure 8A). (Merico et al. 2014; Morrow et al. 2018)

The 22q11.2 deletions are the most common human microdeletion syndromes. Estimated prevalence of the deletion is one in every 6,000 live births. (Botto et al. 2003; Oskarsdottir, Vujic, and Fasth 2004) The prevalence of the duplication in the general population of UK Biobank was of 279 within the 419,861 individuals (1/1,500). (Kendall et al. 2019) In newly diagnosed cases, about 68% of deletions and 25% of duplications are *de novo*. (DECIPHER v9.17: Mapping the Clinical Genome n.d.)

[23] Genomic coordinates under the human genome reference build version 19 (also called genome reference consortium human build 37, or GRCh37)



B. Constrain scores of genes encompassed in the 22q11.2 locus

GENE	pLI	o/e	LOEUF
SEPT5	0.93	0.14	0.36
AC002472.13	0.00	1.23	1.74
AC006547.14	0.00	0.72	1.16
AIFM3	0.00	0.72	0.99
ARVCF	0.00	0.58	0.82
C22orf29	0.00	0.00	0.00
C22orf39	0.00	1.03	1.77
CDC45	0.00	0.38	0.60
CLDN5	0.90	0.00	0.41
CLTCL1	0.00	0.80	0.98
COMT	0.00	1.02	1.70
CRKL	0.45	0.20	0.64
DGCR14	0.00	0.61	0.96
DGCR2	0.00	0.55	0.88
DGCR6L	0.00	0.70	1.27
DGCR8	1.00	0.08	0.21
FAM230A	0.48	0.00	1.45
GGT2	0.00	0.73	1.36
GNB1L	0.00	0.60	1.05
GP1BB	0.51	0.00	1.33
GSC2	0.00	1.24	1.91
HIC2	1.00	0.00	0.18
HIRA	1.00	0.05	0.14
KLHL22	0.42	0.22	0.44
LZTR1	0.00	2.28	1.99

GENE	pLI	o/e	LOEUF
MED15	1.00	0.13	0.24
MRPL40	0.05	0.35	0.79
P2RX6	0.00	0.72	1.10
PI4KA	0.00	0.36	0.46
PRODH	0.00	0.76	1.10
RANBP1	0.82	0.10	0.47
RIMBP3	0.00	0.92	1.46
RIMBP3B	0.41	0.00	1.67
RTN4R	0.96	0.00	0.30
SCARF2	1.00	0.06	0.19
SERPIND	0.00	0.97	1.50
SLC25A1	0.06	0.31	0.65
SLC7A4	0.00	0.76	1.23
SNAP29	0.01	0.41	0.86
TANGO2	0.00	0.51	0.89
TBX1	0.84	0.14	0.43
THAP7	0.01	0.41	0.86
TRMT2A	0.00	0.61	0.92
TSSK2	0.00	0.55	1.15
TXNRD2	0.00	0.69	0.99
UFD1L	1.00	0.05	0.23
USP41	0.00	0.81	1.19
ZDHHC8	0.85	0.18	0.37
ZNF74	0.00	0.56	0.89

Figure 8: 22q11.2 locus

A. 22q11.2 locus with RefSeq genes codes and segmental duplications corresponding to low copy repeat A to D. Coordinates are based on Hg19. **B**. Coding genes encompassed in the 22q11.2 locus and corresponding constraint scores. These scores reflect intolerance to haploinsufficiency (pLI, observed/expected (o/e), and LOEUF defined by gnomAD v.2.1.1, Karczewski *et al.*, 2020). Genes in red are intolerant to haploinsufficiency.

The 22q11.2 deletion is enriched in NDD and is widely considered to be the genetic mutation with the highest effect-size for schizophrenia (OR = 92) (Table 4), with a lifetime prevalence of the disorder estimated to be about 25% in carriers. (Rees et al. 2014; Sanders et al. 2019) A study also suggested a higher prevalence of 22q11.2 deletions in COS compared to adult-onset schizophrenia (Ahn et al. 2014), and several studies reported this deletion in EOP cases (Ivanov et al. 2003; Vorstman et al. 2006) This CNV is also formally associated to ASD (OR = 32). (Sanders et al. 2019) At the contrary, duplications seem to be protective for schizophrenia (OR = 0.2) and is not significantly enriched in ASD (Table 4). (Rees et al. 2014; Sanders et al. 2019)

Cognition, behavior, language and motor disorders

The effect-size of the 22q11.2 deletion on cognition and other phenotypic traits have been extensively studied since its discovery. Its mean effect-size on IQ was estimated at -1.9 SD, with no significant difference between the verbal and the performance IQ. (Aken et al. 2009; Butcher et al. 2012; Vangkilde et al. 2016) Individuals with this deletion also show high impairment in daily functioning and adaptive skills, as measured by the VABS, with a mean effect-size of -2.3 SD. (Butcher et al. 2012) The social skills are affected, with +3.2 SD on the SRS raw score (Vangkilde et al. 2016), and the expressive language is significantly most impaired than the receptive one. (Solot et al. 2000; Glaser et al. 2002) Thus, expressive language and speech disorders dominate early and are persistent, with delay in language milestones. In a study with 305 children, the onset of language was delayed in approximately 70% of cases who did not speak or used only a few words or signs at 24 months of age or older. (Solot et al. 2000) The majority of children with 22q11.2 deletion demonstrate significant motor disorders persisting during the development, with gross and fine motor difficulties, as well as balance and coordination impairments. (Swillen et al. 2005; Sobin et al. 2006; Aken et al. 2009; Boot et al. 2015) Compared to controls, stationary positioning and locomotion were negatively affected by a mean effect-size of -1 and -1.5 z-scores, respectively, of the

corresponding subtest of the Peabody developmental motor scale – 2nd edition. (Swillen et al., 2005)

It is very likely that the duplication may be largely undetected in persons having almost no clear clinical symptoms. Thus, due to the low penetrance, estimated at 14% (Kendall et al. 2019), the exact effect-size of this microduplication is rather difficult to establish, because carriers are not systematically referred to clinics, conversely to the reciprocal deletion. Overall, the clinical phenotype associated with the duplication is mild but has a very heterogeneous expression. IQ seemed to be slightly affected with an intelligence ranging from mild ID to normal. (Campenhout et al. 2012; Yu et al. 2019) In the general population of UK Biobank[24], the mean effect-size was of -0.32 z-scores point of IQ. (Kendall et al. 2019) Behavioral problems are frequent in 22q11.2 duplication carriers, with reported aggressive and impulsive behaviors, as well as social deficits with behavior not adapted to the social demands of the situation. (Campenhout et al. 2012; Yu et al. 2019) Motor and language disorder are also present, with delayed onset for walking and speech, as well as gross and fine motor skills impairments. (Campenhout et al. 2012; Yu et al. 2019)

Anthropometric phenotypes, congenital malformations and neurological symptoms

Height and weight seemed to not be affected in deletion or duplication. Head circumference was either in the range of macrocephalia or microcephalia in the 8 duplication cases reported by Campenhout and colleagues, but it is not a characteristic feature of the 22q11.2 deletion or duplication. (Campenhout et al. 2012; McDonald-McGinn et al. 2015)

Typical congenital physical features associated with the 22q11.2 deletion can be ascribed to problems with the morphogenesis and subsequent abnormal function of pharyngeal arch system derivatives, including the craniofacial structures, thymus,

[24] UK Biobank is a national and international health biomedical research resource, following the health and well-being of 500,000 volunteer participants. This cohort provides extensive phenotypic and genetic data on half a million adults from the general population.

parathyroid, aortic arch and the cardiac outflow tract. (McDonald-McGinn et al. 2015) Thus, congenital facial or skeletal dysmorphia are systematically reported in deletion and duplication carriers (Campenhout et al. 2012; Yu et al. 2019) Notably, structural abnormalities of the palate are extremely common and characteristic of the 22q11.2 deletion, with over 71% of cases with this feature. Cardiovascular defects are often the initial manifestation that leads to diagnosis for deletions, which is not the case of duplications. (Campenhout et al. 2012; McDonald-McGinn et al. 2015; Yu et al. 2019) Overall, congenital heart disease represents the main cause of mortality of children with the 22q11.2 deletion (around 87%). (McDonald-McGinn et al. 2015)

Finally, epilepsy occurs in 15% of patients carrying the deletion and is rare in duplication carriers. (Campenhout et al. 2012; McDonald-McGinn et al. 2015; Wither et al. 2017; Yu et al. 2019)

How gene dosage of the 22q11.2 locus modulates brain structure and functional connectivity

The largest neuroimaging study to date in children and adults with 22q11.2 deletion syndrome (aged 8–50 years) reported a clear cortical phenotype in individuals with 22q11.2 deletion in comparison to controls, with thicker cortex bilaterally in major regions, thinner cortex in the superior temporal, cingulate, and parahippocampal cortex. (Sun et al. 2018) Extending these analyses to surface area showed a global reduction in deletion compared to controls. Analyzing these neuroanatomical patterns with a classifier successfully delineate carriers and controls with 93.8% of accuracy. For the first time, a genetic syndrome - highly associated with schizophrenia - was correlated to anatomical abnormalities converging with those found in idiopathic schizophrenia. Recent follow-up work comparing volume and subcortical morphometry data from deletion carriers and controls showed lower intracranial, thalamus, putamen, hippocampus, and amygdala volumes, as well as a greater lateral ventricle volume in deletion carriers. (Ching et al. 2020) Deletion carriers with psychosis, and carriers of the common deletion (LCR A to D), exhibited a higher effect on volumetric data compared to

carriers without psychosis or with the smaller deletion (LCR A to B). Lower CT and surface area in the same subcortical structures were also observed. Overall, the overlaps found between effect-sizes associated with CNVs and idiopathic psychiatric conditions may highlight the mechanistic homogeneity provided by a genetic-first approach.

Only few studies have investigated brain organization using functional connectivity during rest in 22q11.2 deletion, the reciprocal duplication or their relation. A study of association between functional network connectivity and severity metric of psychosis and CNV status reported an underconnectivity in visual, frontoparietal, and default mode networks in deletion, compared to control. (Mattiaccio et al. 2018) Decreased connectivity in the posterior cingulate and overconnectivity in the precuneus and superior parietal lobule were associated with higher thought disturbance score. Knowing the penetrance of schizophrenia in 22q11.2 deletion carriers and the alterations of the hippocampus and the thalamus in schizophrenia, Schleifer *et al.* performed hypothesis-driven functional analysis in these two regions of interest. (Schleifer et al. 2019) The thalamus displayed overconnectivity with somatomotor regions and underconnectivity with frontoparietal networks in deletion carriers compared to controls. The opposite pattern was reported for the hippocampus.

The 22q11.2 duplication effect on brain structure and connectivity has been far less investigated. In a study comparing deletion and duplication carriers to controls, a negative gene dosage effect on CT and a positive gene dosage effect on intracranial volume were reported (deletion<control<duplication). (Lin et al. 2017) Mirror patterns were extended into subcortical regions volumes for the hippocampus.

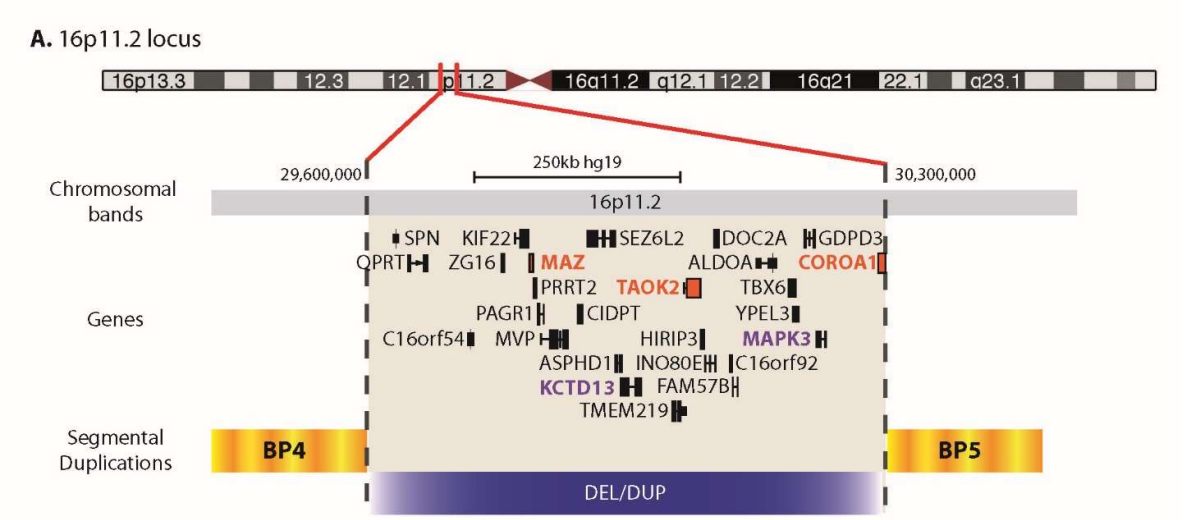
The 16p11.2 deletions and duplications[25]

Recurrent deletions and duplications between BPs 4 and 5 on chromosome 16p11.2 (29.6–30.3 Mb, hg19) were first associated with ASD and neurodevelopmental disorders in 2008 (Figure 9A). (Weiss et al. 2008) This genomic region encompasses 31 unique genes and is flanked by LCRs. 16p11.2 CNVs are the most frequently identified recurrent CNVs in patients with ASD and other neurodevelopmental disorders. Their rather high population frequency (1/1000 individuals from unselected populations carries either a deletion or a duplication) (Kendall et al. 2017) has allowed for the collection of large samples. Deletions and duplications occur *de novo* respectively in 60% and 24% of the cases (DECIPHER v9.17: Mapping the Clinical Genome n.d.) and have been associated with a range of neurodevelopmental disorders. (Sanders et al. 2019) The overrepresentation of deletions and duplications has been demonstrated in ASD cohorts with an OR of 14 (Table 4). (Sanders et al. 2019) The latter would translate into a risk for ASD of 15% based on the ASD population prevalence of 1.5%. 16p11.2 deletions and duplications occur in 0.31% to 1% of ASD cases. (Weiss et al. 2008; D'Angelo et al. 2016) As opposed to deletions, duplications are over-represented in schizophrenia (OR = 9.4) and an association study suggested a higher prevalence in childhood- than in adult-onset schizophrenia. (Ahn et al. 2014; Marshall et al. 2017)

Cognition, behavior, language and motor disorders

Several studies have measured the effect-size of 16p11.2 CNVs on cognition and other clinical traits. Deletions and duplications have a mean effect-size of approximately -1.5 SD on IQ. (D'Angelo et al. 2016) In the general population of UK Biobank, the prediction of the effect-size was smaller, with -0.3 to -0.4 z-score points of IQ for deletion and duplication, respectively. (Kendall et al. 2019) They both affect social responsiveness by approximately +3 SD, and negatively impact gross and fine motor skills.

[25] The following review about the 16p11.2 have been integrated in a chapter from a collaborative Notebook on ASD. Elise Douard and Sébastien Jacquemont are the authors of this chapter.



B. Constrain scores of genes encompassed in the 16p11.2 locus

GENE	pLI	o/e	LOEUF
CORO1A	0.97	0.10	0.32
MAPK3	0.04	0.31	0.61
GDPD3	0.00	1.06	1.51
YPEL3	0.04	0.45	1.15
TBX6	0.00	0.37	0.69
ALDOA	0.00	0.42	0.76
FAM57B	0.66	0.16	0.52
C16orf92	0.03	0.49	1.26
DOC2A	0.01	0.37	0.69
INO80E	0.01	0.45	0.95
HIRIP3	0.00	0.55	0.87
TAOK2	1.00	0.13	0.24
TMEM219	0.00	0.79	1.47
KCTD13	0.00	0.51	0.95
ASPHD1	0.00	0.50	1.00
SEZ6L2	0.12	0.25	0.42
CDIPT	0.13	0.31	0.79
MVP	0.00	0.49	0.73
PAGR1	0.74	0.11	0.54
MAZ	0.94	0.07	0.35
PRRT2	0.58	0.18	0.56
KIF22	0.00	0.59	0.85
ZG16	0.62	0.14	0.65
C16orf54	0.37	0.21	1.01
QPRT	0.00	0.52	1.10
SPN	0.01	1.82	1.94

Figure 9: 16p11.2 locus

A. 16p11.2 locus with RefSeq genes codes and segmental duplications corresponding to breakpoints 4 to 5. Coordinates are based on Hg19. **B.** Coding genes encompassed in the 16p11.2 locus and corresponding constraint scores. These scores reflect intolerance to haploinsufficiency (pLI, observed/expected (o/e), and LOEUF defined by gnomAD v.2.1.1, Karczewski *et al.*, 2020). Genes in red are intolerant to haploinsufficiency.

(Moreno-De-Luca et al. 2015; D'Angelo et al. 2016; Martin-Brevet et al. 2018) Language disorders are also major symptoms in deletion carriers. Measures of phonological memory (non-word repetition task) are decreased by -1.4 SD in deletion carriers whereas duplication carriers show no differences or even outperform intrafamilial controls when adjusting for IQ. (Hippolyte et al. 2016; Martin-Brevet et al. 2018) Overall, deletion carriers show a variety of speech and/or language difficulties. Studies have characterized these speech deficits as meeting diagnostic criteria for childhood apraxia of speech[26] or dysarthria. A majority of carriers demonstrate phonological errors, with final consonant deletion, gliding, weak syllable deletion, cluster reduction and cluster simplification most common. (Fedorenko et al. 2016)

Anthropometric phenotypes, congenital malformations and neurological symptoms

These reciprocal CNVs are also associated with mirror anthropometric phenotypes such as obesity and being underweight, as well as increased and decreased global and regional brain volumes in deletion and duplication carriers, respectively. (Jacquemont et al. 2011; Maillard et al. 2015) A large case-control obesity study showed that deletions were absent from healthy non-obese controls and accounted for 0.7% of morbid obesity cases (body mass index (BMI) ≥ 40 kg/m^2 in adults or ≥ 2 SD in children), demonstrating the potential importance in common disease of rare variants with large effects. (Jacquemont et al. 2011) Conversely, the duplication was associated with being underweight, defined as a BMI < 18.5 kg/m^2 in adults and < 2 SD from the mean in children, as well as failure to thrive, feeding and eating disorders. Effect-sizes on BMI and head circumference were both approximately -1 and +1 z-score for duplications and deletions on both measures when compared to controls. (D'Angelo et al. 2016)

Peripheral nervous system problems have been reported in individuals with 16p11.2 deletions. Height is slightly below average and vertebral anomalies are observed in 20% of carriers. (Zufferey et al. 2012) More infrequent features of the 16p11.2 deletion

[26] Apraxia is a motor speech disorder affecting the production, sequencing and timing of syllables and words.

include congenital anomalies of kidney and urinary tract. (Knoers and Renkema 2019) Although studies have reported various dysmorphic features (Rosenfeld et al. 2010; Reinthaler et al. 2014), larger studies have not confirmed a characteristic pattern of dysmorphic features that would facilitate a clinical diagnosis.

In a large clinical series (D'Angelo et al. 2016), epilepsy was reported in 35 of 180 duplication probands (19.4%) but only in 2 of 90 of their carrier relatives (2.2%). In deletions the reported frequency of epilepsy is similar: 69 of 317 probands (21.8%) and 4 of 73 relatives (5.5%).

How gene dosage of the 16p11.2 locus modulates brain structure and functional connectivity

Brain abnormalities have been reported by structural MRI studies (D'Angelo et al. 2016): out of a subset of 86 duplication carriers, enlarged ventricles and cerebellar hypoplasia were the most frequent findings (13 [15.1%] and 10 [11.6%], respectively). In deletion carriers, posterior fossa abnormalities were observed most frequently (36 of 108 [33.3%]), along with Chiari type I malformations (11 of 36 [30.6%]). Beyond these abnormalities detectable by the naked eye in a diagnostic setting, several quantitative neuroimaging studies reported gene dosage effects on global and regional brain metrics, demonstrating an association between brain anatomy changes and 16p11.2 copy number, with significant negative correlations between number of copies at this locus and both total intracranial volume and global cortical surface area. (Maillard et al. 2015; Martin-Brevet et al. 2018)

Regions affected by CNVs at the 16p11.2 locus include the insula, transverse temporal gyrus, and calcarine cortex (negative gene dosage), as well as the precentral gyrus and superior and middle temporal gyri (positive gene dosage). (Maillard et al. 2015; Martin-Brevet et al. 2018) Reciprocal changes in the language areas (middle, superior temporal gyrus and caudate), may underlie the language deficits reported in deletion but not in duplication carriers. Opposing volume changes in the reward circuitry (striatum, medio-dorsal thalamus, orbito-frontal cortex and insula) which are associated with

eating behavior, may explain the mirror BMI phenotype in 16p11.2 CNVs carriers. These regions overlap with those identified by studies of gray matter (GM) alterations present in individuals with psychosis or at high risk for developing psychosis as well as ASD. These regions include the anterior insular, anterior cingulate cortex and superior temporal gyrus. (Goodkind et al. 2015; Cauda et al. 2017)

A recent resting state MRI study has shown that the 16p11.2 deletion is associated with global overconnectivity which predominantly involved the ventral attention, motor, and frontoparietal networks. (Moreau et al. 2020) Duplication showed mirror effect with a global underconnectivity involving the anterior and lateral default mode network and the limbic network. Regional functional connectivity signatures defined by the 16p11.2 deletions and duplications, in particular, those implicating the thalamus, somatomotor, posterior insula and cingulate regions, contributed to the complex architecture of idiopathic ASD as well as schizophrenia.

Estimation errors related to ascertainment in gene first studies

Clinical and unselected population ascertainment can bias the estimated effect-size on cognitive or behavioral phenotypes. This is exemplified by the difference between 16p11.2 deletion effect-sizes on IQ estimated at -1.5 SD in the clinical cohorts and at -0.3 SD in the general population. (D'Angelo et al. 2016; Kendall et al. 2019) Systematic overestimations of the effect-size are observed in the clinical cohorts, which enroll individuals with sufficiently severe impairment to be referred, in comparison to the general populations, including mostly individuals with typical functioning.

Linking genes within recurrent loci to clinical phenotypes

The 22q11.2 example

The common 22q11.2 locus encompassed 46 protein-coding genes, 27 pseudogenes, 10 non-coding RNAs (including one read-through transcript), and 7 microRNAs (Figure 8A). (Merico et al. 2014; Morrow et al. 2018)

With respect to neuropsychiatric phenotypes, multiple candidate genes within the 22q11.2 region are expressed in the brain. (Guna, Butcher, and Bassett 2015) Many studies have investigated the potential effects of 22q11.2 genes using mouse models. This process was accelerated with the International Mouse Phenotyping Consortium, which screens mouse models of several 22q11.2 genes in translational phenotypes, including working memory, vocalization, and behavioral phenotypes seen in 22q11.2 carriers and idiopathic cases of schizophrenia and other NDDs. (Koscielny et al. 2014) These studies allow to delineate 22q11.2 candidate genes that are likely to contribute to the associated NDDs, such as EOP and ASD.

The most studied gene encompassed by the 22q11.2 region is *TBX1*, which encodes a T-box transcription factor. Heterozygous LoF variants of *TBX1* in mice result in partially penetrant cardiovascular, thymic and parathyroid glands defects. (Lindsay et al. 2001; Merscher et al. 2001) *TBX1* knock-out mice are embryonic lethal with persistent truncus arteriosus, cleft palate, and absence of thymus and parathyroid glands. At the cellular level, reduced proliferation and premature differentiation of progenitor cells expressing *TBX1* were detected in the mouse model. (Caprio and Baldini 2014) *TBX1* was also implicated in brain microvascular development and may play some part in cognitive and behavioral deficits related to schizophrenia. (Paylor et al. 2006; Cioffi et al. 2014)

DGCR8, DiGeorge syndrome critical region gene 8, encodes the *DGCR8* microprocessor complex subunit, a double-stranded RNA-binding protein that mediates the biogenesis of miRNAs[27]. (Stark et al. 2008; Xu et al. 2013) An increase in the global *DGCR8* expression in the superior temporal gyrus and the dorsolateral prefrontal cortex was previously observed in post-mortem study of individuals with schizophrenia. (Beveridge et al. 2010) In mouse models, heterozygosity of *DGCR8* results in neuronal deficits characteristic of the 22q11.2 deletion syndrome, whereas inactivation of both alleles in neural crest cells results in cardiovascular defects. (Stark et al. 2008; Chapnik et al.

[27] miRNAs are small non-coding RNAs that regulate the expression of target genes by binding to specific sites in mRNAs for translational repression or degradation.

2012) Small alterations in miRNA expression levels can have profound effects on brain development and plasticity, notably in synapses. (Beveridge et al. 2010; Follert, Cremer, and Béclin 2014; Petri et al. 2014) Recent studies propose that *DGCR8* may play a role in modulating the expression of genes outside of the 22q11.2 region that contribute to the phenotypes associated with 22q11.2 deletion. (Stark et al. 2008; Brzustowicz and Bassett 2012; Merico et al. 2014)

SCARF2 encodes the scavenger receptor expressed by endothelial cells 2 protein (SREC2). This receptor contains putative epidermal growth factor-like domains and may be involved in intracellular signaling. Homozygous or heterozygous variants of *SCARF2* are associated with Van den Ende-Gupta syndrome, which is characterized by severe contractural arachnodactyly, facial dysmorphism, as well as skeletal anomalies. (Bedeschi et al. 2011)

Recently developed genomic metrics also provide evidence for the contribution of genes to the clinical phenotype. Notably, the "probability of being loss-of-function intolerant" (pLI)[28] and "loss-of-function observed/expected upper bound fraction" (LOEUF)[29] scores are measuring when genes are intolerant to LoF variations. (Lek et al. 2016; Karczewski et al. 2020) Such scores directly reflect the level of negative selection pressure or reproductive success. Based on these scores, 7 genes are classified as highly intolerant to haploinsufficiency[30]: *HIRA*, *HIC2*, *SCARF2*, *DGCR8*, *UFD1L*, *MED15*, *RTN4R* (Figure 8A and 7B).

[28] The pLI measures a gene's intolerance to variation by comparing the expected to the observed amount of LoF variation in a gene using a reference population. (gnomAD v2.1.1). Genes that are significantly depleted of their expected LoF variants ($pLI > 0.9$) are considered intolerant of such variants. (Karczewski et al. 2020)

[29] The LOEUF score is the measure of loss-of-function observed/expected upper bound fraction based on similar principles than the pLI (gnomAD v2.1.1). Genes with a LOEUF < 0.35 are considered intolerant to LoF variants. (Karczewski et al. 2020)

[30] Haploinsufficiency of a gene arises when a single copy of the allele is insufficient to produce the standard phenotype, such that it produces little or no gene product. Such phenomenon may arise from LoF variants.

The 16p11.2 example

Linking genes within the BP4-BP5 region to phenotypes has been hindered by the number of genes (31 known protein-coding genes) at this locus and the diversity of clinical symptoms associated with these CNVs.

Several studies have attempted to identify major genes contributing to the neurodevelopmental and/or anthropometric phenotypes. Mouse models suggested that gene dosage of *TAOK2* (Thousand and one amino-acid kinase 2) increases brain size and impacts synapse development. (Richter et al. 2019) The absolute brain volume of *TAOK2* knock-out mice was significantly enlarged compared with wild type mice derived from absolute and relative volumetric increases in the hindbrain, midbrain, hypothalamus, thalamus, and cerebellum. The somatosensory cortex showed a relative decrease in volume. *TAOK2* Heterozygotic mice (similar to the heterozygous deletions in humans) also showed significant increases in brain volume, but not as dramatic as knock-out mice, consistent with a gene dosage effect. One *de novo* LoF variant and two *de novo* missense variants in *TAOK2* have also been reported in patients with neurodevelopmental disorders but these observations do not yet constitute a significant excess of *de novo* variants. (Coe et al. 2019)

KCTD13 (Potassium Channel Tetramerization Domain containing 13) may also be a major driver of neuroanatomical phenotypes. Overexpression of human *KCTD13* in zebrafish embryo can induce microcephaly, whereas suppression of the zebrafish ortholog yields a macrocephalic phenotype, potentially mimicking the phenotypes seen in 16p11.2 CNV carriers. (Golzio et al. 2012) *KCTD13* has also been linked to alteration of hippocampal synaptic transmission and dendritic complexity. (Golzio et al. 2012)

MAPK3 gene encodes the MAP kinase, *ERK1*. Variants in upstream elements regulating the ERK pathway have been genetically linked to ASD. Specifically, MAP kinases are important for normal cortical development and function. (Pucilowska et al. 2015; Blizinsky et al. 2016) While an excess of *de novo* missense variants has been reported for *MAPK3* in patients with neurodevelopmental disorders (Coe et al. 2019), this gene is also

currently considered as tolerant to haploinsufficiency and has not been associated with any human phenotype.

Of note, *TAOK2* interacts with *KCTD13* in the RhoA signaling pathway (Richter et al. 2019), and with *MAPK3* by activating *MAPK* pathway. (Ultanir et al. 2014)

Genes with the high intolerance to haploinsufficiency, measured by pLI and LOEUF, also provide evidence for several candidate genes including *TAOK2*, *CORO1A* and *MAZ* (Figure 9A and 8B). (Lek et al. 2016; Karczewski et al. 2020) Importantly, an excess of *de novo* LoF variants has not been reported in any of the genes within this region.

Overall, these observations have highlighted potential candidates and they suggest that multiple genes within 22q11.2 or 16p11.2 may contribute to the phenotypes via additive effects or interactions. (Iyer et al. 2018). The genes altered individually in these loci are not enough to explain the clinical presentation, given that none of the associated phenotypes show complete penetrance. For many neurodevelopmental CNVs, multiple genes with smaller individual effects appear to contribute to the overall risk. In addition, even among individuals with similarly sized recurrent CNVs, the clinical phenotypes are highly heterogeneous, suggesting the involvement of other genetic, environmental or stochastic factors. (Rees et al. 2014) The pleiotropic effect associated with these loci could be due to activity of proteins coded by the non-altered copy, or compensatory mechanisms. (Zinkstok et al. 2019) For example, evidence suggests that the 22q11.2 region contains regulatory genes and miRNAs that affect gene expression outside of the locus, and that genetic background variation might affect phenotypic expression. (Stark et al. 2008; Brzustowicz and Bassett 2012; Merico et al. 2014; Zinkstok et al. 2019)

Non-recurrent CNVs points towards a polygenic and additive model for NDDs

The statistical power required to conduct the studies described above limits this approach to most frequent variants. However, beyond the most recurrent CNVs associated with NDDs, genomic burden studies suggest that many more loci contribute to these disorders. (Krumm et al. 2015; Marshall et al. 2017; Wray, Wijmenga, et al. 2018;

Sanders et al. 2019; Sullivan and Geschwind 2019; Zarrei et al. 2019) In the genetic and NDD clinic, most of the CNVs identified in patients are non-recurrent and their effect-size on cognitive and behavioral traits as well as ASD risk remains undocumented. (Huguet et al. 2018) To my knowledge, no study measured the effect-sizes of non-recurrent CNVs on such traits and it remains unclear whether the overrepresentation of these CNVs in NDDs is related to their effect on core symptoms or on DSM-V-defined clinical specifiers (such as intelligence, language or co-occurring conditions in ASD).

Using constraint scores to estimate the effect-size of genome-wide deletions and duplications burden[31]

Statistical models using constraint scores such as the pLI and the LOEUF, trained on deletions and duplications in populations not selected for a clinical condition as well as individuals with ASD can accurately estimate the effect-size of CNVs on IQ. These models explain the mean effect-size of any CNV on IQ with an accuracy close to 80%. (Huguet et al. 2018; Huguet et al. 2021) These additive models suggest that 1) the effect-size of CNVs on IQ is highly associated with constraint sores such as pLI and LOEUF (Huguet et al. 2018; Huguet et al. 2021), 2) the effect-size of a CNV on IQ is the sum of small effects associated with each gene encompassed in the CNV and 3) a large proportion of the genome decrease IQ when deleted or duplicated, consistent with a highly polygenic model. (Weiner et al. 2017; Wray, Wijmenga, et al. 2018; Sanders et al. 2019) Models estimate that one third of the coding genes affected IQ by >1 point, when deleted. (Huguet et al. 2018; Huguet et al. 2021) These statistical models show that among many genetic and functional scores, measures of genetic fitness (*i.e.*, pLI and LOEUF) remain the best variable to explain the effect of CNVs on IQ. This link between genetic fitness and cognitive ability is in line with studies showing an excess of *de novo* variants in ID, ASD and other neurodevelopmental disorders. (Sanders et al. 2015; Deciphering Developmental Disorders Study 2017)

[31] The following review about the constraint scores have been integrated in a chapter from a collaborative Notebook on ASD. Elise Douard and Sébastien Jacquemont are the authors of this chapter.

To our knowledge, no publication to date has used these models to measure the effect-sizes of CNVs on risk for NDDs, such as ASD or EOP.

Knowledge Gap and Hypothesis

Rare deleterious SNVs and CNVs are identified in 15 to 20% of individuals with ASD. (Jiang et al. 2013; Sanders et al. 2015; Tammimies et al. 2015). The largest ASD case-control studies to date identified more than 400 genes and 14 CNVs. (Abrahams et al. 2013; Sanders et al. 2019; Satterstrom et al. 2020). These studies were only able to identify gene with the highest level of *de novo* SNVs and within the most recurrent CNVs. However, there is evidence that a much broader spectrum of rare variants may be implicated, as suggested by the overall increase in CNV burden associated with ASD. (Krumm et al. 2015; Wray, Wijmenga, et al. 2018; Sanders et al. 2019; Sullivan and Geschwind 2019; Zarrei et al. 2019)

Even less is known about rare variants contributing to EOP. To date, no variant was formally associated with this disorder. The lack of genetic data in EOP is reinforced by the absence of clinical guideline recommending genetic screening for these patients.

Knowledge gap: The effect-sizes on ASD risk of non-recurrent CNVs remain undocumented and even less is known about their contribution to EOP.

This is particularly problematic in the neurodevelopmental clinic, where these undocumented CNVs are routinely identified in patients referred for NDDs.

In addition, limited progress has been made in identifying phenotype-genotype relationships in ASD. The effect-size of genome-wide CNVs on the cognitive and behavioral dimensions related to ASD is understudied and have only been characterized for a handful of recurrent CNVs (*e.g.,* 22q11.2, 16p11.2). *De novo* variants have been associated with lower IQ, an atypical ASD profile characterized by less impairment in social communication and language, as well as greater motor delay. (Bishop et al. 2017; Buja et al. 2018; Satterstrom et al. 2020) Overall, the reasons underlying the overrepresentation of CNVs in autistic individuals remains unclear. It may be due to

their effect on core symptoms of ASD, or the clinical specifiers of ASD (*e.g.,* intellectual disabilities or language impairment co-occurring conditions). However, this has not been quantified to date.

To tackle the issue of non-recurrent CNVs, my research group recently developed statistical models, trained on deletions and duplications in the general population and showed that they can accurately estimate the effect-size of deleterious recurrent CNVs on IQ. (Huguet et al. 2018; Huguet et al. 2021)

Hypothesis: Genes with the same genetic and functional characteristics confer the same risk for ASD or EOP. Therefore, statistical models trained on large case-control datasets with CNVs encompassing genes covering a broad range of genetic and functional characteristics can predict ASD or EOP risks conferred by any CNVs.

Overarching Aim: Developing statistical models based on genomic scores to estimate the ASD and EOP risks conferred by any genome-wide CNV.

Specific Objectives

Aim 1: Estimating the effect-size of genome-wide CNVs on ASD risk

Only few recurrent CNVs have been associated with a clear effect-size on ASD risk through the genetic-first approach. To estimate the ASD risk conferred by any genome-wide deletion or duplication, I propose to develop statistical models based on constraint scores (*e.g.,* pLI, LOEUF) of genes encompassed in CNVs.

I will train the models using data on deletions and duplications identified in large ASD cohorts (*i.e.,* SSC, MSSNG) and unselected populations (*i.e.,* Generation Scotland, IMAGEN, Lothian Birth Cohort (LBC), Saguenay Youth Study (SYS)).

Aim 2: Estimating the effect-size of genome-wide CNVs on traits altered in ASD: intelligence, language and motor skills

ID, language impairment, and motor disorder are the rules rather than the exception in ASD; and the overlap between ASD and these disorders complicate the delineation of the phenotype-genotype relationships in ASD. In this aim, I will investigate whether the over-representation of CNVs in ASD is due to their effect on core symptoms or clinical specifiers.

To achieve this, I will apply the same statistical models (aim 1) to measures the effect-size of genome-wide CNVs on IQ, language impairment, and motor disorder in ASD cohorts (*i.e.,* SSC, MSSNG) and unselected population cohorts (*i.e.,* IMAGEN, SYS) when available.

Aim 3: Investigating the contribution of genome-wide CNVs to EOP in a unique sample and deducting substantial arguments to discuss the need of genetic screening for patient with EOP

To my knowledge, no CNVs have been associated with EOP risk and the etiology of this pathology remains understudied. In contrast with ASD, no guidelines for genetic screening in children with EOP exist yet.

To shed the light on the genetic architecture of EOP, I will:

3a) Provide the first rare recurrent CNVs association analysis in EOP, by comparing recurrent CNVs prevalence in EOP sample to the rate in controls from the unselected population (*i.e.,* Generation Scotland, IMAGEN, LBC).

3b) Estimate the EOP risk conferred by any genome-wide deletion or duplication by applying similar statistical models previously presented for ASD based on constraint score (*i.e.,* LOEUF) of genes encompassed in CNVs. I will train the models using genome-wide deletions and duplications identified in the EOP sample referred to the genetic

clinic of the Boston Children's hospital and unselected populations (*i.e.,* Generation Scotland, IMAGEN, LBC).

Effect-sizes of CNVs on EOP risk will be compared to effect-sizes of CNVs on ASD risk.

This work provides a novel framework to model ASD and EOP risk and the phenotypic profile of genomic variants.

II. Paper #I: Effect-Sizes of Deletions and Duplications on Autism Risk Across the Genome

Substantial contributions of the candidate:

1) Conception and design of the project with A. Zeribi (Biomedical M.Sc.), C. Schramm (Statistic Post.-Doc.), G. Huguet and S. Jacquemont.
2) CNV filtering and annotation of the ASD sample - MSSNG and SSC (controls were already published in Huguet et al., 2018); and selection of the phenotypes. (CNVs detection was done by Z. Saci and J.-L. Martineau).
3) Main and supplementary analyses (statistical models were reviewed by our Post.-Doc. in Statistics: C. Schramm and M. A. Loum) with A. Zeribi (preliminary analyses, seizure, regression, ADOS and ADI-R analyses); as well as P. Tamer (SRS analyses) and S. Nowak (CBCL analyses), two students that I supervised during their Master.
4) Interpretation of the statistical analyses with A. Zeribi, C. Schramm, G. Huguet & S. Jacquemont.
5) Creation of all tables and figures (Figure 4 with G. Huguet).
6) Main author when drafting (with C. Schramm, P. Tamer and S. Jacquemont) and answering to the AJP reviewers (with M. A. Loum and S. Jacquemont).

Authors

Elise Douard, M.Sc.[1,2]**; Abderrahim Zeribi, M.D.[1,2]**; Catherine Schramm, Ph.D.[1,2,3]**; Petra Tamer, B.Sc.[1,2]; Mor Absa Loum, Ph.D.[1,2]; Sabrina Nowak, B.Sc.[1,2]; Zohra Saci, Ph.D.[2]; Marie-Pier Lord, M.Sc.[1,2]; Borja Rodríguez-Herreros, Ph.D.[2,4]; Martineau Jean-Louis, M.Sc.[2]; Clara Moreau, M.Sc.[1,2]; Eva Loth, Ph.D.[5]; Gunter Schumann, Ph.D.[6], Zdenka Pausova, Ph.D.[7,8]; Mayada Elsabbagh, Ph.D.[9]; Laura Almasy, Ph.D.[10,11]; David C. Glahn, Ph.D.[12,13]; Thomas Bourgeron, Ph.D.[14]; Aurélie Labbe, Ph.D.[15]; Tomas Paus, Ph.D.[16,17]; Laurent Mottron, M.D., Ph.D.[18,19]; Célia M. T. Greenwood, Ph.D.[3,20,21]; Guillaume Huguet, Ph.D.[1,2]†; and Sébastien Jacquemont, M.D.[1,2]†

**Shared first authorship, † Shared last authorship
1. Université de Montréal, Montreal, Quebec, Canada
2. UHC Sainte-Justine Research Center, Montreal, Quebec, Canada
3. Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada
4. Sensory-Motor Laboratory, Jules-Gonin Eye Hospital, University of Lausanne, Lausanne, Switzerland
5. Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
6. Center for Population Neuroscience and Stratified Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
7. The Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada
8. Departments of Physiology and Nutritional Sciences, University of Toronto, Toronto, Ontario, Canada
9. Departments of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada
10. Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, Pennsylvania, United States
11. Department of Genetics, University of Pennsylvania Perelman School of Medicine, Pennsylvania, United States
12. Departments of Psychiatry, Boston Children's Hospital, Boston, Massachusetts, United States
13. Harvard Medical School, Boston, Massachusetts, United States
14. Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, Paris, France
15. Département de Sciences de la Décision, HEC Montreal, Montreal, Quebec, Canada
16. Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada
17. Departments of Psychology and Psychiatry, University of Toronto, Toronto, Ontario, Canada
18. Centre de Recherche de CIUSSS-NIM, Montreal, Quebec, Canada
19. Département de Psychiatrie, Université de Montréal, Montreal, Quebec, Canada
20. Department of Epidemiology, Biostatistics & Occupational Health, McGill University, Montreal, Quebec, Canada
21. Gerald Bronfman Department of Oncology, and Department of Human Genetics, McGill University, Montreal, Quebec, Canada

Received in American Journal of Psychiatry in August 14, 2019; revisions received January 23 and March 24, 2020; accepted May 11, 2020.

Acknowledgments

Funding/Support: This research was enabled by support provided by Calcul Quebec (http://www.calculquebec.ca) and Compute Canada (http://www.computecanada.ca). Sebastien Jacquemont is a recipient of a Bursary Professor fellowship of the Swiss National Science Foundation, a Canada Research Chair in neurodevelopmental disorders, and a chair from the Jeanne et Jean Louis Levesque Foundation. Catherine Schramm is supported by an Institute for Data Valorization (IVADO) fellowship. Petra Tamer is supported by a Canadian Institute of Health Research (CIHR) Scholarship Program. Guillaume Huguet is supported by the Sainte-Justine Foundation, the Merit Scholarship Program for foreign students, and the Network of Applied Genetic Medicine fellowships. Eva Loth is supported by European Autism Interventions, which receives support from the Innovative Medicines Initiative Joint Undertaking under grant agreement 115300, the resources of which are composed of financial contributions from grant FP7/2007-2013 from the European Union's Seventh Framework Programme, the European Federation of Pharmaceutical Industries and Associations companies' in-kind contributions, and Autism Speaks. Thomas Bourgeron is a recipient of a chair of the Bettencourt-Schueler foundation. Laurent Mottron is a recipient of the Marcel & Rolande Gosselin research chair. This work is supported by a grant from the Brain Canada Multi-Investigator initiative and CIHR grant 159734 (Sebastien Jacquemont, Celia Greenwood, Tomas Paus). The Canadian Institutes of Health Research and the Heart and Stroke Foundation of Canada fund the Saguenay Youth Study (SYS). SYS was funded by the Canadian Institutes of Health Research (Tomas Paus, Zdenka Pausova) and the Heart and Stroke Foundation of Canada (Zdenka Pausova). Funding for the project was provided by the Wellcome Trust. This work was also supported by an NIH award U01 MH119690 granted to Laura Almasy, Sebastien Jacquemont and David Glahn and U01 MH119739. The authors gratefully acknowledge the resources provided by the Autism Speaks MSSNG project and the Autism Genetic Resource Exchange Consortium, as well as the participating families.

We are grateful to all the families who participated in the Simons Variation in Individuals Project (VIP) and the Simons VIP Consortium (data from Simons VIP are available through

SFARI Base). We thank the coordinators and staff at the Simons VIP and SCC sites. We are grateful to all of the families at the participating SSC sites and the principal investigators (A. Beaudet, M.D., R. Bernier, Ph.D., J. Constantino, M.D., E. Cook, M.D., E. Fombonne, M.D., D. Geschwind, M.D., Ph.D., R. Goin-Kochel, Ph.D., E. Hanson, Ph.D., D. Grice, M.D., A. Klin, Ph.D., D. Ledbetter, Ph.D., C. Lord, Ph.D., C. Martin, Ph.D., D. Martin, M.D., Ph.D., R. Maxim, M.D., J. Miles, M.D., Ph.D., O. Ousley, Ph.D., K. Pelphrey, Ph.D., B. Peterson, M.D., J. Piggot, M.D., C. Saulnier, Ph.D., M. State, M.D., Ph.D., W. Stone, Ph.D., J. Sutcliffe, Ph.D., C. Walsh, M.D., Ph.D., Z. Warren, Ph.D., and E. Wijsman, Ph.D.). We appreciate obtaining access to phenotypic data on SFARI base.

Additional Contributions: Julien Buratti (Institute Pasteur), and Vincent Frouin, Ph.D. (Neurospin), acquired data for IMAGEN. Manon Bernard, BSc (database architect, The Hospital for Sick Children), and Helene Simard, MA, and her team of research assistants (Cégep de Jonquière) acquired data for the Saguenay Youth Study. Antoine Main, M.Sc. (UHC Sainte-Justine Research Center, HEC Montreal), Lionel Lemogo, M.Sc. (UHC Sainte-Justine Research Center), and Claudine Passo, Pg.D. (UHC Sainte-Justine Research Center), provided bioinformatical support. Maude Auger, Pg.D.; and Kristian Agbogba, B.Sc. (UHC Sainte-Justine Research Center), provided website development. Dr. Paus is the Tanenbaum Chair in Population Neuroscience at the Rotman Research Institute, University of Toronto, and the Dr. John and Consuela Phelan Scholar at Child Mind Institute, New York.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, or interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

ABSTRACT

Objective: Deleterious copy number variants (CNVs) are identified in up to 20% of individuals with autism. However, levels of autism-risk conferred by most rare CNVs remain unknown. We recently developed statistical models to estimate the effect-size on IQ of all CNVs including undocumented ones. We aimed to extend this model to autism susceptibility.

Methods: We identified CNVs in two autism (Simons Simplex Collection and MSSNG) and two unselected populations (IMAGEN and Saguenay Youth Study). Statistical models tested 9 quantitative variables associated with genes encompassed in CNVs to explain their effect on IQ, autism susceptibility and behavioural domains.

Results: The "probability-of-being loss-of-function intolerant" (pLI) best explains the effect of CNVs on IQ and autism risk. Deleting 1 point of pLI decreases IQ by 2.6 points in autism and unselected populations. The effect of duplications on IQ is three-fold smaller. Autism susceptibility increases when deleting or duplicating any point of pLI. This is true for individuals with high or low IQ and after removing de novo and known recurrent neuropsychiatric CNVs. Once CNV effects on IQ are accounted for, autism susceptibility remains mostly unchanged for duplications but decreases for deletions. Model estimates for autism risk overlap with previously published observations. Deletions and duplications differentially affect social communication, behaviour, and phonological memory, whereas both equally affect motor skills.

Conclusions: Autism risk conferred by duplications is less influenced by IQ compared to deletions. Our model, trained on CNVs encompassing >4,500 genes, suggests highly polygenic properties of gene dosage with respect to autism risk and IQ loss. These models will help to interpret CNVs identified in the clinic.

INTRODUCTION

Autism is a neurodevelopmental condition currently defined by atypical social communication and interaction, intense interests, and repetitive behaviour (American Psychiatric Association 2013). Levels of general intelligence and language are not diagnostic criteria but are recognized as clinical specifiers which have been defined as important features of the heterogeneity of autism. (Ousley and Cermak 2014) Neurodevelopmental and psychiatric comorbidities occur in up to 70% of children with autism. (Simonoff et al. 2008) The heritability of autism has been estimated between 50-80%. (Gaugler et al. 2014; Sandin et al. 2017) Deleterious Single Nucleotide Variants (SNV) and Copy Number Variants (CNVs) are identified in 15 to 20% of individuals with autism. (Jiang et al. 2013; Sanders et al. 2015; Tammimies et al. 2015) The largest rare variant autism case-control association studies to date have formally associated 102 genes and 16 CNVs at 13 genomic loci. (Malhotra and Sebat 2012; Moreno-De-Luca et al. 2013; Satterstrom et al. 2018; Sanders et al. 2019) Many more genomic loci are likely implicated as suggested by the overall increase in CNV burden associated with autism. (Girirajan et al. 2011; Malhotra and Sebat 2012; Krumm et al. 2015; Marshall et al. 2017; Sanders et al. 2019) Therefore, the susceptibility to autism conferred by most CNVs remains undocumented. This is particularly problematic in the neurodevelopmental clinic, where undocumented CNVs are routinely diagnosed in a large proportion of patients.

Even less is known about the effect-size of CNVs on the cognitive and behavioural dimensions related to autism, which have only been characterized for a handful of recurrent CNVs (*e.g.* 22q11.2, 16p11.2, 15q11.2, and 1q21.1 loci). These CNVs show reproducible effect-sizes on cognition, language, socio-communication, and brain structure, suggesting that these alterations drive their over-representation in autism or other neurodevelopmental and psychiatric conditions. (Butcher et al. 2012; D'Angelo et al. 2016; Bernier et al. 2016)

Limited progress has been made in identifying phenotype-genotype relationships in autism. Studies have demonstrated that rare *de novo* variants are associated with lower intelligence quotient (IQ) and are over-represented in females. (Girirajan et al. 2011; Girirajan et al. 2013; Iossifov et al. 2014; Jacquemont et al. 2014; Mottron et al. 2015) *De novo* variants have also been associated with an atypical autism profile characterized by less impairment in social communication and language, as well as greater likelihood of motor delay. (Bishop et al. 2017; Buja et al. 2018) Overall, the reasons underlying the overrepresentation of rare variants in autistic individuals remains unclear. It may be due to their effect on core symptoms of autism, or DSM-5-defined clinical specifiers of autism (intelligence, language, co-occurring conditions). Since CNVs have a strong influence on IQ and behavioural problems, including autism symptoms, it is of interest to examine the effect-size of CNVs on autism risk while accounting for their effect-size on IQ.

We previously reported that statistical models, trained on benign deletions in populations not selected for a clinical condition, can accurately estimate the effect-size of deleterious deletions on non-verbal IQ (NVIQ). (Huguet et al. 2018) These results suggest that 1) the effect-size of deletions on NVIQ can be estimated using constraint scores, such as the "probability of being Loss-of-function Intolerant" (pLI, definition in textbox, Figure 10) (Lek et al. 2016), and 2) the effect of haploinsufficiency on NVIQ applies to a large proportion of the genome, consistent with a highly polygenic model. (Weiner et al. 2017; Wray et al. 2018) Using pLI as an explanatory variable, we estimated that one third of the coding genes affect NVIQ by >1 point, when deleted. (Huguet et al. 2018) Previously, we were unable to establish the effect-size of duplications, likely due to inadequate power with the then-available sample size. Here, we propose to develop similar models to estimate autism susceptibility conferred by undocumented CNVs. We also aim to estimate their effects on cognitive and behavioural dimensions, which may underpin their overrepresentation in autism.

We 1) tested whether the effect-size of gene dosage on NVIQ is the same across unselected populations and autism cohorts, 2) selected models that best explain the

autism risk conferred by any deletions or duplications, while accurately adjusting for their effect on NVIQ established in step 1, and 3) investigated the cognitive, behavioural, and motor phenotypes that may explain the association between gene dosage and autism.

Models integrating genomic and functional scores of genes included in CNVs were trained on all CNVs ≥ 50 kb identified in two autism cohorts and two cohorts recruited from unselected populations. We provide a novel framework to model autism risk and the phenotypic profile of rare variants, regardless of effect-size and inheritance. This approach contrasts with previous genotype-phenotype studies restricted to small groups of individuals with *de novo* or recurrent variants.

METHODS

Cohorts

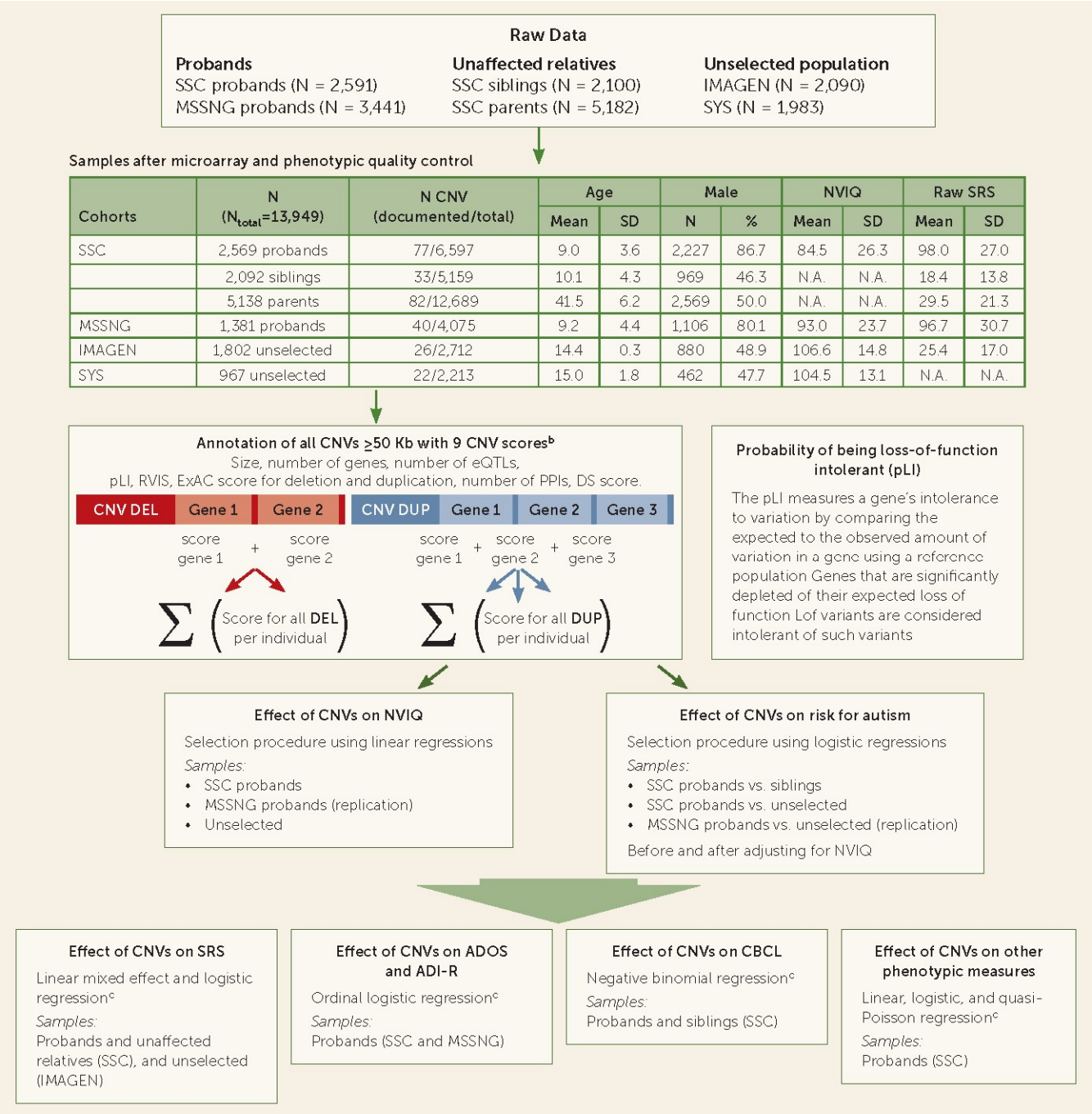
Autism cohorts

We studied two autism samples and intra-familial controls when available (Figure 10 and Table S1 in the online supplement). The Simons Simplex Collection (SSC) (Fischbach and Lord 2010), a cohort of 2,569 simplex families: 2,074 quads (one autistic proband, unaffected parents, and one unaffected sibling) and 495 trios (one autistic proband and unaffected parents). The MSSNG database, used as an independent replication cohort, includes 1,381 probands with autism. (Yuen et al. 2017)

Unselected cohorts

We included 2,769 individuals from two community-based cohorts that we previously studied (Huguet et al. 2018): IMAGEN (N=1,802) (Schumann et al. 2010) and the Saguenay Youth Study (SYS; N=967) (Pausova et al. 2017) (Figure 10 and Table S1 in the online supplement).

Figure 10: Methodological pipeline for a study of effect-sizes of deletions and duplications on autism risk across the genome[a]



Samples after microarray and phenotypic quality control

Cohorts	N (N_{total}=13,949)	N CNV (documented/total)	Age		Male		NVIQ		Raw SRS	
			Mean	SD	N	%	Mean	SD	Mean	SD
SSC	2,569 probands	77/6,597	9.0	3.6	2,227	86.7	84.5	26.3	98.0	27.0
	2,092 siblings	33/5,159	10.1	4.3	969	46.3	N.A.	N.A.	18.4	13.8
	5,138 parents	82/12,689	41.5	6.2	2,569	50.0	N.A.	N.A.	29.5	21.3
MSSNG	1,381 probands	40/4,075	9.2	4.4	1,106	80.1	93.0	23.7	96.7	30.7
IMAGEN	1,802 unselected	26/2,712	14.4	0.3	880	48.9	106.6	14.8	25.4	17.0
SYS	967 unselected	22/2,213	15.0	1.8	462	47.7	104.5	13.1	N.A.	N.A.

[a]Probands from the SSC and the MSSNG database are defined as individuals recruited on the basis of a diagnosis of autism. Siblings and parents from the SSC did not meet diagnostic criteria for autism. Ten individuals from the IMAGEN cohort and none in the SYS met criteria for autism spectrum disorder (as estimated by the Development and Well-Being Assessment). A total of 1,490 unaffected siblings from the SSC, 3,660 unaffected parents from the SSC, and 1,465 individuals from the general population carry at least one CNV >50 kb (see Table S1 in the supplementary material 1). ADI-R=Autism Diagnostic Interview–Revised; ADOS=Autism Diagnostic Observation Schedule; CBCL=Child Behavior Checklist; CNV=copy number variant; DEL=deletions; DS score=differential stability score; DUP=duplications; eQTLs=expression quantitative trait loci; ExAC=Exome Aggregation Consortium; N.A.=not applicable; NVIQ=nonverbal IQ; pLI=probability of being loss-of-function intolerant; PPIs=protein-protein interactions; RVIS=residual variation intolerance score; SRS=Social Responsiveness Scale; SSC=Simons Simplex Collection; SYS=Saguenay Youth Study.

[b]Microarray quality control and CNV selection and annotation were performed as previously described (Girirajan et al. 2011) (see also the Methods section in the supplementary material 1).

[c]The model used and the available data for each phenotype are detailed in Table 5 and Tables S7–S10 in the supplementary material 1.

CNV calling and annotation

We analyzed genotyping data from SSC, IMAGEN, and SYS and whole genome sequencing data from MSSNG. CNV detection, filtering, and annotation are detailed in Methods in the online supplement. We attributed 9 scores to deletions and duplications. These included size, number of genes, number of expression quantitative trait loci regulating genes expressed in the brain (Ramasamy et al. 2014). Each coding gene with all isoforms fully encompassed in CNVs was annotated using 4 constraint scores which reflect genetic fitness. The pLI score (ExAC v1.0), which is available for 18,224 genes and ranges from 0 (the gene is tolerant to haploinsufficiency) to 1 (the gene is intolerant to haploinsufficiency with a 100% probability). (Lek et al. 2016) Genes with 80% or 90% probabilities of being intolerant are considered as intolerant (Lek et al. 2016; Karczewski et al. 2019; Sanders et al. 2019). The 3 other constraint scores included the residual variation intolerance score (RVIS) (Petrovski et al. 2015), the deletion and duplication scores from ExAC (Ruderfer et al. 2016). Coding genes were also scored using the number of protein-protein interactions (Szklarczyk et al. 2015) and the differential stability score (Hawrylycz et al. 2015). We computed the ancestry in the SSC, IMAGEN and SYS cohorts based on HapMap3 reference population. (International HapMap Consortium 2003)

Clinical assessments

NVIQ data were available across all cohorts. (Fischbach and Lord 2010; Schumann et al. 2010; Pausova et al. 2017; Yuen et al. 2017) The assessment methods are detailed in Methods and Table S2 in the online supplement. All other cognitive, behavioural, and motor phenotypes are detailed in Table 5, Methods and Table S1 in the online supplement. Participants underwent age- and development-appropriate standardized cognitive and behavioural tests.

Table 5: Effect-size of gene dosage measured by pLI on phenotypes in autistic probands from the Simons Simplex Collection[a]

			No Adjustment for NVIQ			Adjustment for NVIQ		
Phenotypic Measurements	N	CNV Variable	b or Odds Ratio	SE or 95% CI	p	b or Odds Ratio	SE or 95% CI	p
Autism-related symptoms								
Regression	2,568	pLI DEL	0.86	0.75-0.95	8.4×10^{-3}	0.8	0.70-0.89	1.9×10^{-4}
		pLI DUP	0.99	0.92-1.04	0.65	0.96	0.90-1.02	0.19
Language and phonology								
CTOPP score	1,988	pLI DEL	–0.08	0.02	5.5×10^{-4}	–0.02	0.02	0.24
		pLI DUP	–0.02	0.02	0.25	0.006	0.01	0.66
Word delay	2,567	pLI DEL	1.16	1.07-1.27	5.0×10^{-4}	1.12	1.03-1.22	0.01
		pLI DUP	1.03	0.98-1.09	0.24	1.02	0.96-1.08	0.6
Phrase delay	2,567	pLI DEL	1.04	0.98-1.09	0.42	0.95	0.87-1.04	0.25
		pLI DUP	1.06	1.00-1.14	0.08	1.03	0.96-1.11	0.46
Adaptive skills (VABS-II)								
Total score	2,569	pLI DEL	–0.07	0.02	3.1×10^{-5}	–0.004	0.01	0.72
		pLI DUP	–0.03	0.01	2.6×10^{-3}	–0.01	0.01	0.4
Daily living	2,569	pLI DEL	–0.07	0.02	1.2×10^{-4}	–0.004	0.01	0.8
		pLI DUP	–0.04	0.01	3.4×10^{-3}	–0.01	0.01	0.38
Communication	2,569	pLI DEL	–0.07	0.02	3.4×10^{-4}	0.01	0.01	0.54
		pLI DUP	–0.04	0.01	4.5×10^{-3}	–0.005	0.01	0.6
Socialization	2,569	pLI DEL	–0.06	0.02	4.8×10^{-4}	–0.01	0.01	0.67
		pLI DUP	–0.03	0.01	0.02	–0.004	0.01	0.66
Motor skills								
VABS-II motor skills	919	pLI DEL	–0.11	0.04	4.1×10^{-3}	–0.08	0.03	0.01
		pLI DUP	–0.07	0.02	1.3×10^{-3}	–0.04	0.02	0.02
VABS-II gross motor skills	926	pLI DEL	–0.08	0.03	0.01	–0.07	0.03	0.02
		pLI DUP	–0.05	0.02	4.6×10^{-3}	–0.04	0.02	0.03
VABS-II fine motor skills	923	pLI DEL	–0.1	0.04	0.01	–0.07	0.03	0.04
		pLI DUP	–0.06	0.02	8.8×10^{-3}	–0.03	0.02	0.14
Onset for walking in months	2,550	pLI DEL	1.03	1.02-1.04	2.2×10^{-11}	1.03	1.02-1.04	4.6×10^{-9}
		pLI DUP	1.02	1.01-1.03	7.0×10^{-9}	1.02	1.01-1.03	5.3×10^{-8}
Delayed onset for walking	2,564	pLI DEL	1.16	1.05-1.28	2.0×10^{-3}	1.11	1.00-1.22	0.03
		pLI DUP	1.2	1.11-1.30	6.1×10^{-6}	1.19	1.09-1.29	4.2×10^{-5}
DCDQ score	2,209	pLI DEL	–0.07	0.02	2.5×10^{-3}	–0.03	0.02	0.16
		pLI DUP	–0.03	0.01	0.04	–0.01	0.01	0.33
Associated neurological condition								
Nonfebrile seizure	2,566	pLI DEL	1.12	1.01-1.23	0.02	1.07	0.96-1.17	0.19
		pLI DUP	1.04	0.95-1.11	0.3	1.02	0.94-1.09	0.63

[a]Phenotypic measures were z-scored using normative data when available or computed using the full autistic proband group (see the Methods section in the supplementary material 1). Boldface indicates p values significant bellow the statistical threshold ($p \leq 2.7\text{x}10^{-3}$). Effects in this table represent either the normalized b z-scores with their standard errors or odds ratios with their 95% confidence intervals. CTOPP=Comprehensive Test of Phonological Processing; CNV=copy number variant; DCDQ=Developmental Coordination Disorder Questionnaire; NVIQ=nonverbal IQ; pLI=probability of being loss-of-function intolerant; pLI DEL or pLI DUP=deleted or duplicated point of pLI score; VABS-II=Vineland Adaptive Behavior Rating Scales, Second Edition.

Statistical analyses

Effect-size of gene dosage on general intelligence in probands and the unselected populations

For each individual, we computed the sum of a given score for deletions and duplications separately (Figure 10, Methods in the online supplement). These deletion and duplication scores were used as two independent main effects in the model. We performed a stepwise variable selection procedure based on Bayesian information criteria to identify which score (among the 9 tested) best explain NVIQ for deletions and duplications. This was performed independently for the SSC probands, the unselected populations, and MSSNG as a replication dataset. To investigate the influence of the presence of lower IQ in the SSC, we assessed the effect-size of gene dosage on NVIQ in the SSC probands after performing 1:2 matching with MSSNG probands based on NVIQ (Methods and Figure S2 in the online supplement). Age, sex, ancestry and familial relatedness were used as covariates when applicable (Methods in the online supplement).

Effect-size of gene dosage on autism risk

We performed the same stepwise variable selection procedure to identify CNV scores that best explain the effect-size of deletions and duplications on autism risk. The dependent variable was the binary diagnosis (autism/control) and independent variables were the selected CNV scores. Conditional logistic regression was used when matching SSC probands with their unaffected siblings. Simple logistic regression was used when comparing SSC probands with the unselected populations. We assessed the effect-size of gene dosage on autism risk beyond its effect on NVIQ by adjusting for NVIQ or performing 1:1 matching of probands with individuals from the unselected populations based on NVIQ (Methods and Figure S1D in the online supplement). Replication analyses were performed using the MSSNG dataset. Sex, ancestry and

familial relatedness were used as covariates when applicable (Methods in the online supplement).

To estimate the proportion of autism-risk potentially mediated by NVIQ for deletions and duplication, we performed a counterfactual-based mediation analysis on the pooled dataset.

Sensitivity analyses

For sensitivity analyses, we pooled all samples and excluded individuals with CNVs > 10 points of pLI (deletions with an effect > 2 standard deviations of NVIQ) or recurrent CNVs associated with neurodevelopmental disorders or rare *de novo* CNVs (Tables S3, S4 and S5 in the online supplement).

Estimating and validating the level of autism risk

We compared the autism risk estimated by our model to that previously published for recurrent CNVs. Our literature search identified 16 CNVs with available odds ratios (ORs) (Sanders et al. 2019; Malhotra and Sebat 2012; Moreno-De-Luca et al. 2013; Chaste et al. 2014) (Table S6 in the online supplement). The model was trained using a pooled dataset including SSC and MSSNG probands, unaffected siblings, and unselected populations, excluding these 16 CNVs.

To illustrate the output of our model, we computed the autism risk for each CNV called in both autism cohorts including at least one gene with a pLI annotation. We also computed autism risk for any 1MB CNV across the genome, generating a series of 1Mb deletions and duplications (Human Gene Nomenclature) by moving a sliding window in 50Kb steps across the genome. (Wain et al. 2002) We chose 1Mb CNVs based on thresholds for deleteriousness used in previous studies. (Cooper et al. 2011; Jacquemont et al. 2014)

Effect-size of gene dosage on measures of core symptoms and specifiers of autism

We investigated the effect of the previously selected CNVs score on cognitive, behavioural, and motor phenotypes to understand why they increase susceptibility to autism. The choice of the statistical model depended on the distribution of the phenotypic measure (Methods and Table S7 in the online supplement). The Social Responsiveness Scale (SRS) was investigated using the entire SSC, MSSNG probands and IMAGEN cohorts (Methods and Table S8 in the online supplement). The Autism Diagnostic Observation Schedule (ADOS) and Autism Diagnostic Interview-Revised (ADI-R) were investigated using probands from SSC and MSSNG (Methods and Table S9 in the online supplements). The Child behaviour Checklist (CBCL) was investigated on probands and unaffected siblings from the SSC (Methods and Table S10 in the online supplements). All other phenotypic measurements were analysed using SSC probands alone. For all analyses, age, sex, ancestry and familial relatedness were used as covariates when applicable. Phenotypic measures were also tested with and without adjustment for NVIQ and/or autism diagnosis when available (Methods in the online supplement). Computation of the significance threshold is detailed in Methods in the online supplement.

RESULTS

Effect-size of gene dosage on general intelligence in probands and the unselected populations

As we previously observed in unselected populations (26), the variable selection procedure identified the sum of pLI scores as the variable that best explains the variance of NVIQ in the SSC for deletions (r^2=0.014) and duplications (r^2=0.004), compared to the 8 other scores. The sum of pLI scores per individual ranges from 0 to 18.92 and 35.71 for deletions and duplications respectively. As an example, a CNV scoring 2 points of pLI

may include either 2 genes with a 100% probability of being intolerant or 3 genes with moderate to high probabilities (60 to 90%).

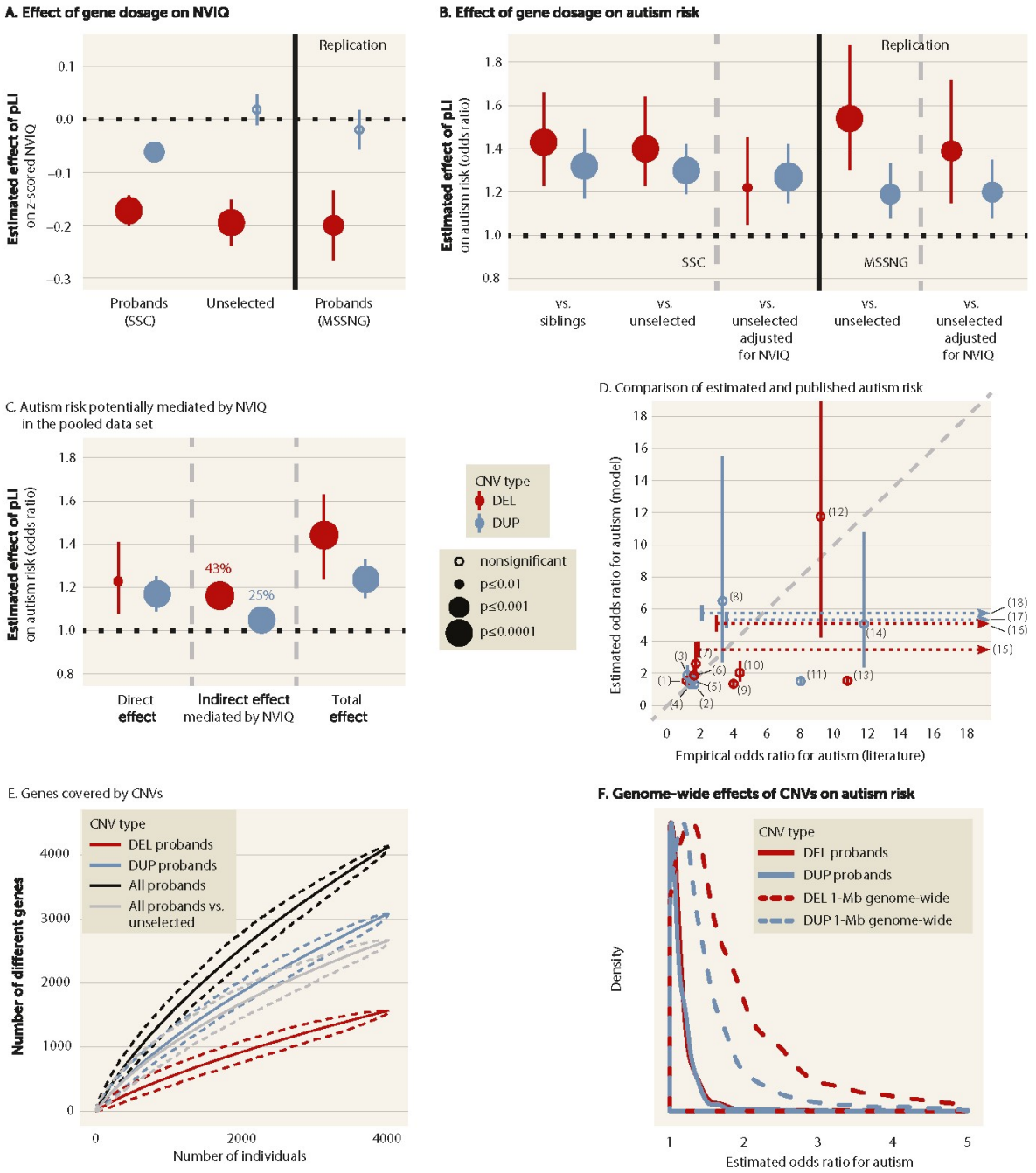
Deleting 1 point of pLI has the same effect-size on z-scored NVIQ in autism probands of both samples (SSC: β=-0.17, SE=0.03, $p=8\times10^{-10}$; MSSNG: β=-0.20, SE=0.07, $p=3\times10^{-3}$) and unselected populations (β=-0.19, SE=0.04, $p=7\times10^{-5}$). The pLI is also the score that best explains the impact of duplications on NVIQ, showing a three-fold smaller effect of pLI points on z-scored NVIQ in the SSC (β=-0.06, SE=0.02, $p=1\times10^{-3}$). No effect of duplications is detected in unselected populations or the MSSNG dataset (Table S11 in the online supplement, Figure 11A).

Matching the SSC and MSSNG based on NVIQ, or removing ratio NVIQ from the SSC, does not influence these effect-sizes (Figure S2 and Table S4 in the online supplement). In the pooled dataset, an autism diagnosis does not influence the effect of deleted or duplicated points of pLI on NVIQ. There is also no interaction with sex. Removing carriers of CNVs with a pLI sum > 10, with a known psychiatric association, or one occurring *de novo*, results in similar effect-sizes for deletions. For duplications, our limited power only allowed us to observe an effect when removing CNVs enriched in neurodevelopmental disorders (Table S4 in the online supplement).

Effect-size of gene dosage on autism risk

The variable selection procedure identified again the sum of pLI scores as the variable that best explains the diagnosis of autism for deletions (r^2=0.004) and duplications (r^2=0.004). Susceptibility to autism increases for each deleted point of pLI and the effect-size is identical when comparing autistic probands with their paired siblings or unselected populations (OR=1.43, 95%CI=1.23-1.66, $p=4\times10^{-6}$; OR=1.40, 95%CI=1.23-1.64, $p=2\times10^{-6}$, respectively). A duplicated point of pLI also increases autism susceptibility (comparing with siblings: OR=1.32, 95%CI=1.17-1.49, $p=5\times10^{-6}$; and the unselected populations: OR=1.30, 95%CI=1.19-1.42, $p=2\times10^{-8}$) (Figure 11B, Table S12 in the online supplement). Of note, there is no difference in pLI burden between intra- and extra-familial controls (unselected populations) (Table S5 in the online supplement).

Figure 11: Effect of gene dosage on nonverbal IQ and autism susceptibility in a study of effect-sizes of deletions and duplications on autism risk across the genome[a]



[a] Panel A shows the effect-size of deleted (DEL) or duplicated (DUP) points of probability of being loss-of-function intolerant (pLI) on nonverbal IQ (NVIQ) in autistic probands and unselected populations. Y-axis values are z-scores for NVIQ (e.g., 0.2 z-score=3 points of NVIQ). Panel B shows autism risk conferred by a deleted or duplicated point of pLI. Y-axis values are odds ratios computed using a logistic regression to explain an autism diagnosis. Control subjects include unaffected siblings or unselected populations. Replication was performed using autistic probands from the MSSNG data set and the unselected populations. Panel C shows autism risk potentially mediated by NVIQ on the pooled data set. Y-axis values are odds ratios computed using a counterfactual-based mediation analysis on a logistic regression. Direct effects of copy number variants (CNVs) are those not mediated by NVIQ. Indirect effects are those potentially mediated by NVIQ. Percentage of effect mediated by NVIQ=(indirect effect/total effect)x100. Total effects are those computed without adjusting for NVIQ. In panel D, we compared the risk of autism estimated by our model and the risk observed in previous published studies on 16 recurrent CNVs (Malhotra et al. 2012; Moreno de Luca et al. 2013; Chaste et al. 2014; Sanders et al. 2019) (see Table S3 in the supplementary material 1). dds ratio estimates from the model overlap with the 95% confidence intervals of odds ratios from previous publications for 14 recurrent CNVs. For three CNVs, the horizontal dotted arrows represent the extreme variability for odds ratios reported in previous publications. Values are detailed in Table S3 in the supplementary material 1. 1=DEL 15q11.2; 2=DUP 15q11.2; 3=DUP 16p11.2 distal; 4=DUP 15q13.3; 5=DUP 16p13.11; 6=DEL 1q21.1; 7=DEL 16p11.2 distal; 8=DUP 22q11.2; 9=DEL 17p12; 10=DEL 16p13.11; 11=DUP 1q21.1; 12=DEL 16p11.2; 13=DEL 15q13.3; 14=DUP 16p11.2; 15=DEL 17q12; 16=DEL 3q29; 17=DUP 7q11.23; 18=DUP 17p11.2. Panel E shows genes covered by deletions, duplications, or both in autistic probands. The gray line represents the excess of genes encompassed in CNVs in autistic probands relative to the unselected populations. The y-axis represents the number of distinct genes encompassed in CNVs (each gene is only counted once). The x-axis represents the number of individuals. The mean and 95% confidence interval were obtained using 1,000 iterations (bootstrap procedure). Panel F shows the distribution of the estimated effects of deletions and duplications. The solid lines indicate the estimated effects computed for all CNVs ≥ 50 kb identified in both autism cohorts. The dotted lines indicate autism risk computed for any CNV of 1Mb across the genome including at least one gene with a pLI annotation.

The risk conferred by deletions measured by pLI decreases substantially but remains borderline significant when the model is adjusted for NVIQ (OR=1.22, 95%CI=1.05-1.45, p=0.01) or when both autism and unselected populations are matched for NVIQ. In contrast, the autism risk conferred by each duplicated point of pLI remains unchanged when adjusting (OR=1.27, 95%CI=1.15-1.42, $p=5\times10^{-6}$) or matching for NVIQ (Figure 11B and Table S12 in the online supplement).

The replication analysis with the MSSNG dataset shows the same effect of deleted or duplicated points of pLI on autism susceptibility. We also replicate the differential effect of NVIQ adjustment on autism risk conferred by deletions and duplications (Figure 11B and Table S12 in the online supplement).

In the pooled dataset mediation analysis suggested that 43% and 25% of the autism risk conferred by deletions and duplications are potentially influenced by NVIQ (Figure 11C and Table S13). However, the effect-size of autism risk for deletions and duplications measured by pLI is the same in both subgroups of individuals above and below median NVIQ (Figure S3 and Table S14). There is no interaction with sex. Autism susceptibility related to gene dosage is unaffected by removing carriers of CNVs with a pLI sum > 10, CNVs with a known association to neurodevelopmental disorder, occurring *de novo,* or individuals from the unselected populations with a suspected diagnosis of autism (n=10) as well as no diagnostic information from the Development and Well-Being Assessment (DAWBA) (N=124) (Table S5 in the online supplement).

Estimating and validating the level of autism risk

ORs have previously been computed for a few recurrent CNVs with broad confidence intervals. The autism risk estimated by our model overlaps with that previously published for 16 recurrent CNVs, except for the 15q13.3 BP4-BP5 deletion and the 1q21.2 duplication, which are discordant (Malhotra and Sebat 2012; Moreno-De-Luca et al. 2013; Chaste et al. 2014; Sanders et al. 2019) (Figure 11D, Table S6 in the online supplement). The results are similar whether we include or exclude the 16 CNVs from the training dataset (Figure S3 in the online supplement). Our model is trained on

deletions and duplications covering over 4,500 different genes in the autism and unselected populations (Figure 11E). The sharply ascending slope of genes encompassed in the CNVs shows no asymptotic effects. Model estimates show that any 1Mb coding deletion or duplication across the genome should increase autism susceptibility, with a median OR of 1.6 and 1.3, respectively (Figure 11F and Table S15 in the online supplement).

Effect-size of gene dosage on measures of core symptoms and specifiers of autism

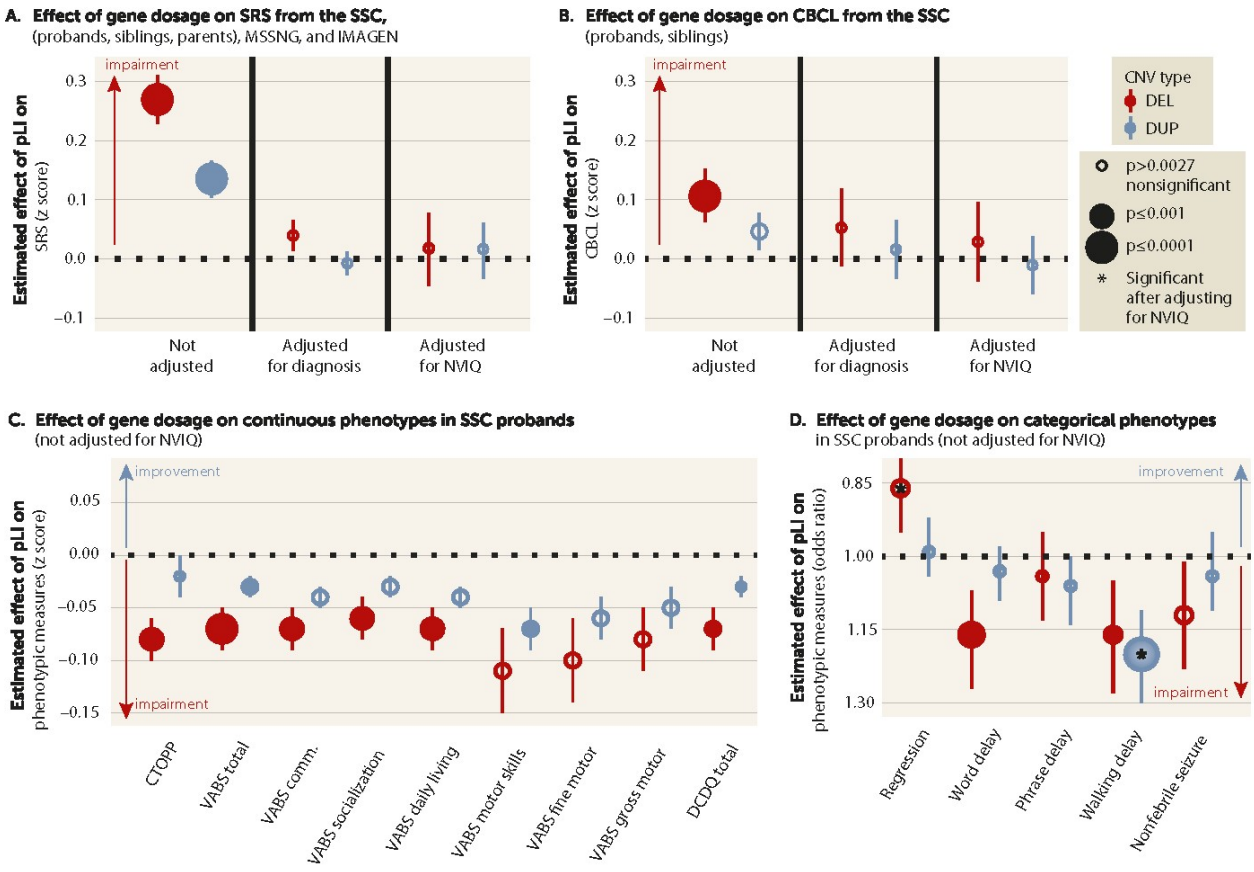
We assessed the cognitive and behavioural symptoms that underlie autism susceptibility conferred by gene dosage.

Autism related symptoms

The pLI increases the SRS, with a 2:1 effect-size ratio for deletions and duplications in the pooled SSC and IMAGEN dataset (deletions: β=3.72 points of raw SRS score per point of pLI,, SE=0.57, $p=5\times10^{-11}$; duplications: β=1.87 points of raw SRS score per point of pLI, SE=0.43, $p=1\times10^{-5}$). The effect-size of pLI on SRS remains the same after adding data from MSSNG (deletions: β=3.68, SE=0.56, $p=4\times10^{-11}$; duplications: β=1.63, SE=0.42, $p=1\times10^{-4}$). This effect of gene dosage is entirely explained by NVIQ and the autism diagnosis (Figure 12A, Figure S5, Table S8 in the online supplement).

Deletions and duplications measured by pLI do not affect the ADOS or ADI-R scores in probands of the SSC and MSSNG datasets, pooled or separately (Table S9 in the online supplement). Moreover, deletions measured by pLI protect against regression in autism and this effect is enhanced after adjusting for NVIQ (OR=0.80, 95%CI=0.70-0.89, $p=2\times10^{-4}$) (Table 5, Figure 12D, Figure S6B in the online supplement).

Figure 12: Effect of gene dosage on phenotypic measurements in a study of effect-sizes of deletions and duplications on autism risk across the genome[a]



[a] Panel A shows the effect-size of a deleted (DEL) or duplicated (DUP) point of probability of being loss-of-function intolerant (pLI) on the z-scored Social Responsiveness Scale (SRS) in autistic probands from the Simons Simplex Collection (SSC) pooled with their unaffected relatives (siblings and parents) and the unselected populations from the IMAGEN data set. Effects were measured with and without adjustment for the diagnosis of autism and for nonverbal IQ (NVIQ). The y-axis represents the estimated effect of pLI on the SRS z-score, computed using the mean and standard deviation of unaffected individuals (0.10 z-score=1.28 SRS raw score point). The analysis adjusting for NVIQ contains only probands from the SSC and individuals from IMAGEN. Panel B shows the effect-size of a deleted or duplicated point of pLI on the z-score of the Child Behavior Checklist (CBCL) in autistic probands from the SSC pooled with their unaffected siblings. Effects were measured with and without adjustment for the diagnosis of autism and for NVIQ. The estimates were originally computed as odds ratios using a negative binomial regression. The y-axis represents the estimated effect of pLI on z-scored CBCL, computed using the mean and standard deviation of unaffected individuals (0.10 z-score=1.52 CBCL raw score point). The analysis adjusting for NVIQ was performed only on probands from the SSC. Panels C and D show the effect-size of a deleted or duplicated point of pLI on continuous and categorical phenotypes in autistic probands from the SSC, unadjusted for NVIQ. Results adjusted for NVIQ are detailed in Figure S4 in the supplementary material 1. The y-axis values in panel C are measures z-scored using normative data (see Table S7 in the supplementary material 1) except for the Developmental Coordination Disorder Questionnaire (DCDQ), which was z-scored using the SSC autistic proband group. The y-axis values in panel D are odds ratios computed by logistic regression. To correct for the number of independent tests, the significance threshold ($p \leq 2.7 \times 10^{-3}$). was computed using the eigenvalues of the correlation coefficients between the measures investigated (see the Methods section in the supplementary material 1). CTOPP=Comprehensive Test of Phonological Processing; VABS-II=Vineland Adaptive Behavior Rating Scales, Second Edition.

Language and phonological memory

There is a clear dissociation between the effect of deletions and duplications on language. Deleted points of pLI are associated with a delay of first-words (OR=1.16, 95%CI=1.07-1.27, $p=5\times10^{-4}$) and negatively affects phonological memory, assessed by the non-word repetition of the Comprehensive Test of Phonological Processing (CTOPP) ($\beta=0.08$, SE=0.02, $p=6\times10^{-4}$). No effects are observed for deletions after adjusting for NVIQ and for duplications with or without adjusting for NVIQ (Table 5, Figure 12C and 12D, Figure S6A and 6B in the online supplement).

Behavioural and emotional symptoms

In the sample pooling probands and unaffected siblings, haploinsufficiency measured by pLI impacts the score of total problems from the CBCL (OR=1.05, 95%CI=1.03-1.08, $p=2\times10^{-6}$). The effect of duplications is weaker (OR=1.02, 95%CI=1.01-1.04, $p=3\times10^{-3}$) (Table S10, Figure 12B). This translates into an increase of 20.63 [95%CI=19.55-21.73] and 7.85 [95%CI=7.28-8.44] points for a deletion or a duplication encompassing 10 points of pLI, respectively. These effects are not observed within SSC probands or unaffected siblings samples.

Adaptive Skills

Adaptive skills measured by the second edition of the Vineland Adaptive behaviour Rating Scales (VABS-II) are negatively affected by the pLI, with a decrease of 2 and 1 point of VABS per deleted or duplicated point of pLI, respectively ($p=3\times10^{-5}$ and $p=3\times10^{-3}$). Total scores and all subscales are equally affected. NVIQ appears to account for most, if not all, of this effect (Table 5, Figure 12C, Figure S6A in the online supplement).

Motor skills and epilepsy

The relationship between the onset of walking measured in months and pLI (deletion: OR=1.03, 95%CI=1.02-1.04, $p=2\times10^{-11}$; duplication: OR=1.02, 95%CI=1.01-1.03, $p=7\times10^{-9}$) translates into a 5.46 [95%CI=5.27-5.65] or 3.58 [95%CI=3.45-3.72] month delay for a deletion or duplication encompassing 10 points of pLI, respectively (Figure

S7 in the online supplement). This remains significant after adjusting for NVIQ for duplications only. The effect-size of gene dosage on motor skills, measured by the VABS-II and the Developmental Coordination Disorder Questionnaire (DCDQ), shows a 2:1 ratio for deletions and duplications with a similar effect for gross and fine motor skills. Gene dosage does not affect the risk of non-febrile seizures (Table 5, Figure 12C and 12D, Figure S6A and 6B in the online supplement).

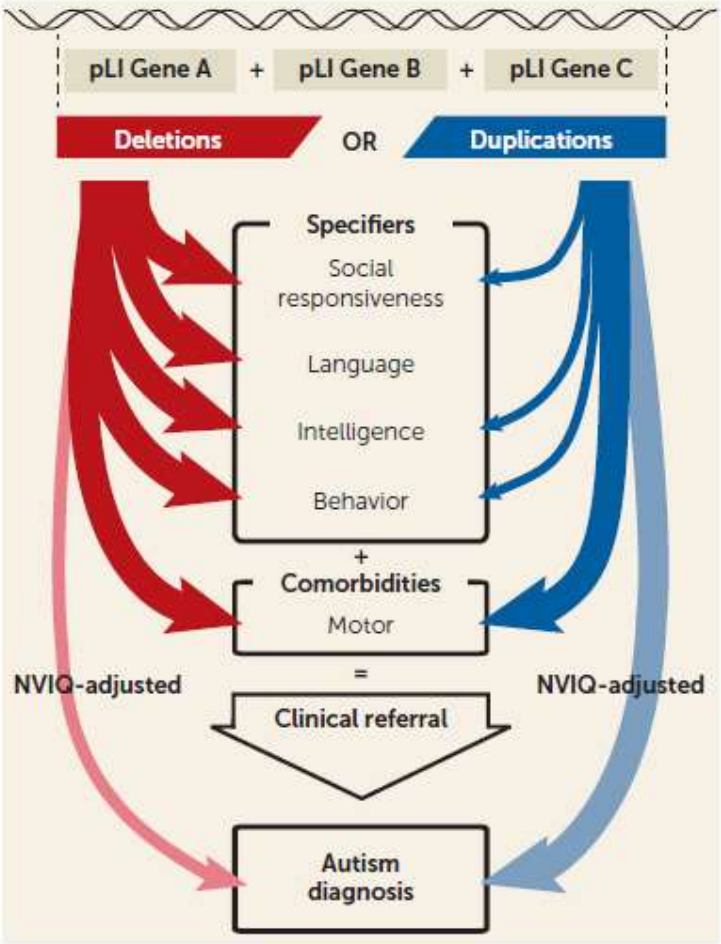
Potential applications in the clinic

We developed a prediction tool available online (https://cnvprediction.urca.ca/) to estimate the effect-size of deletions and duplications on NVIQ, autism risk and the SRS score. As an illustration, our model estimates a decrease in NVIQ of 26.78 [95%CI=26.19-27.37] and 30.89 [95%CI=30.30-31.48] points, an increase in the SRS raw score of 36.93 [95%CI=35.82-38.04] and 42.59 [95%CI=41.48-43.70] points, and an increase in autism risk of 21.05 [95%CI=6.10-72.26] and 33.58 [95%CI=8.05-139.99] for the 16p11.2 and 22q11.2 deletion respectively. We detail the model output for 21 recurrent CNVs in Table S6. Briefly summarized, this tool should be viewed as a translation of gnomAD (Karczewski et al. 2019) information into phenotypic effect-sizes.

DISCUSSION

We propose a model to estimate the effect-size of gene dosage on autism susceptibility, core autism symptoms, general intelligence, and autism specifiers. Haploinsufficiency measured by pLI increases autism susceptibility across the genome but NVIQ drives a large proportion of this effect. Language, motor, social communication, and behavioural problems are also strongly affected by deletions. While these manifestations may increase the probability for deletion carriers of receiving an autism diagnosis, there is no evidence that core symptoms are affected (Figure 13). In contrast, duplicated points of pLI increase autism risk, genome-wide, and the influence of NVIQ is smaller. Increased risk measured by pLI is similar in subgroups of individuals with NVIQ below and above median.

Figure 13: Interpretation of the effect of gene dosage on autism risk[a]



[a] The figure presents a summary interpretation of the differential effects of deletions and duplications on autism risk. Effect-sizes of deletion on most autism specifiers are larger than those of duplications. Duplications and, to a lesser extent, deletions increase the probability of an autism diagnosis after adjusting for their effect on nonverbal IQ (transparent arrows). NVIQ=nonverbal IQ; pLI=probability of being loss-of-function intolerant.

Differential effects of deletions and duplications on autism core symptoms and specifiers

Model estimates show that any 1Mb coding deletion or duplication across the genome should increase autism susceptibility, with a median OR of 1.6 and 1.3, respectively (Figure 11F). GWAS conducted on common variants also showed that the bulk of the heritability for complex conditions (*i.e.* schizophrenia) is spread across the genome and largely driven by genes with no clear relevance to disease. (Boyle, Li, and Pritchard 2017; Wray et al. 2018) Gene dosage affects NVIQ, social communication, and adaptive behaviour, with a deletion:duplication effect-size ratio of 2-3:1. Although both CNVs

equally affect motor skills, phonological memory may be predominantly affected by haploinsufficiency. Similar differential profiles have been reported for 16p11.2 CNVs with phonological memory deficits in deletion but not duplication carriers. (Hippolyte et al. 2016) We posit that general phenotypic profiles may be associated with deletions and duplications irrespective of the genomic loci. Genes included in the CNVs may mostly influence the effect-size but not the profile of symptoms. Consistent with this interpretation, the phenotypic profile of haploinsufficiency delineated by our model has been similarly reported in patients with *de novo* loss of function variants (Bishop et al. 2017; Buja et al. 2018). In addition, excluding large effect-size *de novo* variants from our analyses does not modify the effect-size of gene dosage, measured by pLI, on NVIQ and autism risk. Therefore, molecular functional networks enriched in genes with an excess of *de novo* mutations (chromatin remodelling, synaptic function) (Huguet, Ey, and Bourgeron 2013; Pinto et al. 2014; Krumm et al. 2015) may be related to large effect-sizes rather than specific effects on autism risk. Interestingly, although previous studies have shown lower NVIQ and a higher burden of deleterious CNVs in females from the SSC (Jacquemont et al. 2014), we did not identify any interaction between the effect of pLI and sex. This suggests that deleting or duplicating one point of pLI affects NVIQ and increases autism risk similarly in both sexes.

Potential clinical applications

Our models are implemented in a prediction tool (https://cnvprediction.urca.ca/), which is designed to predict the effect-size of CNVs, not the symptoms of the individual who carries the CNV. If symptoms are discordant, the clinician may conclude that additional factors should be investigated. Discordance may be defined when the estimated effect-size of the CNV is 1 SD (15 IQ points) lower than the IQ loss observed in the carrier (compared to the population mean = 100). If a CNV with an effect-size of -10 IQ points is identified in a carrier with mild intellectual disabilities and an IQ of 60 (-40 compared to population mean) the majority of the cognitive deficits are caused by additional factors. The estimates of autism risk provided by models in this study overlap with risk

computed in previous studies. As an example, our model estimates for 16p11.2 and 22q11.2 deletions are similar to the previously published effect for NVIQ, (loss of 25 (Moreno-De-Luca et al. 2015) and 29 (Vangkilde et al. 2016) points), autism risk (OR of 11.8 (Malhotra and Sebat 2012) and 32.37 (Sanders et al. 2019)) and SRS (gain of 44 (Moreno-De-Luca et al. 2015) and 49 (Vangkilde et al. 2016) points). Overall, the output of these models can help interpret CNVs in the clinic, but estimates should be interpreted with caution.

Limitations

Discordance between autism risk estimated by the model and literature observations allows for the identification of CNVs, which may encompass genes with specific properties. For example, autism susceptibility and deficits associated with the 15q13.3 (*CHRNA7*) deletion appear to be underestimated by our model. 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). The pLI was not developed to measure intolerance to duplications and results should, therefore, be interpreted with caution. Our findings suggest, however that pLI may be a general measure of dosage sensitivity, in line with recent data from gnomAD-SV. (Collins et al. 2019) Since gene dosage is not comparable between sex-linked and autosomal CNVs, we could not pool both types of CNVs. Sex-linked CNVs were excluded from this study because they were too rare in our samples to be studied separately. The effect of gene dosage on SRS was very robust but was mainly explained by the autism diagnosis. This suggests that the SRS may not measure a continuous dimension since this score is unable to provide additional granularity within the autism group or the controls despite large sample size. Some phenotypic measures such as phonological memory and motor skills were only available for autism probands and results may not be generalizable to non-autism samples. Larger samples, with additional intrafamilial controls, novel functional annotations, and more refined

models are required to improve our estimates of CNV effect-sizes on cognitive dimensions.

Of note, although CNV with large effect-sizes have significant impacts on the development of an individual, they only explain a small fraction of the variance of general intelligence (1.4% and 0.4% for deletions and duplications) and liability for autism (0.4 and 0.4% for deletions and duplications) at the population level, which is concordant with previous reports. (Gaugler et al. 2014)

Conclusion

Our study highlights the extreme polygenicity of autism susceptibility conferred by gene dosage. It also delineates cognitive mechanisms which may explain in part the overrepresentation of CNVs in autism. Among mutations over-represented in autism, those truly related to core symptoms may be less common than previously thought. Future large-scale studies simultaneously investigating the effect of genomic variants on categorical diagnoses and continuous dimensions are warranted. This study represents a new framework to study rare variants and can help in the interpretation of the effect-size of undocumented CNVs identified in the neurodevelopmental clinic.

REFERENCES

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

Bernier, Raphael, Kyle J. Steinman, Beau Reilly, et al. 2016 Clinical Phenotype of the Recurrent 1q21.1 Copy-Number Variant. Genetics in Medicine: Official Journal of the American College of Medical Genetics 18(4): 341–349.

Bishop, Somer L., Cristan Farmer, Vanessa Bal, et al. 2017 Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. The American Journal of Psychiatry 174(6): 576–585.

Boyle, Evan A., Yang I. Li, and Jonathan K. Pritchard 2017 An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169(7): 1177–1186.

Buja, Andreas, Natalia Volfovsky, Abba M. Krieger, et al.
2018 Damaging de Novo Mutations Diminish Motor Skills in Children on the Autism Spectrum. Proceedings of the National Academy of Sciences of the United States of America 115(8): E1859–E1866.

Butcher, Nancy J., Eva W. C. Chow, Gregory Costain, et al.
2012 Functional Outcomes of Adults with 22q11.2 Deletion Syndrome. Genetics in Medicine 14(10): 836–843.

Chaste, Pauline, Stephan J. Sanders, Kommu N. Mohan, et al.
2014 Modest Impact on Risk for Autism Spectrum Disorder of Rare Copy Number Variants at 15q11.2, Specifically Breakpoints 1 to 2. Autism Research : Official Journal of the International Society for Autism Research 7(3): 355–362.

Collins, Ryan L., Harrison Brand, Konrad J. Karczewski, et al.
2019 An Open Resource of Structural Variation for Medical and Population Genetics. BioRxiv: 578674.

Cooper, Gregory M., Bradley P. Coe, Santhosh Girirajan, et al.
2011 A Copy Number Variation Morbidity Map of Developmental Delay. Nature Genetics 43(9): 838–846.

D'Angelo, Debra, Sébastien Lebon, Qixuan Chen, et al.
2016 Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA Psychiatry 73(1): 20–30.

Fischbach, Gerald D., and Catherine Lord
2010 The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. Neuron 68(2): 192–195.

Gaugler, Trent, Lambertus Klei, Stephan J. Sanders, et al.
2014 Most Genetic Risk for Autism Resides with Common Variation. Nature Genetics 46(8): 881–885.

Girirajan, Santhosh, Zoran Brkanac, Bradley P. Coe, et al.
2011 Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. PLoS Genetics 7(11): e1002334.

Girirajan, Santhosh, Megan Y. Dennis, Carl Baker, et al.
2013 Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder. The American Journal of Human Genetics 92(2): 221–237.

Hawrylycz, Michael, Jeremy A. Miller, Vilas Menon, et al.
2015 Canonical Genetic Signatures of the Adult Human Brain. Nature Neuroscience 18(12): 1832–1844.

Hippolyte, Loyse, Anne M. Maillard, Borja Rodriguez-Herreros, et al.
2016 The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. Biological Psychiatry 80(2): 129–139.

Huguet, Guillaume, Elodie Ey, and Thomas Bourgeron
2013 The Genetic Landscapes of Autism Spectrum Disorders. Annual Review of Genomics and Human Genetics 14: 191–213.

Huguet, Guillaume, Catherine Schramm, Elise Douard, et al.
2018 Measuring and Estimating the Effect-sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry.

International HapMap Consortium
2003 The International HapMap Project. Nature 426(6968): 789–796.

Iossifov, Ivan, Brian J. O'Roak, Stephan J. Sanders, et al.
2014 The Contribution of *de Novo* Coding Mutations to Autism Spectrum Disorder. Nature 515(7526): 216–221.

Jacquemont, Sébastien, Bradley P. Coe, Micha Hersch, et al.
2014 A Higher Mutational Burden in Females Supports a "Female Protective Model" in Neurodevelopmental Disorders. The American Journal of Human Genetics 94(3): 415–425.

Jiang, Yong-hui, Ryan K. C. Yuen, Xin Jin, et al.
2013 Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing. The American Journal of Human Genetics 93(2): 249–263.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, et al.
2019 Variation across 141,456 Human Exomes and Genomes Reveals the Spectrum of Loss-of-Function Intolerance across Human Protein-Coding Genes. BioRxiv: 531210.

Krumm, Niklas, Tychele N. Turner, Carl Baker, et al.
2015 Excess of Rare, Inherited Truncating Mutations in Autism. Nature Genetics 47(6): 582–588.

Lek, Monkol, Konrad J. Karczewski, Eric V. Minikel, et al.
2016 Analysis of Protein-Coding Genetic Variation in 60,706 Humans. Nature 536(7616): 285–291.

Malhotra, Dheeraj, and Jonathan Sebat
2012 CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell 148(6): 1223–1241.

Marshall, Christian R., Daniel P. Howrigan, Daniele Merico, et al.
2017 Contribution of Copy Number Variants to Schizophrenia from a Genome-Wide Study of 41,321 Subjects. Nature Genetics 49(1): 27–35.

Moreno-De-Luca, Andres, David W. Evans, K. B. Boomer, et al.
2015 The Role of Parental Cognitive, Behavioral, and Motor Profiles in Clinical Variability in Individuals with Chromosome 16p11.2 Deletions. JAMA Psychiatry 72(2): 119–126.

Moreno-De-Luca, D., S. J. Sanders, A. J. Willsey, et al.
2013 Using Large Clinical Data Sets to Infer Pathogenicity for Rare Copy Number Variants in Autism Cohorts. Molecular Psychiatry 18(10): 1090–1095.

Mottron, Laurent, Pauline Duret, Sophia Mueller, et al.
2015 Sex Differences in Brain Plasticity: A New Hypothesis for Sex Ratio Bias in Autism. Molecular Autism 6(1): 33.

Ousley, Opal, and Tracy Cermak
2014 Autism Spectrum Disorder: Defining Dimensions and Subgroups. Current Developmental Disorders Reports 1(1): 20–28.

Pausova, Zdenka, Tomas Paus, Michal Abrahamowicz, et al.
2017 Cohort Profile: The Saguenay Youth Study (SYS). International Journal of Epidemiology 46(2): e19.

Petrovski, Slavé, Ayal B. Gussow, Quanli Wang, et al.
2015 The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. PLoS Genetics 11(9): e1005492.

Pinto, Dalila, Elsa Delaby, Daniele Merico, et al.
2014 Convergence of Genes and Cellular Pathways Dysregulated in Autism Spectrum Disorders. American Journal of Human Genetics 94(5): 677–694.

Ramasamy, Adaikalavan, Daniah Trabzuni, Sebastian Guelfi, et al.
2014 Genetic Variability in the Regulation of Gene Expression in Ten Regions of the Human Brain. Nature Neuroscience 17(10): 1418–1428.

Ruderfer, Douglas M., Tymor Hamamsy, Monkol Lek, et al.
2016 Patterns of Genic Intolerance of Rare Copy Number Variation in 59,898 Human Exomes. Nature Genetics 48(10): 1107–1111.

Sanders, Stephan J., Xin He, A. Jeremy Willsey, et al.
2015 Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87(6): 1215–1233.

Sanders, Stephan J., Mustafa Sahin, Joseph Hostyk, et al.
2019 A Framework for the Investigation of Rare Genetic Disorders in Neuropsychiatry. Nature Medicine.

Sandin, Sven, Paul Lichtenstein, Ralf Kuja-Halkola, et al.
2017 The Heritability of Autism Spectrum Disorder. JAMA 318(12): 1182–1184.

Satterstrom, F. Kyle, Jack A. Kosmicki, Jiebiao Wang, et al.
2018 Novel Genes for Autism Implicate Both Excitatory and Inhibitory Cell Lineages in Risk. BioRxiv: 484113.

Schumann, G., E. Loth, T. Banaschewski, et al.
2010 The IMAGEN Study: Reinforcement-Related Behaviour in Normal Brain Function and Psychopathology. Molecular Psychiatry 15(12): 1128–1139.

Simonoff, Emily, Andrew Pickles, Tony Charman, et al.
2008 Psychiatric Disorders in Children with Autism Spectrum Disorders: Prevalence, Comorbidity, and Associated Factors in a Population-Derived Sample. Journal of the American Academy of Child and Adolescent Psychiatry 47(8): 921–929.

Szklarczyk, Damian, Andrea Franceschini, Stefan Wyder, et al.
2015 STRING V10: Protein–Protein Interaction Networks, Integrated over the Tree of Life. Nucleic Acids Research 43(D1): D447–D452.

Tammimies, Kristiina, Christian R. Marshall, Susan Walker, et al.
2015 Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. JAMA 314(9): 895–903.

Vangkilde, A., J. R. M. Jepsen, H. Schmock, et al.
2016 Associations between Social Cognition, Skills, and Function and Subclinical Negative and Positive Symptoms in 22q11.2 Deletion Syndrome. Journal of Neurodevelopmental Disorders 8(1): 42.

Wain, Hester M., Elspeth A. Bruford, Ruth C. Lovering, et al.
2002 Guidelines for Human Gene Nomenclature. Genomics 79(4): 464–470.

Weiner, Daniel J., Emilie M. Wigdor, Stephan Ripke, et al. 2017 Polygenic Transmission Disequilibrium Confirms That Common and Rare Variation Act Additively to Create Risk for Autism Spectrum Disorders. Nature Genetics 49(7): 978–985.

Wray, Naomi R., Cisca Wijmenga, Patrick F. Sullivan, Jian Yang, and Peter M. Visscher 2018 Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell 173(7): 1573–1580.

Yuen, Ryan K. C., Daniele Merico, Matt Bookman, et al. 2017 Whole Genome Sequencing Resource Identifies 18 New Candidate Genes for Autism Spectrum Disorder. Nature Neuroscience 20(4): 602–611.

III. Paper #2: Rate of Deleterious Copy Number Variants Similar in Early Onset Psychosis and Autism Spectrum Disorders: Implications for Clinical Practice

Substantial contributions of the candidate:

1) CNV filtering and annotation of all the samples (CNVs detection of the EOP sample was done by the cytogenetic laboratory of Boston Hospital, and CNVs detection of the control and ASD samples was done by Z. Saci).
2) Conception and design of all the statistical analyses.
3) Interpretation of the statistical analyses with S. Jacquemont.
4) Creation of all figures, Table 7 and all the supplementary materials.
5) Main contributor when drafting the paper and answering to the AJP reviewers (both with C. Brownstein, D. A. Glahn and S. Jacquemont).

Authors

Catherine A. Brownstein, PhD[1,2,3,4]**, Elise Douard, M Sc[5,6]**, Josephine Mollon, PhD[1,7,8], Richard Smith, PhD[1,2,3,4], Margaret Hojlo, BA[1,7], Ananth Das[1,7], Maria Goldman[7], Emily Garvey, BA[7], Kristin Cabral, BS[2,3], Jianqiao Lil, BS[2,3], Joshua Bowen, BS[2,3], Abhijit S. Rao, BSA [2], Casie Genetti, MS CGC[2,3,4], Devon Carroll, APRN-BC[7,8], Emma E. M. Knowles, PhD[1,7,8], Emma Deaso, BA[1,7], Pankaj B Agrawal, MD[2,3,4,12], Alan H. Beggs, PhD[2,3,4], Eugene D'Angelo, PhD[1,7,8], Laura Almasy, PhD[10], Aaron Alexander-Bloch, MD, PhD[11], Zohra Saci, PhD[6], Clara A. Moreau, PhD[5,6], Guillaume Huguet, PhD[5,6], Anthony Deo, MD PhD[1,13,14], Sébastien Jacquemont, MD[5,6]†, David C. Glahn, PhD[1,2,3,7,8,9]†, Joseph Gonzalez-Heydrich, MD[1,7,8,9]†

**Shared first authorship, † Shared last authorship

1. Early Psychosis Investigation Center (EPICenter), Boston Children's Hospital, Boston, MA, USA
2. Division of Genetics and Genomics, Boston Children's Hospital, Boston, MA, USA
3. The Manton Center for Orphan Disease Research, Boston Children's Hospital, Boston, MA, USA
4. Department of Pediatrics, Harvard Medical School, Boston, MA, USA
5. Department of Medicine, Université de Montréal, Montreal, Quebec, Canada
6. Department of Pediatrics, UHC Sainte-Justine Research Center, Montreal, Quebec, Canada
7. Department of Psychiatry and Behavioral Sciences, Boston Children's Hospital, Boston, MA, USA
8. Department of Psychiatry, Harvard Medical School, Boston, MA, USA
9. Tommy Fuss Center for Neuropsychiatric Disease Research, Boston Children's Hospital, Boston, MA, USA
10. Department of Biomedical and Health Informatics (DBHi), Children's Hospital of Philadelphia, PA, USA
11. Department of Psychiatry, Children's Hospital of Philadelphia, PA, USA
12. Division of Newborn Medicine, Boston Children's Hospital, Boston, MA, USA
13. Department of Psychiatry, Rutgers-Robert Wood Johnson Medical School, Piscataway, NJ
14. Rutgers University Behavioral Health Care, Piscataway, NJ

Under revision in *JAMA Psychiatry*

Acknowledgments

We gratefully acknowledge the patients and families who participated in this study.

Funding/Support

This work was funded by the Tommy Fuss Center for Neuropsychiatric Research, the Boston Children's Hospital Intellectual and Developmental Disabilities Research Center Molecular Genetics Core Facility supported by P50HD105351 from the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development, and NIH/NIMH U01 MH119690, The Stanley Center of the Broad Institute of MIT and Harvard, the Jonathan and Robin Klein and Anne and Paul Marcus families.

Abstract

Background: Copy number variants (CNVs) are strongly associated with neurodevelopmental and psychotic disorders. Early onset psychosis (EOP), where symptoms appear before 18 years of age, is thought to be more strongly influenced by genetic factors than adult-onset psychotic disorders. However, the prevalence and pathogenicity of CNVs in EOP is unclear.

Methods. We documented the prevalence of recurrent CNVs and the pathogenicity of deletions and duplications genome-wide in 137 children and adolescents with EOP compared to 5,540 individuals with autism spectrum disorders (ASD) and 16,504 population controls. Specifically, we first compared the frequency of regions previously associated with neurodevelopmental and neuropsychiatric illnesses. Second, the pathogenicity of an individual's total CNV burden was compared using an aggregate index of the haploinsufficiency of each gene in every CNV across the genome.

Outcomes. Prevalence of recurrent CNVs was higher in EOP than in ASD (OR=2.30, p=0.02) and controls (OR=5.06, $p=3x10^{-5}$). However, the difference between the EOP and ASD attenuated when EOP participants with co-occurring ASD were excluded. CNV pathogenicity was higher in EOP compared to controls for both deletions (OR=1.30, $p=9x10^{-8}$) and duplications (OR=1.09, p=0.02), but not compared to ASD.

Interpretation. Given the similar CNV burdens in EOP and ASD, our findings suggest that all children and adolescents with a psychotic diagnosis should undergo genetic screening, as is recommended in ASD.

Funding. Our work was supported by the Fuss Center for Neuropsychiatric Research, the Boston Children's Hospital Intellectual and Developmental Disabilities Research Center Molecular Genetics Core Facility (P50HD105351) and NIH/NIMH U01MH119690.

Introduction

Text box:

Early-onset psychosis **(EOP),** a broad range of psychiatric illnesses with prominent psychotic symptoms, including schizophrenia, bipolar disorder with psychosis, affective and other non-affective psychotic disorders with symptom onset prior to 18 years of age.

Very early-onset psychosis **(VEOP)**, defined as symptom onset before age 13 years.

Recurrent CNVs, genomic loci flanked by low copy repeat (LCR) sequences that greatly increase the risk of homologous recombination. These non-allelic homologous recombinations result in similar or identical CNVs in unrelated individuals.

Non-recurrent CNVs, structural variants with non-recurrent end-point. The non-recurrent junctions do not coincide with LCRs, but tend to occur in the vicinity of regions that are rich in LCRs resulting in complex regional genomic architecture.

LOEUF score, measures a gene's intolerance to variation by comparing the observed and expected number of LoF variation for a given gene in a reference population. Low LOEUF indicate strong selection against inactivation, and high LOEUF indicate higher tolerance.

Pathogenicity score, Sum of 1/LOEUF scores of genes encompassed in CNVs identified in an individual. Higher pathogenicity score suggests alteration of multiple coding genes which score more or less against inactivation.

Rare copy number variants (CNVs) are deletions and duplications of genomic segments[1] with high relative risk for psychotic disorders like schizophrenia[2]. Individuals with childhood-onset schizophrenia whose symptoms begin prior to age 13 have a nearly 3-fold increase of recurrent CNVs relative to those with adult-onset schizophrenia, suggesting a greater genetic component in the childhood form of the disorder[3]. However, childhood-onset schizophrenia is a rare and understudied illness[4] and only a single, small cohort has been genetically characterized to date[3]. Moreover, only around half of children and adolescents with a psychiatric diagnosis that includes prominent psychotic features meet strict criteria for schizophrenia[5] and childhood diagnoses often change over the course of development[6]. Consequently, there is considerable interest in understanding the genetic underpinnings of the more inclusive early onset psychosis

(EOP) categorization, which captures psychotic symptomatology across the spectrum of diagnostic criteria. EOP, defined as any psychiatric diagnosis with pronounced psychotic symptoms with onset prior to 18 years of age, is associated with lower premorbid psychosocial function, more hospitalizations, poorer cognitive functioning, and worse overall prognosis than adult-onset illness[5,7]. Functional outcomes are highly variable in EOP youth[5,6] and CNV status has been shown to influence these outcomes[8]. Although genomic information could help to disentangle the clinical heterogeneity in EOP, the genetic architecture of early onset psychosis is largely unknown.

Establishing the burden of recurrent CNVs in EOP is an important first step in characterizing the genetic architecture of this extreme phenotype. However, approximately 90% of CNVs identified in the clinic are non-recurrent and are therefore too rare (*i.e.*, insufficient copies) for association studies of individual CNVs to be practical[9]. Recently, we developed a strategy to estimate an individual's genome-wide CNV burden by deriving a single aggregate index of the pathogenicity of each gene encapsulated by every CNV across the genome, regardless of the population prevalence of the mutation[10-12]. To provide an estimate of the pathogenicity of each gene in a deleted or duplicated region, we use the loss-of-function observed/expected upper bound fraction (LOEUF) score[13]. The LOEUF is calculated by comparing the observed and expected number of loss-of-function (LoF) mutations for a given gene in a reference population[13]. Low LOEUF scores indicate strong selection against predicted LoF variation in a gene, while high LOEUF scores indicate relatively higher tolerance to inactivation. Thus, LOEUF scores provide a method for documenting the biological ramifications of individual mutations and inferring pathobiology[14]. Our aggregate pathogenicity score has been successfully used to model autism spectrum disorder (ASD)[12] and general intelligence[10,11]. Applying this method to an EOP sample will enable us to model genome-wide pathogenicity in EOP and directly compare our index with ASD and unselected control cohorts, even in the context where the individual pathogenic CNVs are extremely rare.

Findings that up to 20% of individuals with ASD carry a deleterious genetic mutation[15] have led organizations like the American Academy of Pediatrics to recommend that children presenting with ASD symptoms undergo genomic screening[16,17]. These established guidelines involve routine use of chromosomal microarrays (CMA) to document CNVs and aid diagnosis and treatment[18]. The burden of recurrent CNVs was similar for ASD and a small childhood onset schizophrenia cohort[3], suggesting that these disorders have comparable genetic architectures and should be subject to similar genetic screening approaches. Correspondingly, if the burden of recurrent CNVs is similar among children and adolescents with the broader EOP phenotype, this would strongly support the development of guidelines for genomic screening in this population. Such an approach could help aid diagnosis, therapeutic choices, and clinical staging of individuals with EOP, many of whom do not respond to first line treatments[19]. However, there are currently no genomic screening guidelines for children or adults with psychotic disorders[20].

In the present manuscript we aim (1) to establish the prevalence of recurrent CNVs in our diagnostically heterogeneous EOP cohort (n=137) and, by using bioinformatic tools, (2) to model the pathogenicity of recurrent and non-recurrent CNVs across the genome. Additionally, we will (3) compare the CNV burden and pathogenicity observed among EOP probands to those seen in 5,540 individuals with ASD and 16,504 unselected population controls.

Methods

EOP Samples

EOP participants (N=137) were referred to the Developmental Neuropsychiatry Program at Boston Children's Hospital (BCH). Prior to enrollment into the cohort, clinical diagnoses were ascertained by a board-certified child psychiatrist (JGH) specializing in EOP. Diagnoses were subsequently confirmed via medical record with a DSM-5 checklist

and a consensus diagnosis was reached (see Table 6 for diagnostic breakdown). Inclusion criteria were having a DSM-5 diagnosis for current or lifetime Axis I psychotic disorder with onset prior to age 18. Exclusion criteria were 1) substance- or medication-induced psychosis; 2) psychosis secondary to a brain infection (*e.g.*, encephalitis); 3) psychosis due to a neurodegenerative disorder such as Wilson's disease, Dystonia muscular deformations, Huntington's disease, Friedrich's ataxia, Ataxia Telangiectasia, or Parkinson's disorder; and 4) severe neurodevelopmental disorder or other impairment impacting ability to describe symptoms or provide other information required for this study. All EOP participants or guardians provided written informed consent (and participants <18 years provided assent) on forms approved by the BCH IRB as part of the Manton Center for Orphan Disease Research. After providing written informed consent/assent, each participant provided blood samples.

ASD and Unselected Populations

We compared the EOP participants to two pooled ASD cohorts: 2,585 children from the Simons Simplex Collection (SSC)[21] and 3,171 probands from the MSSNG database[22]. We also compared EOP participants to individuals from three pooled unselected community-based cohorts: IMAGEN (N=1,802)[23], Generation Scotland (GS)[24] (N=14,160), and the Lothian Birth Cohort25 (LBC) (N=554) (see Figure 14). Studies for each cohort were reviewed by local institutional review boards[21-25].

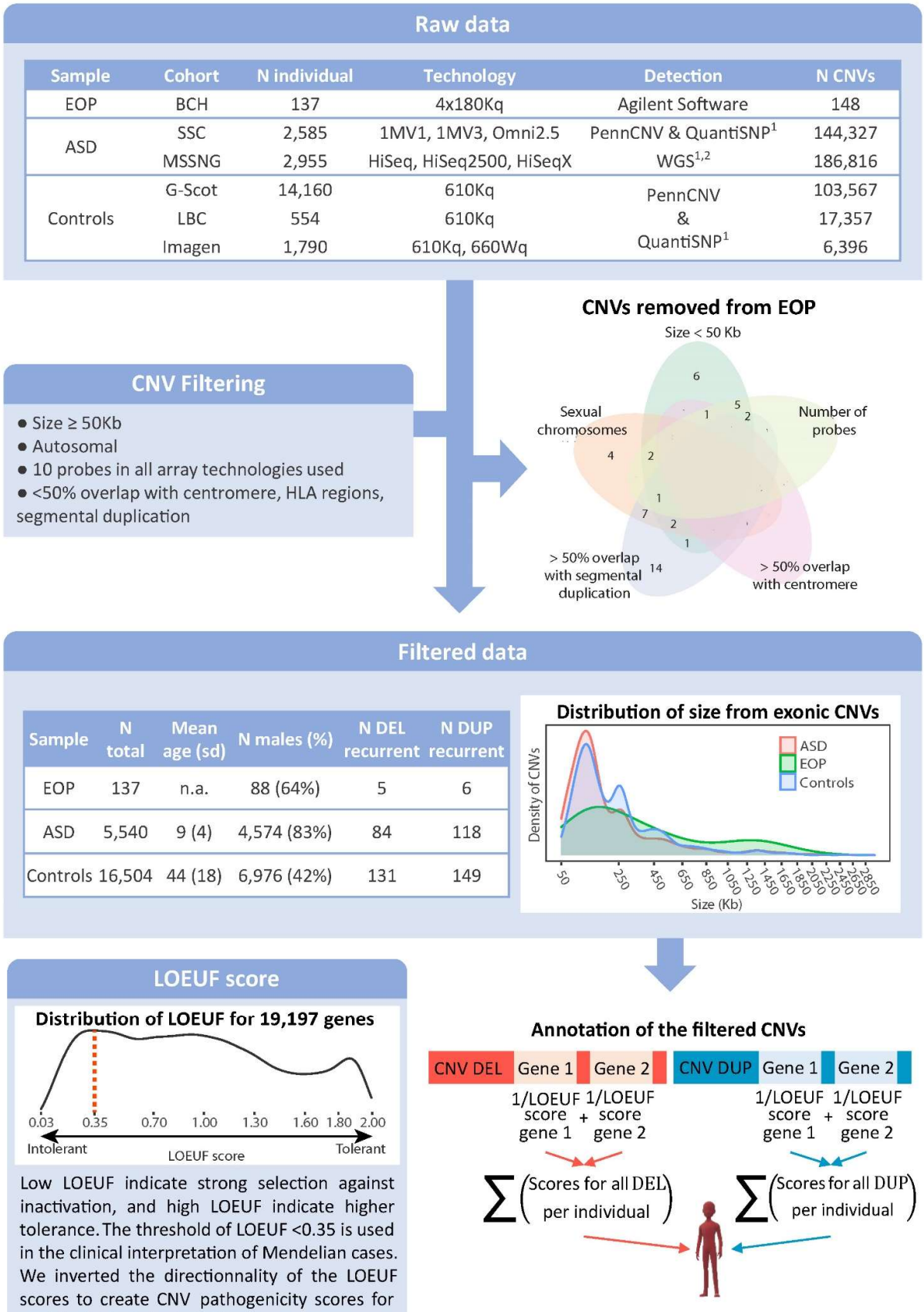
Genotyping and CNV Calling

For the EOP sample, genomic DNA from blood was extracted using standard protocols. Dye-swap array-CGH experiments were performed according to the experimental procedures described by Agilent Technologies (Santa Clara, CA) using standard 4×180K Surescan arrays and analyzed with the Agilent Cytogenomics Software. Probe sequences and locations are based on Genome Reference Consortium build 37 (GRCh37/hg19).

Table 6: Diagnostic and Demographic Breakdown of EOP Cohort

Demographics	Full Sample (n=137)	EOP no ASD (n=99)	EOP with ASD (n=38)
Sex			
Male	87	59	28
Female	50	40	10
Race			
Caucasian	93	69	24
Black/African American	17	12	5
Asian/Pacific Islander	3	2	1
Two or more races	2	2	0
Unknown/Not available	22	14	8
Ethnicity			
Hispanic/Latino	34	27	7
Not Hispanic/Latino	68	46	22
Unknown/Not available	35	26	9
Age at Psychosis Symptom Onset			
<8 years	38	25	13
8-13 years	76	55	21
14-18 years	23	19	4
Primary Psychosis Spectrum Disorder			
Schizophrenia	39	25	14
Affective Psychosis			
Schizoaffective Disorder (Bipolar Type)	11	7	4
Schizoaffective Disorder (Depressed Type)	8	8	0
Major Depressive Disorder Psychotic Features	17	16	1
Bipolar Disorder Psychotic Features	17	14	3
Other Psychotic Disorders			
Schizophreniform Disorder	2	0	2
Brief Psychotic Disorder	1	1	0
Other Specified or Unspecified Schizophrenia Spectrum and Other Psychotic Disorder	43	29	14
Co-Occurring Diagnoses			
Intellectual Disability	17	10	7
History of Seizures	24	16	8

Figure 14: Methodological pipeline for the CNV filtering and annotation



The two tables describe the total CNVs identified in the EOP, ASD and UP populations before and after filtering. The Venn diagram represents distribution of the EOP CNVs discarded in function of the filtering criteria to which they belong: The size is less than 50 Kb; the number of probes is less than 10 in at least one of the technologies used in the cohort included; it overlaps more than 50% with a centromere or a segmental duplication, it is positioned on a sexual chromosome. The first density plot presents the distribution of the size of the CNVs included in the analyses across the different samples (EOP in green, ASD in red, Unselected population in blue). The CNV sizes on the x axis are represented with a square root transformation. The last density plot represents the distribution of LOEUF score across 19,197 coding genes. A LOEUF score ≤ 0.35 is the defined clinical threshold for intolerant genes. For CNV annotation, the coding gene totally encompassed in deletions and duplications were identified and the LOEUF score of each gene was attributed. For each individual, the number of gene encompassed and the 1/LOEUF scores for deletions and duplications were summed separately. CNV: copy number variant; N: number; DEL: deletions; DUP: duplications; EOP: early-onset psychosis; BCH: Boston children hospital; ASD: autism spectrum disorders; SSC: Simons Simplex Collection; UP: unselected populations; G-Scot: Generation Scotland; LBC: Lothian birth cohort; LOEUF: loss-of-function observed/expected upper bound fraction. [1]Huguet *et al*., *Mol Psy* (2021); [2]Trost *et al*., *Am J Hum Genet* (2018).

Three criteria were used to determine the presence of a CNV: 1) at least 7 consecutive probes in the same direction; 2) 1.5-fold average difference between test and reference DNA; and 3) CNV not present in the within-slide control DNA sample. We applied the same pipeline to data from the ASD and unselected cohorts[10-12]. To harmonize the samples, CNVs were filtered by discarding: CNVs <50kb; CNVs that appeared on sex chromosomes; CNVs with >50% overlap with segmental duplication or centromere; and CNVs with <10 probes across all detection technologies used in all included cohorts (see Figure 14).

Recurrent CNV Analyses

We identified recurrent loci and genes previously associated with neurodevelopmental or neuropsychiatric disorders (see Table S1 in supplementary material 2). These loci were defined by >40% overlap with a specific deletion or duplication or if the genes were disrupted by the CNVs. To test for differences in the prevalence of recurrent loci between cohorts, two-sided Fisher's exact tests were employed. When comparing the frequency of individual recurrent CNVs between samples, p-values were adjusted for multiple comparisons using Benjamin-Hochberg correction for false discovery rate (FDR). Secondary analyses excluding EOP participants with co-occurring ASD (n=38) were conducted using the same models. Fisher's exact tests were computed using the fisher.test() function in R.

CNV Pathogenicity Analyses

For each participant, we computed the sum of 1/LOEUF (pathogenicity score) for deletions and duplications separately using our previously published annotation pipeline (see Figure 14)[10]. Briefly, each coding gene with all isoforms fully encompassed in filtered CNVs was identified using ENSEMBL map (Gencode V19 (hg19))[27] and was annotated using the inverse LOEUF (1/LOEUF) score (gnomAD version 2.1.1)[13], which is available for 19,197 genes and ranges from 0.5 (gene tolerant to haploinsufficiency) to 33.3 (gene intolerant to haploinsufficiency). A score of 0 was assigned to individuals with no coding genes encompassed in any CNV. We tested the genome-wide CNV burden with logistic regression models:

$$\ln(odds(Y_i=\text{diagnosis}_i)) \sim \beta_0 + \beta_1\ \text{Pathogenicity Score}_{DELi} + \beta_2\ \text{PathogenicityScore}_{DUPi} + \beta_3\ \text{sex}_i$$

Where $\text{PathogenicityScore}_{DEL/DUP}$ are the sum of 1/LOEUF for deletions (DEL) and duplications (DUP) respectively. β_0, β_1, β_2 and β_3 are the vectors of coefficients for fixed effects. The logistic regression models were computed using the glm() function in R.

Results

EOP Cohort

A total of 137 EOP patients (88 (64.2%) male) were included in this study (see Table 6). Mean age of psychosis symptom onset was 9.8 years (range = 4-17 years old), with 99 (72.3%) patients having psychosis onset before the age of 13 years (*i.e.*, very early-onset psychosis (VEOP)). Thirty-eight (28%) of individuals with EOP had co-occurring ASD,17 (12%) had intellectual disability, and 7 (5%) had both ASD and ID. Among individuals with very early-onset, 34 (34%) had co-occurring ASD. Sixty-nine CNVs meeting quality control criteria were identified in 55 individuals from the EOP cohort (see Figure 14).

Prevalence of Recurrent CNVs in EOP, ASD and Unselected Population Cohorts

When focusing on regions previously associated with neurodevelopmental or neuropsychiatric disorders, we found 11 (8.0% of sample) recurrent CNV carriers in the EOP cohort (see Table S2 in supplementary material 2). In contrast, 202 (3.6%) individuals from the ASD cohort and 280 (1.7%) individuals from the unselected population were recurrent CNV carriers (see Figure 15A). Thus, the prevalence of recurrent CNVs in children and adolescents with EOP was double that observed in ASD (OR=2.30, 95% CI=1.10-4.36, p=0.02) and five times the rate in unselected populations (OR=5.06, 95% CI=2.43-9.50, p=3x10^{-5}). When focusing on EOP participants without co-occurring ASD (n=99), 6 recurrent CNVs carriers remained. In this subgroup of EOP youth without comorbid ASD, enrichment for recurrent CNVs was no longer significant compared to ASD (OR=1.70, 95% CI=0.60-3.92, p=0.18), but remained significant compared to unselected populations (OR=3.74, 95% CI=1.33-8.55, p=7x10^{-3}).

Three recurrent CNVs were individually enriched in EOP participants relative to unselected controls after FDR correction (Table 7): 1q21.1 duplication (OR=52.60, 95% CI=8.69-233.00, p_{FDR}=6x10^{-4}, see Table S3 in supplementary material 2 for a description of 1q21.1 patients), 16p13.11 deletion (OR=30.51, 95% CI=3.13-155.48, p_{FDR}=0.01) and 22q11.2 proximal deletion (OR=∞ [no individuals with the deletion in the unselected cohort], 95%CI 3.09-∞, p_{FDR}=0.02). These same loci were also enriched in the EOP cohort relative to the ASD cohort, but none survived FDR correction (see Table 7).

When EOP participants with co-occurring ASD were removed from the analysis, the same three recurrent CNVs were individually enriched in EOP participants relative to unselected controls after FDR correction (see Table S4 in supplementary material 2): 1q21.1 duplication (p_{FDR}=5x10^{-3}) (see Table S3 in supplementary material 2 for a description of 1q21.1 patients), 16p13.11 deletion (p_{FDR}= 5x10^{-3}) and 22q11.2 proximal deletion (p_{FDR}=0.01). No recurrent CNVs survived the FDR correction when comparing to the ASD cohorts (see Table S4 in supplementary material 2).

Table 7: Enrichment of recurrent CNVs identified in EOP relative to ASD and unselected populations

	EOP (n=137)		ASD (n=5,540)		Controls (n=16,504)		EOP vs. Controls OR [95%CI] p-value		EOP vs. ASD OR [95%CI] p-value	
Loci	**DEL (n=5)**	**DUP (N=6)**	**DEL (n=35)**	**DUP (n=45)**	**DEL (n=83)**	**DUP (n=44)**	**DEL**	**DUP**	**DEL**	**DUP**
1q21.1	--	3	--	19	--	7	--	**52.60 [8.69-233.00] p=6x10^{-5}**	--	6.50 [1.22-22.47] p=0.015
15q11.2 BP1-BP2	1	--	16	--	70	--	n.s.	--	n.s.	--
16p11.2 proximal	1	1	11	14	5	6	24.23 [0.51-218.14] p=0.048	n.s.	n.s.	n.s.
16p13.11	2	2	7	12	8	31	**30.51 [3.13-155.48] p=3x10^{-3}**	7.87 [0.90-31.39] p=0.030	11.69 [1.17-62.20] p=0.019	6.82 [0.73-31.10] p=0.043
22q11.2 proximal	1	--	1	--	0	--	**∞ [3.09-inf] p=8x10^{-3}**	--	40.54 [0.52-inf] p=0.048	--

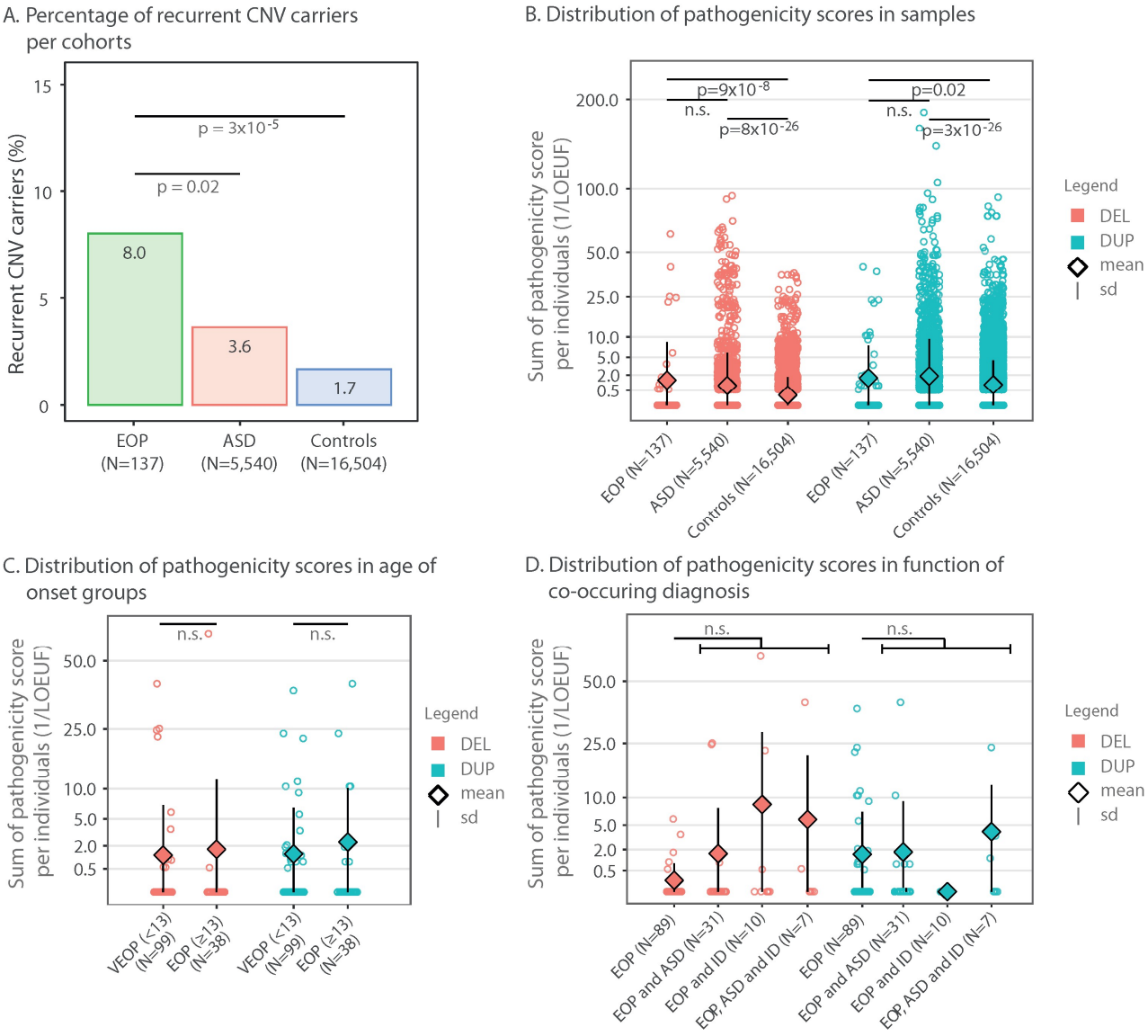
Legend: Odds ratios are computed using Fisher's exact test for deletions and duplications. Significant p-values after FDR correction are in bold (≤ 0.009 when comparing EOP to controls and ≤ 0.005 when comparing EOP to ASD). EOP: early-onset psychosis; Controls: unselected population; ASD: autism spectrum disorder; DEL: deletion; DUP: duplication; OR: Odds ratio; 95% CI: 95% confidence intervals; --: not applicable; n.s.: non-significant

Pathogenicity of CNVs genome-wide in EOP, ASD and Unselected Population Cohorts

The pathogenicity score of genes included in any CNV were higher in EOP participants relative to the unselected cohort for both deletions (1.39 vs. 0.23, OR=1.30, 95% CI=1.26-1.35, p=9x10^{-8}) and duplications (1.63 vs. 0.94, OR=1.09, 95% CI=1.06-1.12, p=0.02) (see Figure 15B). Similar results were found when comparing ASD participants to the unselected cohort (deletions: 0.86 vs. 0.23, OR=1.24, 95% CI=1.22-1.26, p=8x10^{-26}; duplication: 1.88 vs. 0.94, OR=1.12, 95% CI=1.11-1.13, p= 3x10^{-26}). Genome-wide pathogenicity scores did not differ significantly between the EOP and ASD cohorts (deletions: 1.39 vs. 0.86, OR=1.03, 95% CI=1.01-1.06, p=0.33; duplications: 1.63 vs. 1.88 OR=0.98; 95% CI=0.95-1.01 p=0.61; see Figure 15B).

Effect-sizes were similar and remained significant for deletions, but not duplications, after removing individuals with comorbid ASD from the EOP sample (EOP (n=99) vs. Unselected (n=16,504) deletions: 1.00 vs. 0.23, OR=1.24, 95% CI=1.19-1.30, p=3x10^{-4}; duplication: 1.42 vs. 0.94, OR=1.07, 95% CI=1.03-1.11, p=0.17 NS). When comparing very early (<13 years) and early onset (13-18 years) EOP participants, no significant difference in CNV pathogenicity was found, either before or after removing individuals with comorbid ASD (see Figure 15C). Similarly, no significant differences in pathogenicity score were found between EOP individuals without and with comorbid ASD, ID, and both ASD and ID (see Figure 15D). Finally, no sex differences (49 females, 88 males) were observed for pathogenicity scores for deletions (OR=1.01, 95% CI= 0.96-109, p=0.70) or duplications (OR=0.97, 95% CI= 0.91-103, p=0.34).

Figure 15: Measuring the genome-wide CNVs burden in EOP



A. Rates of disease-related recurrent CNVs in individuals with early onset psychosis (EOP), autism spectrum disorder (ASD) probands and controls from the unselected population. B. Distribution of the pathogenicity score in controls, ASD and the EOP samples. Scores for deletions and duplications per individual are represented in red and blue, respectively. Individuals without a CNV or with a non-coding CNV have a score of 0. Coding CNVs have scores ranging from 0.5 to approximately 180. Y axis: the value (root squared) of the sum of 1/LOEUF of all genes encompassed in CNVs identified in each individual. The diamonds and black bars represent the mean and standard deviation of each group. C. Distribution of the pathogenicity score per individual in the VEOP and EOP sub-groups. The VEOP sub-group is defined as the child with an onset before 13 years old, all the other individuals are included in the EOP sub-group. D. Distribution of the pathogenicity score in the EOP samples in function of the co-occurring diagnoses. EOP: early-onset psychosis; ASD: autism spectrum disorder; Controls: unselected population; DEL: deletions; DUP: duplications; n.s.: non-significant; sd: standard deviation; ID: intellectual disability.

Discussion

We found that children and adolescents with various early onset psychosis (EOP) diagnoses showed a similar prevalence of recurrent CNVs as individuals with autism spectrum disorder (ASD)[12]. Prevalence of recurrent CNVs in these EOP individuals was also similar to previous reports of individuals with the more restrictive childhood onset schizophrenia diagnosis[3]. Initially, we selected and analyzed recurrent CNVs that have previously been associated with neurodevelopmental[26] and psychotic[2] illnesses. However, these recurrent CNVs represent only a fraction (~10%) of all CNVs observed in the population. Thus, we also used an index of genome-wide pathogenicity of deletions and duplications[10], finding that the functional impact of CNVs regardless of their population prevalence in EOP youth was comparable to that of ASD cohorts. To our knowledge, this is the first time that prevalence of neurodevelopmentally-associated recurrent CNVs and pathogenicity of total CNV burden have been reported for EOP. Our results indicate that EOP is associated with a substantial CNV burden, strongly suggesting that systematic genetic screening in EOP is clinically warranted.

Given the success of genetic screening in ASD[28], our findings suggest that all children and adolescents with a psychotic diagnosis could substantially benefit from CMA testing, as well as the potential for further testing contingent upon family history and/or clinical features. Universal genetic screening[20] would help disentangle the clinical heterogeneity among EOP youth[6], potentially leading to specific treatment regiments. As detailed in Moreno-De-Luca *et al.*[29], genetic diagnoses allow clinicians to communicate more effectively with patients and families, and allow the potential for genetic counseling. Genetic information would also connect families to additional resources and networks, such as other families with the same CNV. Forming cohorts of patients with the same CNV may also yield valuable information about comorbidities, the range of possible phenotypes, and disease progression. Furthermore, children and adolescents with EOP who carry recurrent CNVs associated with serious non-psychiatric medical conditions (*e.g.*, cardiovascular abnormalities in 22q11.2 deletion syndrome or

the high incidence of hypotonia and epilepsy in 15q11.2 duplication carriers) could be more carefully monitored. Finally, information derived from genetic screening is often invaluable to families of children with ASD[28], helping parents appreciate the biological nature of the illness. Similar genetic information would no doubt be well received by EOP families as well. Overall, genetic screening in EOP has the potential to bring us one step closer to true precision medicine in pediatric psychiatry.

Among the recurrent CNVs previously associated with neurodevelopmental and neuropsychic illness, we documented a significant enrichment for three mutations: 22q11.2 proximal deletion (due to a lack of carries in the population control), 16p13.11 deletion and 1q21.1 duplication. Each of these CNVs was reported in the prior childhood onset schizophrenia study[3] and in ASD cohorts[12]. While these specific CNVs could be particularly informative to the pathobiology of psychosis and neurodevelopment, these mutations were also among the most commonly observed CNVs in a large sample of individuals with idiopathic adult-onset schizophrenia[2]. As the number of CNVs observed is directly related to sample size, it is possible that with a larger EOP cohort, evidence for additional recurrent CNVs might emerge. Given this possibility, it is difficult to speculate about the genetic architecture of EOP and whether child and adolescent psychosis onset is more strongly influenced by an enumerable set of rare mutations or by countless mutations of small effect (*e.g.*, polygenic model), the currently favored model for idiopathic adult-onset psychosis. Further investigations, with much larger EOP samples, are needed.

Since recurrent CNVs reflect only a fraction of deletions and duplications observed in neurodevelopmental and neuropsychiatric illnesses, we also used a pathogenicity score as a global index of the functional genomic impact of CNVs across the genome. Using this pathogenicity score, we demonstrated a higher overall burden of genes intolerant to mutation in EOP compared to unselected samples. Moreover, larger effect-sizes were observed for deletions than duplications. Interestingly, these effect-sizes observed for EOP were in the same range as those found when comparing ASD to unselected populations[12], suggesting a major contribution of haploinsufficiency in both EOP and

ASD. Additional work is needed to document the relative strength of our pathogenicity score in EOP compared to adult-onset psychotic disorders, as well as to other neurodevelopmental disorders.

Our study has several strengths, such as the use of a unique sample of EOP patients with a range of comorbidities commonly reported in psychotic disorders, as opposed to previous studies which have removed all cases with comorbidities (*e.g.*, comorbid ID).[3] Moreover, we were able to compare CNV burden in EOP to ASD and general population controls. However, our EOP sample is small, which is to be anticipated given the rarity of psychosis in children and adolescents. Future studies with larger samples might reveal additional recurrent CNVs or stronger effects of genome-wide duplications, similar to findings in larger ASD or schizophrenia samples[2,12]. We also did not observe significant differences in CNV burden between children and adolescents with early (<18 years) compared to very early (<13 years) psychosis onset, or between EOP youth with and without comorbid disorders. Larger sample sizes are needed to examine whether factors such as illness onset and comorbid illness influence genetic predisposition in EOP. Nonetheless, the high frequency of CNVs in our EOP cohort suggests that routine screening for CNVs should be made available to EOP patients and could have important implications for genetic counseling and patient management. These relatively high penetrance risk alleles are also promising targets for biological research aimed at developing animal and cellular models to identify novel disease mechanisms and drug targets for psychotic disorders.

References

1. Girirajan S, Rosenfeld JA, Coe BP, et al. Phenotypic heterogeneity of genomic disorders and rare copy-number variants. *The New England journal of medicine.* 2012;367(14):1321-1331.

2. Marshall CR, Howrigan DP, Merico D, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. *Nature genetics.* 2017;49(1):27-35.

3. Ahn K, Gotay N, Andersen TM, et al. High rate of disease-related copy number variations in childhood onset schizophrenia. *Mol Psychiatry.* 2014;19(5):568-572.

4. Nicolson R, Rapoport JL. Childhood-onset schizophrenia: rare but worth studying. *Biological psychiatry.* 1999;46(10):1418-1428.

5. Stentebjerg-Olesen M, Pagsberg AK, Fink-Jensen A, Correll CU, Jeppesen P. Clinical Characteristics and Predictors of Outcome of Schizophrenia-Spectrum Psychosis in Children and Adolescents: A Systematic Review. *Journal of child and adolescent psychopharmacology.* 2016;26(5):410-427.

6. Castro-Fornieles J, Baeza I, de la Serna E, et al. Two-year diagnostic stability in early-onset first-episode psychosis. *J Child Psychol Psychiatry.* 2011;52(10):1089-1098.

7. Patel PK, Leathem LD, Currin DL, Karlsgodt KH. Adolescent Neurodevelopment and Vulnerability to Psychosis. *Biol Psychiatry.* 2021;89(2):184-193.

8. Stefansson, H., Meyer-Lindenberg, A., Steinberg, et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*. 2014;*505*(7483), 361-366.

9. Zarrei M, Burton CL, Engchuan W, et al. A large data resource of genomic copy number variation across neurodevelopmental disorders. *NPJ Genom Med.* 2019;4:26.

10. Huguet G, Schramm C, Douard E, et al. Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability. *Molecular psychiatry.* 2021.

11. Huguet G, Schramm C, Douard E, et al. Measuring and Estimating the Effect-sizes of Copy Number Variants on General Intelligence in Community-Based Samples. *JAMA psychiatry.* 2018;75(5):447-457.

12. Douard E, Zeribi A, Schramm C, et al. Effect-sizes of Deletions and Duplications on Autism Risk Across the Genome. *The American journal of psychiatry.* 2021;178(1):87-98.

13. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.

14. Collins RL, Glessner JT, Porcu E, et al. A cross-disorder dosage sensitivity map of the human genome. *medRxiv.* 2021:2021.2001.2026.21250098.

15. Tammimies K, Marshall CR, Walker S, et al. Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. *Jama.* 2015;314(9):895-903.

16. Council on Children With D, Section on Developmental Behavioral P, Bright Futures Steering C, Medical Home Initiatives for Children With Special Needs Project Advisory C. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics.* 2006;118(1):405-420.

17. Hyman SL, Levy SE, Myers SM, Council On Children With Disabilities SOD, Behavioral P. Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. *Pediatrics.* 2020;145(1).

18. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *American journal of human genetics.* 2010;86(5):749-764.

19. Sikich L, Frazier JA, McClellan J, et al. Double-blind comparison of first- and second-generation antipsychotics in early-onset schizophrenia and schizo-affective disorder: findings from the treatment of early-onset schizophrenia spectrum disorders (TEOSS) study. *Am J Psychiatry.* 2008;165(11):1420-1431.

20. Martin CL, Wain KE, Oetjens MT, et al. Identification of Neuropsychiatric Copy Number Variants in a Health Care System Population. *JAMA Psychiatry.* 2020.

21. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. *Neuron.* 2010;68(2):192-195.

22. C Yuen RK, Merico D, Bookman M, et al. Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. *Nature neuroscience.* 2017;20(4):602-611.

23. Schumann G, Loth E, Banaschewski T, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Molecular psychiatry.* 2010;15(12):1128-1139.

24. Smith BH, Campbell A, Linksted P, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *International journal of epidemiology.* 2013;42(3):689-700.

25. Deary IJ, Gow AJ, Pattie A, Starr JM. Cohort profile: the Lothian Birth Cohorts of 1921 and 1936. *International journal of epidemiology.* 2012;41(6):1576-1584.

26. Kendall KM, Bracher-Smith M, Fitzpatrick H, et al. Cognitive performance and functional outcomes of carriers of pathogenic copy number variants: analysis of the UK Biobank. *The British journal of psychiatry : the journal of mental science.* 2019;214(5):297-304.

27. Frankish A, Diekhans M, Ferreira AM, et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic acids research.* 2019;47(D1):D766-d773.

28. Jeste SS, Geschwind DH. Disentangling the heterogeneity of autism spectrum disorder through genetic findings. *Nat Rev Neurol.* 2014;10(2):74-81.

29. Moreno-De-Luca D, Ross ME, Ross DA. Leveraging the Power of Genetics to Bring Precision Medicine to Psychiatry: Too Little of a Good Thing? *Biological psychiatry.* 2018;83(8):e45-e46.

IV. Discussion

Rare pathogenic recurrent CNVs represent only ~10% of the CNVs identified in neurodevelopmental clinics. (Huguet et al. 2018; Zarrei et al. 2019) It is likely that the remaining portion of the genome contributes to the genetic architecture of NDDs such as EOP and ASD. (Wray, Wijmenga, et al. 2018; Sullivan and Geschwind 2019) A reliance on CMA detection has restricted the focus of association studies to CNVs with the largest effect-sizes, and the lack of reliable reference resources has prevented the field from systematically studying ultra-rare variants. (Miller et al. 2010; Collins et al. 2020) Recently, my research group developed a strategy to estimate the effect-sizes of genome-wide CNVs on cognition, regardless of their population prevalence, using the constraint scores (*i.e.*, pLI and LOEUF) of genes encompassed in the identified CNVs. (Huguet et al. 2018; Huguet et al. 2021) Such models were accurately predicting CNV effect-size on IQ at 75% when using pLI score (Huguet et al. 2018) to 78% when using LOEUF score. (Huguet et al. 2021) For the current thesis, similar statistical models were developed to estimate ASD and EOP risk conferred by any genome-wide CNV.

Contributions

Aim 1: Measuring the contribution of genome-wide CNVs to ASD diagnosis

Among 10 genetic and functional scores, the constraint score (*i.e.*, pLI) was identified as the best variable explaining ASD risk variance. Paper #1 and #2 demonstrated a 2-fold higher effect-size of deletions over duplications with both pLI and LOEUF scores ($OR_{deletion}$ = 1.45 vs. $OR_{duplication}$ = 1.25 for 1 point of pLI; $OR_{deletion}$ = 1.24 vs. $OR_{duplication}$ = 1.12 for 0.35 points of LOEUF). I estimated that any 1Mb CNV across the genome (encompassing at least one gene with a pLI value) increased ASD risk with a mean OR = 1.60 for deletions and 1.30 for duplications (paper #1). I also compared the

predicted effect-size to published empirical ASD risk for 16 recurrent CNVs, and only two loci were underestimated by the model (paper #1).

Main message:

- *Effect-size on ASD of genome-wide CNVs can be measured using constraint scores such as pLI or LOEUF scores.*
- *Effect-size of deletions is twice higher than the effect-size of duplications.*

Aim 2: Identifying traits which explain the enrichment of CNVs in ASD populations

In paper #1, I aimed to investigate whether the CNV burden identified above affected core symptoms of ASD diagnosis or ASD specifiers.

Results showed that CNVs measured by pLI did not affect any of the continuous measures of ASD severity (*i.e.*, SRS, ADOS, ADI-R) when adjusting for the diagnosis of ASD. In other words, pLI did not explain the variance of these scores within control groups or ASD populations.

ID is a major specifier of ASD (occurring in up to 48% of cases) (Postorino et al. 2016) and results from paper #1 showed that 45% of the effect of deletions and 23% of the effect of duplications on ASD risk were potentially mediated by IQ. The effect-size of CNVs on general intelligence - measured by IQ and VABS-II - within the ASD probands was 2-fold higher for deletions compared to duplications. Remarkably, these effect-sizes were identical to those previously published in unselected populations. (Huguet et al. 2018; Huguet et al. 2021)

Language and motor disorders are the most prevalent comorbidities in ASD and are observed in up to 30% and 80% of cases respectively. (Frazier et al. 2014; Kopp, Beckung, and Gillberg 2010) Deletions measured by pLI altered measures of language skills (*i.e.*, CTOPP, presence or absence of world delay) and these effects

were driven by IQ. In contrast, duplications showed no effects. Deletions and duplications affected similarly measures of motor skills (*i.e.*, DCDQ, VABS-II, age of onset for walking, presence or absence of walking delay), and remained unchanged after adjusting for IQ. The latter is consistent with several studies demonstrating that ASD individuals with a *de novo* LoF variant presented a delayed age of walking or impaired motor skills. (Bishop et al. 2017; Buja et al. 2018; Satterstrom et al. 2020)

Main message:

- *Deletions and duplications increase ASD risk even after adjusting for their effect on IQ.*
- *However, CNVs do not appear to affect measures of ASD core symptoms.*
- *Deletions contribute to a decrease of language skills driven by general intelligence.*
- *Deletions and duplications similarly contribute to motor impairment, which is not influenced by general intelligence.*

Aim3: Providing knowledge about the genetic architecture of EOP to shed light on its etiology and improve clinical care

Paper #2 is, to my knowledge, the first study to report an association of CNVs with EOP. To date, etiological factors of EOP remained understudied and genetic investigations were limited to CNVs or single nucleotide variants reported case-studies, including 22q11.2 deletion (Ivanov et al. 2003; Vorstman et al. 2006), 16p13.11 deletion and duplication (Brownstein et al. 2016), *RCL1* (Brownstein et al. 2021) and *TRRAP*. (Mavros et al. 2018) Moreover, there are currently no medical guidelines for genetic screening of children with psychotic disorders, which results in missed opportunities for guiding clinical management, and improving treatment outcomes.

Deletions and duplications increase risk for EOP

The paper #2 is the first assessment of the contribution of CNV in a unique sample of 137 EOP patients from the Boston Children's Hospital.

I demonstrated a higher rate of NDD-associated recurrent CNVs in EOP (8.0%) than in the unselected population (1.7%).

I also observed a higher genome-wide burden of deletions and duplications weighted by LOEUF and the effect-sizes of deletions were 2-fold higher than those observed for duplications.

Main message:

- *Recurrent CNVs previously associated with NDDs are enriched in EOP.*
- *The genome-wide burden of deletions and duplications weighted by LOEUF increase risk for EOP.*

Equal contribution of CNVs to EOP and ASD suggests that genetic screening for EOP should be implemented in medical practice

Findings suggesting that up to 20% of individuals with ASD carry a rare pathogenic genetic variation (Jiang et al. 2013; Tammimies et al. 2015) have led to genomic screening recommendations for individuals with a diagnosis of ASD. (Hyman et al. 2020)

Remarkably, the estimated effect-sizes of one deleted or duplicated gene with a LOEUF value of 0.35 was identical for EOP and ASD risk: $OR_{deletions}$ = 1.30 in EOP vs. 1.24 in ASD; $OR_{duplications}$ = 1.09 in EOP vs. 1.12 in ASD. These effect-sizes were also similar to those previously published for schizophrenia when accounting for the number of genes encompassed in all recurrent and non-recurrent CNVs ($OR_{deletions}$ = 1.40; $OR_{duplications}$ = 1.12). (Marshall et al. 2017)

Rates of NDD-related recurrent CNVs in EOP were also in line with those previously reported in the largest CNV study of COS to date (12%). (Ahn et al. 2014)

Given the medical benefits of genetic screening in ASD (Jeste and Geschwind 2014; Hyman et al. 2020), we recommend a systematic genetic screening for EOP in the clinic. Previous studies have demonstrated that genetic screening in NDDs provide valuable

information for genetical counselling, medical comorbidities, disease progression and can help parents appreciate the biological nature of the illness. (Jeste and Geschwind 2014; Moreno-De-Luca, Ross, and Ross 2018) We expect that genetic screening in EOP would provide similar benefits and has the potential to promote personalized medicine in child psychiatry (specific treatment regiments and additional resources for children with EOP and their families).

Main message:

- *The effect-size of deletions and duplications are the same for ASD and EOP risk.*
- *We recommend the implementation of genetic screening in the clinic for children and adolescents with EOP.*

Contributions not directly related to initial aims

In addition to my main aims, these papers provided substantial evidence supporting previously established theories on the genetic architecture of NDDs, such as the polygenic model of psychopathologies and the pleiotropic effect of genomic variants.

Polygenic models: a large proportion of the genome contributes to ASD risk when deleted or duplicated

In paper #1, additive models were trained on CNVs covering over 4,500 different genes with a full spectrum of values measuring intolerance to LoF variants. Results provided by these models were unchanged even after excluding previously known pathogenic CNVs as well as *de novo* CNVs. All of the findings above suggests that a broad spectrum of CNVs, including genes tolerant to haploinsufficiency, also contribute to ASD risk.

In paper #2, such additive models estimated identical effect-sizes of CNVs for on ASD and EOP risk.

Other studies from my laboratory also suggested that approximately half of the genome (all genes with a LOEUF values <1) affects cognitive ability when deleted. (Huguet et al. 2018; Huguet et al. 2021). Preliminary results (described later) suggests that the same findings apply to ASD.

Altogether, these findings support polygenic properties of gene dosage with respect to ASD and EOP, with the contribution of genes with various effect-size. (Wray, Wijmenga, et al. 2018; Sullivan and Geschwind 2019; Myers et al. 2020)

However, variation in one gene alone can also substantially contribute to a disorder (*e.g.*, *FMR1* for fragile X syndrome). Although extreme effect-size genes have been identified in ASD, I expect that the total number of these genes will remain low.

The pleiotropic patterns of genomic variants associated with psychiatric disorders

Results from paper #2 showed a clear genetic overlap of pathogenic recurrent CNVs identified in ASD and EOP. This is in line with several association studies showing shared loci between ASD, COS and adult-onset schizophrenia. (Ahn et al. 2014; Marshall et al. 2017; Sanders et al. 2019; Collins et al. 2021) Moreover, the effect-sizes of recurrent and non-recurrent CNVs on ASD and EOP risk were identical, and were also similar to those previously published for adult-onset schizophrenia. (Marshall et al. 2017) Overall, results suggest pleiotropic effect of a substantial fraction of the genome which may increase risk for both conditions when deleted or duplicated.

My findings are consistent with previous studies reporting pleiotropy for both common and rare variants. (Sanders et al. 2019; Watanabe et al. 2019; Akingbuwa et al. 2021; Lee, Feng, and Smoller 2021; Collins et al. 2021) Previous studies of common variants have demonstrated genetic overlaps in between ASD, COS, adult-onset schizophrenia and psychotic experience. (Ahn et al. 2016; Velthorst et al. 2018; Grove et al. 2019; Perkins et al. 2020) A recent PGC-CDG[32] analysis applied a multivariate genetic association

[32] PGC-CDG (psychiatric genomics consortium-cross-disorder group) is a subdivision of the PGC created in 2008. This group focuses on the study of cross-phenotype genetic influences that transcend diagnostic boundaries.

spanning eight psychiatric disorders (*i.e.,* anorexia nervosa, ADHD, ASD, BD, major depression disorder, obsessive-compulsive disorder, schizophrenia, and Tourette syndrome). (Cross-Disorder Group of the Psychiatric Genomics Consortium 2019) This study revealed three correlated genomic factors that accounted for more than 50% of the common genetic variation underlying these disorders. The first factor comprised compulsive behaviors (*i.e.*, anorexia nervosa, obsessive-compulsive disorder, and Tourette syndrome), the second factor included mood and psychotic disorders (*i.e.*, BD, major depression disorder, and schizophrenia), while the third factor encompassed three early-onset NDDs (*i.e.*, ASD, ADHD, Tourette syndrome) as well as major depression disorder. Interestingly, the pervasive genetic correlation presented above has not been observed between psychiatric and neurological conditions (including stroke, epilepsies, multiple sclerosis, Parkinson disease and migraine). (Brainstorm Consortium et al. 2018)

Overall, genetic risk contributing to specific psychiatric diagnoses are yet to be convincingly demonstrated. (Myers et al. 2020; Rietz et al. 2021)

Altogether, these results provide insights into the underlying structure of psychopathology that could inform the reconceptualization of psychiatric nosology, nourishing the ongoing debate about the validity of categorical diagnoses. (Clark et al. 2017; Lee, Feng, and Smoller 2021)

Limitations

Categorical diagnoses and dimensional measures

Most of the work presented in this thesis relies on categorical diagnosis without taking into account comorbidities as well as continuous cognitive and behavioral traits. In particular, I have failed to accurately delineate the effect of CNVs on either core symptoms of ASD or EOP as opposed to co-occurring psychopathology.

In paper #1, I tried to investigate if the CNV burden in ASD was driven by their effect on core symptoms or on associated specifiers. However, the measures of core symptoms severity used (*i.e.*, SRS, ADOS, ADI-R) did not provide additional granularity once the diagnosis of ASD was accounted for. Although SRS is presented as a continuous measure, its distribution when pooling general population and ASD probands was bimodal. This may suggest that the SRS is mainly categorical as opposed to a continuous trait and it did not provide further information on ASD severity. Alternatively, this may suggest that CNVs increase the risk for ASD without affecting behavioral dimensions captured by the SRS.

I also investigated the ADOS and ADI-R scores (which were available for ASD probands only), but these instruments were designed to establish an ASD diagnosis and are not considered to reflect a dimensional trait.

Limitations related to comorbidities are also present in paper #2 when comparing the rates of CNVs in EOP and ASD. Removing EOP patients with co-occurring ASD attenuated the borderline signal favoring a higher rate in EOP than in ASD. The signal detected before this sensitivity analysis may be driven by the individuals with a more severe and complex clinical phenotype.

The differential diagnoses tend to trap patients into unadapted categories rather than allowing to understand the complexity of the outcomes. (Kanner 1969) The reassessment of the current diagnostic categories towards more dimensional measures is supported by the substantial genetic overlap across disorders previously discussed, which suggests a spectrum of psychopathology instead a clear delineation of the clinical outcomes.

Recent efforts have attempted to provide dimensional measure of psychopathologies, such as the National Institute of Mental Health's Research Domain Criteria (RDoC)[33] from

[33] The RDoC is a research framework for investigating mental disorders. This tool was launched in 2009 by the national institute of mental health to help inform the creation of mental health screening tools, diagnostic systems, and treatments. It is implemented as a matrix that integrates many levels of information

2009 (Insel and Cuthbert 2009; Clark et al. 2017) or *p* factor[34] from 2014. (Caspi et al. 2014; Neumann et al. 2016; Allegrini et al. 2020) These continuous measures can be assessed in both neurotypical and clinical populations. However, studies have shown that dimensional traits systematically show lower heritability than diagnostic categories, and one may question the level of validity of such instruments. (Akingbuwa et al. 2021) On the other hand, this may be due to the small sample size used to compute the heritability of psychiatric traits

Overall, such measures have not yet replaced categorical diagnoses and most genomic datasets lack dimensional traits. Cognitive ability is an exception, because this tool was developed in the late 19th century and was widely studied in both unselected and clinical populations. Therefore, large samples with this measure are available. It allowed us to accurately estimate the effect-size of genome-wide CNVs to general intelligence in both unselected and clinical populations. (Huguet et al. 2018; Huguet et al. 2021)

Sample size

An advantage of the categorical diagnosis is that multiple consortia put effort to merge cumulated samples with a common clinical presentation, such as ASD, schizophrenia or ADHD. These collaborations provided databases with sufficient statistical power to detect signals. Sample size remains the number one limiting factor for investigating the effects of rare variants on traits and risk for disease, and multiple studies failed to detect any significant signal due to the lack of statistical power. This pitfall was emerging in multiple aspects of my analyses.

In the first paper, the sample size after pooling the ASD and unselected populations allowed us to obtain a significant effect-size of CNVs on categorical diagnosis. However, the models using continuous variables (*e.g.*, ADOS, ADI-R, CTOPP) within the ASD

- from genomics to behavior - to explore six major domains of human functioning - from normal to abnormal.

[34] The p factor is a measure of the general psychopathology. It is computed using confirmatory factor analysis across various self-, parent-, and teacher-rated measures of behavioral problems.

probands population were borderline significant or failed to detect any signal. Such difference between estimates for continuous and categorical clinical measures were previously discussed in the paragraph above (Akingbuwa et al. 2021) These results are be likely due to the lack of statistical power when investigating the continuous variables.

Sample size and power was also a limiting factor for the investigation of duplication effects in both papers #1 and #2, because of a two-fold lower effect-sizes than those recorded for deletions.

Sample size limited most analysis in the study of EOP and I expect that, as for cognitive ability and ASD, gene dosage of large proportion of the genome would contribute to EOP risk.

Perspectives

Effect of CNVs on risk for NDDs and on general intelligence measured by intolerance to LoF and GWAS of CNV

Using my models of ASD and EOP risk, I observed an additive effect of genes which score for moderate to high intolerance to LoF variations. Results of this project and of another study from my laboratory demonstrated that deleted genes with a LOEUF ≤ 0.35 have on average very mild effects on ASD risk (OR = 1.24) and cognitive ability (-0.25 to -0.37 z-scored IQ for genes highly intolerant and moderately intolerant respectively,

Figure 16). (Huguet et al. 2021) The second study also suggests that genes with even milder effects on cognitive ability (-0.01 z-score) also seemed to be tolerant to haploinsufficiency (0.35 ≤ LOEUF < 1).

The interpretations of such relationship between ASD risk or cognitive ability, and measures of intolerance to LoF are intuitive for variants with extreme effect-size leading to severe conditions affecting fecundity. However, such reasoning does not apply to CNVs with mild effects and the reasons underlying this tight relationship remain elusive. For future projects of my research group, we aim to investigate the reasons why CNVs with small and moderate effect-sizes on ASD risk and cognitive ability are under negative selection.

We also aim to study the contribution of genome-wide CNVs on ASD and EOP risk in function of the tolerance categories (*e.g.*, highly intolerant, moderately intolerant, tolerant, highly tolerant) such as in Huguet *et al*. (2021). Therefore, similarly sized samples, if not larger, are needed to detect any signal, particularly for the tolerant genes. I recently applied such a method in a sample of 13,639 ASD probands compared to 507,942 individuals from the general population; the preliminary results suggest a decreased effect-size on ASD risk along the axis of tolerance to LoF variations (Figure 17).

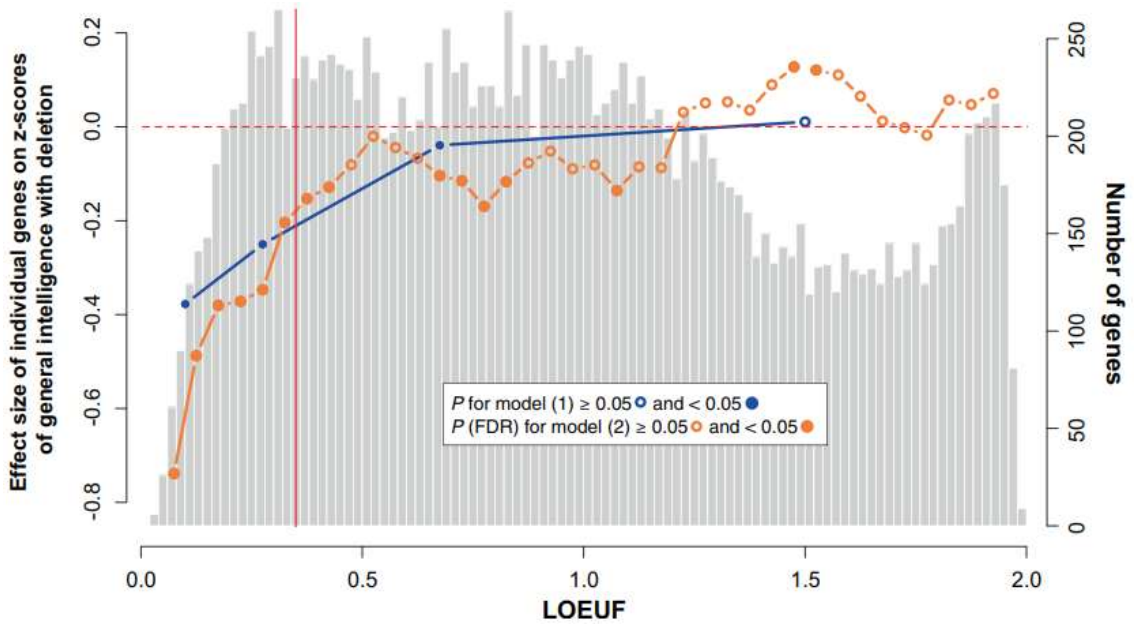


Figure 16: Effect-size of individual genes encompassed in deletions on general intelligence
The light gray histogram represents the distribution of LOEUF values for 18,451 autosomal genes. The blue line represents the estimates for a deleted gene in each of the four categories of LOEUF included in the model: highly intolerant genes (LOEUF < 0.2, n = 980), moderately intolerant genes (0.2 ≤ LOEUF < 0.35, n = 1,762), tolerant genes (0.35 ≤ LOEUF < 1, n = 7,442) and genes highly tolerant to LoF variants (LOEUF ≥ 1, n = 8,267). The orange line represents the estimated effect-size of 37 categories of deleted genes based on their LOEUF values (sliding windows = 0.15). Genes with a LOEUF below 0.35 (vertical red line) are considered as intolerant to LoF variants. Left Y-axis values: z-scored general intelligence (1 z-score is equivalent to 15 points of IQ) for deletion. Right Y-axis values: number of deleted genes represented in the histogram. This Figure is from Huguet *et al.* 2021. (Huguet et al. 2021)

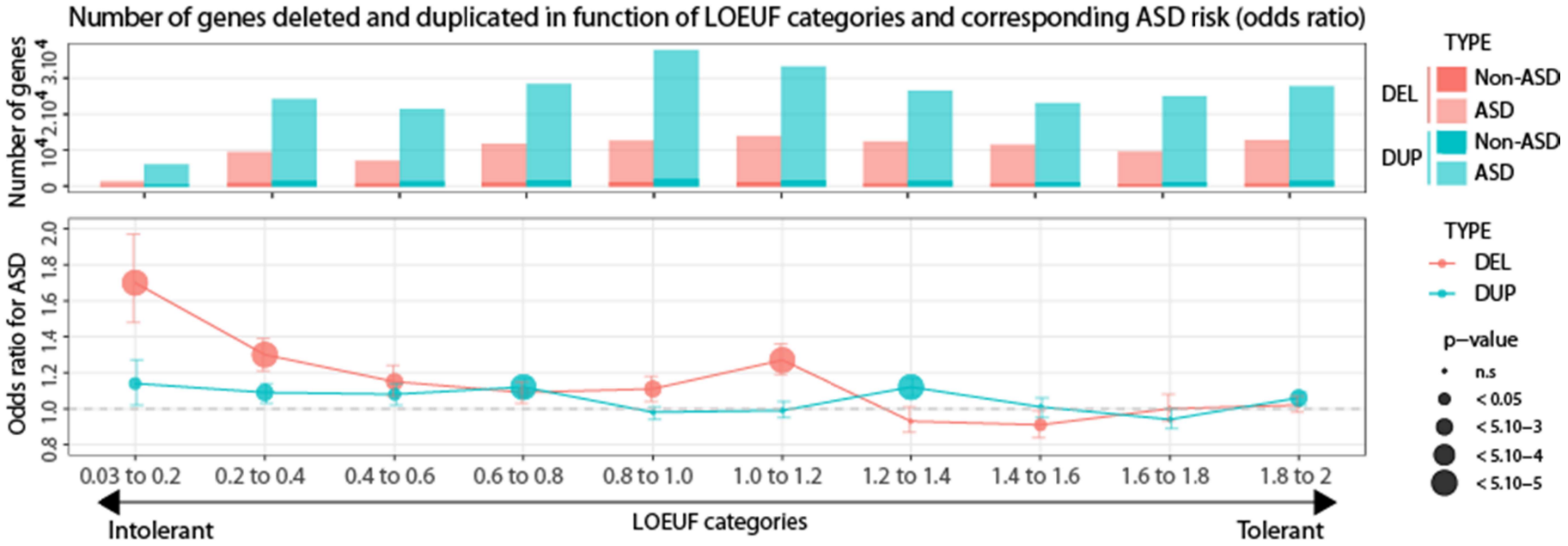


Figure 17: Effect-size of genes deleted and duplicated on ASD risk

The histogram represents the distribution of deleted (red) or duplicated (blue) genes in function of their LOEUF category for individuals with ASD or from the unselected population. The lines represent the estimated odds ratio for a deleted (red) or duplicated (blue) gene in each category of LOEUF included in the model.

With this method and in the current thesis project, we are not able to distinguish variants that may lead preferentially to ASD or EOP. To investigate differences in gene function associated with ASD or EOP, I conducted gene ontology (GO) terms enrichment analyses based on data from the paper #2 using three online tools: *Panther* (http://www.pantherdb.org), *Gorilla* (http://cbl-gorilla.cs.technion.ac.il), and *David* tools (http://david.ncifcrf.gov/tools.jsp.) In these analyses, genes totally encompassed in CNVs from the EOP sample were compared to those in the ASD sample. *Panther* showed a borderline significant enrichment of genes related to the chromatin remodeling/organization[35] in EOP compared to ASD (OR = 7, p-FDR = 0.011). After removing the individuals with comorbid ASD from the EOP sample, results remained significant (OR = 9, p-FDR = 0.002); and two other similar GO terms were enriched in EOP: genes implicated in the regulation of chromatin silencing (OR = 18, p-FDR = 0.037) and in the regulation of gene expression (OR = 18, p-FDR = 0.016). These GO terms were previously enriched in genes linked to childhood-onset schizophrenia (Guo et al. 2021), ASD (Satterstrom et al. 2020), NDDs (Coe et al. 2019), and general intelligence. (Huguet et al. 2021) This is in line with the putative contribution of epigenic changes to the etiology of psychotic spectrum disorders. (Pidsley and Mill 2011) However, no enrichment of these functions was identified in the most recent and largest association studies in adult-onset schizophrenia (Ripke et al. 2014; Marshall et al. 2017), and no significant results or similar trend were found with the two other tools. Much larger EOP samples are required to confirm robust signal.

To further investigate this knowledge gap, an ongoing thesis project aims to identify CNVs and CNV-genes which modulate risk for ASD, schizophrenia/EOP and general intelligence using a GWAS of CNVs inspired from the methodology of Marshall *et al.* 2017. (Marshall et al. 2017) Such analyses may allow to identify CNVs that preferentially modulate one phenotype over the other. Additional GO terms enrichment analyses will be performed with the set of genes significantly associated with each phenotype to explore functions that may explain preferential risk.

[35] Gene ontology category regrouping the genes implicated in nucleosome assembly/nucleosome modeling/histone chaperone which are related to the aggregation, arrangement and bonding together of a nucleosome

Including brain expression scores to understand the relationship between CNVs, brain architecture and NDD risk

Indeed, the constraint scores selected for my models are epidemiological measures of the genetic fitness in human populations, without any consideration of gene function. I acknowledge that it is critical to identify gene functions that underly risk for NDD, but none of those tested in this project (differential stability score[36] and expression quantitative trait loci regulating genes expressed in the brain) outperformed pLI or LOEUF.

In Satterstrom *et al.*, the rare disrupting variants associated with ASD were found to be enriched in genes involved in neuronal communication and expression regulation. (Satterstrom et al. 2020) These genes were also highly enriched in the cortex and cerebellar hemisphere, which is in keeping with prior analyses where ASD-associated genes were expressed at high levels in the human cortex and early in development. (Bourgeron 2007; Parikshak et al. 2013; Chang et al. 2015) Grove *et al.* identified a similar signal with a significant enrichment of common variations in regions with regulatory elements predicted to be active in the fetal brain, specifically in human corticogenesis. (Grove et al. 2019) Altogether, these findings highlight a potential spatiotemporal convergence of genetic risk on this specific developmental epoch.

However, we argue that these enrichment analyses were applied mainly to a small group of genes with the largest effect-size on ASD risk and do not take into account the contribution of genome-wide risk factors. In a recent study showing that almost half of the coding genome affects cognition when deleted, enrichment in neuronal communication and gene regulation was replicated for genes intolerant to LoF, but most functional GO terms were implicated for the other more tolerant contributing genes. (Huguet et al. 2021) As we showed that some CNVs that decrease IQ also increase ASD risk, we expect that this broad functional landscape will also apply to ASD.

[36] The differential stability scores (Hawrylycz et al. 2015), is the mean pairwise correlation between gene expression patterns in brain areas using six adult human brains from the Allen human brain atlas project. (Arnatkevičiūtė, Fulcher, and Fornito 2019)

To make sense of this broad diversity of cellular and molecular functions implicated in cognition and likely ASD, we propose to investigate the macroscopic brain function of these genes while considering their tolerance to LoF variants. We define the macroscopic brain function by the implication of genes in the genesis and maintenance of large-scale brain networks. Indeed, cognition and behavior are conceptualized as an emerging property of large-scale brain networks and is quite removed from cellular and molecular phenotypes.

These macroscopic functions can be measured by spatial patterns of gene expression, which are understudied in psychiatric genomics. Newly developed gene scores, measuring the level of expression of genes in the adult brain, may be potentially informative. For example, the differential stability scores quantify the consistency of gene expression patterns across brain structures (Hawrylycz et al. 2015). The cerebral expression modules are derived from this score and allow to classify genes enriched in certain categories of brain cell ontology (*e.g.*, neurons, astrocytes, oligodendrocytes) (Table 8). Finally, the measures of the cortical gradient of gene expression[37] allow to score genes based on their expression along a gradient from sensory-motor areas to higher-order transmodal areas (*i.e.,* prefrontal cortex) of the human cortex (Figure 18) (Burt et al. 2018; Arnatkevičiūtė, Fulcher, and Fornito 2019)

These scores investigating patterns of brain expression are currently tested in another ongoing thesis project which hypothesize that large groups of genes with similar macroscopic brain functions and similar tolerance to LoF variants may be associated with the same effects on cognition, behavior and NDD risk. This project also aims to distinguish the brain expression pattern of genes modulating NDDs as well as general intelligence.

[37] As example, the first principal component (PC1) of spatial variation per gene scales the dominant spatial expression variation of genes across cortical brain areas. It results from a principal component analysis used to reduce the dimensions when investigating the expression pattern of genes across the adult human brain.

Table 8: Cerebral expression Modules

Module (Hub gene)	Anatomy	Ontology/Pathway
ModuleO1 (*GABRB3*)	Telecephalon	Synaptic transmission, regulation of synaptic plasticity
Module03 (*KCNAB2*)	Hippocampus, thalamus, pons, medulla	Neurotransmitter transport, axon part
Module04 (*GABARAPL1*)	Thalamocortical	Synaptic vesicle cycle, insulin receptor recycling
Module06 (*MEF2C*)	Neocortex, claustrum	Postsynaptic membrane, cell signaling
Module07(*NGEF*)	Striatum, neocortex, amygdala	Calcium signaling pathway, dendritic spine membrane
Module09 (*PGAP1*)	Hippocampus, amygdala, hypothalamus	Synaptic membrane
Module10 (*ADORA2A*)	Striatum	Dopamine receptor signaling
Module11 (*NTNG1*)	Dorsal thalamus	Cadherin signaling pathway, cholinergic synapse
Module12 (*SLC6A3*)	Substantia nigra, ventral tegmental area	Adrenaline, noradrenaline, and dopamine biosynthesis
Module14 (*TLE6*)	Hypothalamus	Neuropeptide signaling
Module15(*NEFH*)	Deep cerebellar nuclei, brainstem	Neuron projection, neurofilament
Module16 (*SLC47A1*)	Dentate gyrus	Protocadherin genes
Module17 (*CBLN3*)	Cerebellar cortex	Spinal cord development
Module19 (*VDAC2*)	Thalamus, cerebellar nuclei, brainstem	Vasculature development, mitochondrial
Module20 (*B3GAT1*)	White matter, neocortex, basal ganglia, ventral thalamus	Eukaryotic translation, ribosomal nucleolus
Module21 (*GBP4*)	Sensory-motor nuclei, choroid	Vasculature development
Module24 (*POGZ*)	Cerebellar cortex, dentate gyrus, white matter, basal ganglia	Chromatin organization
Module25 (*RGS10*)	Ependyma, white matter, substantia nigra	Immune system regulation, inflammatory response
Module26 (*MYCBP*)	Ependyma	Cilium organization, Axoneme
Module28 (*SERPINA6*)	Interbrain-hindbrain nuclei	G-protein coupled receptors, olfactory receptors
Module29 (*GAS5*)	White matter, substantia nigra, globus pallidus	Cytosolic ribosome, translation activity
Module30 (*VAMP3*)	White matter, ventral thalamus, globus pallidus	Myelination, neuron ensheathment
Module32 (*SLC25A18*)	Striatum, amygdala, substantia nigra	Glial cell differentiation, astrocyte differentiation

Each module is given with a representative hub gene, anatomic description, ontology and pathway associations. Adapted from Hawrylycz et al., (2015).

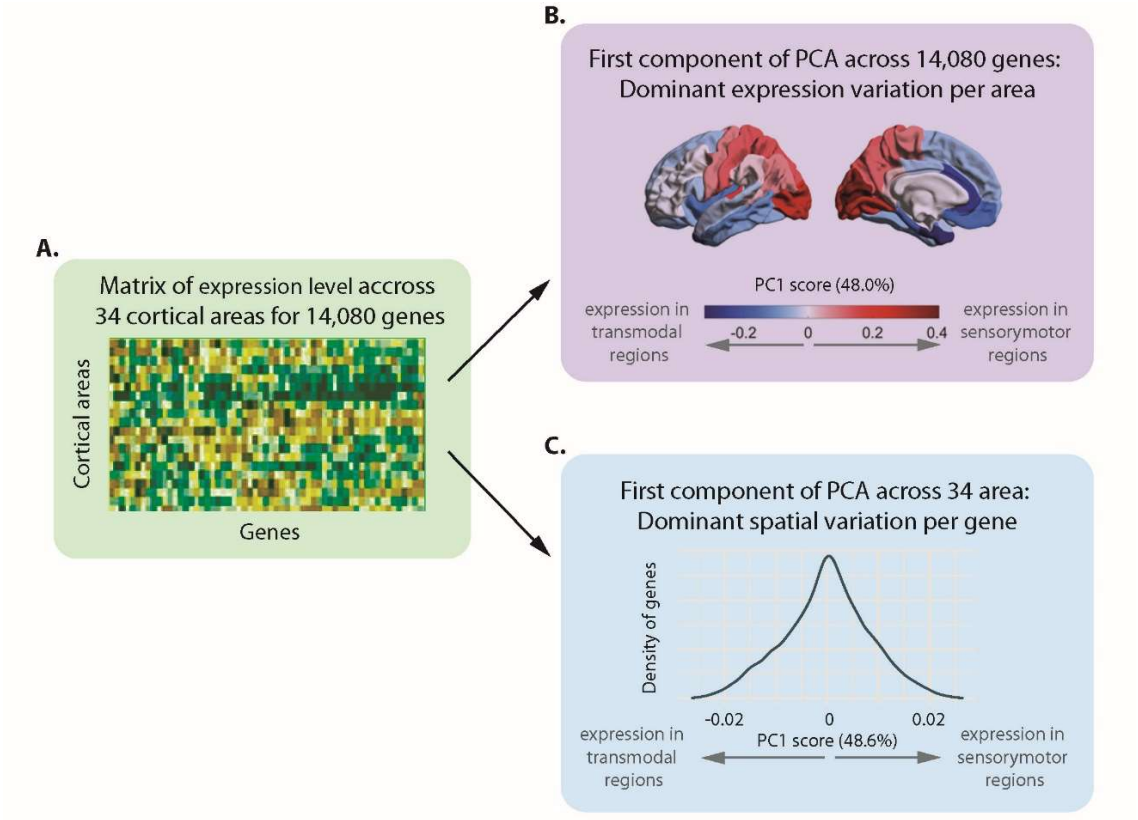


Figure 18: Cortical gradient of gene expression

A. Matrix representing the expression level of 14,080 genes across 34 brain cortical areas. This matrix is based on the data from the 6 human adult brains of the Allen human brain atlas project. (Arnatkevičiūtė et al., 2019) The cortical brain parcellation in 34 areas was defined using the Desikan brain atlas. (Desikan et al. 2006) Putting in relation the expression of sampled genes with defined cortical areas is done in two ways, using principal component analyses (PCA) to reduce the dimensionality. B. Studying the dominant expression variation pattern per brain area (PC1 explaining 48.0% of the expression variance), with a clear gradient from the sensory-motor areas (positive values in red) to the transmodal areas (negative values in blue). C. Studying the dominant spatial variation pattern per gene (PC1 explaining 48.6% of spatial variance), with a smooth transition along a gradient from sensory-motor areas (positive scores) to higher-order transmodal areas (negative scores). The latter score can be used to investigate the contribution of CNVs to NDD risk.

The paradox between the contribution of rare and common variants to ASD risk

Despite evidence that rare deleterious variants in the same genes are implicated across multiple neurodevelopmental and neuropsychiatric disorders, there has been considerable interest in identifying genes that, when mutated, confer predominant risk for ASD. (Satterstrom et al. 2020) These efforts have led to unconvincing results and the debate over the specificity of potential ASD-genes remains controversial. (Wray, Wijmenga, et al. 2018; Coe et al. 2019; Myers et al. 2020)

Indeed, rare variants which increase ASD risk also decrease IQ. (Bishop et al. 2017; Myers et al. 2020; Satterstrom et al. 2020) My results suggested that a portion of CNVs effect-sizes on ASD risk was mediated by their effect on IQ. On the other spectrum of the frequency and effect-size, several studies using linkage disequilibrium and PRS demonstrated that common variants increasing ASD risk also increase IQ (Plomin and von Stumm 2018; Savage et al. 2018; Grove et al. 2019; Yahya 2020) These paradoxical observations suggest that common variant associated with ASD risk may load on cognitive dimensions that are distinct from those affected by rare variants and may be more specific to ASD core symptoms.

Currently, no study has investigated the risk for ASD conferred by rare and common variants simultaneously while adjusting for their effects on general intelligence. Davies *et al.* recently evaluated the use of PRS for individual risk prediction in 22q11 deletion syndrome, and demonstrated that polygenic scores allowed risk stratification among individuals with this highly, but incompletely, penetrant genetic variants. (Davies et al. 2020) Therefore, we hypothesize that we can improve our predictive models by taking into account the effect of both rare and common variants measured by constraint score and PRS for IQ, respectively.

Clinical applications

Some of my models are currently implemented in an online prediction tool (https://cnvprediction.urca.ca/) to help in the interpretation of the effect-size of undocumented CNVs identified in the neurodevelopmental clinic. This tool translates information provided by gnomAD (https://gnomad.broadinstitute.org/), into effect-sizes on IQ, ASD risk and other traits conferred by the CNV.

This tool is designed to predict the effect-size of CNVs, not the symptoms of the individual who carries the CNV. It provides effect-sizes (and confidence intervals) on several traits, which allows clinician to have an idea of the developmental trajectory the patient may encounter. Recently, we demonstrated that the estimated effect-size of CNVs on cognitive ability are close to 80% accurate with empirical findings. (Huguet et al. 2021) For ASD risk, estimates provided by my models were overlapping with risk computed in previous studies. As an example, the model estimates for 16p11.2 and 22q11.2 deletions were similar to the previously published effect for NVIQ (losses of 25 points (Moreno-De-Luca et al. 2015) and 29 points (Vangkilde et al. 2016)), ASD risk (odds ratios of 11.8 (Malhotra and Sebat 2012) and 32.37 (Sanders et al. 2019)), and SRS score (gains of 44 points (Moreno-De-Luca et al. 2015) and 49 points (Vangkilde et al. 2016)).

If the effect-size of the CNV on traits is concordant with the severity of the patient, the clinician may conclude that the CNV contributes to most of the NDD diagnosis of the patient. However, if symptoms are discordant from the estimations, the clinician may conclude that additional factors should be investigated. Discordance may be defined when the estimated effect-size of the CNV is one standard deviation lower than the phenotype observed in the carrier compared with the population mean. For example, if a CNV with an effect-size of -10 IQ points is identified in a carrier with an IQ of 60 (-40 points compared with the population mean), the majority of the cognitive deficits are caused by additional factors.

Overall, the output of these models can help interpret CNVs in the clinic, but estimates should be interpreted with caution.

Conclusion

This project highlights the extreme polygenicity conferred by gene dosage to ASD and EOP risk, as well as the pleiotropic properties of these variants. It also delineates cognitive mechanisms that may explain in part the overrepresentation of CNVs in ASD. Large-scale studies simultaneously investigating the effect of genomic variants on categorical diagnoses and continuous dimensions are warranted. This study represents a new framework to study rare variants and can help in the interpretation of the effect-size of undocumented CNVs identified in the neurodevelopmental clinic. In the future, models combining rare and common variants will likely improve the predictive value of these methods. Precision medicine is coming...

V. References

Abrahams, Brett S., Dan E. Arking, Daniel B. Campbell, et al.
2013 SFARI Gene 2.0: A Community-Driven Knowledgebase for the Autism Spectrum Disorders (ASDs). Molecular Autism 4(1): 36.

Ahn, K., S. S. An, Y. Y. Shugart, and J. L. Rapoport
2016 Common Polygenic Variation and Risk for Childhood-Onset Schizophrenia. Molecular Psychiatry 21(1): 94–96.

Ahn, K., N. Gotay, T. M. Andersen, et al.
2014 High Rate of Disease-Related Copy Number Variations in Childhood Onset Schizophrenia. Molecular Psychiatry 19(5): 568–572.

Aken, Katrijn Van, Karen Caeyenberghs, Bouwien Smits-Engelsman, and Ann Swillen
2009 The Motor Profile of Primary School-Age Children with a 22q11.2 Deletion Syndrome (22q11.2DS) and an Age- and IQ-Matched Control Group. Child Neuropsychology 15(6). Routledge: 532–542.

Akingbuwa, Wonuola A., Anke R. Hammerschlag, Meike Bartels, and Christel M. Middeldorp
2021 Systematic Review: Molecular Studies of Common Genetic Variation in Child and Adolescent Psychiatric Disorders. Journal of the American Academy of Child & Adolescent Psychiatry. https://www.sciencedirect.com/science/article/pii/S0890856721002380, accessed September 16, 2021.

Allegrini, Andrea G., Rosa Cheesman, Kaili Rimfeld, et al.
2020 The p Factor: Genetic Analyses Support a General Dimension of Psychopathology in Childhood and Adolescence. Journal of Child Psychology and Psychiatry, and Allied Disciplines 61(1): 30–39.

Ambalavanan, Amirthagowri, Simon L. Girard, Kwangmi Ahn, et al.
2016 De Novo Variants in Sporadic Cases of Childhood Onset Schizophrenia. European Journal of Human Genetics 24(6): 944–948.

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

An, Joon-Yong, Kevin Lin, Lingxue Zhu, et al.
2018 Genome-Wide de Novo Risk Score Implicates Promoter Variation in Autism Spectrum Disorder. Science 362(6420): eaat6576.

Anney, Richard J. L., Stephan Ripke, Verneri Anttila, et al.
2017 Meta-Analysis of GWAS of over 16,000 Individuals with Autism Spectrum Disorder Highlights a Novel Locus at 10q24.32 and a Significant Overlap with Schizophrenia. Molecular Autism 8(1): 21.

Arango, Celso, Carmen Moreno, Salvador Martínez, et al.
2008 Longitudinal Brain Changes in Early-Onset Psychosis. Schizophrenia Bulletin 34(2): 341–353.

Arciniegas, David B.
2015 Psychosis. Continuum : Lifelong Learning in Neurology 21(3 Behavioral Neurology and Neuropsychiatry): 715–736.

Arnatkevičiūtė, Aurina, Ben D. Fulcher, and Alex Fornito
2019 A Practical Guide to Linking Brain-Wide Gene Expression and Neuroimaging Data. NeuroImage 189: 353–367.

Bai, Dan, Benjamin Hon Kei Yip, Gayle C. Windham, et al.
2019 Association of Genetic and Environmental Factors With Autism in a 5-Country Cohort. JAMA Psychiatry.

Barger, Brian D., Jonathan M. Campbell, and Jaimi D. McDonough
2013 Prevalence and Onset of Regression within Autism Spectrum Disorders: A Meta-Analytic Review. Journal of Autism and Developmental Disorders 43(4): 817–828.

Baumeister, D., R. Akhtar, S. Ciufolini, C. M. Pariante, and V. Mondelli
2016 Childhood Trauma and Adulthood Inflammation: A Meta-Analysis of Peripheral C-Reactive Protein, Interleukin-6 and Tumour Necrosis Factor-α. Molecular Psychiatry 21(5): 642–649.

Bedeschi, M.F., L. Colombo, F. Mari, et al.
2011 Unmasking of a Recessive SCARF2 Mutation by a 22q11.12 de Novo Deletion in a Patient with Van Den Ende-Gupta Syndrome. Molecular Syndromology 1(5): 239–245.

Bedford, Saashi A., Min Tae M. Park, Gabriel A. Devenyi, et al.
2020 Large-Scale Analyses of the Relationship between Sex, Age and Intelligence Quotient Heterogeneity and Cortical Morphometry in Autism Spectrum Disorder. Molecular Psychiatry 25(3). Nature Publishing Group: 614–628.

Bell, Carl C.
1994 DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. JAMA 272(10): 828–829.

Bennett, T. A., P. Szatmari, K. Georgiades, et al.
2014 Language Impairment and Early Social Competence in Preschoolers with Autism Spectrum Disorders: A Comparison of DSM-5 Profiles. Journal of Autism and Developmental Disorders 44(11): 2797–2808.

Bernier, Raphael, Kyle J. Steinman, Beau Reilly, et al.
2016 Clinical Phenotype of the Recurrent 1q21.1 Copy-Number Variant. Genetics in Medicine: Official Journal of the American College of Medical Genetics 18(4): 341–349.

Beveridge, N. J., E. Gardiner, A. P. Carroll, P. A. Tooney, and M. J. Cairns

2010 Schizophrenia Is Associated with an Increase in Cortical MicroRNA Biogenesis. Molecular Psychiatry 15(12). Nature Publishing Group: 1176–1189.

Bhat, Anjana N.
2020 Is Motor Impairment in Autism Spectrum Disorder (ASD) Distinct From Developmental Coordination Disorder (DCD)? A Report From the SPARK Study. Physical Therapy. https://academic.oup.com/ptj/advance-article/doi/10.1093/ptj/pzz190/5801997, accessed March 28, 2020.

Bhat, Anjana N., Rebecca J. Landa, and James C. (Cole) Galloway
2011 Current Perspectives on Motor Functioning in Infants, Children, and Adults With Autism Spectrum Disorders. Physical Therapy 91(7). Oxford Academic: 1116–1129.

Binet, A., and T. Simon
1905 Methodes Nouvelles Pour Le Diagnostique Du Niveau Intellectual Des Anormaux. Studies in Individual Differences, the Search for Intelligence II. Appleton-Century-Croft: 191–244.

Bishop, D. V. M.
2010 Overlaps Between Autism and Language Impairment: Phenomimicry or Shared Etiology? Behavior Genetics 40(5): 618–629.

Bishop, Somer L., Cristan Farmer, Vanessa Bal, et al.
2017 Identification of Developmental and Behavioral Markers Associated With Genetic Abnormalities in Autism Spectrum Disorder. The American Journal of Psychiatry 174(6): 576–585.

Bishop, Somer L., Audrey Thurm, Cristan Farmer, and Catherine Lord
2016 Autism Spectrum Disorder, Intellectual Disability, and Delayed Walking. Pediatrics 137(3): e20152959.

Blank, Rainer, Bouwien Smits-Engelsman, Helene Polatajko, and Peter Wilson
2012 European Academy for Childhood Disability (EACD): Recommendations on the Definition, Diagnosis and Intervention of Developmental Coordination Disorder (Long Version)*. Developmental Medicine & Child Neurology 54(1). John Wiley & Sons, Ltd: 54–93.

Bleuler, Eugen
1911 Dementia praecox, oder Gruppe der Schizophrenien. Leipzig: Deuticke.

Blizinsky, Katherine D., Blanca Diaz-Castro, Marc P. Forrest, et al.
2016 Reversal of Dendritic Phenotypes in 16p11.2 Microduplication Mouse Model Neurons by Pharmacological Targeting of a Network Hub. Proceedings of the National Academy of Sciences of the United States of America 113(30): 8520–8525.

Boat, Thomas F., Joel T. Wu, Committee to Evaluate the Supplemental Security Income Disability Program for Children with Mental Disorders, et al.
2015 Clinical Characteristics of Intellectual Disabilities. Mental Disorders and Disabilities Among Low-Income Children. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK332877/, accessed March 19, 2020.

Boeing, Leonie, Val Murray, Anthony Pelosi, et al.
2007 Adolescent-Onset Psychosis: Prevalence, Needs and Service Provision. The British Journal of Psychiatry 190(1). Cambridge University Press: 18–26.

Boomsma, Dorret, Andreas Busjahn, and Leena Peltonen
2002 Classical Twin Studies and Beyond. Nature Reviews. Genetics 3(11): 872–882.

Boot, Erik, Nancy J. Butcher, Thérèse A. M. J. van Amelsvoort, et al.
2015 Movement Disorders and Other Motor Abnormalities in Adults with 22q11.2 Deletion Syndrome. American Journal of Medical Genetics. Part A 167A(3): 639–645.

Boterberg, Sofie, Tony Charman, Peter B. Marschik, Sven Bölte, and Herbert Roeyers
2019 Regression in Autism Spectrum Disorder: A Critical Overview of Retrospective Findings and Recommendations for Future Research. Neuroscience and Biobehavioral Reviews 102: 24–55.

Botto, Lorenzo D., Kristin May, Paul M. Fernhoff, et al.
2003 A Population-Based Study of the 22q11.2 Deletion: Phenotype, Incidence, and Contribution to Major Birth Defects in the Population. Pediatrics 112(1). American Academy of Pediatrics: 101–107.

Bourgeron, T.
2007 The Possible Interplay of Synaptic and Clock Genes in Autism Spectrum Disorders. Cold Spring Harbor Symposia on Quantitative Biology 72: 645–654.

Boycott, Kym M., Megan R. Vanstone, Dennis E. Bulman, and Alex E. MacKenzie
2013 Rare-Disease Genetics in the Era of next-Generation Sequencing: Discovery to Translation. Nature Reviews. Genetics 14(10): 681–691.

Brainstorm Consortium, Verneri Anttila, Brendan Bulik-Sullivan, et al.
2018 Analysis of Shared Heritability in Common Disorders of the Brain. Science (New York, N.Y.) 360(6395): eaap8757.

Braun, Kim Van Naarden, Deborah Christensen, Nancy Doernberg, et al.
2015 Trends in the Prevalence of Autism Spectrum Disorder, Cerebral Palsy, Hearing Loss, Intellectual Disability, and Vision Impairment, Metropolitan Atlanta, 1991–2010. PLOS ONE 10(4): e0124120.

Brown, Alan S., and Paul H. Patterson
2011 Maternal Infection and Schizophrenia: Implications for Prevention. Schizophrenia Bulletin 37(2): 284–290.

Brownstein, Catherine A., Robin J. Kleiman, Elizabeth C. Engle, et al.
2016 Overlapping 16p13.11 Deletion and Gain of Copies Variations Associated with Childhood Onset Psychosis Include Genes with Mechanistic Implications for Autism Associated Pathways: Two Case Reports. American Journal of Medical Genetics. Part A 170A(5): 1165–1173.

Brownstein, Catherine A., Richard S. Smith, Lance H. Rodan, et al.
2021 RCL1 Copy Number Variants Are Associated with a Range of Neuropsychiatric Phenotypes. Molecular Psychiatry 26(5): 1706–1718.

Brunetti-Pierri, Nicola, Jonathan S. Berg, Fernando Scaglia, et al.
2008 Recurrent Reciprocal 1q21.1 Deletions and Duplications Associated with Microcephaly or Macrocephaly and Developmental and Behavioral Abnormalities. Nature Genetics 40(12): 1466–1471.

Bruni, Teryn P.
2014 Test Review: Social Responsiveness Scale–Second Edition (SRS-2). Journal of Psychoeducational Assessment 32(4): 365–369.

Brzustowicz, Linda M., and Anne S. Bassett
2012 MiRNA-Mediated Risk for Schizophrenia in 22q11.2 Deletion Syndrome. Frontiers in Genetics 3. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3521194/, accessed July 29, 2020.

Buja, Andreas, Natalia Volfovsky, Abba M. Krieger, et al.
2018 Damaging de Novo Mutations Diminish Motor Skills in Children on the Autism Spectrum. Proceedings of the National Academy of Sciences of the United States of America 115(8): E1859–E1866.

Buka, Stephen L., Tyrone D. Cannon, E. Fuller Torrey, Robert H. Yolken, and Collaborative Study Group on the Perinatal Origins of Severe Psychiatric Disorders
2008 Maternal Exposure to Herpes Simplex Virus and Risk of Psychosis among Adult Offspring. Biological Psychiatry 63(8): 809–815.

Bulik-Sullivan, Brendan, Hilary K. Finucane, Verneri Anttila, et al.
2015 An Atlas of Genetic Correlations across Human Diseases and Traits. Nature Genetics 47(11): 1236–1241.

Burks, Barbara Stoddard, Dortha Williams Jensen, and Lewis Madison Terman
1930 The Promise of Youth: Follow-up Studies of a Thousand Gifted Children. Stanford University Press.

Burt, Joshua B., Murat Demirtaş, William J. Eckner, et al.
2018 Hierarchy of Transcriptomic Specialization across Human Cortex Captured by Structural Neuroimaging Topography. Nature Neuroscience 21(9). Nature Publishing Group: 1251–1259.

Butcher, Nancy J., Eva W. C. Chow, Gregory Costain, et al.
2012 Functional Outcomes of Adults with 22q11.2 Deletion Syndrome. Genetics in Medicine 14(10): 836–843.

Caeyenberghs, Karen, Tom Taymans, Peter H. Wilson, et al.
2016 Neural Signature of Developmental Coordination Disorder in the Structural Connectome Independent of Comorbid Autism. Developmental Science 19(4): 599–612.

Campenhout, S, K Devriendt, Jeroen Breckpot, et al.
2012 Microduplication 22q11.2: A Description of the Clinical, Developmental and Behavioral Characteristics during Childhood. Genetic Counseling (Geneva, Switzerland) 23: 135–48.

Canetta, Sarah, Andre Sourander, Heljä-Marja Surcel, et al.
2014 Elevated Maternal C-Reactive Protein and Increased Risk of Schizophrenia in a National Birth Cohort. The American Journal of Psychiatry 171(9): 960–968.

Caprio, Cinzia, and Antonio Baldini
2014 P53 Suppression Partially Rescues the Mutant Phenotype in Mouse Models of DiGeorge Syndrome. Proceedings of the National Academy of Sciences of the United States of America 111(37): 13385–13390.

Caspi, Avshalom, Renate M. Houts, Daniel W. Belsky, et al.
2014 The p Factor: One General Psychopathology Factor in the Structure of Psychiatric Disorders? Clinical Psychological Science : A Journal of the Association for Psychological Science 2(2): 119–137.

Cauda, Franco, Tommaso Costa, Andrea Nani, et al.
2017 Are Schizophrenia, Autistic, and Obsessive Spectrum Disorders Dissociable on the Basis of Neuroimaging Morphological Findings?: A Voxel-Based Meta-Analysis. Autism Research 10(6): 1079–1095.

Chang, Jonathan, Sarah R. Gilman, Andrew H. Chiang, Stephan J. Sanders, and Dennis Vitkup
2015 Genotype to Phenotype Relationships in Autism Spectrum Disorders. Nature Neuroscience 18(2). Nature Publishing Group: 191–198.

de la Chapelle, Albert, Riitta Herva, Maila Koivisto, and Pertti Aula
1981 A Deletion in Chromosome 22 Can Cause Digeorge Syndrome. Human Genetics 57(3): 253–256.

Chapnik, Elik, Vered Sasson, Robert Blelloch, and Eran Hornstein
2012 Dgcr8 Controls Neural Crest Cells Survival in Cardiovascular Development. Developmental Biology 362(1): 50–56.

Chiang, Hsu-Min, Luke Y. Tsai, Ying Kuen Cheung, Alice Brown, and Huacheng Li
2014 A Meta-Analysis of Differences in IQ Profiles between Individuals with Asperger's Disorder and High-Functioning Autism. Journal of Autism and Developmental Disorders 44(7): 1577–1596.

Ching, Christopher R.K., Boris A. Gutman, Daqiang Sun, et al.
2020 Mapping Subcortical Brain Alterations in 22q11.2 Deletion Syndrome: Effects of Deletion Size and Convergence With Idiopathic Neuropsychiatric Illness. American Journal of Psychiatry 177(7). American Psychiatric Publishing: 589–600.

Christensen, Deborah L.
2018 Prevalence and Characteristics of Autism Spectrum Disorder Among Children Aged 8

Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2012. MMWR. Surveillance Summaries 65. https://www.cdc.gov/mmwr/volumes/65/ss/ss6513a1.htm, accessed January 28, 2021.

Cioffi, Sara, Stefania Martucciello, Filomena Gabriella Fulcoli, et al.
2014 Tbx1 Regulates Brain Vascularization. Human Molecular Genetics 23(1): 78–89.

Clark, Lee Anna, Bruce Cuthbert, Roberto Lewis-Fernández, William E. Narrow, and Geoffrey M. Reed
2017 Three Approaches to Understanding and Classifying Mental Disorder: ICD-11, DSM-5, and the National Institute of Mental Health's Research Domain Criteria (RDoC). Psychological Science in the Public Interest: A Journal of the American Psychological Society 18(2): 72–145.

Coe, Bradley P., Holly A. F. Stessman, Arvis Sulovari, et al.
2019 Neurodevelopmental Disease Genes Implicated by de Novo Mutation and Copy Number Variation Morbidity. Nature Genetics 51(1): 106.

Collins, Ryan L., Harrison Brand, Konrad J. Karczewski, et al.
2019 An Open Resource of Structural Variation for Medical and Population Genetics. BioRxiv: 578674.
2020 A Structural Variation Reference for Medical and Population Genetics. Nature 581(7809). Nature Publishing Group: 444–451.

Collins, Ryan L., Joseph T. Glessner, Eleonora Porcu, et al.
2021 A Cross-Disorder Dosage Sensitivity Map of the Human Genome. https://www.medrxiv.org/content/10.1101/2021.01.26.21250098v1, accessed October 21, 2021.

Colvert, Emma, Beata Tick, Fiona McEwen, et al.
2015 Heritability of Autism Spectrum Disorder in a UK Population-Based Twin Sample. JAMA Psychiatry 72(5): 415–423.

Constantino, John N.
2011 The Quantitative Nature of Autistic Social Impairment. Pediatric Research 69(8): 55–62.
2019 Early Behavioral Indices of Inherited Liability to Autism. Pediatric Research 85(2): 127–133.

Constantino, John N., Sandra A. Davis, Richard D. Todd, et al.
2003 Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. Journal of Autism and Developmental Disorders 33(4): 427–433.

Constantino, John N., and Christian P. Gruber
2012 Social Responsivness Scale-Second Edition (SRS-2). Torrance, CA: Western Psychological Services.

Croen, Lisa A., Ousseny Zerbo, Yinge Qian, et al.
2015 The Health Status of Adults on the Autism Spectrum: Autism. Sage UK: London, England,

SAGE PublicationsSage UK: London, England. https://journals.sagepub.com/doi/10.1177/1362361315577517, accessed March 7, 2020.

Cross-Disorder Group of the Psychiatric Genomics Consortium
2019 Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179(7): 1469-1482.e11.

van Daalen, Emma, Chantal Kemner, Nienke E. Verbeek, et al.
2011 Social Responsiveness Scale-Aided Analysis of the Clinical Impact of Copy Number Variations in Autism. Neurogenetics 12(4): 315–323.

Dale Philip S., Price Thomas S., Bishop Dorothy V. M., and Plomin Robert
2003 Outcomes of Early Language Delay. Journal of Speech, Language, and Hearing Research 46(3). American Speech-Language-Hearing Association: 544–560.

D'Angelo, Debra, Sébastien Lebon, Qixuan Chen, et al.
2016 Defining the Effect of the 16p11.2 Duplication on Cognition, Behavior, and Medical Comorbidities. JAMA Psychiatry 73(1): 20–30.

Davies, Robert W., Ania M. Fiksinski, Elemi J. Breetvelt, et al.
2020 Using Common Genetic Variation to Examine Phenotypic Expression and Risk Prediction in 22q11.2 Deletion Syndrome. Nature Medicine 26(12): 1912–1918.

De Rubeis, Silvia, Xin He, Arthur P. Goldberg, et al.
2014 Synaptic, Transcriptional and Chromatin Genes Disrupted in Autism. Nature 515(7526): 209–215.

DECIPHER v9.17: Mapping the Clinical Genome
N.d. https://decipher.sanger.ac.uk/disorders#syndromes/overview, accessed August 24, 2017.

Deciphering Developmental Disorders Study
2017 Prevalence and Architecture of de Novo Mutations in Developmental Disorders. Nature 542(7642): 433–438.

Delgado-Lobete, Laura, Sonia Pértega-Díaz, Sergio Santos-del-Riego, and Rebeca Montes-Montes
2020 Sensory Processing Patterns in Developmental Coordination Disorder, Attention Deficit Hyperactivity Disorder and Typical Development. Research in Developmental Disabilities 100: 103608.

Desikan, Rahul S., Florent Ségonne, Bruce Fischl, et al.
2006 An Automated Labeling System for Subdividing the Human Cerebral Cortex on MRI Scans into Gyral Based Regions of Interest. NeuroImage 31(3): 968–980.

Dewey, Deborah, Marja Cantell, and Susan G. Crawford
2007 Motor and Gestural Performance in Children with Autism Spectrum Disorders, Developmental Coordination Disorder, and/or Attention Deficit Hyperactivity Disorder. Journal of the International Neuropsychological Society: JINS 13(2): 246–256.

Di Martino, A., C.-G. Yan, Q. Li, et al.
2014 The Autism Brain Imaging Data Exchange: Towards a Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. Molecular Psychiatry 19(6). Nature Publishing Group: 659–667.

DiGuiseppi, Carolyn, Susan Hepburn, Jonathan M. Davis, et al.
2010 Screening for Autism Spectrum Disorders in Children with Down Syndrome: Population Prevalence and Screening Test Characteristics. Journal of Developmental and Behavioral Pediatrics: JDBP 31(3): 181–191.

Downs, Johnny M., Harry Dean, Suzannah Lechler, et al.
2019 Negative Symptoms in Early-Onset Psychosis and Their Association With Antipsychotic Treatment Failure. Schizophrenia Bulletin 45(1): 69–79.

Downs, Johnny M., Suzannah Lechler, Harry Dean, et al.
2017 The Association Between Comorbid Autism Spectrum Disorders and Antipsychotic Treatment Failure in Early-Onset Psychosis: A Historical Cohort Study Using Electronic Health Records. The Journal of Clinical Psychiatry 78(9). Physicians Postgraduate Press, Inc.: 0–0.

Duncan, Amie W, and Somer L Bishop
2015 Understanding the Gap between Cognitive Abilities and Daily Living Skills in Adolescents with Autism Spectrum Disorders with Average Intelligence. Autism 19(1). SAGE Publications Ltd: 64–72.

Durand, Christelle M., Catalina Betancur, Tobias M. Boeckers, et al.
2007 Mutations in the Gene Encoding the Synaptic Scaffolding Protein SHANK3 Are Associated with Autism Spectrum Disorders. Nature Genetics 39(1): 25–27.

Edelmann, L., R. K. Pandita, and B. E. Morrow
1999 Low-Copy Repeats Mediate the Common 3-Mb Deletion in Patients with Velo-Cardio-Facial Syndrome. American Journal of Human Genetics 64(4): 1076–1086.

Eicher, John D., and Jeffrey R. Gruen
2015 Language Impairment and Dyslexia Genes Influence Language Skills in Children With Autism Spectrum Disorders. Autism Research 8(2): 229–234.

El-Sayed, Mohamed, R. Grant Steen, Michele D. Poe, et al.
2010 Brain Volumes in Psychotic Youth with Schizophrenia and Mood Disorders. Journal of Psychiatry & Neuroscience: JPN 35(4): 229–236.

Emck, Claudia, Ruud Bosscher, Peter Beek, and Theo Doreleijers
2009 Gross Motor Performance and Self-Perceived Motor Competence in Children with Emotional, Behavioural, and Pervasive Developmental Disorders: A Review. Developmental Medicine and Child Neurology 51(7): 501–517.

Endicott, J., and R. L. Spitzer

1978 A Diagnostic Interview: The Schedule for Affective Disorders and Schizophrenia. Archives of General Psychiatry 35(7): 837–844.

van Erp, T. G. M., D. P. Hibar, J. M. Rasmussen, et al.
2016 Subcortical Brain Volume Abnormalities in 2028 Individuals with Schizophrenia and 2540 Healthy Controls via the ENIGMA Consortium. Molecular Psychiatry 21(4): 585.

Falconer, D. S.
1967 The Inheritance of Liability to Diseases with Variable Age of Onset, with Particular Reference to Diabetes Mellitus. Annals of Human Genetics 31(1): 1–20.

Fedorenko, Evelina, Angela Morgan, Elizabeth Murray, et al.
2016 A Highly Penetrant Form of Childhood Apraxia of Speech Due to Deletion of 16p11.2. European Journal of Human Genetics: EJHG 24(2): 302–306.

Feliciano, Pamela, Amy M. Daniels, LeeAnne Green Snyder, et al.
2018 SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. Neuron 97(3): 488–493.

Fernandez, Arnaud, Malgorzata Marta Drozd, Susanne Thümmler, et al.
2019 Childhood-Onset Schizophrenia: A Systematic Overview of Its Genetic Heterogeneity From Classical Studies to the Genomic Era. Frontiers in Genetics 10. Frontiers. https://www.frontiersin.org/articles/10.3389/fgene.2019.01137/full, accessed June 23, 2021.

Feuk, Lars, Andrew R. Carson, and Stephen W. Scherer
2006 Structural Variation in the Human Genome. Nature Reviews Genetics 7(2). Nature Publishing Group: 85–97.

Fineberg, Anna M., Lauren M. Ellman, Catherine A. Schaefer, et al.
2016 Fetal Exposure to Maternal Stress and Risk for Schizophrenia Spectrum Disorders among Offspring: Differential Influences of Fetal Sex. Psychiatry Research 236: 91–97.

First, Michael B., Janet B. W. Williams, Rhonda S. Karg, and Robert L. Spitzer
2015 Structured Clinical Interview for DSM-5 Disorders-Clinician Version (SCID-5-CV). Arlington, VA: American Psychiatric Association. https://www.appi.org/Products/Interviewing/Structured-Clinical-Interview-for-DSM-5-Disorders, accessed September 14, 2021.

Fischbach, Gerald D., and Catherine Lord
2010 The Simons Simplex Collection: A Resource for Identification of Autism Genetic Risk Factors. Neuron 68(2): 192–195.

Follert, Philipp, Harold Cremer, and Christophe Béclin
2014 MicroRNAs in Brain Development and Function: A Matter of Flexibility and Stability. Frontiers in Molecular Neuroscience 7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3916726/, accessed July 29, 2020.

Folstein, Susan, and Michael Rutter

1977 Infantile Autism: A Genetic Study of 21 Twin Pairs. Journal of Child Psychology and Psychiatry 18(4): 297–321.

Forsyth, Jennifer K., and Robert F. Asarnow
2020 Genetics of Childhood-Onset Schizophrenia: 2019 Update. Child and Adolescent Psychiatric Clinics of North America 29(1): 157–170.

Fountain, Christine, Alix S. Winter, and Peter S. Bearman
2012 Six Developmental Trajectories Characterize Children With Autism. Pediatrics 129(5): e1112–e1120.

Fournier, Kimberly A., Chris J. Hass, Sagar K. Naik, Neha Lodha, and James H. Cauraugh
2010 Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis. Journal of Autism and Developmental Disorders 40(10): 1227–1240.

Frazier, J. A., J. N. Giedd, S. D. Hamburger, et al.
1996 Brain Anatomic Magnetic Resonance Imaging in Childhood-Onset Schizophrenia. Archives of General Psychiatry 53(7): 617–624.

Frazier, Thomas W., Stelios Georgiades, Somer L. Bishop, and Antonio Y. Hardan
2014 Behavioral and Cognitive Characteristics of Females and Males With Autism in the Simons Simplex Collection. Journal of the American Academy of Child and Adolescent Psychiatry 53(3): 329-340.e3.

Fromer, Menachem, Andrew J. Pocklington, David H. Kavanagh, et al.
2014 De Novo Mutations in Schizophrenia Implicate Synaptic Networks. Nature 506(7487). Nature Publishing Group: 179–184.

Galton, Francis
1870 Hereditary Genius: An Inquiry Into Its Laws and Consequences. D. Appleton.

Gandal, Michael J., Jillian R. Haney, Neelroop N. Parikshak, et al.
2018 Shared Molecular Neuropathology across Major Psychiatric Disorders Parallels Polygenic Overlap. Science 359(6376). American Association for the Advancement of Science: 693–697.

Gaugler, Trent, Lambertus Klei, Stephan J. Sanders, et al.
2014 Most Genetic Risk for Autism Resides with Common Variation. Nature Genetics 46(8): 881–885.

Geller, Barbara, Betsy Zimerman, Marlene Williams, et al.
2001 Reliability of the Wahsington University in St. Louis Kiddie Schedule for Affective Disorders and Schizophrenia (WASH-U-KSADS) Mania and Rapid Cycling Sections. Journal of the American Academy of Child & Adolescent Psychiatry 40(4). US: Lippincott Williams & Wilkins: 450–455.

Girirajan, Santhosh, Zoran Brkanac, Bradley P. Coe, et al.
2011 Relative Burden of Large CNVs on a Range of Neurodevelopmental Phenotypes. PLoS Genetics 7(11): e1002334.

Girirajan, Santhosh, Megan Y. Dennis, Carl Baker, et al.
2013 Refinement and Discovery of New Hotspots of Copy-Number Variation Associated with Autism Spectrum Disorder. The American Journal of Human Genetics 92(2): 221–237.

Glaser, Bronwyn, Donna Mumme, Christine Blasey, et al.
2002 Language Skills in Children with Velocardiofacial Syndrome (Deletion 22q11.2). The Journal of Pediatrics 140: 753–8.

Golzio, Christelle, Jason Willer, Michael E. Talkowski, et al.
2012 KCTD13 Is a Major Driver of Mirrored Neuroanatomical Phenotypes of the 16p11.2 Copy Number Variant. Nature 485(7398): 363–367.

Goodkind, Madeleine, Simon B. Eickhoff, Desmond J. Oathes, et al.
2015 Identification of a Common Neurobiological Substrate for Mental Illness. JAMA Psychiatry 72(4): 305–315.

Green, Dido, Tony Charman, Andrew Pickles, et al.
2009 Impairment in Movement Skills of Children with Autistic Spectrum Disorders. Developmental Medicine and Child Neurology 51(4): 311–316.

Grove, Jakob, Stephan Ripke, Thomas D. Als, et al.
2019 Identification of Common Genetic Risk Variants for Autism Spectrum Disorder. Nature Genetics 51(3): 431.

Guna, Alina, Nancy J. Butcher, and Anne S. Bassett
2015 Comparative Mapping of the 22q11.2 Deletion Region and the Potential of Simple Model Organisms. Journal of Neurodevelopmental Disorders 7(1). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4487986/, accessed July 29, 2020.

Guo, Suqin, Jiewei Liu, Wenqiang Li, et al.
2021 Genome Wide Association Study Identifies Four Loci for Early Onset Schizophrenia. Translational Psychiatry 11(1): 1–11.

Gurholt, Tiril P., Vera Lonning, Stener Nerland, et al.
2020 Intracranial and Subcortical Volumes in Adolescents with Early-Onset Psychosis: A Multisite Mega-Analysis from the ENIGMA Consortium. Human Brain Mapping n/a(n/a). https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.25212, accessed June 14, 2021.

Gurillo, Pedro, Sameer Jauhar, Robin M. Murray, and James H. MacCabe
2015 Does Tobacco Use Cause Psychosis? Systematic Review and Meta-Analysis. The Lancet. Psychiatry 2(8): 718–725.

Hajebi, Ahmad, Homayoun Amini, Leila Kashani, and Vandad Sharifi
2018 Twelve-Month Course and Outcome of Methamphetamine-Induced Psychosis Compared with First Episode Primary Psychotic Disorders. Early Intervention in Psychiatry 12(5): 928–934.

Hawks, Zoë W., and John N. Constantino

2020 Neuropsychiatric “Comorbidity” as Causal Influence in Autism. Journal of the American Academy of Child & Adolescent Psychiatry 59(2). Elsevier: 229–235.

Hawrylycz, Michael, Jeremy A. Miller, Vilas Menon, et al.
2015 Canonical Genetic Signatures of the Adult Human Brain. Nature Neuroscience 18(12): 1832–1844.

Hibar, D. P., L. T. Westlye, T. G. M. van Erp, et al.
2016 Subcortical Volumetric Abnormalities in Bipolar Disorder. Molecular Psychiatry 21(12): 1710–1716.

Hilker, Rikke, Dorte Helenius, Birgitte Fagerlund, et al.
2018 Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register. Biological Psychiatry 83(6). Novel Mechanisms in Schizophrenia Pathophysiology: 492–498.

Hippolyte, Loyse, Anne M. Maillard, Borja Rodriguez-Herreros, et al.
2016 The Number of Genomic Copies at the 16p11.2 Locus Modulates Language, Verbal Memory, and Inhibition. Biological Psychiatry 80(2): 129–139.

Hoekstra, Rosa A., Francesca Happé, Simon Baron-Cohen, and Angelica Ronald
2010 Limited Genetic Covariance between Autistic Traits and Intelligence: Findings from a Longitudinal Twin Study. American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics 153B(5): 994–1007.

Holiga, Štefan, Joerg F. Hipp, Christopher H. Chatham, et al.
2019 Patients with Autism Spectrum Disorders Display Reproducible Functional Connectivity Alterations. Science Translational Medicine 11(481). American Association for the Advancement of Science. https://stm.sciencemag.org/content/11/481/eaat9223, accessed August 11, 2020.

Huguet, Guillaume, Catherine Schramm, Elise Douard, et al.
2018 Measuring and Estimating the Effect Sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry.
2021 Genome-Wide Analysis of Gene Dosage in 24,092 Individuals Estimates That 10,000 Genes Modulate Cognitive Ability. Molecular Psychiatry. Nature Publishing Group: 1–14.

Hyman, Susan L., Susan E. Levy, Scott M. Myers, and Section on Developmental and Behavioral Pediatrics Council on Children with Disabilities
2020 Executive Summary: Identification, Evaluation, and Management of Children With Autism Spectrum Disorder. Pediatrics 145(1). American Academy of Pediatrics. https://pediatrics.aappublications.org/content/145/1/e20193448, accessed September 10, 2021.

Insel, Thomas R., and Bruce N. Cuthbert
2009 Endophenotypes: Bridging Genomic Complexity and Disorder Heterogeneity. Biological Psychiatry 66(11): 988–989.

Iossifov, Ivan, Dan Levy, Jeremy Allen, et al.
2015 Low Load for Disruptive Mutations in Autism Genes and Their Biased Transmission. Proceedings of the National Academy of Sciences of the United States of America 112(41): E5600-5607.

Iossifov, Ivan, Brian J. O'Roak, Stephan J. Sanders, et al.
2014 The Contribution of *de Novo* Coding Mutations to Autism Spectrum Disorder. Nature 515(7526): 216–221.

Ivanov, D., G. Kirov, N. Norton, et al.
2003 Chromosome 22q11 Deletions, Velo-Cardio-Facial Syndrome and Early-Onset Psychosis: Molecular Genetic Study. The British Journal of Psychiatry 183(5). Cambridge University Press: 409–413.

Iyer, Janani, Mayanglambam Dhruba Singh, Matthew Jensen, et al.
2018 Pervasive Genetic Interactions Modulate Neurodevelopmental Defects of the Autism-Associated 16p11.2 Deletion in Drosophila Melanogaster. Nature Communications 9(1): 1–19.

Jacquemont, Sébastien, Bradley P. Coe, Micha Hersch, et al.
2014 A Higher Mutational Burden in Females Supports a "Female Protective Model" in Neurodevelopmental Disorders. The American Journal of Human Genetics 94(3): 415–425.

Jacquemont, Sébastien, Alexandre Reymond, Flore Zufferey, et al.
2011 Mirror Extreme BMI Phenotypes Associated with Gene Dosage at the Chromosome 16p11.2 Locus. Nature 478(7367): 97–102.

Jamain, Stéphane, Hélène Quach, Catalina Betancur, et al.
2003 Mutations of the X-Linked Genes Encoding Neuroligins NLGN3 and NLGN4 Are Associated with Autism. Nature Genetics 34(1): 27–29.

Jeste, Shafali S., and Daniel H. Geschwind
2014 Disentangling the Heterogeneity of Autism Spectrum Disorder through Genetic Findings. Nature Reviews. Neurology 10(2): 74–81.

Jiang, Yong-hui, Ryan K. C. Yuen, Xin Jin, et al.
2013 Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing. The American Journal of Human Genetics 93(2): 249–263.

Jønch, Aia Elise, Elise Douard, Clara Moreau, et al.
2019 Estimating the Effect Size of the 15Q11.2 BP1–BP2 Deletion and Its Contribution to Neurodevelopmental Symptoms: Recommendations for Practice. Journal of Medical Genetics 56(10). BMJ Publishing Group Ltd: 701–710.

Kallmann, F. J., and B. Roth
1956 Genetic Aspects of Preadolescent Schizophrenia. The American Journal of Psychiatry 112(8): 599–606.

Kanner, L.
1943 Autistic Disturbances of Affective Contact. Nervous Child 2: 217–250.

Kanner, Leo
1969 The Children Haven't Read Those Books: Reflections on Differential Diagnosis. Acta Paedopsychiatrica: International Journal of Child & Adolescent Psychiatry 36(1). Germany: Springer: 2–11.

Karczewski, Konrad J., Laurent C. Francioli, Grace Tiao, et al.
2020 The Mutational Constraint Spectrum Quantified from Variation in 141,456 Humans. Nature 581(7809). Nature Publishing Group: 434–443.

Karras, Heather C., Danita N. Morin, Kamaldeep Gill, Sara Izadi-Najafabadi, and Jill G. Zwicker
2019 Health-Related Quality of Life of Children with Developmental Coordination Disorder. Research in Developmental Disabilities 84. Developmental Coordination Disorder: From Knowledge into Practice: 85–95.

Kaufman, Joan, Boris Birmaher, David Brent, Uma Rao, and Neal Ryan
1996 Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version (K-SADS-PL). Journal of the American Academy of Child & Adolescent Psychiatry.

Kaufmann, Walter E., Sharon A. Kidd, Howard F. Andrews, et al.
2017 Autism Spectrum Disorder in Fragile X Syndrome: Cooccurring Conditions and Current Treatment. Pediatrics 139(Supplement 3). American Academy of Pediatrics: S194–S206.

Kelley, Mary E., Claire Ramsay Wan, Beth Broussard, et al.
2016 Marijuana Use in the Immediate 5-Year Premorbid Period Is Associated with Increased Risk of Onset of Schizophrenia and Related Psychotic Disorders. Schizophrenia Research 171(1–3): 62–67.

Kendall, Kimberley M., Matthew Bracher-Smith, Harry Fitzpatrick, et al.
2019 Cognitive Performance and Functional Outcomes of Carriers of Pathogenic Copy Number Variants: Analysis of the UK Biobank. The British Journal of Psychiatry 214(5). Cambridge University Press: 297–304.

Kendall, Kimberley M., Elliott Rees, Valentina Escott-Price, et al.
2017 Cognitive Performance Among Carriers of Pathogenic Copy Number Variants: Analysis of 152,000 UK Biobank Subjects. Biological Psychiatry 82(2): 103–110.

Kim, So Hyun, Vanessa H. Bal, and Catherine Lord
2018 Longitudinal Follow-up of Academic Achievement in Children with Autism from Age 2 to 18. Journal of Child Psychology and Psychiatry 59(3): 258–267.

Kirov, G., A. J. Pocklington, P. Holmans, et al.
2012 De Novo CNV Analysis Implicates Specific Abnormalities of Postsynaptic Signalling Complexes in the Pathogenesis of Schizophrenia. Molecular Psychiatry 17(2). Nature Publishing Group: 142–153.

Knoers, Nine V. A. M., and Kirsten Y. Renkema
2019 The Genomic Landscape of CAKUT; You Gain Some, You Lose Some. Kidney International 96(2): 267–269.

Kopp, Svenny, Eva Beckung, and Christopher Gillberg
2010 Developmental Coordination Disorder and Other Motor Control Problems in Girls with Autism Spectrum Disorder and/or Attention-Deficit/Hyperactivity Disorder. Research in Developmental Disabilities 31(2): 350–361.

Koscielny, Gautier, Gagarine Yaikhom, Vivek Iyer, et al.
2014 The International Mouse Phenotyping Consortium Web Portal, a Unified Point of Access for Knockout Mice and Related Phenotyping Data. Nucleic Acids Research 42(Database issue): D802–D809.

Krumm, Niklas, Tychele N. Turner, Carl Baker, et al.
2015 Excess of Rare, Inherited Truncating Mutations in Autism. Nature Genetics 47(6): 582–588.

Larsen, Eric, Idan Menashe, Mark N. Ziats, et al.
2016 A Systematic Variant Annotation Approach for Ranking Genes Associated with Autism Spectrum Disorders. Molecular Autism 7: 44.

Lau, Way K. W., Mei-Kei Leung, and Benson W. M. Lau
2019 Resting-State Abnormalities in Autism Spectrum Disorders: A Meta-Analysis. Scientific Reports 9(1). Nature Publishing Group: 3892.

Law, J., J. Boyle, F. Harris, A. Harkness, and C. Nye
2000 Prevalence and Natural History of Primary Speech and Language Delay: Findings from a Systematic Review of the Literature. International Journal of Language & Communication Disorders 35(2): 165–188.

Le Couteur, Ann, Michael Rutter, Catherine Lord, et al.
1989 Autism Diagnostic Interview: A Standardized Investigator-Based Instrument. Journal of Autism and Developmental Disorders 19(3): 363–387.

Ledbetter, D. H., V. M. Riccardi, S. D. Airhart, et al.
1981 Deletions of Chromosome 15 as a Cause of the Prader-Willi Syndrome. The New England Journal of Medicine 304(6): 325–329.

Lee, Phil H., Verneri Anttila, Hyejung Won, et al.
2019 Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell 179(7): 1469-1482.e11.

Lee, Phil H., Yen-Chen A. Feng, and Jordan W. Smoller
2021 Pleiotropy and Cross-Disorder Genetics Among Psychiatric Disorders. Biological Psychiatry 89(1). Elsevier: 20–31.

Lejeune, J., R. Turpin, and M. Gautier
1959 [Chromosomic diagnosis of mongolism]. Archives Francaises De Pediatrie 16: 962–963.

Lek, Monkol, Konrad J. Karczewski, Eric V. Minikel, et al.
2016 Analysis of Protein-Coding Genetic Variation in 60,706 Humans. Nature 536(7616): 285–291.

Li, Dongyun, Hans-Otto Karnath, and Xiu Xu
2017 Candidate Biomarkers in Children with Autism Spectrum Disorder: A Review of MRI Studies. Neuroscience Bulletin 33(2): 219–237.

Lin, Amy, Christopher R. K. Ching, Ariana Vajdi, et al.
2017 Mapping 22q11.2 Gene Dosage Effects on Brain Morphometry. Journal of Neuroscience 37(26). Society for Neuroscience: 6183–6199.

Lindsay, E. A., F. Vitelli, H. Su, et al.
2001 Tbx1 Haploinsufficieny in the DiGeorge Syndrome Region Causes Aortic Arch Defects in Mice. Nature 410(6824): 97–101.

Lingam, Raghu, Jean Golding, Marian J. Jongmans, et al.
2010 The Association Between Developmental Coordination Disorder and Other Developmental Traits. Pediatrics 126(5). American Academy of Pediatrics: e1109–e1118.

Loomes, Rachel, Laura Hull, and William Polmear Locke Mandy
2017 What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. Journal of the American Academy of Child and Adolescent Psychiatry 56(6): 466–474.

Lord, Catherine, Susan Risi, Linda Lambrecht, et al.
2000 The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. Journal of Autism and Developmental Disorders 30(3): 205–223.

Lord, Catherine, Michael Rutter, and Ann Le Couteur
1994 Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. Journal of Autism and Developmental Disorders 24(5): 659–685.

Lord, Catherine, Michael Rutter, Susan Goode, et al.
1989 Austism Diagnostic Observation Schedule: A Standardized Observation of Communicative and Social Behavior. Journal of Autism and Developmental Disorders 19(2): 185–212.

Lord, Catherine, Michael L. Rutter, Pamela C. DiLavore, et al.
2012 Autism Diagnostic Observation Schedule–2nd Edition (ADOS-2). Los Angeles, CA: Western Psychological Corporation.

Lubs, H A
1969 A Marker X Chromosome. American Journal of Human Genetics 21(3): 231–244.

Lubs, Herbert A., Roger E. Stevenson, and Charles E. Schwartz
2012 Fragile X and X-Linked Intellectual Disability: Four Decades of Discovery. American Journal of Human Genetics 90(4): 579–590.

Lyall, Kristen, Lisa Croen, Julie Daniels, et al.
2017 The Changing Epidemiology of Autism Spectrum Disorders. Annual Review of Public Health 38: 81–102.

MacKenzie, Lynn E., Sabina Abidi, Helen L. Fisher, et al.
2016 Stimulant Medication and Psychotic Symptoms in Offspring of Parents With Mental Illness. Pediatrics 137(1). American Academy of Pediatrics. https://pediatrics.aappublications.org/content/137/1/e20152486, accessed September 14, 2021.

Magenis, R. E., M. G. Brown, D. A. Lacy, S. Budden, and S. LaFranchi
1987 Is Angelman Syndrome an Alternate Result of Del(15)(Q11q13)? American Journal of Medical Genetics 28(4): 829–838.

Maillard, A. M., A. Ruef, F. Pizzagalli, et al.
2015 The 16p11.2 Locus Modulates Brain Structures Common to Autism, Schizophrenia and Obesity. Molecular Psychiatry 20(1): 140–147.

Malaspina, D., C. Corcoran, KR Kleinhaus, et al.
2008 Acute Maternal Stress in Pregnancy and Schizophrenia in Offspring: A Cohort Prospective Study. BMC Psychiatry 8(1): 71.

Malhotra, Dheeraj, and Jonathan Sebat
2012 CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell 148(6): 1223–1241.

Marconi, Arianna, Marta Di Forti, Cathryn M. Lewis, Robin M. Murray, and Evangelos Vassos
2016 Meta-Analysis of the Association Between the Level of Cannabis Use and Risk of Psychosis. Schizophrenia Bulletin 42(5): 1262–1269.

Marshall, Christian R., Daniel P. Howrigan, Daniele Merico, et al.
2017 Contribution of Copy Number Variants to Schizophrenia from a Genome-Wide Study of 41,321 Subjects. Nature Genetics 49(1): 27–35.

Martin-Brevet, Sandra, Borja Rodríguez-Herreros, Jared A. Nielsen, et al.
2018 Quantifying the Effects of 16p11.2 Copy Number Variants on Brain Structure: A Multisite Genetic-First Study. Biological Psychiatry 84(4). Genetic and Environmental Factors Influencing Developmental Cognitive Impairments: 253–264.

Matson, Michael L., Johnny L. Matson, and Jennifer S. Beighley

2011 Comorbidity of Physical and Motor Problems in Children with Autism. Research in Developmental Disabilities 32(6): 2304–2308.

Mattiaccio, Leah M., Ioana L. Coman, Carlie A. Thompson, et al.
2018 Frontal Dysconnectivity in 22q11.2 Deletion Syndrome: An Atlas-Based Functional Connectivity Analysis. Behavioral and Brain Functions 14(1): 2.

Maulik, Pallab K., Maya N. Mascarenhas, Colin D. Mathers, Tarun Dua, and Shekhar Saxena
2011 Prevalence of Intellectual Disability: A Meta-Analysis of Population-Based Studies. Research in Developmental Disabilities 32(2): 419–436.

Mavros, Chrystal F., Catherine A. Brownstein, Roshni Thyagrajan, et al.
2018 De Novo Variant of TRRAP in a Patient with Very Early Onset Psychosis in the Context of Non-Verbal Learning Disability and Obsessive-Compulsive Disorder: A Case Report. BMC Medical Genetics 19(1): 197.

McDonald-McGinn, Donna M., Kathleen E. Sullivan, Bruno Marino, et al.
2015 22q11.2 Deletion Syndrome. Nature Reviews. Disease Primers 1: 15071.

McGrath, John J., Rosa Alati, Alexandra Clavarino, et al.
2016 Age at First Tobacco Use and Risk of Subsequent Psychosis-Related Outcomes: A Birth Cohort Study. The Australian and New Zealand Journal of Psychiatry 50(6): 577–583.

McGuire, Dana Olzenak, Lin H. Tian, Marshalyn Yeargin-Allsopp, Nicole F. Dowling, and Deborah L. Christensen
2019 Prevalence of Cerebral Palsy, Intellectual Disability, Hearing Loss, and Blindness, National Health Interview Survey, 2009–2016. Disability and Health Journal 12(3): 443–451.

McKetin, Rebecca, Jonathon Gardner, Amanda L. Baker, et al.
2016 Correlates of Transient versus Persistent Psychotic Symptoms among Dependent Methamphetamine Users. Psychiatry Research 238: 166–171.

Mefford, Heather C., Andrew J. Sharp, Carl Baker, et al.
2008 Recurrent Rearrangements of Chromosome 1q21.1 and Variable Pediatric Phenotypes. New England Journal of Medicine 359(16): 1685–1699.

Merico, Daniele, Gregory Costain, Nancy J. Butcher, et al.
2014 MicroRNA Dysregulation, Gene Networks, and Risk for Schizophrenia in 22q11.2 Deletion Syndrome. Frontiers in Neurology 5: 238.

Merscher, S., B. Funke, J. A. Epstein, et al.
2001 TBX1 Is Responsible for Cardiovascular Defects in Velo-Cardio-Facial/DiGeorge Syndrome. Cell 104(4): 619–629.

Miller, David T., Margaret P. Adam, Swaroop Aradhya, et al.
2010 Consensus Statement: Chromosomal Microarray Is a First-Tier Clinical Diagnostic Test for Individuals with Developmental Disabilities or Congenital Anomalies. The American Journal of Human Genetics 86(5): 749–764.

Mills, Kathryn L., Anne-Lise Goddings, Megan M. Herting, et al.
2016 Structural Brain Development between Childhood and Adulthood: Convergence across Four Longitudinal Samples. NeuroImage 141: 273–281.

Ming, Xue, Michael Brimacombe, and George C. Wagner
2007 Prevalence of Motor Impairment in Autism Spectrum Disorders. Brain and Development 29(9): 565–570.

Mistry, Sumit, Judith R. Harrison, Daniel J. Smith, Valentina Escott-Price, and Stanley Zammit
2018 The Use of Polygenic Risk Scores to Identify Phenotypes Associated with Genetic Risk of Schizophrenia: Systematic Review. Schizophrenia Research 197: 2–8.

Miyahara, Motohide
2013 Meta Review of Systematic and Meta Analytic Reviews on Movement Differences, Effect of Movement Based Interventions, and the Underlying Neural Mechanisms in Autism Spectrum Disorder. Frontiers in Integrative Neuroscience 7. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3607787/, accessed March 25, 2020.

Modabbernia, Amirhossein, Eva Velthorst, and Abraham Reichenberg
2017 Environmental Risk Factors for Autism: An Evidence-Based Review of Systematic Reviews and Meta-Analyses. Molecular Autism 8: 13.

Moreau, Clara A., Sebastian G. W. Urchs, Kumar Kuldeep, et al.
2020 Mutations Associated with Neuropsychiatric Conditions Delineate Functional Brain Connectivity Dimensions Contributing to Autism and Schizophrenia. Nature Communications 11(1). Nature Publishing Group: 5272.

Moreno-De-Luca, Andres, David W. Evans, K. B. Boomer, et al.
2015 The Role of Parental Cognitive, Behavioral, and Motor Profiles in Clinical Variability in Individuals with Chromosome 16p11.2 Deletions. JAMA Psychiatry 72(2): 119–126.

Moreno-De-Luca, Daniel, Jennifer G. Mulle, Erin B. Kaminsky, et al.
2010 Deletion 17q12 Is a Recurrent Copy Number Variant That Confers High Risk of Autism and Schizophrenia. The American Journal of Human Genetics 87(5): 618–630.

Moreno-De-Luca, Daniel, Michael E. Ross, and David A. Ross
2018 Leveraging the Power of Genetics to Bring Precision Medicine to Psychiatry: Too Little of a Good Thing? Biological Psychiatry 83(8): e45–e46.

Morris, Tim T., Neil M. Davies, Gibran Hemani, and George Davey Smith
2020 Population Phenomena Inflate Genetic Associations of Complex Social Traits. Science Advances 6(16). American Association for the Advancement of Science: eaay0328.

Morrow, Bernice E., Donna M. McDonald-McGinn, Beverly S. Emanuel, Joris R. Vermeesch, and Peter J. Scambler
2018 Molecular Genetics of 22q11.2 Deletion Syndrome. American Journal of Medical Genetics Part A 176(10): 2070–2081.

Moss, Jo, Caroline Richards, Lisa Nelson, and Chris Oliver
2013 Prevalence of Autism Spectrum Disorder Symptomatology and Related Behavioural Characteristics in Individuals with Down Syndrome. Autism: The International Journal of Research and Practice 17(4): 390–404.

Myers, Scott M., Thomas D. Challman, Raphael Bernier, et al.
2020 Insufficient Evidence for "Autism-Specific" Genes. The American Journal of Human Genetics 106(5). Elsevier: 587–595.

Neumann, Alexander, Irene Pappa, Benjamin B. Lahey, et al.
2016 Single Nucleotide Polymorphism Heritability of a General Psychopathology Factor in Children. Journal of the American Academy of Child & Adolescent Psychiatry 55(12). Elsevier: 1038-1045.e4.

Niemi, Mari E. K., Hilary C. Martin, Daniel L. Rice, et al.
2018 Common Genetic Variants Contribute to Risk of Rare Severe Neurodevelopmental Disorders. Nature 562(7726): 268–271.

Norbury, Courtenay Frazier, Debbie Gooch, Charlotte Wray, et al.
2016 The Impact of Nonverbal Ability on Prevalence and Clinical Presentation of Language Disorder: Evidence from a Population Study. Journal of Child Psychology and Psychiatry, and Allied Disciplines 57(11): 1247–1257.

Nudel, Ron, Camilla A. J. Christiani, Jessica Ohland, et al.
2020 Language Deficits in Specific Language Impairment, Attention Deficit/Hyperactivity Disorder, and Autism Spectrum Disorder: An Analysis of Polygenic Risk. Autism Research 13(3): 369–381.

O'Roak, Brian J., Laura Vives, Santhosh Girirajan, et al.
2012 Sporadic Autism Exomes Reveal a Highly Interconnected Protein Network of de Novo Mutations. Nature 485(7397): 246–250.

Oskarsdottir, S, M Vujic, and A Fasth
2004 Incidence and Prevalence of the 22q11 Deletion Syndrome: A Population-Based Study in Western Sweden. Archives of Disease in Childhood 89(2): 148–151.

Pagnozzi, Alex M., Eugenia Conti, Sara Calderoni, Jurgen Fripp, and Stephen E. Rose
2018 A Systematic Review of Structural MRI Biomarkers in Autism Spectrum Disorder: A Machine Learning Perspective. International Journal of Developmental Neuroscience: The Official Journal of the International Society for Developmental Neuroscience 71: 68–82.

Pain, Oliver, Frank Dudbridge, Alastair G. Cardno, et al.
2018 Genome-wide Analysis of Adolescent Psychotic-like Experiences Shows Genetic Overlap with Psychiatric Disorders. American Journal of Medical Genetics 177(4): 416–425.

Pan, Chien-Yu, Chia-Liang Tsai, and Chia-Hua Chu
2009 Fundamental Movement Skills in Children Diagnosed with Autism Spectrum Disorders

and Attention Deficit Hyperactivity Disorder. Journal of Autism and Developmental Disorders 39(12): 1694–1705.

Parikshak, Neelroop N., Rui Luo, Alice Zhang, et al.
2013 Integrative Functional Genomic Analyses Implicate Specific Molecular Pathways and Circuits in Autism. Cell 155(5): 1008–1021.

Paylor, Richard, Beate Glaser, Annalisa Mupo, et al.
2006 Tbx1 Haploinsufficiency Is Linked to Behavioral Disorders in Mice and Humans: Implications for 22q11 Deletion Syndrome. Proceedings of the National Academy of Sciences of the United States of America 103(20): 7729–7734.

Pearson, Niamh, Tony Charman, Francesca Happé, Patrick F. Bolton, and Fiona S. McEwen
2018 Regression in Autism Spectrum Disorder: Reconciling Findings from Retrospective and Prospective Research. Autism Research: Official Journal of the International Society for Autism Research 11(12): 1602–1620.

Perkins, Diana O., Loes Olde Loohuis, Jenna Barbee, et al.
2020 Polygenic Risk Score Contribution to Psychosis Prediction in a Target Population of Persons at Clinical High Risk. The American Journal of Psychiatry 177(2): 155–163.

Petri, Rebecca, Josephine Malmevik, Liana Fasching, Malin Åkerblom, and Johan Jakobsson
2014 MiRNAs in Brain Development. Experimental Cell Research 321(1): 84–89.

Petruzzelli, Maria Giuseppina, Lucia Margari, Andrea Bosco, et al.
2018 Early Onset First Episode Psychosis: Dimensional Structure of Symptoms, Clinical Subtypes and Related Neurodevelopmental Markers. European Child & Adolescent Psychiatry 27(2): 171–179.

Phelan, Katy, R. Curtis Rogers, and Luigi Boccuto
1993 Phelan-McDermid Syndrome. *In* GeneReviews®. Margaret P. Adam, Holly H. Ardinger, Roberta A. Pagon, et al., eds. Seattle (WA): University of Washington, Seattle. http://www.ncbi.nlm.nih.gov/books/NBK1198/, accessed August 5, 2020.

Pidsley, Ruth, and Jonathan Mill
2011 Epigenetic Studies of Psychosis: Current Findings, Methodological Approaches, and Implications for Postmortem Research. Biological Psychiatry 69(2). Postmortem Studies of Psychosis: Status, Opportunities, and Challenges: 146–156.

Pieretti, M., F. P. Zhang, Y. H. Fu, et al.
1991 Absence of Expression of the FMR-1 Gene in Fragile X Syndrome. Cell 66(4): 817–822.

Pinborough-Zimmerman Judith, Satterfield Robert, Miller Judith, et al.
2007 Communication Disorders: Prevalence and Comorbid Intellectual Disability, Autism, and Emotional/Behavioral Disorders. American Journal of Speech-Language Pathology 16(4). American Speech-Language-Hearing Association: 359–367.

Plomin, Robert, and Sophie von Stumm

2018 The New Genetics of Intelligence. Nature Reviews Genetics 19(3). Nature Publishing Group: 148–159.

Postema, Merel C., Daan van Rooij, Evdokia Anagnostou, et al.
2019 Altered Structural Brain Asymmetry in Autism Spectrum Disorder in a Study of 54 Datasets. Nature Communications 10(1). Nature Publishing Group: 4958.

Postorino, Valentina, Laura Maria Fatta, Veronica Sanges, et al.
2016 Intellectual Disability in Autism Spectrum Disorder: Investigation of Prevalence in an Italian Sample of Children and Adolescents. Research in Developmental Disabilities 48: 193–201.

Printzlau, Frida, Jeanne Wolstencroft, and David H. Skuse
2017 Cognitive, Behavioral, and Neural Consequences of Sex Chromosome Aneuploidy. Journal of Neuroscience Research 95(1–2): 311–319.

Pucilowska, Joanna, Joseph Vithayathil, Emmanuel J. Tavares, et al.
2015 The 16p11.2 Deletion Mouse Model of Autism Exhibits Altered Cortical Progenitor Proliferation and Brain Cytoarchitecture Linked to the ERK MAPK Pathway. Journal of Neuroscience 35(7): 3190–3200.

Puig-Antich, J, and W. J. Chambers
1978 Schedule for Affective Disorders and Schizophrenia for School-Age Children (Present Episode Version)(K-SADS-P). Biometrics Research.

Purcell, Shaun M., Jennifer L. Moran, Menachem Fromer, et al.
2014 A Polygenic Burden of Rare Disruptive Mutations in Schizophrenia. Nature 506(7487). Nature Publishing Group: 185–190.

Rapin, Isabelle
1996 Preschool Children with Inadequate Communication: Developmental Language Disorder, Autism, Low IQ. Clinics in Developmental Medicine, No. 139. Cambridge University Press, 40 West 20th Street, New York, NY 1011-4211 ($69.

Rapoport, J. L., J. N. Giedd, and N. Gogtay
2012 Neurodevelopmental Model of Schizophrenia: Update 2012. Molecular Psychiatry 17(12): 1228–1238.

Rattaz, C., C. Michelon, and A. Baghdadli
2015 Symptom Severity as a Risk Factor for Self-Injurious Behaviours in Adolescents with Autism Spectrum Disorders. Journal of Intellectual Disability Research 59(8): 730–741.

Rees, E., G. Kirov, A. Sanders, et al.
2014 Evidence That Duplications of 22q11.2 Protect against Schizophrenia. Molecular Psychiatry 19(1). Nature Publishing Group: 37–40.

Reindal, Lise, Terje Nærland, Bernhard Weidle, et al.
2019 Age of First Walking and Associations with Symptom Severity in Children with Suspected

or Diagnosed Autism Spectrum Disorder. Journal of Autism and Developmental Disorders. https://doi.org/10.1007/s10803-019-04112-y, accessed March 28, 2020.

Reinthaler, Eva M., Dennis Lal, Sebastien Lebon, et al.
2014 16p11.2 600 Kb Duplications Confer Risk for Typical and Atypical Rolandic Epilepsy. Human Molecular Genetics 23(22): 6069–6080.

Rescorla, Leslie, and Ellen Schwartz
1990 Outcome of Toddlers with Specific Expressive Language Delay. Applied Psycholinguistics 11(4). Cambridge University Press: 393–407.

Richter, Melanie, Nadeem Murtaza, Robin Scharrenberg, et al.
2019 Altered TAOK2 Activity Causes Autism-Related Neurodevelopmental and Cognitive Abnormalities through RhoA Signaling. Molecular Psychiatry 24(9): 1329–1350.

Rietz, Ebba Du, Erik Pettersson, Isabell Brikell, et al.
2021 Overlap between Attention-Deficit Hyperactivity Disorder and Neurodevelopmental, Externalising and Internalising Disorders: Separating Unique from General Psychopathology Effects. The British Journal of Psychiatry 218(1). Cambridge University Press: 35–42.

Rijsdijk, Frühling V., Irving I. Gottesman, Peter McGuffin, and Alastair G. Cardno
2011 Heritability Estimates for Psychotic Symptom Dimensions in Twins with Psychotic Disorders. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics 156(1): 89–98.

Ripke, Stephan, Benjamin M. Neale, Aiden Corvin, et al.
2014 Biological Insights from 108 Schizophrenia-Associated Genetic Loci. Nature 511(7510). Nature Publishing Group: 421–427.

Robinson, Elise B., Kaitlin E. Samocha, Jack A. Kosmicki, et al.
2014 Autism Spectrum Disorder Severity Reflects the Average Contribution of de Novo and Familial Influences. Proceedings of the National Academy of Sciences of the United States of America 111(42): 15161–15165.

Robinson, W. P., J. Waslynka, F. Bernasconi, et al.
1996 Delineation of 7q11.2 Deletions Associated with Williams-Beuren Syndrome and Mapping of a Repetitive Sequence to within and to Either Side of the Common Deletion. Genomics 34(1): 17–23.

van Rooij, Daan, Evdokia Anagnostou, Celso Arango, et al.
2018 Cortical and Subcortical Brain Morphometry Differences Between Patients With Autism Spectrum Disorder and Healthy Individuals Across the Lifespan: Results From the ENIGMA ASD Working Group. The American Journal of Psychiatry 175(4): 359–369.

Rosenfeld, Jill A., Justine Coppinger, Bassem A. Bejjani, et al.
2010 Speech Delays and Behavioral Problems Are the Predominant Features in Individuals with Developmental Delays and 16p11.2 Microdeletions and Microduplications. Journal of Neurodevelopmental Disorders 2(1): 26–38.

Rudolph, Johanna M., and Laurence B. Leonard
2016 Early Language Milestones and Specific Language Impairment. Journal of Early Intervention 38(1). SAGE Publications Inc: 41–58.

Rutter, Michael, Ann Le Couteur, and Catherine Lord
2003 Autism Diagnostic Interview-Revised (ADI-R).

Rydzewska, Ewelina, Laura Anne Hughes-McCormack, Christopher Gillberg, et al.
2018 Prevalence of Long-Term Health Conditions in Adults with Autism: Observational Study of a Whole Country Population. BMJ Open 8(8). British Medical Journal Publishing Group: e023945.

Sagar, Angela, Jeffrey R. Bishop, D. Clare Tessman, et al.
2013 Co-Occurrence of Autism, Childhood Psychosis, and Intellectual Disability Associated with a de Novo 3q29 Microdeletion. American Journal of Medical Genetics. Part A 161A(4): 845–849.

Sanders, Stephan J., Xin He, A. Jeremy Willsey, et al.
2015 Insights into Autism Spectrum Disorder Genomic Architecture and Biology from 71 Risk Loci. Neuron 87(6): 1215–1233.

Sanders, Stephan J., Mustafa Sahin, Joseph Hostyk, et al.
2019 A Framework for the Investigation of Rare Genetic Disorders in Neuropsychiatry. Nature Medicine.

Sandin, Sven, Paul Lichtenstein, Ralf Kuja-Halkola, et al.
2017 The Heritability of Autism Spectrum Disorder. JAMA 318(12): 1182–1184.

Santen, Gijs W. E., Emmelien Aten, Yu Sun, et al.
2012 Mutations in SWI/SNF Chromatin Remodeling Complex Gene ARID1B Cause Coffin-Siris Syndrome. Nature Genetics 44(4): 379–380.

Satterstrom, F. Kyle, Jack A. Kosmicki, Jiebiao Wang, et al.
2020 Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 180(3): 568-584.e23.

Savage, Jeanne E., Philip R. Jansen, Sven Stringer, et al.
2018 Genome-Wide Association Meta-Analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence. Nature Genetics 50(7). Nature Publishing Group: 912–919.

Schimmelmann, Benno G., Philippe Conus, Sue Cotton, Patrick D. McGorry, and Martin Lambert
2007 Pre-Treatment, Baseline, and Outcome Differences between Early-Onset and Adult-Onset Psychosis in an Epidemiological Cohort of 636 First-Episode Patients. Schizophrenia Research 95(1–3): 1–8.

Schleifer, Charles, Amy Lin, Leila Kushan, et al.
2019 Dissociable Disruptions in Thalamic and Hippocampal Resting-State Functional

Connectivity in Youth with 22q11.2 Deletions. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 39(7): 1301–1319.

Sebat, Jonathan, B. Lakshmi, Dheeraj Malhotra, et al.
2007 Strong Association of De Novo Copy Number Mutations with Autism. Science 316(5823): 445–449.

Sharp, Andrew J., Heather C. Mefford, Kelly Li, et al.
2008 A Recurrent 15q13.3 Microdeletion Syndrome Associated with Mental Retardation and Seizures. Nature Genetics 40(3). Nature Publishing Group: 322–328.

Shen, Mark D., Christine W. Nordahl, Gregory S. Young, et al.
2013 Early Brain Enlargement and Elevated Extra-Axial Fluid in Infants Who Develop Autism Spectrum Disorder. Brain: A Journal of Neurology 136(Pt 9): 2825–2835.

Sieradzka, Dominika, Robert A. Power, Daniel Freeman, et al.
2014 Are Genetic Risk Factors for Psychosis Also Associated with Dimension-Specific Psychotic Experiences in Adolescence? PLOS ONE 9(4). Public Library of Science: e94398.
2015 Heritability of Individual Psychotic Experiences Captured by Common Genetic Variants in a Community Sample of Adolescents. Behavior Genetics 45(5): 493–502.

Singh, Tarjinder, James T. R. Walters, Mandy Johnstone, et al.
2017 The Contribution of Rare Variants to Risk of Schizophrenia in Individuals with and without Intellectual Disability. Nature Genetics 49(8): 1167–1173.

Smeland, Olav B., Shahram Bahrami, Oleksandr Frei, et al.
2020 Genome-Wide Analysis Reveals Extensive Genetic Overlap between Schizophrenia, Bipolar Disorder, and Intelligence. Molecular Psychiatry 25(4). Nature Publishing Group: 844–853.

Smelror, Runar E., Cecilie Johannessen, Kirsten Wedervang-Resell, et al.
2021 Cognitive Impairment Profile in Adolescent Early-Onset Psychosis Using the MATRICS Battery: Age and Sex Effects. Neuropsychology 35(3). US: American Psychological Association: 300–309.

Smith, A. C., L. McGavran, J. Robinson, et al.
1986 Interstitial Deletion of (17)(P11.2p11.2) in Nine Patients. American Journal of Medical Genetics 24(3): 393–414.

Smith, J.C.
1930 Das Ursachenverhältnis Des Schwachsinns Beleuchtet Durch Untersuchungen von Zweilligen 125: 678–692.

Sobin, Christina, Samantha H. Monk, Karen Kiley-Brabeck, Jananne Khuri, and Maria Karayiorgou
2006 Neuromotor Deficits in Children With the 22q11 Deletion Syndrome. Movement Disorders : Official Journal of the Movement Disorder Society 21(12): 2082–2089.

Solot, Cynthia B, Carol Knightly, Steven D Handler, et al.
2000 Communication Disorders in the 22Q11.2 Microdeletion Syndrome. Journal of Communication Disorders 33(3): 187–204.

Sparrow, Sara S., David A. Balla, and Domenic Cicchetti
1984 The Vineland Adaptive Behavior Scales:Interview Edition, Survey Form. *In* . Circle Pines. American Guidance Service.
2005 Vineland Adaptive Behavior Scales, Second Edition. Bloomington, MN: Pearson Assessments. https://www.pearsonclinical.com/psychology/products/100000668/vineland-adaptive-behavior-scales-second-edition-vineland-ii-vineland-ii.html.

St Pourcain, B., E. B. Robinson, V. Anttila, et al.
2018 ASD and Schizophrenia Show Distinct Developmental Profiles in Common Genetic Overlap with Population-Based Social Communication Difficulties. Molecular Psychiatry 23(2). Nature Publishing Group: 263–270.

Stark, Kimberly L., Bin Xu, Anindya Bagchi, et al.
2008 Altered Brain MicroRNA Biogenesis Contributes to Phenotypic Deficits in a 22q11-Deletion Mouse Model. Nature Genetics 40(6): 751–760.

Stentebjerg-Olesen, Marie, Anne K. Pagsberg, Anders Fink-Jensen, Christoph U. Correll, and Pia Jeppesen
2016 Clinical Characteristics and Predictors of Outcome of Schizophrenia-Spectrum Psychosis in Children and Adolescents: A Systematic Review. Journal of Child and Adolescent Psychopharmacology 26(5): 410–427.

Stern, William, and G Whipple
1912 The Psychological Methods of Intelligence Testing. Trans.). Baltimore: Warwick and York.

Sullivan, Patrick F., and Daniel H. Geschwind
2019 Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell 177(1): 162–183.

Sumner, Emma, Hayley C. Leonard, and Elisabeth L. Hill
2016 Overlapping Phenotypes in Autism Spectrum Disorder and Developmental Coordination Disorder: A Cross-Syndrome Comparison of Motor and Social Skills. Journal of Autism and Developmental Disorders 46(8): 2609–2620.

Sun, Daqiang, Christopher R. K. Ching, Amy Lin, et al.
2018 Large-Scale Mapping of Cortical Alterations in 22q11.2 Deletion Syndrome: Convergence with Idiopathic Psychosis and Effects of Deletion Size. Molecular Psychiatry. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6292748/, accessed July 28, 2020.

Swillen, Ann, Hilde Feys, Tamara Adriaens, et al.
2005 Early Motor Development in Young Children with 22q.11 Deletion Syndrome and a Conotruncal Heart Defect. Developmental Medicine & Child Neurology 47(12): 797–802.

Tager-Flusberg, Helen, and Robert M. Joseph
2003 Identifying Neurocognitive Phenotypes in Autism. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 358(1430): 303–314.

Tager-Flusberg, Helen, Rhea Paul, and Catherine Lord
2005 Language and Communication in Autism. *In* Handbook of Autism and Pervasive Developmental Disorders: Diagnosis, Development, Neurobiology, and Behavior, Vol. 1, 3rd Ed Pp. 335–364. Hoboken, NJ, US: John Wiley & Sons Inc.

Tammimies, Kristiina, Christian R. Marshall, Susan Walker, et al.
2015 Molecular Diagnostic Yield of Chromosomal Microarray Analysis and Whole-Exome Sequencing in Children With Autism Spectrum Disorder. JAMA 314(9): 895–903.

Taylor, Mark J., Mina A. Rosenqvist, Henrik Larsson, et al.
2020 Etiology of Autism Spectrum Disorders and Autistic Traits Over Time. JAMA Psychiatry. https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2765489, accessed August 4, 2020.

Terman, Lewis M.
1916 Binet Scale and the Diagnosis of Feeble-Minded. Journal of the American Institute of Criminal Law and Criminology 7: 530.

Thapar, Anita, Miriam Cooper, and Michael Rutter
2017 Neurodevelopmental Disorders. The Lancet Psychiatry 4(4): 339–346.

Thurm, Audrey, Cristan Farmer, Emma Salzman, Catherine Lord, and Somer Bishop
2019 State of the Field: Differentiating Intellectual Disability From Autism Spectrum Disorder. Frontiers in Psychiatry 10. Frontiers. https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00526/full, accessed August 12, 2020.

Tomasi, Dardo, and Nora D. Volkow
2019 Reduced Local and Increased Long-Range Functional Connectivity of the Thalamus in Autism Spectrum Disorder. Cerebral Cortex (New York, N.Y.: 1991) 29(2): 573–585.

Torrey, E. Fuller, John J. Bartko, and Robert H. Yolken
2012 Toxoplasma Gondii and Other Risk Factors for Schizophrenia: An Update. Schizophrenia Bulletin 38(3): 642–647.

Trost, Brett, Susan Walker, Zhuozhi Wang, et al.
2018 A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. American Journal of Human Genetics 102(1): 142–155.

Trotta, Antonella, Marta Di Forti, Conrad Iyegbe, et al.
2015 Familial Risk and Childhood Adversity Interplay in the Onset of Psychosis. BJPsych Open 1(1): 6–13.

Trotta, Antonella, Marta Di Forti, Valeria Mondelli, et al.
2013 Prevalence of Bullying Victimisation amongst First-Episode Psychosis Patients and Unaffected Controls. Schizophrenia Research 150(1): 169–175.

Ultanir, Sila K., Smita Yadav, Nicholas T. Hertz, et al.
2014 MST3 Kinase Phosphorylates TAO1/2 to Enable Myosin Va Function in Promoting Spine Synapse Development. Neuron 84(5): 968–982.

Valenti, Marco, Maria Chiara Pino, Monica Mazza, et al.
2020 Abnormal Structural and Functional Connectivity of the Corpus Callosum in Autism Spectrum Disorders: A Review. Review Journal of Autism and Developmental Disorders 7(1): 46–62.

Vangkilde, A., J. R. M. Jepsen, H. Schmock, et al.
2016 Associations between Social Cognition, Skills, and Function and Subclinical Negative and Positive Symptoms in 22q11.2 Deletion Syndrome. Journal of Neurodevelopmental Disorders 8(1): 42.

Varese, Filippo, Feikje Smeets, Marjan Drukker, et al.
2012 Childhood Adversities Increase the Risk of Psychosis: A Meta-Analysis of Patient-Control, Prospective- and Cross-Sectional Cohort Studies. Schizophrenia Bulletin 38(4): 661–671.

Velthorst, Eva, Sean Froudist-Walsh, Eli Stahl, et al.
2018 Genetic Risk for Schizophrenia and Autism, Social Impairment and Developmental Pathways to Psychosis. Translational Psychiatry 8(1): 1–11.

Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, et al.
1991 Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome. Cell 65(5): 905–914.

Vissers, Lisenka E. L. M., Christian Gilissen, and Joris A. Veltman
2016 Genetic Studies in Intellectual Disability and Related Disorders. Nature Reviews Genetics 17(1): 9–18.

Vorstman, Jacob A. S., Monique E. J. Morcus, Sasja N. Duijff, et al.
2006 The 22q11.2 Deletion in Children: High Rate of Autistic Disorders and Early Onset of Psychotic Symptoms. Journal of the American Academy of Child and Adolescent Psychiatry 45(9): 1104–1113.

Wagner, RK, JK Torgesen, and CA Rashotte
1999 Comprehensive Test of Phonological Processing. PRO-ED.

Walsh, Tom, Jon M. McClellan, Shane E. McCarthy, et al.
2008 Rare Structural Variants Disrupt Multiple Genes in Neurodevelopmental Pathways in Schizophrenia. Science (New York, N.Y.) 320(5875): 539–543.

Wang, Quan, Rui Chen, Feixiong Cheng, et al.

2019 A Bayesian Framework That Integrates Multi-Omics Data and Gene Networks Predicts Risk Genes from Schizophrenia GWAS Data. Nature Neuroscience 22(5): 691–699.

Wang, Weidi, Roser Corominas, and Guan Ning Lin
2019 De Novo Mutations From Whole Exome Sequencing in Neurodevelopmental and Psychiatric Disorders: From Discovery to Application. Frontiers in Genetics 10. Frontiers. https://www.frontiersin.org/articles/10.3389/fgene.2019.00258/full, accessed August 4, 2020.

Warrier, Varun, Roberto Toro, Hyejung Won, et al.
2019 Social and Non-Social Autism Symptoms and Trait Domains Are Genetically Dissociable. Communications Biology 2(1). Nature Publishing Group: 1–13.

Wassink, Thomas H., Joseph Piven, Veronica J. Vieland, et al.
2002 Evaluation of FOXP2 as an Autism Susceptibility Gene. American Journal of Medical Genetics 114(5): 566–569.

Watanabe, Kyoko, Sven Stringer, Oleksandr Frei, et al.
2019 A Global Overview of Pleiotropy and Genetic Architecture in Complex Traits. Nature Genetics 51(9): 1339–1348.

Wechsler, David
1939 The Measurement of Adult Intelligence. Williams & Wilkins.

Weiner, Daniel J., Emilie M. Wigdor, Stephan Ripke, et al.
2017 Polygenic Transmission Disequilibrium Confirms That Common and Rare Variation Act Additively to Create Risk for Autism Spectrum Disorders. Nature Genetics 49(7): 978–985.

Weiss, Lauren A., Yiping Shen, Joshua M. Korn, et al.
2008 Association between Microdeletion and Microduplication at 16p11.2 and Autism. The New England Journal of Medicine 358(7): 667–675.

Westerinen, H., M. Kaski, L. J. Virta, et al.
2017 The Nationwide Register-Based Prevalence of Intellectual Disability during Childhood and Adolescence. Journal of Intellectual Disability Research 61(8): 802–809.

WHO, Multicentre Growth Reference Study Group, and Mercedes de Onis
2006 WHO Motor Development Study: Windows of Achievement for Six Gross Motor Development Milestones. Acta Paediatrica 95(S450): 86–95.

Willatt, Lionel, James Cox, John Barber, et al.
2005 3q29 Microdeletion Syndrome: Clinical and Molecular Characterization of a New Syndrome. American Journal of Human Genetics 77(1): 154–160.

Wilson, Brenda N., Bonnie J. Kaplan, Susan G. Crawford, Anne Campbell, and Deborah Dewey
2000 Reliability and Validity of a Parent Questionnaire on Childhood Motor Skills. American Journal of Occupational Therapy 54(5). American Occupational Therapy Association: 484–493.

Wither, Robert G., Felippe Borlot, Alex MacDonald, et al.

2017 22q11.2 Deletion Syndrome Lowers Seizure Threshold in Adult Patients without Epilepsy. Epilepsia 58(6): 1095–1101.

Wolke, D., S. T. Lereya, H. L. Fisher, G. Lewis, and S. Zammit
2014 Bullying in Elementary School and Psychotic Experiences at 18 Years: A Longitudinal, Population-Based Cohort Study. Psychological Medicine 44(10): 2199–2211.

Woodward, Neil D., Monica Giraldo-Chica, Baxter Rogers, and Carissa J. Cascio
2017 Thalamocortical Dysconnectivity in Autism Spectrum Disorder: An Analysis of the Autism Brain Imaging Data Exchange. Biological Psychiatry. Cognitive Neuroscience and Neuroimaging 2(1): 76–84.

Wray, Naomi R., Stephan Ripke, Manuel Mattheisen, et al.
2018 Genome-Wide Association Analyses Identify 44 Risk Variants and Refine the Genetic Architecture of Major Depression. Nature Genetics 50(5). Nature Publishing Group: 668–681.

Wray, Naomi R., Cisca Wijmenga, Patrick F. Sullivan, Jian Yang, and Peter M. Visscher
2018 Common Disease Is More Complex Than Implied by the Core Gene Omnigenic Model. Cell 173(7): 1573–1580.

Xie, Fagen, Morgan Peltier, and Darios Getahun
2016 Is the Risk of Autism in Younger Siblings of Affected Children Moderated by Sex, Race/Ethnicity, or Gestational Age? Journal of Developmental and Behavioral Pediatrics: JDBP 37(8): 603–609.

Xu, Bin, Pei-Ken Hsu, Kimberly L. Stark, Maria Karayiorgou, and Joseph A. Gogos
2013 Derepression of a Neuronal Inhibitor Due to MiRNA Dysregulation in a Schizophrenia-Related Microdeletion. Cell 152(1–2): 262–275.

Yahya, Reema
2020 The Hidden Correlation between Intelligence and Autism. The British University in Dubai (BUiD). https://bspace.buid.ac.ae/handle/1234/1617, accessed August 12, 2020.

Yu, Alexander, Donald Turbiville, Fangling Xu, et al.
2019 Genotypic and Phenotypic Variability of 22q11.2 Microduplications: An Institutional Experience. American Journal of Medical Genetics Part A 179(11): 2178–2189.

Yuen, Ryan K. C., Bhooma Thiruvahindrapuram, Daniele Merico, et al.
2015 Whole-Genome Sequencing of Quartet Families with Autism Spectrum Disorder. Nature Medicine 21(2): 185–191.

Zarrei, Mehdi, Christie L. Burton, Worrawat Engchuan, et al.
2019 A Large Data Resource of Genomic Copy Number Variation across Neurodevelopmental Disorders. NPJ Genomic Medicine 4. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6779875/, accessed February 20, 2020.

Zavos, Helena M. S., Daniel Freeman, Claire M. A. Haworth, et al.
2014 Consistent Etiology of Severe, Frequent Psychotic Experiences and Milder, Less Frequent

Manifestations: A Twin Study of Specific Psychotic Experiences in Adolescence. JAMA Psychiatry 71(9): 1049–1057.

Zinkstok, Janneke R, Erik Boot, Anne S Bassett, et al. 2019 Neurobiological Perspective of 22q11.2 Deletion Syndrome. The Lancet. Psychiatry 6(11): 951–960.

Zufferey, Flore, Elliott H. Sherr, Noam D. Beckmann, et al. 2012 A 600 Kb Deletion Syndrome at 16p11.2 Leads to Energy Imbalance and Neuropsychiatric Disorders. Journal of Medical Genetics 49(10): 660–668.

Zweier, Christiane, Eiko K. de Jong, Markus Zweier, et al. 2009 CNTNAP2 and NRXN1 Are Mutated in Autosomal-Recessive Pitt-Hopkins-like Mental Retardation and Determine the Level of a Common Synaptic Protein in Drosophila. The American Journal of Human Genetics 85(5): 655–666.

Zwicker, Alyson, Eileen M. Denovan-Wright, and Rudolf Uher 2018 Gene–Environment Interplay in the Etiology of Psychosis. Psychological Medicine 48(12). Cambridge University Press: 1925–1936.

VI. Supplementary material

All contributions

During this thesis, I participated to the following projects.

As 1st and 2nd co-author

2021

Huguet G., Schramm C., **Douard E. A.** et al., *Genome-wide analysis of gene dosage in 24,092 individuals estimates that 10,000 genes modulate cognitive ability* (2021) Molecular Psychiatry. DOI: https://doi.org/10.1038/s41380-020-00985-z

Gagnon D., Zeribi A., **Douard, E. A.** et al., *Bayonet-shaped language development in autism with regression: a retrospective study* (2021) Molecular Autism. DOI: https://doi.org/10.1186/s13229-021-00444-8

2020

Douard E. A. et al., *Effects-sizes of deletions and duplications on autism risk across the genome* (2020) American Journal of Psychiatry. DOI: https://doi.org/10.1176/appi.ajp.2020.19080834

Douard E. A. and Jacquemont S., *16p11.2 and other recurrent CNVs associated with autism susceptibility* (2020) Chapter accepted for publication

Douard E. A. and Huguet G., *La génétique de l'autisme : Contribution de multiples gènes / The genetics of autism: contributions of multiple genes* (2020) Scientific communication project. http://www.autismresearchgroupmontreal.ca/SurLeSpectre/CIUSSS_Magazine_printemps20_FR_PP_LR.pdf

2019

Jonch A. E., **Douard E. A.** et al., *Estimating the effect-size of the 15Q11.2 BP1–BP2 deletion and its contribution to neurodevelopmental symptoms: recommendations for practice* (2019) J. Med. Genet. DOI: https://doi.org/10.1136/jmedgenet-2018-105879

2018

Huguet G., Schramm C., **Douard E. A.** et al., *Measuring and Estimating the Effect-sizes of Copy Number Variants on General Intelligence in Community-Based Samples* (2018) JAMA Psychiatry. DOI: https://doi.org/10.1001/jamapsychiatry.2018.0039

Other co-authorship:

2021

Modenato C., Kumar K., Moreau C., Martin-Brevet S., Huguet G., Schramm C., Jean-Louis M., Martin C.-O., Younis N., Tamer P., **Douard E. A.**, Thébault-Dagher F., Côté V, Charlebois A.-R., Deguire F., M Maillard A., Rodriguez-Herreros B., Pain A., Richetin S., Melie-Garcia L., Kushan L., Silva A. I., van den Bree M. B. M., Linden D. E. J., Owen M. J., Hall J., Lippé S., Chakravarty M., Bzdok D., Bearden C. E., Draganski B. and Jacquemont S. *Effects of eight neuropsychiatric copy number variants on human brain structure* (2021) Translational psychiatry. DOI: https://doi.org/10.1038/s41398-021-01490-9

Schmilovich Z., Huguet G., He Q., Musa-Johnson A., **Douard E. A.**, Loum M. A., Liao C., Ross J. P., Dionne-Laporte A., Spiegelman D., Hayward C., Banaschewski T., Bokde A., Desrivieres S., Lemaitre H., Schumann G., Xiong L., Dion P. A., Jacquemont S., Chaumette B. and Rouleau G. A., *Copy-Number Variants in The Contactin-5 Gene Are a Potential Risk Factor for Autism Spectrum Disorder* (2021) Preprint on Research Square. DOI: https://doi.org/10.21203/rs.3.rs-660740/v1

Tesfaye R., Huguet G., Smiliovich Z., Absa Loum M., **Douard E. A.**, Jean-Louis M., Martinot J. L., Whelan R, Desrivieres S., Heinz A., Schumann G., Hayward C., Elsabbagh M. and Jacquemont S., *Investigating the contributions of circadian pathway and insomnia risk genes to autism and sleep disturbances* (2021) Preprint on medRxiv. DOI: https://doi.org/10.1101/2021.06.24.21259489

Moreau C. A., Kumar K., Harvey A., Huguet G., Urchs S., **Douard E. A.**, Schultz L. M., Sharmarke H., Jizi K., Martin C.-O., Younis N., Tamer P., Rolland T., Martineau J.-L., Orban P., Shin D., Silva A. I., Hall J., van den Bree M. B. M., Owen M. J., Linden D. E. J., Labbe A., Maillard A. M., Nowakowski T. J., Lippé S., Bearden C. E., Almasy L., Glahn D. C., Thompson P. M., Bourgeron T., Bellec P. and Jacquemont S.. *Atlas of functional connectivity relationships across rare and common genetic variants, traits, and psychiatric conditions* (2021) Preprint on medRxiv. DOI: https://doi.org/10.1101/2021.05.21.21257604

Rolland T., Cliquet F., Anney R. J. L., Traut N., Mathieu A., Huguet G., Leblond C. S., **Douard E. A.**, Amsellem F., Malesys S., Maruani A., Toro R., Packer A., Chung W. K., Jacquemont S., Delorme R. and Bourgeron T. *Towards a gene-level map of resilience to genetic variants associated with autism* (2021) Preprint on medRxiv. DOI: https://doi.org/10.1101/2021.02.12.21251621

Klein M., Shanta O., Hong O., MacDonald J., Thiruvahindrapuram B., de Pins A., Charney A., Letovsky S., Humphrey J., **Douard E. A.**, BD ASD, Jacquemont S., Scherer S., Sebat J., G2MH Consortium and CNV Working Group. *Analysis of genomic copy number variation across psychiatric disorders* (2021) Under review

2020

Moreau C. A., Urchs S. G. W., Kuldeep K., Orban P., Schramm C., Dumas G., Labbe A., Huguet G., **Douard E.**, Quirion P.-O., Lin A., Kushan L., Grot S., Luck D., Mendrek A., Potvin S., Stip E., Bourgeron T., Evans A. C., Bearden C. E., Bellec P. and Jacquemont S. *Mutations associated with neuropsychiatric conditions delineate functional brain connectivity dimensions contributing to autism and schizophrenia* (2020) Nature communications. DOI: https://doi.org/10.1038/s41467-020-18997-2

2019

Sanders S. J., Sahin M., Hostyk J., Thurm A., Jacquemont S., Avillach P., **Douard E. A.**, Martin C. L., Modi M. E., Moreno-De-Luca A., Raznahan A., Anticevic A., Dolmetsch R., Feng G., Geschwind D. H., Glahn D. C., Goldstein D. B., Ledbetter D. H., Mulle J. G., Pasca S. P., Samaco R., Sebat J., Pariser A., Lehner T., Gur R. E. and Bearden C. E., *A framework for the investigation of rare genetic disorders in neuropsychiatry* (2019) Nature Medecine. DOI: https://doi.org/10.1038/s41591-019-0581-5

Saitovitch A., Rechtman E., Lemaitre H., Tacchella J.-M., Vinçon-Leite A., **Douard E. A.**, Calmon R., Grévent D., Philippe A., Chabane N., Brunelle F., Boddaert N. and Zilbovicius M. *Superior temporal sulcus hypoperfusion in children with autism spectrum disorder: an arterial spin-labeling magnetic resonance study* (2019) Preprint on BioRxiv. DOI: https://doi.org/10.1101/771584

Supplementary material 1 (paper #1)

SUPPLEMENTARY METHODS

Definition of autism

Autism spectrum disorder (ASD) refers to autism as defined in DSM-V (1). As the participants involved in the current study were diagnosed following DSM-IV criteria which uses a subtyping strategy (autistic disorder, Asperger syndrome, and PDDNOS) (2), we will use the generic term ‘’autism” to avoid confusion.

CNV detection, annotation and filtering

Genotyping and whole genome sequencing

Genotyping data

CNV detections and standard filtering strategies were previously published (3). CNV calling was performed using the same pipeline for individuals from the Simons Simplex Collection (SSC) (4), IMAGEN (5), and Saguenay Youth Study (SYS) (6) to obtain a harmonized dataset.
In the IMAGEN cohort (5), 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 (CNG; Paris, France).
In SYS cohort (6), 1,994 Illumina SNP arrays were analyzed using Illumina 610Kq (N probes=620,901; N arrays=599) and the HumanOmniExpress BeadChip - V12 (HOE-V12) (N probes=730,525; N arrays=1,395). The genotyping was performed at CNG for 610Kq and at the Genome Analysis Centre of Helmholtz Zentrum München (Munich, Germany) for HOE-V12.
In the SSC (4), 10,032 individuals were genotyped at Yale University using Illumina SNP genotyping arrays 1Mv1, 1Mv3 Duo, or Omni2.5M.

Whole genome sequencing data

In the MSSNG database (7), 7,233 individuals were sequenced at multiple sites using Illumina sequencing HiSeq, HiSeq 2,500, or HiSeqX.
Next generation sequencing data were analysed using Broadinstitute Genome Analysis Toolkit (GATK) best practice (8).

Call of CNVs

CNVs from SSC, IMAGEN and SYS were called using PennCNV (9) and QuantiSNP (10) with the following parameters:

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

Then, we merged detected CNVs from both algorithms with CNVision (11).
For MSSNG, read alignment data were used to compute CNV calling following the workflow of Trost et al. (12).

Filtering of micro-array

To ensure good quality of CNVs, we kept only micro-arrays without too much noise.

- For IMAGEN and SYS cohorts:
  - Wave Factor (WF) < |0.05|
  - Standard deviation of the Log-R-Ratio (LRR-SD) < 0.35
  - Standard deviation of the B allele frequency (BAF-SD) < 0.08
  - Call Rate > 0.99
- For SSC cohort: all micro-array 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) 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 ≥ 500Kb and ≥ 100 SNPs) that spanned known large assembly gaps (greater than 150Kb).

CNV filtering

CNVs with the following criteria were selected for analysis:

- Size ≥ 50 Kb
- Autosomal (Since deletions and duplications are not comparable between sex-linked and autosomal CNVs, we could not pool both types of CNVs. Sex-linked CNVs were therefore excluded from the analysis.)
- 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.
- And additional filters were applied for CNVs which are not 40% overlapping with recurrent CNVs from TableS3: Overlap with Segmental duplicates or centromeric regions < 50%.

CNV annotation and scoring

The annotation of CNVs was performed with a R package developed by our team. This package was developed using RefSeq (https://genome.ucsc.edu/) and ANNOVAR (13) coding and non-coding genes. For these analyses, we excluded the pseudo-genes (UCSC site: https://genome.ucsc.edu/).
Deletions and duplication were annotated for size, number of genes, number of expression quantitative trait loci regulating genes expressed in the brain (14), and each coding gene with all isoforms fully encompassed in CNVs was annotated using 6 constraint scores. For each individual, we computed the sum of these annotations for deletions and for duplications.
Coding genes were annotated using the following constraint scores and transformations:

- the probability of being loss-of-function intolerant (pLI) (15) between 0 and 1 where 1 means that the gene is completely intolerant;
- the residual variation intolerance score (16) between 0 and 100 was transformed with 100-RVIS such as 100 represents the more intolerant;
- the ExAC CNV score for deletions (17) between -2.62 and 3.81 was transformed with deletion score + min(deletion score) such that it becomes between 0 and 6.43 where 6.43 represents the more intolerant.
- the ExAC CNV score for duplication (17) between -2.53 and 2.86 was transformed with duplication score + min(duplication score) such that it becomes between 0 and 5.39 where 5.39 represents the more intolerant.
- the number of protein-protein interactions defined as the number of proteins interacting with each protein coding gene according to STRINGs Protein v10 for Human database (18) (9606.protein.links.v10.txt.gz; http://string-db.org/) where protein networks were defined based on high confidence (> 0.7) interactions;
- the differential stability (DS) score (19), a correlation-based metric which assess reproducibility of regional patterns of gene expression in the brain between -0.057 and 0.97 was transformed with DS + min(DS) such that it becomes between 0 and 1.027 where a higher score means high specific expression in brain.

For the six scores detailed above, the default value associated to gene without available score was 0.

De novo CNVs identification

De novo CNVs in the SSC were identified in probands, unaffected siblings, and unselected population from SYS using two previously published datasets (11, 20), combined with our own algorithm developed in R. A CNV was considered as *de novo* only if it was defined as such by all three approaches.

Frequency

We used two sources to calculate the frequency. First frequency (DGV frequency) was annotated using the database of genomic variants (DGV) on the basis of a minimal overlap of 70% between the CNV of interest and its more closely resembling CNV displayed in DGV (DGV hg19, http://www.dgv.tcag.ca). If CNV was seen in several cohorts, we selected the maximal frequency conditionally to the fact that the sample size of the DGV cohort is ≥ 100; otherwise, frequency was considered as null.
Second frequency was computed on the CNVs of 1,804 adolescents in IMAGEN and 893 parents in SYS from the general population cohort (GP-Cohort frequency). It was computed on the basis of a minimal overlap of 70% between CNVs. Since CNVs with a frequency < 0.1% were visualized to ensure the CNVs veracity, the GP-Cohort frequency was recomputed after excluding false positive CNVs. This process was done iteratively until no more CNV is excluded. Then the CNVs of the entire database (SSC and general population cohorts) were annotated with GP-Cohort frequency on the basis of a minimal overlap of 70% between the CNV of interest and its more closely resembling CNV displayed in GP-Cohort. The frequency of a CNV that is not seen in GP-Cohort was considered as null.

Definition of a rare CNV

Throughout the paper, a rare CNV is either a known recurrent CNV (Table S3) with a DGV frequency < 0.1% or a non-recurrent CNVs with the following characteristics: (i) DGV frequency < 0.1%; (ii) < 50% of the CNV is contained in regions present at > 1% in DGV (11, 21, 22); (iii) unselected population frequency 1/1000. All CNVs annotated as rare were manually curated by visual inspection (SnipPeep, http://snippeep.sourceforge.net) and false positives were excluded.

Genetic analysis of pairwise ancestry and population stratification

Classical multidimensional scaling (MDS) was used to identify ancestry based on the identity by state (IBS) matrices of genetic distances (D) in the IMAGEN, the SYS and the SSC cohorts, based on the reference population HapMap3 (23) with 993 individuals (Hapmap Consortium 2003 (www.hapmap.ncbi.nlm.nih.gov). PLINK (24) (pngu.mgh.harvard.edu/purcell/plink/) was used to do these calculations. SNPs were filtered to keep only autosomal SNPs with minor allele frequency (MAF) > 5% and with good quality, significance threshold for a test of Hardy-Weinberg equilibrium < 1.10^{-6} and missing genotype rates < 10%. Related individuals were identified based on D defined by the following formula:

$$D = \frac{IBS2 + 0.5\, IBS1}{N\ SNP\ pairs}$$

with *IBS1* and *IBS2* being the number of loci at which a pair of individuals share either 1 or 2 alleles identical by state, respectively, and *N SNP pairs* is the number of loci tested. Pairs of individuals were defined as related when $D \geq 0.8$.
Ancestry was estimated using Admixture (25) (http://www.genetics.ucla.edu/software/admixture) with reference populations from HapMap3 (23) allowing for 4 ancestry components (Africa, Asia, European and India). Results show a strong European ancestry component in the three datasets. We then performed a principal components analysis based on the variance-standardized relationship matrix and displayed the 3 first ancestry dimensions and associated eigenvalues.

Clinical assessments

Non-verbal Intelligence Quotient (IQ)

Intellectual ability was measured using standardized tests according to the cognitive level of the participant (Table S2).
For SSC autistic probands, non-verbal intelligence quotient (NVIQ) scores were obtained from the Differential Ability Scales, 2nd Edition (DAS-II) (26) for early years (N=1,031) and school age children (N=1,213), the Wechsler Intelligence Scale for Children, 4th Edition (WISC-IV) (27) (N=45), the Wechsler Abbreviated Scale of Intelligence – First Edition (WASI-I) (28) (N=63) or the Mullen Scales of Early Learning (MSEL) (29) (N=213) (see density distribution of NVIQ by test used in Figure S1B). Norm-referenced standard scores (deviation NVIQ) were available for most of the participants (85.10%). However, for individuals from SSC who were not able to obtain a deviation NVIQ due to their age and/or developmental level, ratio NVIQ were derived by dividing mental age by chronological age and multiplying by 100. See Bishop et al., 2011 for more details concerning convergence between ratio and deviation NVIQ (30).
For MSSNG autistic probands, NVIQ scores were obtained from the Leiter international performance scale – Original and revised (31, 32) (N=372), the raven progressive matrices (33) (N=214), the Stanford-Binet intelligence scale (N=281), the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV) (27) (N=46), the Wechsler Abbreviated Scale of Intelligence – First and Second Editions (WASI-I, WASI-II) (28, 34) (N=338) or the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition (WPPSI-IV) (35) (N=128) (see density distribution of NVIQ by test used in Figure S1C). Deviation NVIQ were available for all participants.
Individuals in IMAGEN undertook the fourth edition of the Wechsler intelligence scale for children (WISC-IV) whereas SYS cohort used the third edition (WISC-III) to assess children. Deviation NVIQ were available for all participants.
As a preeminent test of fluid intelligence, the Raven's progressive matrices, has been reported as a better assessment of general cognitive abilities in autistic individuals than common NVIQ tests (36), we also investigated the matrices subtests measured in SSC probands using the Differential Ability Scales – second edition (DAS-II) (26).

Social Responsiveness Scale

For all the individuals from the SSC and for the unselected population from IMAGEN, severity of social deficits was ascertained with scores from the social responsiveness scale (SRS) (37, 38).

Severity of autism main domains

For probands from the SSC and MSSNG, the severity of autism main domains were assessed by domain-calibrated scores from the Autism Diagnostic Observation Schedule (ADOS) (39, 40) and the Autism Diagnostic Interview-Revised (ADI-R) (41).

Child behaviour Checklist

For SSC probands and their unaffected siblings, behavioural problems were assessed on the Total Problems Score, as well as the Internalizing and Externalizing domains of the Preschool and School-aged Child behaviour Checklist (CBCL) (42, 43).

Phenotypes only available for probands from the SSC

Language and phonology

Phonological short-term memory was measured using scaled scores from the non-word repetition subtest of the Comprehensive Test of Phonological Processing (CTOPP) (44).
Language level was measured using age of first words, age of first phrases and the overall level of expressive language from the ADI-R (45) (question 09, question 10 and question 30 respectively).
Words and phrases delay variables in probands have been derived using both items of age of first words and age of first phrases from the ADI-R score as following:
We considered that autistic probands have word or phrase delay if parents reported an age of first words greater than 24 months on question 9 and the first phrases after 33 months on question 10. Also, if autistic probands scored 993, 994, or 997 on question 9 or 10, they were considered as delayed in the corresponding variable. At the contrary, if they scored of 996 on question 9 or 10, they were considered as non-delayed in the corresponding variable.

Adaptative skills

Standard scores of total adaptive skills, communication, interaction and daily living skills domains were measured using the Vineland Adaptive behaviour Rating Scales - Second Edition (VABS-II). (46)

Motor skills

Motor skills were measured using the corresponding main domain of the VAB-II, as well as gross and fine motor skills subdomains.
We also used the age of onset for walking reported in the ADI-R (question 5) as continuous and categorical measures of gross motor skills. (45)

Walking delay variables in probands have been derived using item of age of onset for walking from the ADI-R score (question 5).
We considered that autistic probands have walking delay if parents reported an age of first walk greater than 18 months on question 5, as well as if they scored 997 at the same item. At the contrary, if they scored of 996 on question 5, they were considered as non-delayed.
Motor coordination was assessed by the Developmental Coordination Disorder Questionnaire (DCDQ).

Associated neurological condition

Finally, the presence of non-febrile seizures was assessed from the ADI-R (question 85) and the Medical History Form.

Statistical analyses

Stepwise variable selection procedure based on Bayesian information criteria

The stepwise variable selection procedure was used to choose the best model with the smallest Bayesian Information Criterion (BIC). We allowed pairwise interactions conditionally upon main effects to also be included in the model. Since variable selection and model fitting steps were performed on the same dataset, this could induce a selection bias in the estimates. We used a bootstrap (1,000 iterations) estimation of the bias in our final models to correct for this bias. (47)
Genetic variable selection procedure was performed using the boot.stepAIC() function from the R package 'bootStepAIC'. (48)

Effect of gene dosage on general intelligence

The model selected in SSC cohort was applied in the unselected population sample. Both were adjusted for NVIQ test used, sex and ancestry. Moreover, the unselected population sample includes related individuals, thus we included a random effect in the model to take into account the familial relationship.
The model adjusted for familial relationship with a random effect could be written as:
$NVIQ \sim \boldsymbol{\alpha}X + \gamma Z + \beta_1 CNV_{DEL} + \beta_2 CNV_{DUP}$
where X represents the adjustments covariates (NVIQ test used, sex and ancestry) and Z is the familial relatedness; $CNV_{DUP/DEL}$: CNV scoring selected as best genetic explanatory variable in SSC cohort; (α, β_1, β_2) and γ are respectively the vectors of coefficients for fixed and random effects. The mixed effect model was computed using lme() function from the 'nlme' R package. (49)

Replication in MSSNG

The model selected in SSC cohort was applied in MSSNG cohort, except that the ancestry is not available for MSSNG dataset and therefore we could not adjust for it. The MSSNG dataset also includes related individuals, thus we included a random effect in the model to take into account the familial relationships.

Matching autistic probands from SSC and from MSSNG based on NVIQ

We used a matching procedure to obtain identical NVIQ distributions in both cohorts (1MSSNG:2SSC) (Figure S2B). Matching was made by searching for the nearest neighbor within ± 5 points of NVIQ. We used the Match() function from the R package 'Matching'. (51) To avoid errors due to randomness, we performed the matching 500 times, obtaining 500 matched cohorts and 500 corresponding estimates.

Model with pooled dataset to test the interaction between CNV scoring and diagnosis

NVIQ ~ $\boldsymbol{\alpha}X + \gamma Z + \beta_1\ CNV_{DEL} + \beta_2\ CNV_{DUP} + \beta_3\ CNV_{DEL}$ * diagnosis + $\beta_4\ CNV_{DUP}$ * diagnosis
where X represents the adjustments covariates (NVIQ test used, sex, and diagnosis) and Z is the familial relatedness; $CNV_{DUP/DEL}$: CNV scoring selected as best genetic explanatory variable is SSC cohort; (α, β_1, β_2, β_3, β_4) and γ are respectively the vectors of coefficients for fixed and random effects. The mixed effect model was computed using lme() function from the 'nlme' R package. (49)

Effect of gene dosage on autism risk

All enrichment analyses were performed by excluding related individuals of MSSNG and general population cohorts, *i.e.* only one subject by family was included in the analyses.
Conditional logistic regression was performed for the comparison of probands paired to their unaffected siblings using the clogit() function from the R package 'survival'. (50)
All models for this section included sex and ancestry as covariates when available.

Matching of autistic probands and general population regarding the NVIQ

We used a matching procedure to pairs probands from the SSC with individuals from the unselected populations (ratio 1:1) with similar NVIQ (± 5 points of NVIQ). We used the Match() function from the R package 'Matching'. (51)
To avoid errors due to randomness, we performed the matching 500 times, obtaining 500 matched cohorts including 1,411 to 1,469 pairs, and 500 corresponding estimates. The Figure S1A and S1D represent the NVIQ distributions in initial cohorts and in an example of matched cohorts.

Contribution of NVIQ in autism risk effect

To estimate the proportion of autism-risk potentially mediated by NVIQ for deletions and duplication, we performed a counterfactual-based mediation analysis on the pooled dataset. We used the neImpute(), neModel(), and neEffdecomp() functions from the R package 'medflex' (52) on two logistic regression including sum of pLI in deletions or duplications as the main explanatory variable. We also applied a logistic regression including sum of pLI in deletions and duplications as the two mains explanatory variables to estimate the autism risk conferred by deletions and duplications in both subgroups of individuals above and below median NVIQ.

Effect of gene dosage on phenotypic measures

SRS as a continuous variable in the SSC probands, unaffected siblings and parents, and in the unselected population from IMAGEN

We used a linear mixed effect model to quantify the effect of gene dosage measured by pLI scores on SRS total raw score after pooling probands and their unaffected relatives (siblings and parents). A kinship matrix was generated to model the genetic covariance between related individuals using the kinship() function from the R package 'kinship2' (53) and this covariance was used as a random effect in the model performed with the function lmekin() from the R package 'coxme' (54).
We further explored a potential effect of gene dosage on the SRS within the autism group, unaffected siblings, parents and unselected population from IMAGEN using a linear regression and adjusting for the abnormal distributions with a square root transformation of the SRS scores when necessary (Table S8). All models used were adjusted for age, sex, ancestry, and in a second time for NVIQ and/or for the diagnosis of autism.

SRS as a categorical variable in the SSC probands, unaffected siblings and unselected population from IMAGEN

We also investigated the SRS scores based on the previously published *T*-score categorization (55) 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 deficiencies in social behaviour;
- ***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 categorical SRS: clinical (obtained after merging the moderate, mild and clinically significant categories) and normal (Table S8, Figure S5C and S5D). This logistic regression model took into account the family relatedness as random factor using the glmer() function from the R package 'lme4'.(56)
A cumulative ordinal regression model was also performed on SRS coding for 4 different levels of social deficits (normal, moderate, mild and clinically significant) (Table S8, Figure S5E and S5F). This model was applied using the function vglm() from the R package 'VGAM'. (57) All models used were adjusted for age, ancestry, and in a second time for the diagnosis of autism.

Autism core symptoms in probands from the SSC and MSSNG

A cumulative ordinal regression model was used to assess the effect of gene dosage on ADOS and ADI-R domains in autistic probands from the SSC and MSSNG separately, and in the pooled dataset (probands from the SSC and MSSNG).
Ordinal regression (cumulative, parallel slopes):
$ln(P(Y>k)/P(Y<=k)) \sim \alpha_k X+\beta_1 pLI_{DEL} + \beta_2 pLI_{DUP}$

Models assessing the ADI-R were adjusted for age, sex and in a second time for NVIQ. When exploring the ADOS, models were corrected for sex and in a second time for NVIQ (Table S9). Additional correction for the population was applied when the model was run on the pooled dataset.

CBCL in probands and unaffected siblings from the SSC

Early analysis using T scores generated by the CBCL demonstrated that despite T scores being normed for sex and age, sex and age emerged as significant factors affecting scores on the school-aged tests. For all analysis afterwards, raw scores were used and corrected for age and sex in the analysis. Several possible models, including an ordinal model testing binning based on "pre-clinical" and "clinical" thresholds based on T-scores, were discarded with the switch to raw scores. Pre-school and school-aged tests were analyzed separately.
Since raw scores have an oversampling of zeros and are overdispersed, it was necessary to use a negative binomial distribution function to adequately assess the relationship between CBCL scores and gene dosage measured by pLI. Probands and unaffected siblings were assessed separately and together. In the pooled sample, a negative binomial mixed effects model was used, with family as a random variable to account for familial relatedness. The function used for probands and siblings alone was glm.nb() from the R package 'MASS' (58), and for pooled samples, glmmTMB() was used from the package 'glmmTMB'. (59) All models used were adjusted for age, sex, ancestry, and in a second time for NVIQ and/or for the diagnosis of autism (Table S10).

See Table S7 for detail of the models used for the other phenotypical measures.

Computation of significance threshold

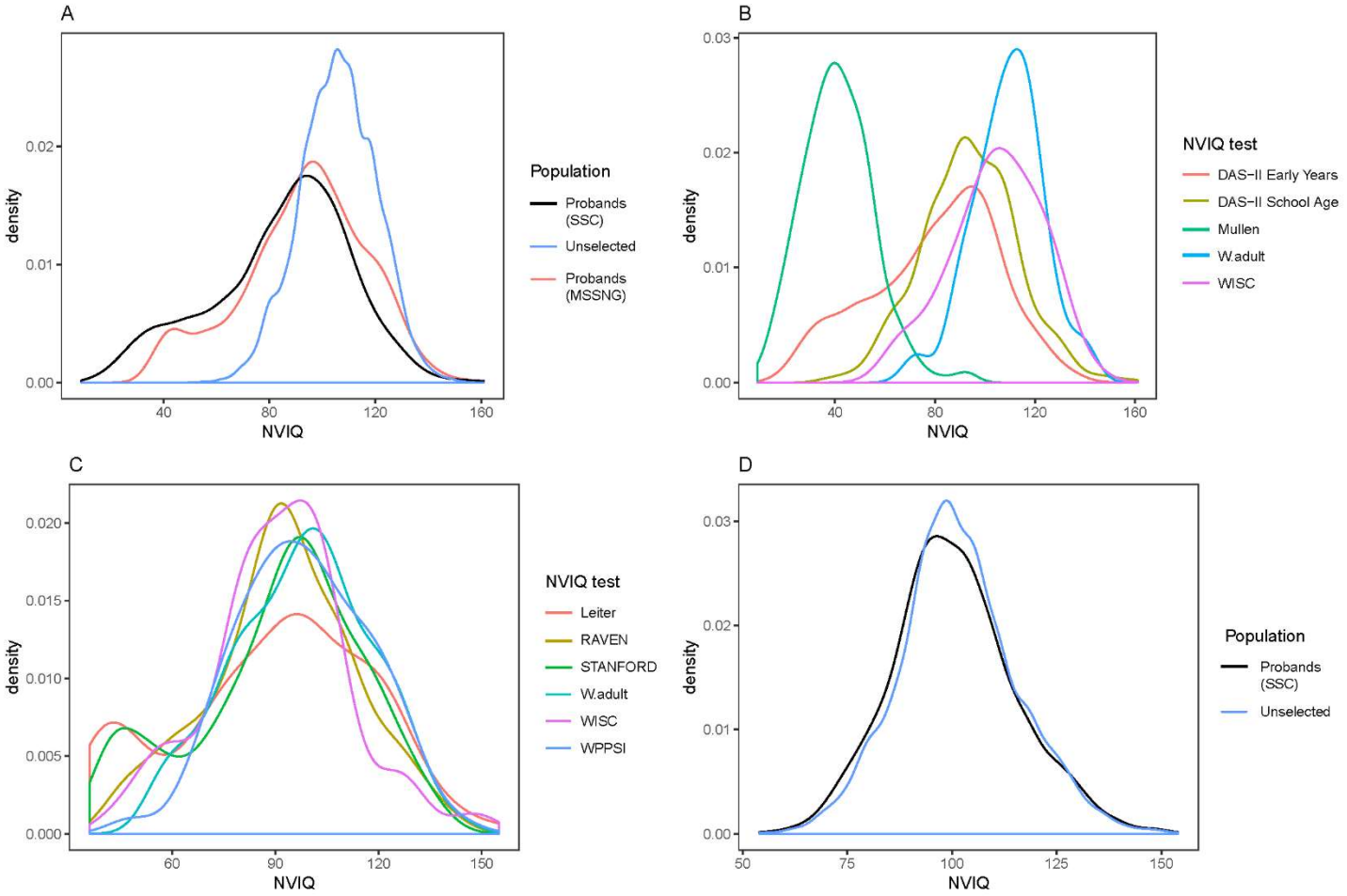
To control for multiple testing, we computed our significance threshold by adapting the methodology of Cheverud (2001) previously applied to control the FWER when analyzing many SNPs (60). When computing the matrix of correlations between phenotypes, Cheverud argued that the eigenvalues of this matrix could be used to estimate the effective number of independent tests. Thus, we calculated an effective number of independent tests, m_e, and then used this number in a Bonferroni-style correction.
m_e was calculated as follow:

$$m_e = 1 + (m - 1)\left(1 - \frac{1}{m}\,\mathrm{Var}(\lambda)\right)$$

$$Var(\lambda) = \frac{1}{m-1}\sum_{i=1}^{m}(\lambda_i - 1)^2$$

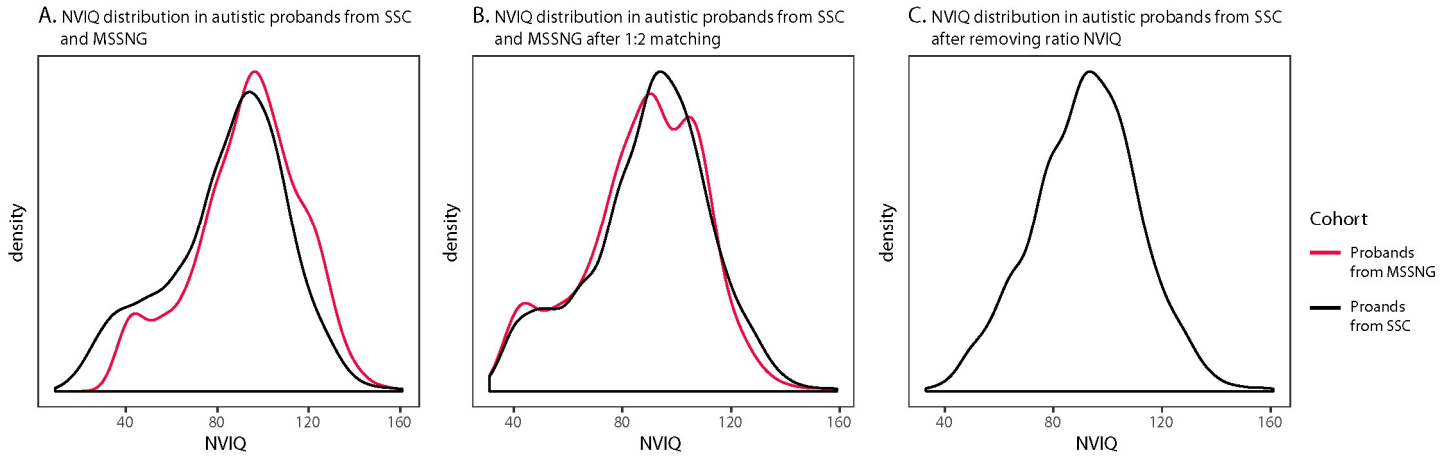
SUPPLEMENTARY FIGURES

Figure S19: NVIQ distributions.



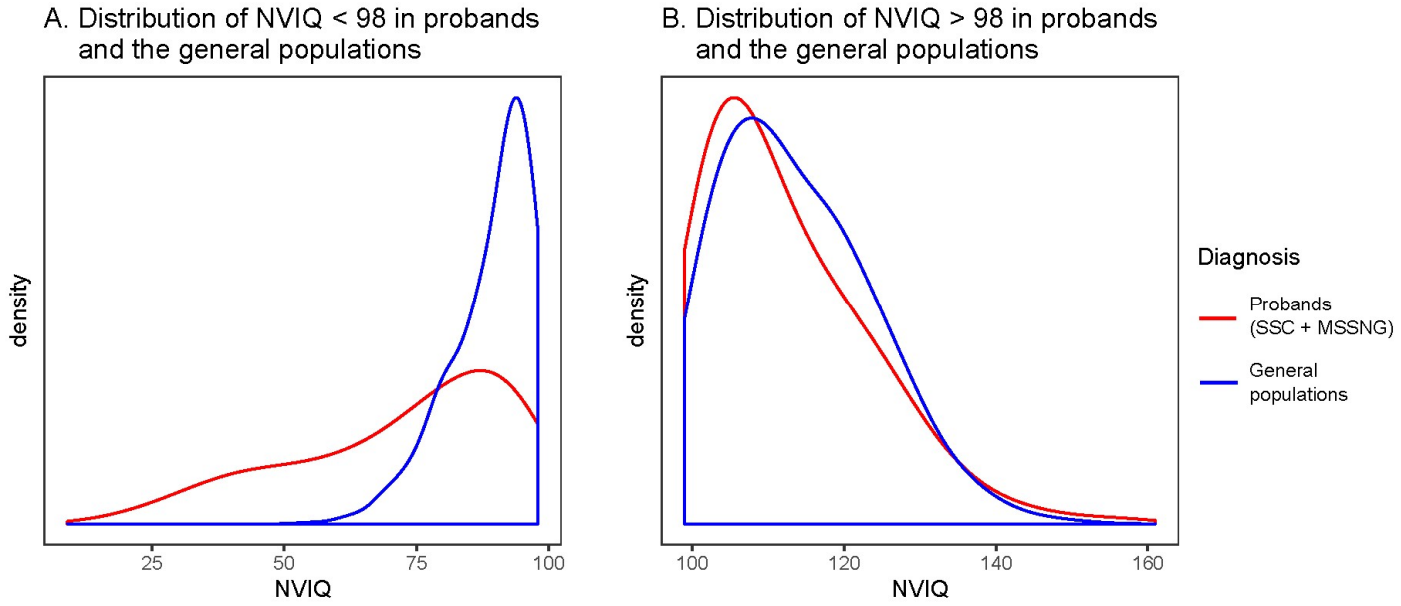
A) Density distribution of NVIQ in the autistic probands from SSC (black) and MSSNG (red), and from the unselected population (green: IMAGEN and SYS pooled). B) Density distribution of NVIQ in the autistic probands from SSC in function of the test used. C) Density distribution of NVIQ in the autistic probands from MSSNG in function of the test used. D) Density distribution of NVIQ in the autistic probands from SSC (black) and from the unselected population (green: IMAGEN and SYS pooled) after 1:1 matching procedure. NVIQ: Non-verbal IQ; DAS-II: Differential Ability Scales, 2nd Edition; WASI: Wechsler Abbreviated Scale Intelligence, WISC: Wechsler Intelligence Scale for Children; WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

Figure S20: NVIQ distributions after selection for sensitivity analyses.



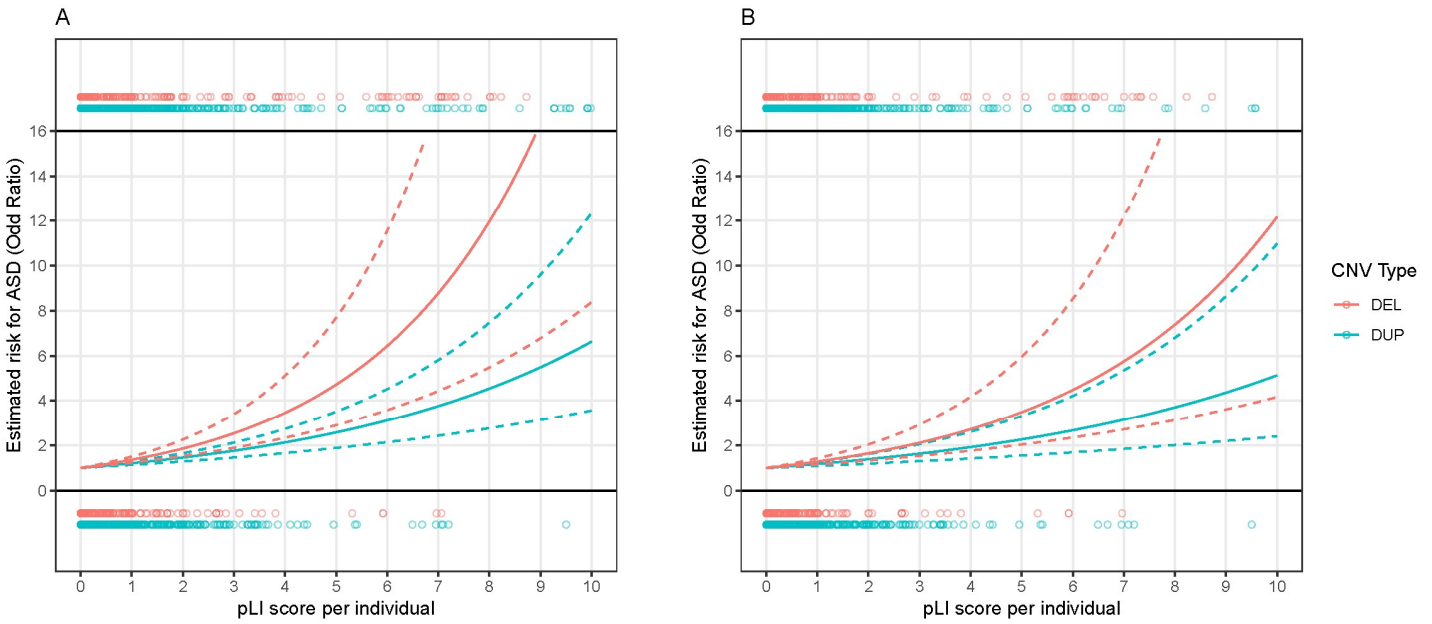
A) NVIQ distribution of autistic probands from SSC (black) and MSSNG (red) before matching. B) NVIQ distribution of autistic probands from SSC (black) and MSSNG (red) after 1:2 matching on NVIQ (example of 1 among 500 iterations). C) NVIQ distribution of autistic probands from SSC (black) after removing 382 individuals with a ratio NVIQ.

Figure S21: NVIQ distributions of individual below or above the median (98), using probands from SSC and MSSNG, and the unselected populations.



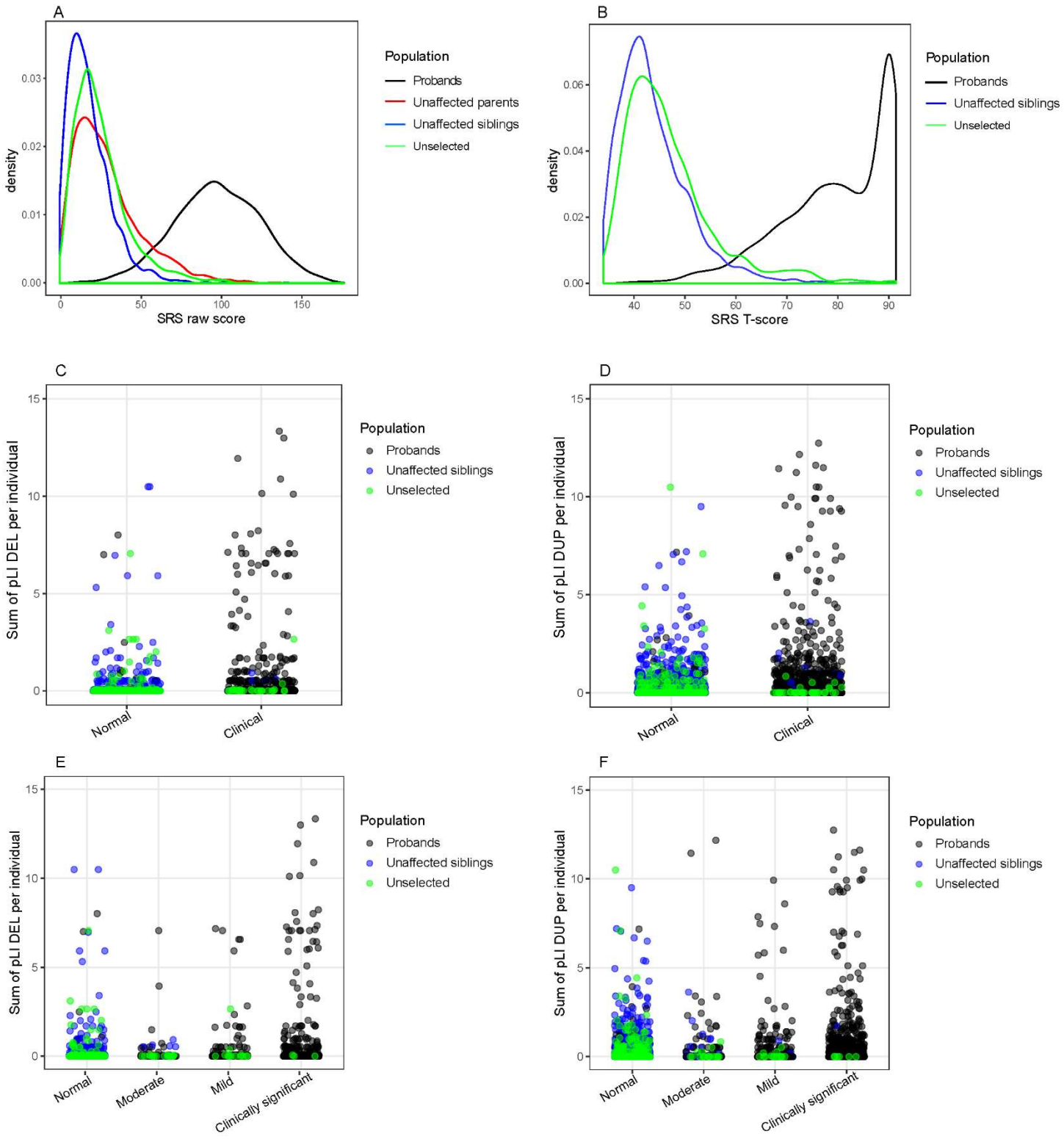
A) NVIQ distribution of autistic probands from SSC and MSSNG (red) and unselected populations (blue) with a NVIQ bellow 98. B) NVIQ distribution of autistic probands from SSC and MSSNG (red) and unselected populations (blue) with a NVIQ over 98.

Figure S22: Effect of gene dosage measured by pLI on autism risk.



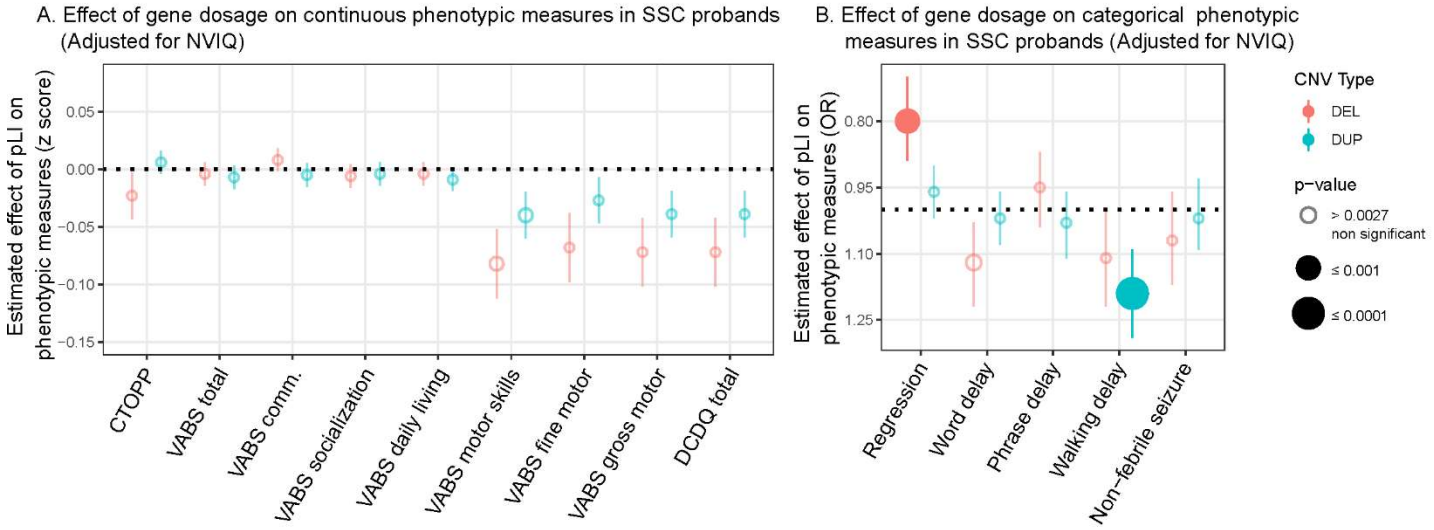
Representation of the non-linear effect of deleted (red) or duplicated (blue) point of pLI on autism risk estimated by logistic regression model when including (A) or excluding (B) the 16 recurrent CNVs detailed in Table S6. X-axis represents the sum of pLI score per individual and y-axis represents the risk of autism diagnosis estimated by our model in odds ratio.

Figure S23: Effect of gene dosage measured by pLI on SRS score.



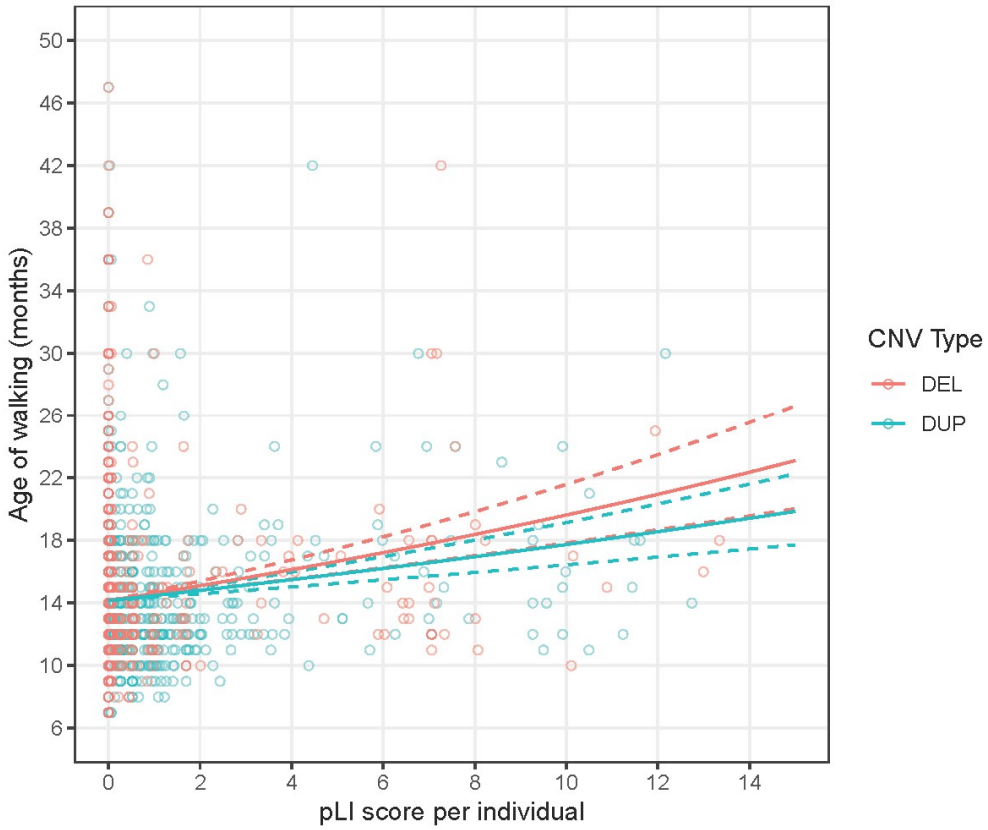
A) SRS raw score distribution in the autistic probands from SSC (black), their unaffected siblings (blue), their parents (red) and the unselected population from IMAGEN (green). B) SRS T-scores distribution in the autistic probands from SSC (black), their unaffected siblings (blue), and from the unselected population from IMAGEN (green). C to F) Representation of the distribution of Sum of pLI deleted (C,E) or duplicated (D,F) per individual in function of SRS categories (binary (C,D) or all assessed categories (E,F)) in the autistic probands from SSC (black), their unaffected siblings (blue) or unselected population from IMAGEN (green).

Figure S24: Effect of gene dosage measured by pLI on phenotypic measures adjusted for NVIQ in SSC probands.



(A, B) Effect-size of a deleted (red) or duplicated (blue) point of pLI on continuous (A) and categorical (B) phenotypes in autistic probands from the SSC adjusted for NVIQ. Y-axis values of panel (A) are measures z-scored using normative data (Table S7) except for DCDQ which was z-scored using the SSC autistic proband group. Y-axis values of panel (B) are odds ratios computed by logistic regression. The significance threshold was computed (60) to account for multiple testing: 0.0027. CTOPP: Comprehensive Test of Phonological Processing; VABS-II: Vineland Adaptive behaviour Rating Scales - Second Edition; DCDQ: Developmental Coordination Disorder Questionnaire.

Figure S25: Effect of gene dosage measured by pLI on age of onset for walking.



Representation of the non-linear effect of deleted (red: OR, 1.03 [95% CI, 1.02-1.04]; $P=5\times10^{-12}$) or duplicated (blue: OR, 1.02 [95% CI, 1.02-1.03]; $P=2\times10^{-9}$) point of pLI on age of onset for walking estimated using quasi-Poisson model adjusted for sex. X-axis represents the sum of pLI score per individual and y-axis represents the age of onset for walking in month.

SUPPLEMENTARY TABLES

Table S1: Description of cohorts for demographic, genotypical characteristics and phenotypical measures.

Variables	IMAGEN	SYS	SSC probands	SSC siblings	SSC parents	MSSNG
Demographic characteristics						
N individuals	1,802	967	2,569	2,092	5,138	1,381
Age, mean (SD)[(b)]	14.45 (0.37)	14.99 (1.84)	9.03 (3.58)	10.06 (4.33)	41.48 (6.18)	9.21 (4.44)
N male (%)	880 (48.91)	462 (47.72)	2,227 (86.66)	969 (46.31)	2,569 (50.00)	1,106 (80.09)
Genotypical characteristics (Chr 1-22)						
Detection technology	Illumina 610Kq or 660Wq array	Illumina 610Kq or HumanOmniExpress12 array	Illumina 1Mv1, 1Mv3 or Omni 2.5 array			Illumina HiSeq, HiSeq 2,500 or HiSeqX sequencing
N carriers of CNVs	855	610	1,835	1,490	2,040	932
N carriers of documented CNVs[(a)]	26	22	77	33	82	40
N total CNVs	2,712	2,213	6,597	5,159	12,689	4,075
N deletions (total)	1,574	1,212	3,323	2,562	6,410	2,435
N deletions, mean (SD)	0.87 (0.91)	1.25 (1.06)	1.29 (1.05)	1.22 (1.05)	1.24 (1.04)	1.76 (1.39)
N duplications (total)	1,138	1,001	3,274	2,597	6,279	1,640
N duplications, mean (SD)	0.63 (0.79)	1.03 (1.04)	1.27 (1.13)	1.24 (1.12)	1.22 (1.11)	1.19 (1,15)
pLI deletions, sum	94.97	55.16	416.00	118.08	232.37	171.61
pLI deletions per individual, mean (SD)	0.05 (0.34)	0.06 (0.46)	0.16 (1.00)	0.06 (0.46)	0.04 (0.39)	0.12 (0.63)
pLI duplications, sum	226.27	248.75	816.38	375.74	891.78	350.23
pLI duplications per individual, mean (SD)	0.13 (0.53)	0.26 (0.73)	0.32 (1.44)	0.18 (0.63)	0.17 (0.60)	0.25 (1.12)
Phenotypical measures						
N NVIQ	1,744	966	2,569	-	-	1,381
NVIQ, mean (SD)	106.62 (14.77)	104.50 (13.09)	84.47 (26.27)	-	-	92.97 (23.72)
Autism related symptoms						
N DAWBA DSM	1,303	-	-	-	-	-
N SRS	977	-	2,556	2,078	4,838	598
SRS raw score,	25.42 (16.99)	-	97.98 (26.97)	18.36 (13.82)	29.53 (21.34)	-

mean (SD)						
SRS T-score, mean (SD)	49.86 (36.51)	-	79.52 (10.45)	43.80 (7.04)	-	-
N ADOS-overall css	-	-	2,498	-	-	679
ADOS-overall css, mean (SD)	-	-	7.45 (1.68)	-	-	7.30 (2.11)
N ADOS-social affect css	-	-	2,372	-	-	325
ADOS-social css, mean (SD)	-	-	7.25 (1.76)	-	-	6.86 (2.00)
N ADOS-rrsb css	-	-	2,449	-	-	330
ADOS-rrsb css, mean (SD)	-	-	7.78 (1.92)	-	-	8.52 (1.51)
N ADIR reciprocal social interactions	-	-	2,567	-	-	403
ADIR reciprocal social interactions, mean (SD)	-	-	20.35 (5.69)	-	-	18.26 (7.98)
N ADIR rrb	-	-	2,567	-	-	696
ADIR rrb, mean (SD)	-	-	6.52 (2.50)	-	-	6.43 (2.50)
N ADIR verbal communication	-	-	2,254	-	-	374
ADIR verbal communication, mean (SD)	-	-	16.48 (4.29)	-	-	14.42 (5.82)
N ADIR non-verbal communication	-	-	2,567	-	-	71
ADIR non-verbal communication, mean (SD)	-	-	9.26 (3.45)	-	-	10.76 (3.03)
N with/without regression	-	-	911/1,657	-	-	-
Language and phonology						
N ADIR overall level of language	-	-	2,568	-	-	1,039
N age of first word	-	-	2,467	-	-	-
age of first word in months, mean (SD)	-	-	24.38 (14.97)	-	-	-
N with/without word delay	-	-	935/1,632	-	-	-
N age of first phrase	-	-	2,311	-	-	-

age of first phrase in months, mean (SD)	-	-	39.18 (18.52)	-	-	-
N with/without phrase delay	-	-	1,567/1,000	-	-	-
N CTOPP	-	-	1,988	-	-	-
CTOPP, mean (SD)	-	-	7.76 (2.86)	-	-	-
behavioural problems						
N CBCL total score	-	-	1,945	1,596	-	-
CBCL total score, mean (SD)	-	-	50.77 (23.93)	17.99 (15.18)	-	-
N CBCL internalizing	-	-	1,945	1,596	-	-
CBCL internalizing, mean (SD)	-	-	11.06 (8.46)	4.71 (5.25)	-	-
N CBCL externalizing	-	-	1,945	1,596	-	-
CBCL externalizing, mean (SD)	-	-	11.56 (7.91)	5.06 (5.05)	-	-
Adaptative skills						
N VABS-II total	-	-	2,569	-	-	-
VABS-II total, mean (SD)	-	-	73.08 (12.13)	-	-	-
N VABS-II daily living	-	-	2,569	-	-	-
VABS-II daily living, mean (SD)	-	-	76.33 (13.92)	-	-	-
N VABS-II communication	-	-	2,569	-	-	-
VABS-II communication, mean (SD)	-	-	76.98 (14.63)	-	-	-
N VABS-II socialization	-	-	2,569	-	-	-
VABS-II socialization, mean (SD)	-	-	70.91 (12.59)	-	-	-
Motor skills						
N VABS-II motor	-	-	919	-	-	-
VABS-II motor, mean (SD)	-	-	81.75 (12.60)	-	-	-
N VABS-II gross motor	-	-	926	-	-	-
VABS-II gross motor, mean (SD)	-	-	12.25 (2.23)	-	-	-
N VABS-II fine motor	-	-	923	-	-	-

VABS-II fine motor, mean (SD)	-	-	11.78 (2.69)	-	-	-
N age of onset of walking	-	-	2,550	-	-	-
age of onset of walking in months, mean (SD)	-	-	13.56 (4.00)	-	-	-
N with/without onset of walking delay	-	-	159/2,405	-	-	-
N DCDQ	-	-	2,209	-	-	-
DCDQ, mean (SD)	-	-	38.50 (12.44)	-	-	-
Associated neurological condition						
N with/without non-febrile seizure	-	-	233/2,333	-	-	-

(a) Number of carriers of a CNV which overlap ≥ 30% with a documented CNV from the Table S6; SD: Standard deviation; Sum pLI deletions or duplications: sum of score of pLI for the entire population; NVIQ: Non-verbal intelligence quotient; DAWBA: Development and Well-Being Assessment; SRS: Social Responsiveness Scale; ADOS: Autism Diagnostic Observation Schedule; css: calibrated severity score; rrsb: repetitive, restricted and stereotyped behaviours; ADI-R: Autism Diagnostic Interview-Revised; CTOPP: Comprehensive Test of Phonological Processing; CBCL: Child behaviour Checklist; VABS-II: Vineland Adaptive behaviour Rating Scales - Second Edition; DCDQ: Developmental Coordination Disorder Questionnaire.

Table S2: NVIQ available in autistic probands from SSC and MSSNG, and individuals from unselected population.

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
SYS	966	14.99	1.84	462	47.72	WISC-III	104.50	13.09
SSC autistic probands	2,564	9.03	3.58	2,227	86.66	DAS-II, MSEL, WASI-I, WISC-IV	84.47	26.27
MSSNG autistic probands	1,381	9.21	4.44	1,106	80.09	Leiter, Raven, Stanford-Binet, WASI-I, WASI-II, WISC-IV, WPPSI-IV	92.97	23.72

SD: Standard deviation; NVIQ: Non-verbal intelligence quotient; DAS-II: Differential Ability Scales - Second Edition; MSEL: Mullen Scale of Early Learning; WASI-I or II: Wechsler Abbreviated Scale of Intelligence – First or Second Edition; WISC-IV: Wechsler Intelligence Scale for Children, Fourth Edition; WPPSI-IV: Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition.

Table S3: Breakpoints used to detect recurrent CNVs associated to neurodevelopmental disorders.

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	1	10077413	DEL	*GABRD*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	710137	9977413	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	860137	3660140	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	145288643	145628643	DEL	*HFE2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	145338643	149783376	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	145338643	147883376	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	146573376	147393376	DEL	*GJA5*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	146573376	147393376	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	168733376	173733377	DEL	*FMO* and *DNM3*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr1	245033377	248833377	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	50146496	51256496	DEL	*NRXN1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	57746496	61736496	DEL	*VRK2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	59646496	63146496	DUP	*PEX13* to *AHSA2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	96726273	97676273	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	100693568	108443568	DEL	*NCK2* and *FHL2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	111333937	113233529	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	111383531	113093529	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	111383531	113093529	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	200161755	200511755	DEL	*SATB2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	235735261	243102476	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr2	239705243	242471327	DEL	*HDAC4*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	9525000	11025000	DUP	*JAGN1* to *TATDN2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	87237310	87557310	DEL	*CHMP2B* to *POU1F1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	115237310	115647310	DEL	*GAP43*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	191517306	193017306	DEL	*FGF12*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	195715603	197355603	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	195715603	197355603	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	195745603	197355603	DEL	*DLG1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr3	195745603	197355603	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr4	110000	7049099	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr4	1870202	2010202	DEL	*WHSC1* and *WHSC2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr4	80780976	83280976	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr5	1	11727000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr5	87964244	88224244	DEL	*MEF2C*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr5	175717394	177057394	DEL	*NSD1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr5	180117394	180817394	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr6	92043279	104693307	DEL	*FOXP1* and *SIM1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr6	100813279	100943279	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr6	165330010	170908075	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	10239	3833474	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	66482565	72272064	DEL	*AUTS2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	66482565	72272064	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	72662064	74262064	DUP	

Table S3 continued

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	72662064	74262064	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	72742064	74142064	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	72742064	74142064	DEL	*ELN* and *GTF2I*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	74962064	76662064	DEL	*RHBDD2* and *HIP1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr7	74962064	76662064	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr8	160000	11912591	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr8	8092590	11892591	DEL	*SOX7* and *CLDN23*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr8	8092590	11892591	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr8	8212590	11912591	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	32010000	39010000	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	131060179	141080179	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	137810179	141080179	DEL	*EHMT1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	137810179	141080179	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	137860179	141080179	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr9	137860179	141080179	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr10	46929994	48429994	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr10	49389994	52389994	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr10	81690020	88940020	DEL	*SFTPD* to *GLUD1*, *NRG3*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr10	81960020	88800020	DEL	*NRG3* and *GRID1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr10	127760010	135400010	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr11	310000	3443424	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr11	43983424	46063424	DEL	*EXT2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr11	67753424	71282352	DEL	*SHANK2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr11	128044790	134844790	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr12	6469739	6809739	DUP	*SCNN1A* to *PIANP*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr12	65073733	68643733	DEL	*GRIP1* and *HMGA2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr14	104480247	106378955	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	22648636	28626405	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	22648636	31912708	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	22798636	23088559	DEL	*NIPA1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	24818907	28426405	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	24818907	28426405	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	30862708	32962708	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	31132708	32482708	DEL	*CHRNA7*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	72912946	75792945	DEL	*PML*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	72912946	74412947	DEL	*BBS4, NPTN, NEO1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	72962947	75532947	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	72962947	76012945	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	74012947	75532947	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	74012947	76012945	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	74012947	78132945	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	74012947	75532947	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	74412947	75592947	DEL	*CLK3, CSK*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	75592947	75792945	DEL	*SIN3A*

Table S3 continued

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	75972945	78202945	DEL	*FBXO22* and *TPSAN3*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	83182945	84738996	DEL	*HOMER2* and *BNC1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	85138996	85698996	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	99362477	102521392	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr15	99362477	102521392	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	160000	5159999	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	3779999	3859999	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	14892499	16892499	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	14892499	18292499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	15502499	16292499	DEL	*MYH11*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	15502499	16292499	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	21352499	29442499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	21612499	29042499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	21892499	22492499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	21942499	22462499	DEL	*EEF2K* and *CDR2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	28442499	30342499	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	28442499	30342499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	28772499	29112499	DEL	*SH2B1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	29652499	30202499	DEL	*TBX6*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	29652499	30202499	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr16	83792499	90222499	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	50000	2593250	DEL	*YWHAE* and *PAFAH1B1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	50000	2593250	DUP	*YWHAE* and *PAFAH1B1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	100000	4153251	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	553250	1353250	DEL	*PAFAH1B1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	553250	1353250	DUP	*PAFAH1B1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	2363250	2923250	DEL	*YWHAE*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	2363250	2923250	DUP	*YWHAE*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	16709275	20479408	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	16709275	20309408	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	16709275	20479408	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	29165874	30215887	DEL	*NF1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	34815887	36205887	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	34815887	36205887	DEL	*TCF2*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	43644217	44144178	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	43704217	44184217	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	43704217	44184217	DEL	*MAPT*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	57655218	58075218	DEL	*TUBD1* and *TMEM49*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	58065218	60305218	DEL	*TBX2* and *TBX4*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr17	72088405	81060000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr18	110000	5310000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr18	70949020	77899009	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr19	199000	5899000	DUP	

Table S3 continued

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr19	199000	8789000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr21	42478130	47975572	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	17470000	25020000	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	18820000	22270000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	19020000	20290000	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	19020000	20290000	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	21910000	23650000	DEL	*BCR* and *MAPK1*
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	21910000	23650000	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	44268667	51244566	DEL	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	47021336	51244566	DUP	
Cooper et al. 2012 (61) and Coe et al. 2014 (62)	chr22	51113134	51173134	DEL	*SHANK3*
Marshall et al 2017 (63)	chr1	145430996	148237104	DEL	
Marshall et al 2017 (63)	chr1	145430996	148237104	DUP	
Marshall et al 2017 (63)	chr2	49920350	51032536	DEL	*NRXN1*
Marshall et al 2017 (63)	chr3	196018732	197628732	DEL	
Marshall et al 2017 (63)	chr7	65373855	65401085	DEL	*ZNF92,* Protective
Marshall et al 2017 (63)	chr7	65373855	65401085	DUP	*ZNF92,* Protective
Marshall et al 2017 (63)	chr7	73328061	74727726	DUP	
Marshall et al 2017 (63)	chr7	158660506	159179546	DEL	*WDR60* and *VIPR2*
Marshall et al 2017 (63)	chr7	158660506	159179546	DUP	*WDR60* and *VIPR2*
Marshall et al 2017 (63)	chr8	100025494	100889814	DEL	*VPS13B*
Marshall et al 2017 (63)	chr9	841690	969090	DEL	*DMRT1*
Marshall et al 2017 (63)	chr9	841690	969090	DUP	*DMRT1*
Marshall et al 2017 (63)	chr13	20397624	20437776	DUP	*ZMYM5,* Protective
Marshall et al 2017 (63)	chr15	22784509	23074432	DEL	
Marshall et al 2017 (63)	chr15	30840505	32190507	DEL	
Marshall et al 2017 (63)	chr16	28811178	29041178	DEL	
Marshall et al 2017 (63)	chr16	29641178	30191178	DUP	
Marshall et al 2017 (63)	chr22	19032487	21065711	DEL	
Marshall et al 2017 (63)	chr22	19032487	21065711	DUP	Protective
Marshall et al 2017 (63)	chrX	148793685	148798928	DUP	*MAGEA11*, Protective
Marshall et al 2017 (63)	chrX	154918531	155342497	DUP	
Moreno del luca et al 2013 (64)	chr1	144000000	144340000	DUP	
Moreno del luca et al 2013 (64)	chr1	144000000	144340000	DEL	
Moreno del luca et al 2013 (64)	chr1	145040000	145860000	DEL	
Moreno del luca et al 2013 (64)	chr1	145040000	145860000	DUP	
Moreno del luca et al 2013 (64)	chr1	145040000	145860000	DEL	
Moreno del luca et al 2013 (64)	chr1	145040000	145860000	DUP	
Moreno del luca et al 2013 (64)	chr3	197230000	198840000	DUP	
Moreno del luca et al 2013 (64)	chr3	197230000	198840000	DEL	
Moreno del luca et al 2013 (64)	chr5	175650000	176990000	DUP	
Moreno del luca et al 2013 (64)	chr5	175650000	176990000	DEL	
Moreno del luca et al 2013 (64)	chr7	72380000	73780000	DEL	
Moreno del luca et al 2013 (64)	chr7	72380000	73780000	DUP	

Table S3 continued

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Moreno del luca et al 2013 (64)	chr8	8130000	11930000	DEL	
Moreno del luca et al 2013 (64)	chr8	8130000	11930000	DUP	
Moreno del luca et al 2013 (64)	chr10	81950000	88790000	DUP	
Moreno del luca et al 2013 (64)	chr15	22370000	26100000	DEL	
Moreno del luca et al 2013 (64)	chr15	22370000	26100000	DUP	
Moreno del luca et al 2013 (64)	chr15	22370000	26100000	DUP	
Moreno del luca et al 2013 (64)	chr15	28920000	30270000	DEL	
Moreno del luca et al 2013 (64)	chr15	28920000	30270000	DUP	
Moreno del luca et al 2013 (64)	chr15	28920000	30270000	DEL	
Moreno del luca et al 2013 (64)	chr15	28920000	30270000	DUP	
Moreno del luca et al 2013 (64)	chr16	15410000	16200000	DUP	
Moreno del luca et al 2013 (64)	chr16	15410000	16200000	DEL	
Moreno del luca et al 2013 (64)	chr16	21850000	22370000	DUP	
Moreno del luca et al 2013 (64)	chr16	21850000	22370000	DEL	
Moreno del luca et al 2013 (64)	chr16	28680000	29020000	DEL	
Moreno del luca et al 2013 (64)	chr16	28680000	29020000	DUP	
Moreno del luca et al 2013 (64)	chr16	28680000	29020000	DEL	
Moreno del luca et al 2013 (64)	chr16	28680000	29020000	DUP	
Moreno del luca et al 2013 (64)	chr16	29560000	30110000	DEL	
Moreno del luca et al 2013 (64)	chr16	29560000	30110000	DUP	
Moreno del luca et al 2013 (64)	chr16	29560000	30110000	DEL	
Moreno del luca et al 2013 (64)	chr16	29560000	30110000	DUP	
Moreno del luca et al 2013 (64)	chr17	16650000	20420000	DEL	
Moreno del luca et al 2013 (64)	chr17	16650000	20420000	DUP	
Moreno del luca et al 2013 (64)	chr17	26190000	27240000	DUP	
Moreno del luca et al 2013 (64)	chr17	31890000	33280000	DEL	
Moreno del luca et al 2013 (64)	chr17	31890000	33280000	DUP	
Moreno del luca et al 2013 (64)	chr17	31890000	33280000	DEL	
Moreno del luca et al 2013 (64)	chr17	41060000	41540000	DUP	
Moreno del luca et al 2013 (64)	chr22	17400000	18670000	DEL	
Moreno del luca et al 2013 (64)	chr22	17400000	18670000	DUP	
Moreno del luca et al 2013 (64)	chr22	17400000	18670000	DEL	
Moreno del luca et al 2013 (64)	chr22	17400000	18670000	DUP	
Moreno del luca et al 2013 (64)	chr22	20240000	21980000	DEL	
Moreno del luca et al 2013 (64)	chr22	20240000	21980000	DUP	
Stefansson et al. (65)	chr1	146089254	147859944	DEL	
Stefansson et al. (65)	chr1	146089254	147858944	DUP	
Stefansson et al. (65)	chr2	1792885	2335045	DUP	*MYT1L*
Stefansson et al. (65)	chr2	50145643	51259674	DEL	*NRXN1*
Stefansson et al. (65)	chr3	195766737	197216349	DEL	
Stefansson et al. (65)	chr7	157860945	159119486	DEL	*WDR60* and *VIPR2*
Stefansson et al. (65)	chr7	158726462	158947294	DUP	*WDR60* and *VIPR2*
Stefansson et al. (65)	chr10	46508694	51912781	DEL	

Table S3 continued

Reference	Chr	Start hg19	Stop hg19	Type	Note protective CNV
Stefansson et al. (65)	chr10	47543322	51912781	DUP	
Stefansson et al. (65)	chr13	93879078	95060273	DUP	*GPC6*
Stefansson et al. (65)	chr15	22750305	23272733	DEL	
Stefansson et al. (65)	chr15	22770994	28535266	DUP	
Stefansson et al. (65)	chr15	28973396	30556183	DUP	
Stefansson et al. (65)	chr15	29562640	30689724	DEL	
Stefansson et al. (65)	chr15	30936285	32515849	DEL	
Stefansson et al. (65)	chr15	32018731	32620127	DEL	*CHRNA7*
Stefansson et al. (65)	chr16	14989844	16291983	DUP	
Stefansson et al. (65)	chr16	15125441	16291983	DEL	
Stefansson et al. (65)	chr16	21947230	22423698	DEL	
Stefansson et al. (65)	chr16	28814098	29043450	DEL	
Stefansson et al. (65)	chr16	29595483	30192561	DEL	
Stefansson et al. (65)	chr16	29624247	30198151	DUP	
Stefansson et al. (65)	chr17	14101029	15471179	DEL	
Stefansson et al. (65)	chr17	34815551	36249430	DEL	
Stefansson et al. (65)	chr17	34815551	36249430	DUP	
Stefansson et al. (65)	chr22	20718116	21465780	DUP	
Stefansson et al. (65)	chr22	20733495	21465780	DEL	
Huguet et al. (3)	chr1	145430996	148237104	DEL	
Huguet et al. (3)	chr1	145430996	148237104	DUP	
Huguet et al. (3)	chr2	49920350	51032536	DEL	*NRXN1*
Huguet et al. (3)	chr3	196018732	197628732	DEL	
Huguet et al. (3)	chr7	73328061	74727726	DUP	
Huguet et al. (3)	chr7	65373855	65401085	DEL	*ZNF92*
Huguet et al. (3)	chr7	65373855	65401085	DUP	*ZNF92*
Huguet et al. (3)	chr7	158660506	159179546	DEL	*VIPR2* and *WDR60*
Huguet et al. (3)	chr7	158660506	159179546	DUP	*VIPR2* and *WDR60*
Huguet et al. (3)	chr8	99013266	99877580	DEL	*VPS13B*
Huguet et al. (3)	chr9	841690	969090	DEL	*DMRT1*
Huguet et al. (3)	chr9	841690	969090	DUP	*DMRT1*
Huguet et al. (3)	chr13	19837453	19863633	DUP	*ZMYM5*
Huguet et al. (3)	chr15	30840505	32190507	DEL	
Huguet et al. (3)	chr15	22784509	23074432	DEL	
Huguet et al. (3)	chr16	29641178	30191178	DUP	
Huguet et al. (3)	chr16	28811178	29041178	DEL	
Huguet et al. (3)	chr22	19032487	21065711	DEL	
Huguet et al. (3)	chr22	19032487	21065711	DUP	

Chr: Chromosome; hg19: Homo sapiens (human) *genome* assembly GRCh37 (*hg19*) from *Genome* Reference Consortium; DEL: deletion; DUP: duplication.

Table S4: Sensitivity analysis of the effect of gene dosage measured by pLI on NVIQ in autistic probands from SSC and MSSNG, and in the unselected population.

Phenotype	Models	N	pLI DEL			pLI DUP		
			β	SE	p	β	SE	p
NVIQ	All SSC probands	2,564	**-0.17**	**0.03**	**8.29×10^{-10}**	**-0.06**	**0.02**	**1.50×10^{-3}**
			No significant interaction with the type of NVIQ test			No significant interaction with the type of NVIQ test		
	SSC probands after 1:2 matching with MSSNG probands	2,245	**-0.14**	**0.03**	**100%***	**-0.06**	**0.02**	**96.8%***
	SSC probands after removing ratio NVIQ[a]	2,182	**-0.15**	**0.03**	**2.08×10^{-6}**	**-0.07**	**0.02**	**2.10×10^{-3}**
	All individuals (SSC probands + MSSNG probands + unselected population)	6,656	**-0.18**	**0.02**	**1.44×10^{-16}**	**-0.04**	**0.01**	**3.20×10^{-3}**
			Interaction pLI deletions and diagnosis p=0.86			Interaction pLI duplications and diagnosis p=0.16		
			Interaction pLI deletions and sex p=0.62			Interaction pLI duplications and sex p=0.51		
	Remove carriers of CNVs with a pLI > 10	6,629	**-0.18**	**0.03**	**2.43×10^{-11}**	-0.02	0.02	0.31
	Remove carriers of recurrent CNVs previously associated with NDD	6,484	**-0.20**	**0.03**	**3.42×10^{-11}**	**-0.05**	0.02	**9.53×10^{-3}**
	Remove carriers of rare *de novo* CNVs[b]	4,126	**-0.11**	**0.05**	**0.04**	-0.04	0.02	0.23

[a] Ratio NVIQ were provided for 382 very impaired probands (see supplementary Methods). [b] Information about the transmission of each CNV is not available in IMAGEN, the selection of rare CNVs are described in supplementary Methods. *percentage of the 500 matched samples providing an estimate with a p-value ≤ 0.05. Significative p-values are in bold (≤ 0.05). SE: Standard error; NVIQ: Non-verbal intelligence quotient; NDD: Neurodevelopmental disorders; pLI: probability of being Loss-of-function Intolerant; pLI DEL or pLI DUP: deleted or duplicated point of pLI score; CNV: Copy number variants.

Table S5: Sensitivity analysis of the effect of gene dosage measured by pLI on autism risk using autistic probands from SSC and MSSNG, unaffected siblings from SSC, and the unselected population.

Populations	Models	N individuals		CNV score	Not adjusted for NVIQ			Adjusted for NVIQ		
		Probands	Controls		OR	95%CI	p	OR	95%CI	p
Probands (SSC-MSSNG) Vs. Unselected population	All individuals	3,703	2,224	pLI DEL	**1.43**	**1.26-1.67**	**6.38×10^{-7}**	**1.28**	**1.11-1.53**	**2.14×10^{-3}**
				pLI DUP	**1.23**	**1.14-1.34**	**7.19×10^{-7}**	**1.21**	**1.11-1.34**	**3.56×10^{-5}**
				pLI DEL*sex	1.01	0.76-1.37	0.96	1.02	0.74-1.44	0.89
				pLI DUP*sex	1.06	0.90-1.26	0.51	1.07	0.89-1.29	0.50
	Remove carriers of CNVs with a pLI >10	3,680	2,222	pLI DEL	**1.42**	**1.24-1.66**	**2.05×10^{-6}**	**1.28**	**1.10-1.52**	**2.95×10^{-3}**
				pLI DUP	**1.26**	**1.15-1.39**	**3.01×10^{-6}**	**1.24**	**1.12-1.39**	**8.43×10^{-5}**
	Remove carriers of recurrent CNVs associated with NDD	3,578	2,193	pLI DEL	**1.40**	**1.19-1.73**	**4.19×10^{-4}**	1.21	0.99-1.56	0.10
				pLI DUP	**1.23**	**1.11-1.38**	**2.41×10^{-4}**	**1.21**	**1.08-1.38**	**2.09×10^{-3}**
	Remove carriers of rare *de novo* CNVs[a,b]	3,070	480	pLI DEL	1.22	0.91-1.84	0.26	1.20	0.86-1.90	0.34
				pLI DUP	0.92	0.82-1.05	0.22	0.89	0.78-1.02	0.08
	Removing the 10 individuals with autism risk from Imagen[c]	3,703	2,091	pLI DEL	**1.41**	**1.24-1.64**	**1.49×10^{-6}**	**1.26**	**1.09-1.50**	**3.46×10^{-3}**
				pLI DUP	**1.22**	**1.13-1.33**	**1.98×10^{-6}**	**1.20**	**1.10-1.33**	**7.99×10^{-5}**
SSC probands Vs. Siblings	Remove carriers of rare *de novo* CNVs[b,d]	1,950	1,950	pLI DEL	**1.44**	**1.03-2.01**	**0.03**	N.A.	N.A.	N.A.
				pLI DUP	**1.21**	**1.03-1.41**	**0.02**	N.A.	N.A.	N.A.
		Unaffected siblings	Imagen + SYS							
SSC unaffected siblings vs. Unselected population	All individuals	2,074	2,224	pLI DEL	1.03	0.88-1.21	0.69	N.A.	N.A.	N.A.
				pLI DUP	1.09	0.98-1.21	0.10	N.A.	N.A.	N.A.

[a] Information about the transmission of each CNV was not available in IMAGEN, this analysis was underpowered because of the lack of information on *de novo* CNVs in the unselected population; [b] the selection of rare CNVs are described in supplementary Methods, [c] 10 individuals from IMAGEN met criteria for Autism as estimated by the DAWBA (Development and Well-Being Assessment), we also excluded 124 individual without diagnostic information (DAWBA), [d] NVIQ was not available in unaffected siblings for the adjustment. OR: Odds ratio; 95%CI: 95% Confidence interval; CNV: Copy number variant; NVIQ: Non-verbal IQ; NDD: Neurodevelopmental disorders; pLI: probability of being Loss-of-function Intolerant; pLI DEL or pLI DUP: deleted or duplicated point of pLI score; pLI DEL*sex or DUP*sex: interaction between pLI and sex; N.A.: Not applicable.

Table S6: Breakpoints used to detect recurrent CNVs associated to autism and corresponding empirical and estimated odds ratios in autistic population.

Locus	Type	Chr	Start - Stop hg19 (Mb)	Sum of pLI	N autism cases[a]	N Controls [a]	Published autism risk [a]		Ref	Estimated autism risk[b]		Estimated loss of NVIQ points		Estimated gain of SRS points	
							OR	95%CI		OR	95%CI		95%CI		95%CI
1q21.1 (class I)	DEL	1	146.57-147.50	2.49	1/3,032	16/75,505	**1.56**	**0.21-11.74**	(66)	**2.15**	**1.57-2.93**	6.72	6.13-7.31	9.27	8.16-10.38
1q21.1 (class I)	DUP	1	146.57-147.50	2.49	8/3,032	19/57,730	8.03	3.51-18.37	(66)	1.44	1.19-1.74	1.49	1.20-1.78	4.65	3.81-5.49
3q29	DEL	3	195.73-197.34	6.56	1/2,120	1/63,649	**30.04**	**1.88-480.40**	(66)	**7.50**	**3.31-17.01**	17.71	17.12-18.30	24.42	23.31-25.53
5q35	DEL	5	175.65-176.99	11.92	1/3,955	0/13,696	∞	N.S.	(64)	38.91	8.79-172.24	32.18	31.59-32.77	44.38	43.27-45.49
7q11.23 (WBS)	DUP	7	72.72-74.15	10.27	4/2,120	1/16,257	**30.73**	**3.43-275.07**	(66)	**4.45**	**2.03-9.72**	6.16	5.87-6.45	19.16	18.32-20.00
15q11.2 (BP1-BP2)	DEL	15	22.75-23.27	1.70	8/2,525	19/7,086	**1.30**	**0.42-3.96**	(67)	**1.69**	**1.36-2.08**	4.59	4.00-5.18	6.33	5.22-7.44
15q11.2 (BP1-BP2)	DUP	15	22.75-23.27	1.70	20/2,525	38/7,086	**1.80**	**0.82-3.97**	(67)	**1.28**	**1.12-1.46**	1.02	0.73-1.31	3.17	2.33-4.01
15q13.3 (BP4-BP5)	DEL	15	30.92-32.51	1.71	4/2,120	13/74,106	10.77	3.51-33.07	(66)	1.69	1.37-2.09	4.62	4.03-5.21	6.37	5.26-7.48
15q13.3 (BP4-BP5)	DUP	15	30.92-32.51	1.71	2/3,955	5/13,696	**1.39**	**0.27-7.14**	(64)	**1.28**	**1.13-1.46**	1.03	0.74-1.32	3.19	2.35-4.03
16p11.2 (BP4-BP5)	DEL	16	29.60-30.30	9.92	18/4,315	25/56,752	**9.50**	**5.18-17.43**	(66)	**21.05**	**6.10-72.26**	26.78	26.19-27.37	36.93	35.82-38.04
16p11.2 (BP4-BP5)	DUP	16	29.60-30.30	9.92	17/4,315	19/56,752	**11.81**	**6.13-22.74**	(66)	**4.23**	**1.99-9.00**	5.95	5.66-6.24	18.51	17.67-19.35
16p11.2 distal	DEL	16	28.81-29.04	3.82	1/3,955	2/13,696	**1.73**	**0.16-19.10**	(64)	**3.23**	**2.01-5.21**	10.31	9.72-10.90	14.22	13.11-15.33
16p11.2 distal	DUP	16	28.81-29.04	3.82	1/3,955	3/13,696	**1.15**	**0.12-11.10**	(64)	**1.74**	**1.30-2.33**	2.29	2.00-2.58	7.13	6.29-7.97
16p13.11	DEL	16	15.12-16.29	2.84	5/3,955	4/13,696	**4.34**	**1.16-16.14**	(64)	**2.39**	**1.68-3.41**	7.67	7.08-8.26	10.57	9.46-11.68
16p13.11	DUP	16	15.12-16.29	2.84	4/2,120	81/62,973	**1.47**	**0.54-4.01**	(66)	**1.51**	**1.22-1.88**	1.70	1.41-1.99	5.30	4.46-6.14
17p11.2 (SMS)	DEL	17	16.58-20.33	12.21	2/4,687	0/151,619	∞	N.S.	(68)	42.54	9.27-195.22	32.97	32.38-33.56	45.46	44.35-46.57
17p11.2 (SMS)	DUP	17	16.58-20.33	12.21	1/4,687	1/151,619	**32.30**	**2.02-517.39**	(68)	**5.90**	**2.33-14.94**	7.33	7.04-7.62	22.79	21.95-23.63
17p12	DEL	17	14.04-15.41	1.17	2/2,120	14/59,086	**4.00**	**0.9-17.5**	(66)	**1.43**	**1.24-1.66**	3.16	2.57-3.75	4.36	3.25-5.47
17q12	DEL	17	34.81-36.25	5.07	2/2,120	4/68,131	**16.08**	**2.94-87.86**	(66)	**4.75**	**2.52-8.94**	13.69	13.10-14.28	18.87	17.76-19.98
22q11.2	DEL	22	18.89-21.90	11.44	5/4,687	5/151,619	**32.37**	**9.37-111.87**	(68)	**33.58**	**8.05-139.99**	30.89	30.30-31.48	42.59	41.48-43.70

22q11.2	DUP	22	18.89-21.90	11.44	12/4,315	23/27,133	**3.28**	**1.63-6.61**	(66)	**5.27**	**2.21-12.60**	6.86	6.57-7.15	21.35	20.51-22.19

[(a)]Based on number of carriers in autistic probands and controls reported in Malhotra et al. (2012) (66) Moreno DeLuca et al. (2013) (64), Chaste et al. (2014) (67), and Sanders et al. (2019) (68). In bold: CNVs for which previously published data and our estimation are overlapping. BP: Break points; WBS: William Beuren Syndrome; SMS: Smith Magenis Syndrome; DEL: deletion; DUP: duplication; chr: chromosome; hg19: Homo sapiens (human) *genome* assembly GRCh37 (*hg19*) from *Genome* Reference Consortium; pLI: probability of being Loss-of-function Intolerant; Sum of pLI: sum of score of pLI for the corresponding region; OR: Odds ratio; 95% CI: 95% confidence intervals; N.S.: Non-significant; Ref: reference.

Table S7: Description of models used for the investigation of the effect of gene dosage on phenotypical measures of autistic probands from SSC.

Phenotype	N	Type of normalization	Regression model	covariates
Autism related symptoms				
Regression	2,568	N.A.	Logistic	Sex, ancestry
Language and phonology				
CTOPP	1,988	z-scored with normative data: mean=10, SD=3	Linear	Sex, ancestry
Word delay	2,567	N.A.	Logistic	Sex, ancestry
Phrase delay	2,567	N.A.	Logistic	Sex, ancestry
Adaptive skills (VABS-II)				
Total score	2,569	z-scored with normative data: mean=100, SD=15	Linear	Ancestry
Daily living	2,569	z-scored with normative data: mean=100, SD=15	Linear	Ancestry
Communication	2,569	z-scored with normative data: mean=100, SD=15	Linear	Ancestry
Socialization	2,569	z-scored with normative data: mean=100, SD=15	Linear	Ancestry
Motor skills				
Motor VABS-II	919	z-scored with normative data: mean=100, SD=15	Linear	Ancestry
Gross motor VABS-II	926	z-scored with normative data: mean=15, SD=3	Linear	Ancestry
Fine motor VABS-II	923	z-scored with normative data: mean=15, SD=3	Linear	Ancestry
Delayed onset for walking	2,564	N.A.	Logistic	Sex, ancestry
Age of onset for walking	2,564	N.A.	Quasi-Poisson	Sex, ancestry
DCDQ score	2,209	z-scored with probands data: mean=38.5, SD=12.4	Linear	Age, sex, ancestry
Associated neurological condition				
Non-febrile seizure	2,566	N.A.	Logistic	Sex, ancestry

SD: Standard deviation; N.A.: Not applicable; CTOPP: Comprehensive Test of Phonological Processing; VABS-II: Vineland Adaptive behaviour Rating Scales - Second Edition; DCDQ: Developmental Coordination Disorder Questionnaire.

Table S8: Effect of gene dosage measured by pLI on SRS using autistic probands from SSC, their unaffected siblings and parents and the unselected population from IMAGEN.

Population	N	SRS-score	Model	CNV score	Effect-size (β or OR)	SE or 95%CI	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
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 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
MSSNG probands	598	Total-raw	Linear not adjusted for NVIQ	pLI DEL	0.71	1.78	0.69
				pLI DUP	-0.33	1.37	0.81
			Linear adjusted for NVIQ	pLI DEL	-0.70	1.76	0.69
				pLI DUP	-0.45	1.34	0.73
				NVIQ	**-0.30**	**0.06**	**$1.32x10^{-7}$**
IMAGEN	977	√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 probands + MSSNG probands	3,154	Total-raw	Linear not adjusted for NVIQ	pLI DEL	0.44	0.51	0.39
				pLI DUP	-0.27	0.45	0.54
			Linear adjusted for NVIQ	pLI DEL	-0.41	0.50	0.41
				pLI DUP	-0.52	0.43	0.23
				NVIQ	**-0.27**	**0.02**	**$1.40x10^{-46}$**
SSC probands + MSSNG probands + IMAGEN	4,131	Total-raw	Linear not adjusted for NVIQ or autism diagnosis	pLI DEL	**3.66**	**0.56**	**$5.53x10^{-11}$**
				pLI DUP	**1.64**	**0.42**	**$1.12x10^{-3}$**
			Linear adjusted for autism diagnosis	pLI DEL	0.56	0.35	0.11
				pLI DUP	-0.12	0.27	0.66
			Linear adjusted for NVIQ	pLI DEL	0.42	0.59	0.46
				pLI DUP	0.17	0.49	0.75
				NVIQ	**-0.52**	**0.02**	**$1.26x10^{-135}$**
			Linear adjusted for autism diagnosis and NVIQ	pLI DEL	-0.35	0.46	0.45
				pLI DUP	-0.45	0.39	0.25
				NVIQ	**-0.26**	**0.02**	**$2.63x10^{-52}$**
Unaffected siblings + Unaffected 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
	9,473	Total-raw		pLI DEL	**3.47**	**0.58**	**$2.40x10^{-9}$**

SSC probands + Unaffected siblings + Unaffected Parents			Linear mixed-effect not adjusted for autism diagnosis	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
All 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
All SSC + MSSNG + IMAGEN	11,081	Total-raw	Linear mixed-effect not adjusted for autism diagnosis	pLI DEL	**3.68**	**0.56**	**$4.30x10^{-11}$**
				pLI DUP	**1.63**	**0.42**	**$1.20x10^{-4}$**
			Linear mixed-effect adjusted for autism diagnosis	pLI DEL	0.56	0.35	0.11
				pLI DUP	-0.12	0.27	0.66
Probands + Unaffected siblings + IMAGEN	5,189	SRS categories (normal, clinical) [b]	Logistic regression not adjusted for autism diagnosis	pLI DEL	**1.20**	**1.10-1.33**	**$8.46x10^{-5}$**
				pLI DUP	**1.13**	**1.06-1.22**	**$2.00x10^{-4}$**
			Logistic regression adjusted for autism diagnosis	pLI DEL	0.96	0.84-1.15	0.59
				pLI DUP	0.97	0.86-1.12	0.66
Probands + Unaffected siblings + IMAGEN	5,189	SRS categories (normal, moderate, mild, clinically significant) [b]	Ordinal cumulative not adjusted for autism diagnosis	pLI DEL	**1.20**	**1.11-1.30**	**$3.92x10^{-6}$**
				pLI DUP	**1.11**	**1.05-1.18**	**$3.00x10^{-4}$**
			Ordinal cumulative adjusted for autism diagnosis	pLI DEL	1.05	0.96-1.15	0.26
				pLI DUP	0.99	0.93-1.06	0.87

All linear, logistic, or ordinal regression models used were adjusted for age, sex and ancestry. Models take into account family as random-effect when including related individuals (Methods). Effect-size are presented as β for linear regression models and as odds ratio for logistic and ordinal regression models. [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 S5); [b] Based on the previously published *T*-score categorization (55) (Methods). The statistical threshold after correction for multiple testing is $p \leq 2.7.10^{-3}$. Significant results are in bold. SE: Standard error; OR: Odds ratio; 95%CI: 95% Confidence interval; NVIQ: Non-verbal intelligence quotient; DEL: deletion; DUP: duplication; pLI: probability of being Loss-of-function Intolerant; pLI DEL or pLI DUP: deleted or duplicated point of pLI score; √Total-raw: square root transformation of the total SRS raw.

Table S9: Effect of gene dosage measured by pLI on autism severity scores (main domains of ADI-R and ADOS-calibrated severity scores) using autistic probands from SSC and MSSNG.

Phenotype	N	CNV score	Not adjusted for NVIQ			Adjusted for NVIQ		
			OR	95%CI	p	OR	95%CI	p
Probands from SSC								
ADI-R reciprocal social interactions	2,567	pLI DEL	1.00	0.94-1.07	0.89	0.93	0.87-0.99	0.04
		pLI DUP	1.06	1.01-1.11	0.02	1.03	0.98-1.08	0.20
ADI-R rrsb	2,567	pLI DEL	0.95	0.89-1.02	0.14	0.94	0.88-1.01	0.10
		pLI DUP	1.01	0.97-1.06	0.54	1.01	0.97-1.06	0.61
ADI-R verbal communication	2,254	pLI DEL	1.00	0.93-1.08	0.91	0.95	0.88-1.02	0.18
		pLI DUP	1.06	1.00-1.12	0.05	1.04	0.98-1.10	0.17
ADI-R non-verbal communication	2,567	pLI DEL	1.01	0.95-1.08	0.73	0.94	0.88-1.00	0.07
		pLI DUP	1.04	0.99-1.09	0.09	1.01	0.97-1.06	0.55
ADOS overall css[(a)]	2,499	pLI DEL	0.94	0.88-1.01	0.09	0.92	0.86-0.98	0.02
		pLI DUP	1.02	0.97-1.07	0.41	1.01	0.96-1.06	0.73
ADOS social affect css[(a)]	2,371	pLI DEL	0.96	0.92-1.03	0.25	0.94	0.87-1.01	0.08
		pLI DUP	1.04	0.99-1.10	0.08	1.03	0.98-1.08	0.19
ADOS rrsb css[(a)]	2,450	pLI DEL	0.98	0.91-1.05	0.55	0.95	0.88-1.02	0.12
		pLI DUP	1.01	0.96-1.06	0.67	0.99	0.98-1.10	0.85
ADI-R Overall level of language	2,568	pLI DEL	1.08	0.98-1.19	0.13	0.93	0.83-1.04	0.18
		pLI DUP	1.06	1.00-1.13	0.04	1.03	0.97-1.10	0.36
Probands from MSSNG								
ADI-R reciprocal social interactions	397	pLI DEL	1.26	0.85-1.85	0.25	1.20	0.82-1.77	0.35
		pLI DUP	0.98	0.77-1.25	0.85	0.99	0.77-1.26	0.93
ADI-R rrsb	695	pLI DEL	1.15	0.98-1.37	0.09	1.15	0.97-1.36	0.11
		pLI DUP	1.00	0.86-1.13	0.87	0.99	0.86-1.13	0.83
ADI-R verbal communication	370	pLI DEL	1.08	0.82-1.42	0.59	1.04	0.79-1.37	0.77
		pLI DUP	1.01	0.79-1.29	0.95	0.97	0.76-1.24	0.80
ADI-R non-verbal communication	73	pLI DEL	1.12	0.77-1.64	0.49	1.10	0.75-1.61	0.63
		pLI DUP	0.90	0.75-1.07	0.27	0.89	0.74-1.07	0.21
ADOS overall css[(a)]	733	pLI DEL	0.94	0.75-1.18	0.58	0.92	0.73-1.16	0.50
		pLI DUP	0.86	0.75-0.99	0.04	0.86	0.75-0.99	0.04
ADOS social affect css[(a)]	373	pLI DEL	0.74	0.50-1.11	0.15	0.74	0.50-1.11	0.15
		pLI DUP	0.90	0.76-1.07	0.23	0.90	0.76-1.07	0.23
ADOS rrsb css[(a)]	388	pLI DEL	0.91	0.61-1.36	0.64	0.91	0.61-1.36	0.63
		pLI DUP	0.95	0.83-1.12	0.60	0.96	0.82-1.11	0.57
ADI-R Overall level of language	1,267	pLI DEL	1.27	1.03-1.21	$7.19.10^{-3}$	1.13	0.94-1.37	0.19
		pLI DUP	1.05	1.03-1.21	0.49	1.05	0.92-1.20	0.44

Pooled probands (SSC + MSSNG)								
ADI-R reciprocal social interactions	2,966	pLI DEL	1.01	0.95-1.08	0.69	0.94	0.88-1.01	0.08
		pLI DUP	1.05	1.00-1.10	0.03	1.03	0.98-1.08	0.24
ADI-R rrsb	3,264	pLI DEL	0.98	0.92-1.04	0.52	0.97	0.91-1.04	0.40
		pLI DUP	1.01	0.97-1.06	0.58	1.01	0.97-1.06	0.66
ADI-R verbal communication	2,626	pLI DEL	1.01	0.94-1.08	0.80	0.95	0.89-1.03	0.20
		pLI DUP	1.06	0.99-1.12	0.05	1.04	0.98-1.10	0.20
ADI-R non-verbal communication	2,642	pLI DEL	1.02	0.95-1.09	0.65	0.94	0.88-1.01	0.09
		pLI DUP	1.03	0.99-1.08	0.18	1.01	0.96-1.05	0.82
ADOS overall css[(a)]	3,122	pLI DEL	0.94	0.88-1.01	0.08	0.92	0.86-0.98	0.01
		pLI DUP	0.99	0.95-1.04	0.88	0.99	0.94-1.03	0.56
ADOS social affect css[(a)]	2,675	pLI DEL	0.95	0.89-1.02	0.16	0.93	0.87-1.00	0.05
		pLI DUP	1.03	0.98-1.08	0.25	1.02	0.97-1.07	0.44
ADOS rrsb css[(a)]	2,766	pLI DEL	0.98	0.91-1.05	0.50	0.94	0.88-1.01	0.11
		pLI DUP	1.00	0.96-1.05	0.91	0.99	0.95-1.04	0.67
ADI-R Overall level of language	3,607	pLI DEL	1.11	1.02-1.21	0.01	0.97	0.88-1.06	0.50
		pLI DUP	1.06	1.00-1.12	0.03	1.04	0.98-1.10	0.23

All ordinal regression models used for each severity score were adjusted for age and sex, and ancestry when available (for autistic probands from SSC only). [(a)]Calibrated severity score were computed based on previously published methodology from Hus et al. (2014) (40). The statistical threshold after correction for multiple testing is $p \leq 2.7.10^{-3}$. NVIQ: Non-verbal intelligence quotient; OR: Odds ratio; 95%CI: 95% Confidence interval; ADI-R: Autism Diagnostic Interview-Revised; ADOS: Autism Diagnostic Observation Schedule; css: calibrated severity score; rrsb: repetitive, restricted and stereotyped behaviours; DEL: deletion; DUP: duplication; pLI: probability of being Loss-of-function Intolerant; pLI DEL or pLI DUP: deleted or duplicated point of pLI score.

Table S10: Effect of gene dosage measured by pLI on CBCL using autistic probands from SSC and unaffected siblings.

Population	N	Model	CNV score	OR	95%CI	p
CBCL total Problems raw score						
Probands	1,945	Negative Binomial not adjusted for NVIQ	pLI DEL	1.01	0.99-1.03	0.43
			pLI DUP	1.00	0.98-1.01	0.64
		Negative Binomial adjusted for NVIQ	pLI DEL	1.01	0.99-1.03	0.40
			pLI DUP	1.00	0.98-1.01	0.67
Unaffected Siblings	1,596	Negative Binomial not adjusted for NVIQ	pLI DEL	1.07	0.98-1.20	0.13
			pLI DUP	1.05	0.99-1.13	0.10
Probands + Unaffected siblings	3,541	Negative Binomial mixed-effect not adjusted for autism diagnosis	pLI DEL	**1.05**	**1.03-1.08**	**$1.94x10^{-6}$**
			pLI DUP	1.02	1.01-1.04	$3.04x10^{-3}$
		Negative Binomial mixed-effect adjusted for autism diagnosis	pLI DEL	1.01	1.00-1.03	0.11
			pLI DUP	1.00	1.00-1.02	0.53
CBCL externalizing Problems raw score						
Probands	1,945	Negative Binomial not adjusted for NVIQ	pLI DEL	1.01	0.98-1.05	0.37
			pLI DUP	1.00	0.97-1.02	0.91
		Negative Binomial adjusted for NVIQ	pLI DEL	1.01	0.98-1.05	0.41
			pLI DUP	1.00	0.97-1.02	0.87
Unaffected Siblings	1,596	Negative Binomial not adjusted for NVIQ	pLI DEL	1.10	0.97-1.28	0.12
			pLI DUP	1.04	0.95-1.13	0.40
Probands + Unaffected siblings	3,541	Negative Binomial mixed-effect not adjusted for autism diagnosis	pLI DEL	**1.05**	**1.02-1.09**	**$4.94x10^{-4}$**
			pLI DUP	1.02	1.00-1.05	0.02
		Negative Binomial mixed-effect adjusted for autism diagnosis	pLI DEL	1.02	0.99-1.05	0.14
			pLI DUP	1.01	0.98-1.03	0.56
CBCL internalizing Problems raw score						
Probands	1,945	Negative Binomial not adjusted for NVIQ	pLI DEL	0.99	0.97-1.02	0.63
			pLI DUP	0.98	0.96-1.00	0.07
		Negative Binomial adjusted for NVIQ	pLI DEL	1.01	0.98-1.04	0.57
			pLI DUP	0.99	0.96-1.01	0.19
Unaffected Siblings	1,596	Negative Binomial not adjusted for NVIQ	pLI DEL	1.08	0.97-1.22	0.15
			pLI DUP	1.06	0.99-1.14	0.10
Probands + Unaffected siblings	3,541	Negative Binomial mixed-effect not adjusted for autism diagnosis	pLI DEL	**1.04**	**1.01-1.07**	**$2.12x10^{-3}$**
			pLI DUP	1.01	0.99-1.03	0.19
		Negative Binomial mixed-effect adjusted for autism diagnosis	pLI DEL	1.01	0.98-1.03	0.61
			pLI DUP	0.99	0.97-1.01	0.52

All negative binomial models used were adjusted for age, sex and ancestry. Models take into account family as random-effect when including related individuals (Methods). Effect-size are presented as odds ratio. The statistical threshold after correction for multiple testing is $p \leq 2.7.10^{-3}$. Significant results are in bold. OR: Odds ratio; 95%CI: 95% Confidence interval; CBCL: Child behaviour Checklist; NVIQ: Non-verbal intelligence quotient; DEL: deletion; DUP: duplication; pLI: probability of being Loss-of-function Intolerant; pLI DEL or pLI DUP: deleted or duplicated point of pLI score.

Table S11: Effect of gene dosage measured by pLI on general intelligence in autistic probands from SSC and MSSNG, and in the unselected population.

Population	Measure	N	Covariates	CNV score	β	SE	p
Probands (SSC)	NVIQ	2,564	Sex, type of test used, ancestry	pLI DEL	**-0.17**	**0.03**	**8.29×10^{-10}**
				pLI DUP	**-0.06**	**0.02**	**1.50×10^{-3}**
	NVIQ-DAS	2,244	Sex, ancestry	pLI DEL	**-0.16**	**0.03**	**1.20×10^{-7}**
				pLI DUP	**-0.07**	**0.02**	**1.10×10^{-3}**
	NVR	1,958	Sex, ancestry	pLI DEL	**-0.10**	**0.03**	**4.60×10^{-4}**
				pLI DUP	**-0.06**	**0.02**	**5.90×10^{-3}**
	Matrices	1,958	Sex, ancestry	pLI DEL	**-0.09**	**0.03**	**9.30×10^{-4}**
				pLI DUP	**-0.04**	**0.02**	**4.80×10^{-2}**
Unselected	NVIQ	2,710	Sex, type of test used, ancestry	pLI DEL	**-0.19**	**0.04**	**6.90×10^{-5}**
				pLI DUP	0.02	0.03	0.52
Probands (MSSNG)	NVIQ	1,381	Sex, type of test used	**pLI DEL**	**-0.20**	**0.07**	**3.10×10^{-3}**
				pLI DUP	-0.02	0.04	0.61

All linear regression models were performed using a z-scored dependant variable. These z-scores are computed using normative data (*e.g.* mean NVIQ=100, SD NVIQ=15). CNV: Copy number variant; SE: Standard error; NVIQ: Non-verbal IQ; DAS: Differential Ability Scales; NVR: Non-verbal reasoning; pLI DEL or pLI DUP: deleted or duplicated point of pLI score.

Table S12: Effect of gene dosage measured by pLI on autism risk.

Group comparison	N		pLI DEL					pLI DUP				
	probands	controls	Sum pLI in probands	Sum pLI in controls	OR	95%CI	p	Sum pLI in probands	Sum pLI in controls	OR	95%CI	p
Not adjusted for NVIQ												
Probands vs. Unaffected Siblings	2,074	2,074	346.58	117.67	**1.43**	**1.23-1.66**	**3.78×10^{-6}**	640.32	371.62	**1.32**	**1.17-1.49**	**4.63×10^{-6}**
Probands vs. Unselected population	2,569	2,223	416.00	114.11	**1.40**	**1.23-1.64**	**2.33×10^{-6}**	816.99	330.31	**1.30**	**1.19-1.42**	**1.88×10^{-8}**
Adjusted for NVIQ												
Probands vs. unselected population	2,569	2,223	416.00	114.11	**1.22**	**1.05-1.45**	**0.01**	816.99	330.31	**1.27**	**1.15-1.42**	**4.89×10^{-6}**
Probands vs. unselected population	1,438	1,438	153.80	93.51	1.21	1.01-1.45	60.4 %*	357.01	219.40	**1.25**	**1.09-1.43**	**98.2 %***
Replication with probands from MSSNG												
Not adjusted												
Probands vs. Unselected population	1,139	2,223	156.41	114.11	**1.54**	**1.30-1.88**	**4.56×10^{-6}**	298.32	330.31	**1.19**	**1.08-1.33**	**8.30×10^{-4}**
Adjusted for NVIQ												
Probands vs. unselected population	1,139	2,223	156.41	114.11	**1.39**	**1.15-1.72**	**1.44×10^{-3}**	298.32	330.31	**1.20**	**1.08-1.35**	**1.44×10^{-3}**

Odds ratios are computed using a logistic regression including sum of pLI in deletions and duplications as the two main explanatory variables. OR represents the autism risk conferred for each deleted or duplicated point of pLI. Sum of pLI represents the sum of all genes deleted or duplicated in all individuals for the group. *percentage of the 500 matched samples providing an estimate with a p-value ≤ 0.05. Significative Significative p-values are in bold (≤ 0.05). pLI: probability of being Loss-of-function Intolerant; sum of pLI: sum of score of pLI in the entire population; pLI DEL or pLI DUP: deleted or duplicated point of pLI; NVIQ: Non-verbal intelligence quotient; OR: Odds ratio; 95%CI: 95% Confidence interval.

Table S13: Autism risk potentially mediated by NVIQ in the pooled dataset (SSC, MSSNG, unselected populations).

Populations	N individuals probands	N individuals controls	Effect	Variable	OR	95% CI	p
Probands (SSC-MSSNG) *vs.* Unselected populations	3,703	2,224	Direct effect	pLI DEL	**1.23**	**[1.08-1.41]**	**2.28×10^{-3}**
				pLI DUP	**1.18**	**[1.10-1.26]**	**1.00×10^{-6}**
			Indirect effect	pLI DEL	**1.17**	**[1.13-1.21]**	**2.00×10^{-16}**
				pLI DUP	**1.06**	**[1.03-1.08]**	**6.56×10^{-7}**
			Total effect	pLI DEL	**1.45**	**[1.26-1.65]**	**7.33×10^{-8}**
				pLI DUP	**1.25**	**[1.16-1.34]**	**2.92×10^{-9}**

Odds ratios are computed using a counterfactual-based mediation analysis on two different logistic regression including sum of pLI in deletions and duplications as the main explanatory variables. OR represents the autism risk conferred for each deleted or duplicated point of pLI. Direct effects of CNVs are those not mediated by NVIQ. Indirect are those potentially mediated by NVIQ. Total effects are those computed without adjusting for NVIQ. Significative p-values are in bold (≤ 0.05). pLI: probability of being Loss-of-function Intolerant. pLI DEL or pLI DUP: deleted or duplicated point of pLI. NVIQ: Non-verbal intelligence quotient; OR: Odds ratio; 95%CI: 95% Confidence interval.

Table S14: Autism risk measured by pLI in subgroups of individual below or above the median (98) in the pooled dataset (SSC, MSSNG, unselected populations).

Populations	N individuals probands	N individuals controls	IQ subgroup	Variable	OR	95% CI	p
Probands (SSC-MSSNG) *vs.* Unselected populations	2,363	667	Below median NVIQ	pLI DEL	**1.27**	**[1.10-1.53]**	**3.41×10^{-3}**
				pLI DUP	**1.34**	**[1.16-1.61]**	**5.57×10^{-4}**
	1,340	1,557	Above median NVIQ	pLI DEL	**1.53**	**[1.19-2.07]**	**2.17×10^{-3}**
				pLI DUP	**1.16**	**[1.04-1.31]**	**1.10×10^{-2}**

Odds ratios are computed using a logistic regression including sum of pLI in deletions and duplications as the two main explanatory variables. OR represents the autism risk conferred for each deleted or duplicated point of pLI. Significative p-values are in bold (≤ 0.05). pLI: probability of being Loss-of-function Intolerant. pLI DEL or pLI DUP: deleted or duplicated point of pLI. NVIQ: Non-verbal intelligence quotient; OR: Odds ratio; 95%CI: 95% Confidence interval.

Table S15: Estimated Genome wide effects of gene dosage on autism risk.

	ALL CNVs (including pLI=0)				CNVs with pLI > 0			
	>50Kb in autistic sample		1MB genome wide		>50Kb in autistic sample		1MB genome wide	
	OR DEL	OR DUP	OR DEL	OR DUP	OR DEL	OR DUP	OR DEL	OR DUP
N CNVs	5,319	4,614	5,586	5,586	2,136	2,045	4,377	4,377
Median	1.00	1.00	1.36	1.21	1.02	1.05	1.58	1.32
Mean	1.15	1.13	3.05	1.58	1.36	1.31	3.62	1.75
70%	1.00	1.00	1.85	1.45	1.02	1.17	2.13	1.58
75%	1.01	1.02	2.04	1.54	1.09	1.2	2.38	1.69
80%	1.02	1.05	2.35	1.68	1.09	1.21	2.66	1.82
90%	1.09	1.21	3.72	2.23	1.35	1.33	4.43	2.48
95%	1.31	1.33	6.65	3.17	1.60	1.63	8.28	3.63

pLI: probability of being Loss-of-function Intolerant; NVIQ: Non-verbal intelligence quotient; OR DEL/DUP: Estimated odds ratio for deletions or duplications based on pLI score. Odds ratio represents the estimated autism risk conferred by deleted or duplicated points of pLI in the autistic sample (probands from SSC and MSSNG) and in a genome of reference (Human Gene Nomenclature) (69) partitioned in CNV of 1Mb.

SUPPLEMENTARY REFERENCES

1. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub, 2013
2. Bell CC: DSM-IV: Diagnostic and Statistical Manual of Mental Disorders. JAMA 1994; 272:828–829
3. Huguet G, Schramm C, Douard E, et al.: Measuring and Estimating the Effect-sizes of Copy Number Variants on General Intelligence in Community-Based Samples. JAMA Psychiatry 2018;
4. Fischbach GD, Lord C: The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron 2010; 68:192–195
5. Schumann G, Loth E, Banaschewski T, et al.: The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. Mol Psychiatry 2010; 15:1128–1139
6. Pausova Z, Paus T, Abrahamowicz M, et al.: Cohort Profile: The Saguenay Youth Study (SYS) [Internet]. Int J Epidemiol [cited 2017 May 1] Available from: https://academic.oup.com/ije/article/doi/10.1093/ije/dyw023/2617159/Cohort-Profile-The-Saguenay-Youth-Study-SYS
7. Yuen RKC, Merico D, Bookman M, et al.: Whole genome sequencing resource identifies 18 new candidate genes for autism spectrum disorder. Nat Neurosci 2017; 20:602–611
8. Van der Auwera GA, Carneiro MO, Hartl C, 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
9. Wang K, Li M, Hadley D, 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; 17:1665–1674
10. Colella S, Yau C, Taylor JM, 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; 35:2013–2025
11. Sanders SJ, Ercan-Sencicek AG, Hus V, et al.: Multiple Recurrent De Novo CNVs, Including Duplications of the 7q11.23 Williams Syndrome Region, Are Strongly Associated with Autism. Neuron 2011; 70:863–885
12. Trost B, Walker S, Wang Z, et al.: A Comprehensive Workflow for Read Depth-Based Identification of Copy-Number Variation from Whole-Genome Sequence Data. Am J Hum Genet 2018; 102:142–155
13. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38:e164
14. Ramasamy A, Trabzuni D, Guelfi S, et al.: Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014; 17:1418–1428
15. Lek M, Karczewski KJ, Minikel EV, et al.: Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016; 536:285–291
16. Petrovski S, Gussow AB, Wang Q, et al.: The Intolerance of Regulatory Sequence to Genetic Variation Predicts Gene Dosage Sensitivity. PLoS Genet 2015; 11:e1005492
17. Ruderfer DM, Hamamsy T, Lek M, et al.: Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. Nat Genet 2016; 48:1107–1111

18. Szklarczyk D, Franceschini A, Wyder S, et al.: STRING v10: protein–protein interaction networks, integrated over the tree of life. Nucleic Acids Res 2015; 43:D447–D452
19. Hawrylycz M, Miller JA, Menon V, et al.: Canonical genetic signatures of the adult human brain. Nat Neurosci 2015; 18:1832–1844
20. Krumm N, Turner TN, Baker C, et al.: Excess of rare, inherited truncating mutations in autism. Nat Genet 2015; 47:582–588
21. Pinto D, Delaby E, Merico D, et al.: Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. Am J Hum Genet 2014; 94:677–694
22. Pinto D, Pagnamenta AT, Klei L, et al.: Functional impact of global rare copy number variation in autism spectrum disorders. Nature 2010; 466:368–372
23. Hapmap Consortium: The International HapMap Project. Nature 2003; 426:789–96
24. Purcell S, Neale B, Todd-Brown K, et al.: PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81:559–75
25. Alexander DH, Novembre J, Lange K: Fast model-based estimation of ancestry in unrelated individuals [Internet]. Genome Res 2009; [cited 2017 Mar 14] Available from: http://genome.cshlp.org/content/early/2009/07/31/gr.094052.109
26. Elliott CD, Salerno JD, Dumont R, et al.: The Differential Ability Scales—Second Edition, inContemporary intellectual assessment: Theories, tests, and issues, 4th ed. New York, NY, US, The Guilford Press, 2007, pp 360–382.
27. Wechsler D: Wechsler Intelligence Scale for Children - Fourth Edition2003;
28. Wechsler D: Wechsler Abbreviated Scale of Intelligence1999;
29. Mullen EM: Mullen Scales of Early Learning1995;
30. Bishop SL, Guthrie W, Coffing M, et al.: 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; 116:331–343
31. Leiter RG: Leiter international performance scale1979;
32. Gale H. Roid, Miller LJ: Leiter international performance scale-revised1997;
33. Raven JC, Court JH, Raven J: Raven's Progressive Matrices1998;
34. Wechsler D: Wechsler Abbreviated Scale of Intelligence - Second Edition2011;
35. Wechsler D: Wechsler Preschool and Primary Scale of Intelligence - Fourth Edition2012;
36. Dawson M, Soulières I, Gernsbacher MA, et al.: The level and nature of autistic intelligence. Psychol Sci 2007; 18:657–662
37. Constantino J, Gruber C: Social Responsiveness Scale (SRS). Los Angel West Psychol Serv 2005;
38. Constantino JN, Davis SA, Todd RD, 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; 33:427–433
39. Lord C, Risi S, Lambrecht L, et al.: The Autism Diagnostic Observation Schedule—Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. J Autism Dev Disord 2000; 30:205–223
40. Hus V, Gotham K, Lord C: Standardizing ADOS Domain Scores: Separating Severity of Social Affect and Restricted and Repetitive Behaviors. J Autism Dev Disord 2014; 44:2400–2412
41. Lord C, Rutter M, Couteur AL: Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. J Autism Dev Disord 1994; 24:659–685

42. Achenbach TM, Rescorla LA: Manual for the ASEBA Preschool Forms & Profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families., 2000
43. Achenbach TM, Rescorla LA: Manual for the ASEBA school-age forms & profiles: an integrated system of mult-informant assessment. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families., 2001
44. Wagner R, Torgesen J, Rashotte C: Comprehensive test of phonological processing1999;
45. Bishop SL, Thurm A, Farmer C, et al.: Autism Spectrum Disorder, Intellectual Disability, and Delayed Walking. Pediatrics 2016; 137:e20152959
46. Sparrow SS, Balla DA, Cicchetti D: Vineland Adaptive Behavior Scales, Second Edition [Internet]. Bloomington, MN, Pearson Assessments, 2005Available from: https://www.pearsonclinical.com/psychology/products/100000668/vineland-adaptive-behavior-scales-second-edition-vineland-ii-vineland-ii.html
47. Efron B: Bootstrap Methods: Another Look at the Jackknife. Ann Stat 1979; 7:1–26
48. Rizopoulos D: Bootstrap stepAIC [Internet]. 2009Available from: https://CRAN.R-project.org/package=bootStepAIC
49. Pinheiro J, Bates D, DebRoy S, et al.: _nlme: Linear and Nonlinear Mixed Effects Models_. [Internet]. 2018Available from: https://CRAN.R-project.org/package=nlme
50. Therneau TM: _A Package for Survival Analysis in S_ [Internet]. 2015Available from: https://CRAN.R-project.org/package=survival
51. Jasjeet S, Sekhon: Multivariate and Propensity Score Matching Software with Automated Balance Optimization: The Matching Package for R. J Stat Softw 2011; 1–52
52. Loeys T, Moerkerke B, Smet OD, et al.: Flexible Mediation Analysis in the Presence of Nonlinear Relations: Beyond the Mediation Formula. Multivar Behav Res 2013; 48:871–894
53. Therneau TM, Sinnwell J: kinship2: Pedigree Functions [Internet]. 2015Available from: https://CRAN.R-project.org/package=kinship2
54. Therneau TM: coxme: Mixed Effects Cox Models [Internet]. 2018Available from: https://CRAN.R-project.org/package=coxme
55. Bruni TP: Test Review: Social Responsiveness Scale–Second Edition (SRS-2). J Psychoeduc Assess 2014; 32:365–369
56. Bates D, Maechler M, Walker S: Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015; 1–48
57. Yee TW: The VGAM Package for Categorical Data Analysis. J Stat Softw 2010; 1–34
58. Ripley B, Venables B, Bates D, et al.: MASS: Support Functions and Datasets for Venables and Ripley's MASS [Internet]. 2019Available from: http://www.stats.ox.ac.uk/pub/MASS4/
59. Magnusson A, Skaug H, Nielsen A, et al.: glmmTMB: Generalized Linear Mixed Models using Template Model Builder [Internet]. 2019Available from: https://github.com/glmmTMB
60. Cheverud JM: A simple correction for multiple comparisons in interval mapping genome scans. Heredity 2001; 87:52–58
61. Cooper GM, Coe BP, Girirajan S, et al.: A Copy Number Variation Morbidity Map of Developmental Delay. Nat Genet 2011; 43:838–846
62. Coe BP, Witherspoon K, Rosenfeld JA, et al.: Refining analyses of copy number variation identifies specific genes associated with developmental delay. Nat Genet 2014; 46:1063–1071

63. CNV and Schizophrenia Working Groups of the Psychiatric Genomics Consortium: Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet 2017; 49:27–35
64. Moreno-De-Luca D, Sanders SJ, Willsey AJ, et al.: Using large clinical data sets to infer pathogenicity for rare copy number variants in autism cohorts. Mol Psychiatry 2013; 18:1090–1095
65. Stefansson H, Meyer-Lindenberg A, Steinberg S, et al.: CNVs conferring risk of autism or schizophrenia affect cognition in controls. Nature 2014; 505:361–366
66. Malhotra D, Sebat J: CNVs: Harbingers of a Rare Variant Revolution in Psychiatric Genetics. Cell 2012; 148:1223–1241
67. Chaste P, Sanders SJ, Mohan KN, et al.: Modest impact on risk for autism spectrum disorder of rare copy number variants at 15q11.2, specifically breakpoints 1 to 2. Autism Res Off J Int Soc Autism Res 2014; 7:355–362
68. Sanders SJ, Sahin M, Hostyk J, et al.: A framework for the investigation of rare genetic disorders in neuropsychiatry. Nat Med 2019;
69. Wain HM, Bruford EA, Lovering RC, et al.: Guidelines for human gene nomenclature. Genomics 2002; 79:464–470

Supplementary material 2 (paper #2)

SUPPLEMENTARY TABLES

Table S1: Recurrent neurodevelopmental or neuropsychiatric-associated loci and genes investigated

Locus	CHR	START	STOP	TYPE	References for the association with NPD
1p36	1	1	2500000	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
1q21.1 TAR	1	145394955	145807817	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
1q21.1 distal+TAR	1	145394955	147394444	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
1q21.1 distal	1	146527987	147394444	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
NRXN1	2	50145643	51259674	Gene	Satterstrom et al. 2020
2q11.2	2	96742409	97677516	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
2q13	2	111394040	112012649	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
2q21.1	2	131481308	131930677	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
2q37	2	239716679	243199373	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
3q29	3	195720167	197354826	DEL	Sanders et al. 2019
				DUP	Coe et al. 2014
4p16.3	4	1552030	2091303	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
5q35	5	175720924	177052594	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
SIM1	6	100836750	100911811	Gene	Coe et al. 2014
7q11.23 WBS	7	72744915	74142892	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
7q11.23 distal	7	75138294	76064412	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
8p23.1	8	8098990	11872558	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
9q34	9	140513444	140730578	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
10q11.21q11.23	10	49390199	51058796	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
10q22q23	10	82045472	88931651	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
11p11.2	11	43940000	46020000	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
CRYL1	13	20977806	21100012	Gene	Coe et al. 2014
13q12.12	13	23555358	24884622	DEL	Coe et al. 2014

				DUP	Coe et al. 2014
15q11.2	15	22805313	23094530	DEL	Sanders et al. 2019
15q11.2q12	15	22805313	28390339	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
15q13.1q13.2 BP3-BP4	15	29161368	30375967	DEL	Stefansson et al. 2014
				DUP	Stefansson et al. 2014
15q13.1q13.3 BP3-BP5	15	29161368	32462776	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
15q13.3 BP4-BP5	15	31080645	32462776	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
15q24	15	72900171	78151253	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
15q25.2	15	83219735	85722039	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
CREBBP	16	3775056	3930121	Gene	Satterstrom et al. 2020
16p13.11	16	15511655	16293689	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
16p12.1	16	21950135	22431889	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
16p11.2 distal	16	28823196	29046783	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
16p11.2 distal+proximal	16	28823196	30200773	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
16p11.2 proximal	16	29650840	30200773	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
16p11.2p12.1	16	21596415	28347808	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
YWHAE	17	1247834	1303556	Gene	Coe et al. 2014
PAFAH1B1	17	2496923	2588909	Gene	Coe et al. 2014
17p12	17	14141387	15426961	DEL	Stefansson et al. 2014
				DUP	Stefansson et al. 2014
17p11.2	17	16812771	20211017	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
17q11.2 *NF1*	17	29107491	30265075	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
17q12	17	34815904	36217432	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
17q21.31	17	43705356	44164691	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
17q23.1q23.2	17	58302389	60289141	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
22q11.2 distal	22	21920127	23653646	DEL	Coe et al. 2014
				DUP	Coe et al. 2014
22q11.2 proximal	22	19037332	21466726	DEL	Sanders et al. 2019
				DUP	Sanders et al. 2019
SHANK3	22	51113070	51171640	Gene	Satterstrom et al. 2020

Legend: List of recurrent loci and genes previously associated with neuropsychiatric disorders in Coe et al. 2014[1], Stefansson et al. 2014[2], Sanders et al. 2019[3], Satterstrom et al. 2020[4]. Coordinates are presented in hg19 (Homo sapiens (human) genome assembly GRCh37) from Genome Reference Consortium; CHR: chromosome; BP: break points; WBS: William-Beuren syndrome.

Table S2: List of the recurrent neurodevelopmental or neuropsychiatric-associated CNVs identified in the EOP sample

Locus	Start (Hg19)	Stop (Hg19)	Type	Genes totally encompassed	Other CNV≥50Kb	Sex	Age of onset for EOP	ASD	ID
1q21.1 distal	146618800	147820000	DUP	PRKAB2, GJA5, FMO5, CHD1L, ACP6, NBPF24, BCL9, GJA8, GPR89B, AC242628.1	no	Male	13	no	no
1q21.1 distal	146618988	147825855	DUP	PRKAB2, ACP6, FMO5, BCL9, AC242628.1, CHD1L, GJA5, GPR89B, NBPF24, GJA8	no	Male	4	no	no
1q21.1 distal	146535353	147857135	DUP	GPR89B, GJA8, ACP6, AC242628.1, BCL9, PRKAB2, NBPF24, GJA5, CHD1L, FMO5	no	Male	15	yes	no
15q11.2 BP1-BP2	22762571	23080867	DEL	CYFIP1, NIPA2, TUBGCP5	no	Male	5	no	no
16p11.2 proximal	29592843	30264892	DEL	PAGR1, SEZ6L2, SLX1A, TBX6, MVP, TAOK2, BOLA2B, AC093512.2, HIRIP3, SULT1A3, YPEL3, MAZ, CDIPT, PRRT2, INO80E, C16orf54, SPN, NPIPB13, KIF22, C16orf92, CORO1A, ALDOA, TMEM219, ASPHD1, QPRT, KCTD13, ZG16, DOC2A, PPP4C, MAPK3, GDPD3, TLCD3B	no	Male	4	yes	yes
16p11.2 proximal	29652999	30357820	DUP	MVP, C16orf92, MAPK3, SLX1A, KCTD13, ZG16, C16orf54, AC093512.2, TBX6, CORO1A, PAGR1, CDIPT, PRRT2, ASPHD1, PPP4C, ALDOA, KIF22, YPEL3, SULT1A3, TAOK2, TMEM219, GDPD3, QPRT, MAZ, BOLA2B, SEZ6L2, NPIPB13, HIRIP3, SPN, TLCD3B, INO80E, DOC2A	no	Male	13	yes	no
16p13.11	14968855	16267250	DEL	MARF1, BMERB1, RRN3, ABCC1, NTAN1, CEP20, PDXDC1, MYH11, NDE1, MPV17L, NPIPA5, NPIPA1	yes (2)	Male	12	no	yes
16p13.11	14906734	16388596	DEL	RRN3, MPV17L, BMERB1, NOMO1, ABCC1, ABCC6, NPIPA5, NDE1, NTAN1, CEP20, MYH11, NPIPA1, PDXDC1, MARF1	no	Female	6	yes	--
16p13.11	14897788	16293305	DUP	MPV17L, MARF1, BMERB1, NPIPA5, CEP20, RRN3, NPIPA1, MYH11, NTAN1, NDE1, NOMO1, ABCC1, PDXDC1	no	Female	13	no	no
16p13.11	14897761	16276117	DUP	NTAN1, PDXDC1, NPIPA1, RRN3, MPV17L, MYH11, CEP20, NOMO1, ABCC1, NDE1, MARF1, BMERB1, NPIPA5	no	Female	4	yes	yes
22q11.2 proximal	18894835	20311763	DEL	ARVCF, PRODH, TANGO2, COMT, DGCR6L, DGCR8, ZDHHC8, TBX1, CLDN5, CLTCL1, SEPTIN5, MRPL40, RANBP1, HIRA, RTN4R, GSC2, UFD1, TXNRD2, DGCR2, CDC45, TRMT2A, SLC25A1, GP1BB, RTL10, ESS2, C22orf39, GNB1L, CCDC188, TSSK2	yes (1)	Male	13	no	yes

Legend: A CNV was considered as recurrent only if it overlapped at more than 40% with a loci described by Kendall *et al.* (2019).[1] Coordinates are based on Hg19 map of the genome (Homo sapiens (human) genome assembly GRCh37 from Genome Reference Consortium). CNV: Copy number variant; DEL: deletion; DUP: duplication; ASD: autism spectrum disorder; ID: intellectual disability; --: unknown.

Table S3: Phenotypes of 1q21 duplication patients in comparison to Rosenfeld, *et al*. 2012[5]

Feature	Distal (BP3–BP4) duplications (Rosenfeld, *et al.)*	1440-01	1464-01	1468-01
Short stature	9.5%	-	-	-
Failure to thrive/feeding problems	20%	?	?	?
Microcephaly	8.7%	-	-	-
Macrocephaly	43.5%	-	-	-
Developmental delay/intellectual disability	77.8%	-	+	-
Hypotonia	14.8%	-	-	-
Seizures	17.9%	-	-	-
Autistic features	41.2%	-	-	+
Other behavioral problems	4%	+	+	+
Hearing loss	0%	-	-	-
Brain abnormalities	75%	-	(left temporal slowing on EEG)	-
Dysmorphic features	51.9%	-	-	-
Cataracts	0%	-	-	-
Other ophthalmologic abnormalities	14.8%	-	-	-
Craniosynostosis	0%	-	-	-
Skeletal limb abnormalities	0%	-	-	-
Other skeletal anomalies	7.4%	-	-	-
Clinodactyly	3.7%	-	-	-
Ligamentous laxity	3.7%	-	-	-
Cardiac anomalies	28.6%	-	-	-
Lung abnormalities	0%	-	-	-
Renal anomalies	3.7%	-	-	-
Genital anomalies	11.1%	-	-	-
Blood disorders	0%	-	-	-

Table S4: Sensitivity analysis of the enrichment of recurrent CNVs identified in EOP relative to ASD and unselected populations after removing individuals with co-occurring ASD from the EOP cohort

Locus	EOP (N=99)		Controls (LBC, GS, Imagen) (N=16,504)		ASD (SSC, MSSNG) (N=5,540)		EOP vs. Controls OR [95%CI]		EOP vs. ASD OR [95%CI]	
	DEL (N=4)	DUP (N=4)	DEL (N=159)	DUP (N=119)	DEL (N=39)	DUP (N=61)	DEL	DUP	DEL	DUP
1q21.1	0	2	4	7	1	19	--	**48.50 [4.86-259.74] ($p=10^{-3}$)**	--	n.s.
15q11.2 BP1-BP2	1	0	70	50	16	6	--	n.s.	n.s.	--
16p13.11	1	1	8	31	7	12	**42.43 [4.34-217.84] ($p=10^{-3}$)**	n.s.	16.27 [1.63-86.92] (p=0.01)	n.s.
22q11.2 proximal	1	0	0	7	1	7	**inf [4.27-inf] ($p=6x10^{-3}$)**	--	56.31 [0.71-inf] (p=0.03)	--

Legend: Odds ratios are computed using Fisher's exact test for deletions and duplications. Significant p-values after FDR correction are in bold (≤ 0.008). As sensitivity analysis, we removed from the EOP cohort individuals with co-occurring ASD, as well as individuals for whom a co-occurring diagnosis of ASD was unknown (N = 38). EOP: early onset psychosis; Controls: unselected population; ASD: autism spectrum disorder; DEL: deletion; DUP: duplication; OR: Odds ratio; 95% CI: 95% confidence intervals; --: not applicable; n.s.: non-significant.

Table S5: Sensitivity analysis of the effect-size of CNVs pathogenicity on EOP risk after removing individuals with co-occurring ASD from the EOP cohort

Mean score (sum 1/LOEUF) per individual						OR [95%CI] p-value			
EOP (n=99)		ASD (n=5,540)		Controls (n=16,504)		EOP vs. controls		EOP vs. ASD	
DEL	DUP	DEL	DUP	DEL	DUP	DEL	DUP	DEL	DUP
1.00	1.42	0.86	1.88	0.23	0.94	**1.24 [1.19-1.30] p=3x10^{-04}**	1.07 [1.03-1.11] p=0.17	n.s.	n.s

Legend: Effect of gene dosage on EOP risk. Odds ratios are computed using logistic regressions including the sum of 1/LOEUF score (pathogenicity score) for genes totally encompassed in deletions and duplications as the two main explanatory variables. OR represents the mean risk conferred by a deletion or a duplication with including 1 intolerant gene (a LOEUF ≤ 0.35). All models were adjusted for sex. Significant p-values are in bold (≤ 0.05). All the models were adjusted for sex. Significative p-values are in bold (≤ 0.05). As sensitivity analysis, we removed from the EOP cohort the individuals with co-occurring ASD, as well as individuals for whom a co-occurring diagnosis of ASD was unknown (N excluded = 38). EOP: early onset psychosis; Controls: unselected population; ASD: autism spectrum disorder; LOEUF: Loss-of-function observed/expected upper fraction; DEL: deletion; DUP: duplication; OR: Odds ratio; 95%CI: 95% Confidence interval; n.s.: non-significant.

SUPPLEMENTARY REFERENCES

1 Coe, B. P., Witherspoon, K., Rosenfeld, J. A., et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nature genetics*, 2014; *46*(10), 1063-1071.

2 Stefansson, H., Meyer-Lindenberg, A., Steinberg, S., et al. CNVs conferring risk of autism or schizophrenia affect cognition in controls. *Nature*, 2014; *505*(7483), 361-366.

3 Sanders, S. J., Sahin, M., Hostyk, J., et al. A framework for the investigation of rare genetic disorders in neuropsychiatry. *Nature medicine*, 2019; *25*(10), 1477-1487.

4 Satterstrom, F. K., Kosmicki, J. A., Wang, J., et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. *Cell*, 2020; *180*(3), 568-584.

5 Rosenfeld J. A., Traylor R. N., Schaefer G. B., et al. Proximal microdeletions and microduplications of 1q21.1 contribute to variable abnormal phenotypes. *Eur J Hum Genet,* 2012; 20: 754–61.